Predicting Future Cognitive Impairment in *De Novo* Parkinson's Disease Using Clinical Data and Structural MRI

Nicola Smith^a, Owen Williams^b, Lucia Ricciardi^c, Francesca Morgante^{c,d}, Thomas R Barrick^c, Mark Edwards^c, Christian Lambert^a,

Affiliations:

- a. Wellcome Centre for Human Neuroimaging, 12 Queen Square, London WC1N 3AR, United Kingdom
- b. University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG
- c. Neurosciences Research Centre, Molecular and Clinical Sciences Research Institute, St George's University of London, SW17 0RE, UK
- d. Department of Clinical and Experimental Medicine, University of Messina, Messina, Italy

Corresponding Author:

Name: Dr Christian Lambert

Address: Wellcome Centre for Human Neuroimaging, 12 Queen Square, London WC1N 3AR, United Kingdom

Telephone: [+44\(0\)20 8266 6468](tel:+44(0)20 8266 6468)

Email: christian.lambert@ucl.ac.uk

RUNNING TITLE:

Preclinical Cognitive Impairment in PD

ABSTRACT

BACKGROUND

Parkinson's disease is the second most common neurodegenerative condition and associated with increasing cognitive dysfunction as the disease progresses. However, subtle cognitive deficits can be detected at diagnosis in 42% of individuals, suggesting that damage may already be present. Our aim was to determine clinical and structural differences in those recently diagnosed with PD who later develop cognitive impairment, and whether these changes predict future cognitive decline.

METHODS

Clinical and imaging data was acquired from the Parkinson's Progression Markers Initiative for 318 individuals with a diagnosis of Parkinson's disease and baseline 3T T1-weighted MRI. The cohort was divided according to cognitive status over follow-up, with 9 individuals developing Parkinson's disease dementia, 102 developing mild cognitive impairment and 207 remaining cognitively unaffected.

FINDINGS

At baseline, those who went on to develop cognitive impairment (mild cognitive impairment or dementia) were **older with more severe motor and non-motor symptoms (anosmia, rapid eye movement sleep behaviour disorder, depression)**. Grey matter loss was present in those destined for Parkinson's disease dementia in the **precuneus, hippocampi, primary olfactory cortex, lingual gyrus, temporal cortex and cerebellum**. Those who later developed mild cognitive impairment had an attenuated but similar pattern of **grey matter loss in the temporal lobe, lingual gyrus and cerebellum**. Using support vector machines with a feature selection step, future cognitive impairment could be predicted using 11 clinical variables (AUC = 0.81), structural imaging (AUC = 0.72) or a combination of these two modalities (AUC = 0.85). These models more accurately predicted those who developed dementia (subgroup sensitivity 100%).

INTERPRETATION

Significant abnormalities in cortical structure is present at least three years before dementia manifests in Parkinson's disease, with associated differences in clinical profiles. Combining this data provides a technique to accurately identify future cognitive impairment, providing a non-invasive way to stratify individuals early on.

KEYWORDS:

1. Parkinson's Disease Dementia
2. Mild Cognitive Impairment
3. Voxel Based Morphometry
4. Machine Learning
5. Pre-clinical diagnosis.

INTRODUCTION

Cognitive impairment is a common and disabling feature of Parkinson's disease, particularly as the disease progresses (Compta *et al.*, 2013). The prevalence of dementia in PD is 30%, with at least 75% of patients being affected after 10 years of disease (Aarsland & Kurz, 2010), causing a major impact on health and social care costs and quality of life (Soh *et al.*, 2011). The spectrum of Parkinson's disease cognitive impairment (PD-CI) can be divided in two broad categories: Parkinson's disease with mild cognitive impairment (PD-MCI) and Parkinson's disease with Dementia (PDD). While dementia is typically considered a feature of advanced disease (Aarsland *et al.*, 2003), evidence of PD-MCI (Litvan *et al.*, 2012) can be detected in 42% of newly diagnosed PD patients (Yarnall *et al.*, 2014).

Demographic, clinical and morphological predictors of PDD have been extensively studied, mostly through cohort studies of treated Parkinson's disease patients with long disease durations. Identified associations include older age, male gender, lower education, more severe parkinsonism, depression, Rapid eye movement sleep behaviour disorder (RBD), autonomic dysfunction, presence of MCI and visual hallucinations (Compta *et al.*, 2013; Hu *et al.*, 2014; Zhu *et al.*, 2014). However, only a few studies prospectively assessed risk factors in newly diagnosed PD patients (Anang *et al.*, 2014; Mak *et al.*, 2015; Schrag *et al.*, 2017). Older age appears to be the most consistent demographic risk factor in newly diagnosed Parkinson's disease, whereas the contribution of male gender (Anang *et al.*, 2014; Cholerton *et al.*, 2018) and fewer years of education (Mak *et al.*, 2015; Nicoletti *et al.*, 2019) are less consistent. Clinical risk factors for development of MCI and dementia included presence of RBD at baseline (either assessed by polysomnography (Anang *et al.*, 2014) or by RBD screening questionnaire (Schrag *et al.*, 2017)), as well as autonomic dysfunction, more severe motor impairment, gait impairment and falls (Anang *et al.*, 2014). Recently, studies analysing the Parkinson's Progression Marker Initiative (PPMI) cohort have also shown an associations with low cerebrospinal fluid amyloid β 42 levels, striatal region of interest dopamine transporter (DAT)-imaging uptake with the development of cognitive impairment in newly diagnosed Parkinson's disease (Schrag *et al.*, 2017; Pagano *et al.*, 2018).

Widespread cortical atrophy in established PDD has been well described, particularly in medial temporal structures, posterior cingulate, precuneus, lingual gyrus and prefrontal cortex (Gee *et al.*, 2017; Melzer *et al.*, 2012). In contrast, the reported abnormalities in PD-MCI are variable,

with a possible trend to small regions of grey matter loss in the pre-frontal and lateral-temporal cortices (Hanganu et al., 2013; Sasikumar et al., 2020; Yarnall *et al.*, 2014). More recently studies have examined the structural changes in the early phase of PD cognitive impairment (PD-CI), demonstrating significant atrophy in fronto-temporal, thalamus, accumbens and caudate compared to those without cognitive impairment (Foo *et al.*, 2017; Zhou *et al.*, 2019). In a recent study using the PPMI cohort, individuals who convert to PD-MCI were found to have significantly greater right temporal pole atrophy than nonconverters at baseline (Zhou *et al.*, 2019).

Notably, few neuroimaging studies have combined clinician and structural metrics to predict conversion to cognitive impairment in newly diagnosed Parkinson's disease. Predicting cognitive impairment could potentially inform clinical discussions with patients and families regarding prognosis, could influence decisions on prescribing or selecting candidates for advanced treatments such as deep brain stimulation surgery, and help identify cohorts for clinical trials directed towards preserving cognitive function.

Our aim was to determine whether there are clinical or structural changes already present at diagnosis in individuals with Parkinson's disease destined to develop cognitive impairment, and whether these abnormalities can be used to predict future cognitive decline. To investigate this, we used the MRI and clinical data from individuals diagnosed with Parkinson's disease participating in the Parkinson's Progression Markers Initiative (PPMI, Parkinson's Progression Markers Initiative 2014).

METHODS

Data Acquisition and Cohort Selection

All available data were downloaded from the PPMI website (<http://www.ppmi-info.org/>) on the 22nd November 2016. Individuals classified as “excluded”, “declined” or “pending” were removed. Sub-selecting those with a diagnosis of Parkinson’s disease, we divided the group into three categories using the most severe documented cognitive category for an individual across all time-points: Parkinson’s disease without dementia (PD-NON), PD-MCI or PDD. We removed any participant where the confidence in the cognitive category (documented by the PPMI investigators) was less than 50%. We then selected only those who had 3T T1-weighted MRI available (Tardif *et al.*, 2010). This provided a final dataset consisting of 207 PD-NON, 102 PD-MCI and 9 PDD (Fig. 1).

The diagnosis of dementia or mild cognitive impairment (MCI) was defined via clinical assessment by the PPMI investigators. Full details are available on the PPMI website. The determination of dementia was based on the following criteria: (1) impaired cognitive function in more than one domain more than 1.5 standard deviations below the standardised mean, (2) decline from pre-morbid function as determined by investigator or reports from informants and (3) significant impact on daily function. The determination of MCI was based on (1) at least one cognitive domain impaired, more than one standard deviations below the standardised mean, (2) decline from pre-morbid function, and (3) lack of significant impact on daily function. Combined, the available cognitive and clinical measures provided enough evidence to satisfy the criteria for probable PDD or level 1 PD-MCI (Emre *et al.*, 2007; Litvan *et al.*, 2012). To calculate the time from diagnosis to cognitive impairment, we took the first PPMI visit where the cognitive category of interest was assigned as the time of cognitive conversion. To calculate the number of individuals who had abnormal cognitive tests at baseline, we defined abnormal as more than one standard deviation from the age-corrected norms. Single domain MCI was defined as abnormalities on two tests of one cognitive domain, and multiple-domain MCI was defined as at least one deficit measured in more than one cognitive domain.

No patient was on any dopaminergic medication at baseline. We extracted the following demographic and clinical data from the baseline PPMI visit: age, gender, years of education, date of symptom onset and date of diagnosis, Movement Disorders Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) (Goetz *et al.*, 2008) parts I-III and total (sum of parts I-

III), Postural Instability and Gait Disorder (PIGD) score (Stebbins *et al.*, 2013), tremor score, tremor-PIGD ratio. We also retrieved the following ratings for non-motor symptoms: REM sleep disorder questionnaire (RBDQ) (Stiasny-Kolster *et al.*, 2007); Geriatric depression scale (GDS) and State-trait anxiety inventory (STAI); Epworth sleepiness scale. Autonomic function was rated by the “Scales for Outcomes in Parkinson's disease – Autonomic” (SCOPA-AUT), and olfaction by the University of Pennsylvania Smell Identification Test (UPSIT) (Doty *et al.*, 1984).

Global cognitive function was measured using the Montreal cognitive assessment (MoCA). In addition to this, the following age-norm corrected measures for four cognitive domains were also extracted: 1) Episodic memory: Hopkins verbal learning test revised (HVLT); 2) visuo-spatial function: Benton judgment of line orientation (JLO); 3) attention and executive function: Letter number sequencing (LNS) and Symbol digit modalities test (SDM). 4) Language - Semantic fluency (SF).

Neuroimaging Pre-Processing

Using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>), the raw DICOMS were imported and each image manually checked. The original image resolution was between 1mm and 1.1mm isotropic, so the MRI data was re-aligned to MNI orientation and re-sliced to 1mm isotropic voxel resolution in subject space (Friston *et al.*, 1995). All re-orientated T1-weighted images were segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) tissue classes (Ashburner & Friston, 2005). These segmentations were used to generate cortical thickness maps using the Voxel-Based Cortical-Thickness toolbox (Hutton *et al.*, 2008) in SPM8 (this toolbox is currently incompatible with SPM12) with a sampling resolution of 0.5mm, CSF smoothness 3mm, CSF thinness 0.65 and number of dilations was set to 1.

The baseline GM, WM and CSF segmentations were then used to generate a group average template using the diffeomorphic warping “Shoot” algorithm in SPM12 (Ashburner & Friston, 2011). The segmentations were then warped using the resulting deformation fields, modulated with the Jacobian determinant data and smoothed with a 6mm full width half maximum (FWHM) Gaussian kernel. The CT maps were also warped using the corresponding deformation fields. The warped CT maps were then used to generate FWHM 6mm smoothed warped weighted images produced as described by Draganski *et al.*, 2011 (Hutton *et al.*, 2009;

Draganski et al, 2011). Total intracranial volume (TIV) was calculated by integrating the Jacobians within the three tissue classes, with the segmentations binarised at a threshold of 0.2 (Malone *et al.*, 2015).

Statistical Analysis

All analyses were carried out using MATLAB 2016b. For each demographic and clinical variable, three group comparisons (PD-NON vs PD-MCI, PD-NON vs PDD and PD-MCI vs PDD) were assessed using the Kruskal-Wallis test, and if significant, tested further using a Mann–Whitney U test. For categorical data, the χ^2 test was used. Significance was set at a multiple comparison Bonferroni corrected $P < 0.001$.

For neuroimaging data, all analyses were performed using Analysis of Variance (ANOVA) in SPM12 controlling for age, handedness, total intracranial volume, gender, research centre and years of education. Peak voxel Family Wise Error (FWE) corrected $P < 0.05$ was considered significant. Significant results are displayed at both $P < 0.001$ uncorrected and FWE corrected $P < 0.05$.

Prediction Analysis

All prediction analyses used binary SVMs in MATLAB 2016b to identify those destined for PD-CI, defined as either PD-MCI or PDD. All data classification used support vector machines and a feature selection step designed in this study to identify the most important features (detailed below and summarised in Supplementary Figure 2). Three potential feature sets were tested: Clinical data alone, MRI data alone and combined clinical and MRI data. Prediction accuracy metrics were calculated as the average over 10-fold cross validation training loops. Each of the final SVM classifiers were tested against an independent dataset that included 20% of the sample that was randomly selected from each group (40 PD-NON, 22 PD-CI consisting of 20 PD-MCI and 2 PDD). Statistical significance, using permutation testing with 1000 permutations, was set at $P < 0.05$.

The feature selection for the SVM was performed as follows (Supplementary Figure 2). Initially the complete dataset of potential features, X , was divided into training, X_t , and independent, X_i , testing datasets. A four-step process was then applied to the training data, X_t , to perform feature selection and identify the best SVM classifier. At each point designated X_t in steps 1-3 (Supplementary Figure 2), a random number generator was seeded using the

timestamp, and the training-test splits re-allocated. The final classifier was then applied to the independent test data, X_i :

1) **Feature weighting:** Using an iterative process, initially ten-fold cross validation was used and average absolute values for the feature weights (W) were extracted. The feature with the highest absolute weight (W_{max}) was identified and removed from $X_{t'}$. The position (i.e. relative importance) of this feature was stored in vector, F . This process was repeated until no potential features remained in $X_{t'}$. The entire process was repeated to construct a matrix of feature order positions for different ten-fold cross validations and was used to calculate a vector containing average feature positions.

2) **Feature extraction:** Each identified feature was then sequentially added to matrix $X_{t'}$ according to their average position from lowest (i.e. most important) to highest (i.e. least important). At each iteration, ten-fold cross validation was used to calculate the average area under the curve of the SVM classification and this was stored in vector, A . The features that were selected for the prediction analysis were those that provided the maximum value in A (i.e. the best model corresponding to the maximum area under the curve).

3) **Classification of training data:** The selected features were applied in a SVM classification to the training data, $X_{t'}$, in a ten-fold cross validation. Classification accuracy metrics for this final classification were calculated.

4) **Independent testing of the classifier:** The final SVM classification model was applied to the independent test data to extract independent classification accuracy metrics.

Slight adaptations to this feature identification and selection steps were required in application to the clinical data, imaging data, and combined clinical and imaging data. These are detailed below.

Predicting Cognitive Decline: Clinical

The prediction method considered 29 clinical variables as constituents of X (summarised in Supplementary Material 1). To enable inclusion of the tremor/PIGD ratio, the value was converted to a standardised scale. This was achieved by normalising each of the three

phenotype scores to have zero mean and standard deviation one (i.e. a z-statistic) and recombining these three rescaled scores such that values for the PIGD phenotype were between 0 and 1, indeterminate between 1 and 2 and tremor dominant between 2 and 3. Each of the 29 clinical metrics were converted to z-statistics and separated into training, X_t , and test data, X_i . The SVM classifier was trained on the X_t as follows:

- 1) The SVM was trained using 10-fold cross validation, with each group containing identical proportions of PD-NON, PD-MCI and PDD. At the end of each fold, the average absolute weight vector was calculated. The feature corresponding to the maximum weight vector (i.e. contributing most to the model estimation) was removed from matrix X_t , and then the process repeated. This allowed the clinical features to be ordered according to the amount they contribute to predicting future PD-CI. To ensure a stable model, this entire process was repeated 500 times, and the order of the features set according to their average position over each permutation. 2) An SVM was trained using 10-fold cross validation, adding the reordered features one at a time and calculating the area under the receiver operator curve for each. 3) The final SVM and clinical features were selected using the one that achieved the maximum AUC, and 10-fold cross validation used to calculate the overall model accuracy metrics. 4) This final model was then applied to the independent test data, X_i , and the overall accuracy metrics extracted.

Predicting Cognitive Decline: Neuroimaging

A similar SVM technique was applied to the neuroimaging data. Initially, the group average grey matter mask was thresholded at 0.1 and used to construct a Gram matrix (X) from the smoothed, modulated GM voxel data over all subjects. The independent test set was removed, X_i , and the SVM classifier trained on the remaining data, X_t , as follows:

- 1) 10-fold cross validation was used. At each loop, the average weight vector was calculated. To obtain a voxel-wise representation of these weights, each feature weight was multiplied by the corresponding subjects native data vector in X_t . This was repeated over all subjects, and summed to provide a final weight vector in voxel space. Voxels containing the largest absolute weight values corresponding to 0.01% of the total GM volume were then removed from X_t , and a new Gram matrix calculated from the remaining data. This process was continued until the average classification error exceeded 40%. This entire process was iterated 100 times, and the order in which the voxels were removed was stored, from which an average feature map was

identified consisting of 980 unique features ordered by importance. 2) An SVM was then trained using 10-fold cross validation, by including the reordered features at each loop in order of importance (high to low), and calculating the area under the receiver operator curve for the resultant SVM classification. The best (final) SVM classification was considered to include the features that achieved the maximum AUC. 3) The selected features were input to a final SVM classification and 10-fold cross validation was used to calculate the overall model accuracy metrics. 4) This final model applied to the independent test data (X_i) and the independent classification accuracy metrics were extracted.

To compare the final feature map to the reported literature (PD-MCI or preclinical PD-CI (Melzer *et al.*, 2012; Hanganu *et al.*, 2013; Anang *et al.*, 2014; Pereira *et al.*, 2014; Mak *et al.*, 2015)), we warped the group average final weight map to 1mm MNI space using FSL FNIRT. We then manually mapped the reported peak-statistical coordinates into MNI space using ITK-SNAP, creating an image. The distance between each reported statistical peak and all non-zero elements of the group average final weight map (Fig. 6) were then calculated, allowing the minimum distance to be extracted and reported.

Predicting Cognitive Decline: Clinical and Neuroimaging Combined

The independent and training data were separated. Using the previously defined optimal features, a Gram matrix, X_t , was generated for the clinical and imaging data separately, and then combined to form a matrix representing a weighted average of the two types of optimised features (i.e. clinical and imaging). To identify the optimal weighting parameters, a grid search through 100 weights between 0 and 1 was performed using the training data, X_t . At each weighting pair, the average AUC over a 10-fold cross validation was calculated (as per Step 2 above). The weighting parameters that provided the maximum AUC were selected, and then Steps (3) and (4) were then applied to determine the SVM classification accuracy metrics for X_t and X_i , as above.

RESULTS

Baseline Demographics and Clinical Profile

Out of 318 Parkinson's disease subjects at baseline without any cognitive impairment, 102 (32.1%) developed PD-MCI and 9 (2.8%) PDD. Table 1 summarises demographical and clinical differences among the three groups at baseline visit.

Years of education and gender distribution were comparable among the three groups. At diagnosis, PD-MCI and PDD were significantly older compared to PD-NON. There were no significant group differences between time of symptom onset to PD diagnosis. The PDD group did have a longer amount of follow-up in the PPMI study (average 4.57 years). Nevertheless, there were no significant differences in the time to MCI conversion between PDD and PD-MCI groups; moreover, the time to PDD conversion (3.45y) was less than the average amount of follow-up in the PD-MCI group (3.62y). This suggests that the disease progression in the PDD group is more aggressive, rather than a consequence of longer follow-up in the PPMI, and is reflected by the five-year progression of their clinical variables (Fig. 2).

Groups destined for cognitive impairment (either PD-MCI or PDD) had worse motor symptoms as per MDS-UPDRS III, and more severe non-motor features in the form of hyposmia, RBDQ score, depression and autonomic symptoms compared to the PD-NON group. Between the groups who later develop cognitive impairment, those destined for PDD had significantly worse hyposmia compared to PD-MCI.

On average, we found at least one abnormal cognitive test in 18.10% of all PD patients at baseline (Table 2). Patients who developed PD-MCI and PDD at follow-up had more abnormal scores that crossed different cognitive domains at baseline compared to PD-NON ($P < 0.001$, Supplementary Table 1). However there were no differences in any of the individual cognitive scores that survived Bonferroni correction (Table 1).

Baseline Voxel-Based Morphometry

Widespread regions of reduced GM density were found in those destined for cognitive impairment that appears to form a progressive spectrum of atrophy between PD-NON, PD-MCI and PDD (Fig. 3). Comparing PD-NON and PD-MCI, the latter group was associated with reduced GM in the cerebellar association zone (Crus I/II), the left angular gyrus, left

ventral pre-motor cortex and lingual gyrus. Progressing from PD-MCI to PDD, revealed additional areas of reduced GM bilaterally in the medial precuneus, posterior cingulate, hippocampi, primary olfactory cortex, and cerebellum (lobules VI, VIII). Comparing PD-NON to PDD demonstrates the full extent of the atrophy in PDD. Whilst additional areas are seen between PD-NON versus PDD that do not appear in the PD-MCI analyses, these may represent regions of evolving atrophy in PD-MCI that fall midway between the PD-NON and PDD spectrum, but are not sufficiently different to achieve significance in either comparison.

No regions of reduced WM were observed in PD-MCI compared to PD-NON. Reduced WM in PDD was observed compared to PD-MCI in the fronto-parietal WM bilaterally, the isthmus of the corpus callosum and the anterior temporal WM (Fig. 4).

Voxel-based cortical thickness was also evaluated between the three groups at the time of PD diagnosis. No regions survived the FWE corrected $P < 0.05$. Given previous reports of differences (Mak *et al.*, 2015), we inspected $P < 0.001$ uncorrected. This revealed several regions of thinning that were similar to the VBM analysis (Supplementary Figure 1).

Predictors of Cognitive Decline: Clinical Features

The optimal SVM model consisted of 11 clinical variables (Fig. 5a and 5b) which were: Age at Parkinson's disease diagnosis, HVLT-R recognition discrimination index, gender, letter number sequencing, SCOPA-AUT, RBDQ total, GDS, HVLT-R total recall, MoCA total, Epworth sleepiness scale, and symbol digit modalities.

The average classification metrics (Supplementary Table 2) were: balanced prediction accuracy 73%, AUC 0.81, sensitivity 62%, specificity 85%, positive predictive value 77% and negative predictive value 76%. Applying the final model to independent test data resulted in a balanced accuracy of 76%, AUC 0.89, sensitivity 63%, specificity 88%, positive predictive value 75% and negative predictive value 82% (permutation testing $P < 0.001$). These results imply that our clinical model can significantly predict future cognitive impairment, and that this generalises well to previously unseen data.

Despite the small numbers of PDD, all training loops were balanced allowing prediction accuracies for the three cognitive sub-groups to be extracted. Over the entire cross validation,

those with positive predictions were more likely to ultimately develop PDD (sensitivity 100%), compared to PD-MCI (sensitivity 53%).

Predictors of Cognitive Decline: Neuroimaging Features

The optimal MRI SVM model consisted of 269 features (Fig. 5c and 5d), corresponding to 27% of all available grey matter. The distribution of these features overlap with VBM findings, but included more of the cerebellum and dorso-lateral temporal lobe, and additional regions such as the bilateral striatum, orbitofrontal cortex and entorhinal cortex (Fig. 6). This map corresponds to the previously reported abnormalities in PD-MCI or preclinical PD-CI, with 31% of published peak voxel coordinates completely overlapping and the remainder within 4.72mm (+/-3.26mm).

The average classification metrics (Supplementary Table 2) were: balanced prediction accuracy 69%, AUC 0.72, sensitivity 46%, specificity 93%, positive predictive value 82% and negative predictive value 71%. Applying the final model to independent test data resulted in a balanced accuracy of 61%, AUC 0.66, sensitivity 37%, specificity 86%, positive predictive value 58% and negative predictive value 71% (permutation testing $P < 0.01$). These results imply that the imaging model can also significantly predict future cognitive impairment, though not as accurately as clinical data alone. Again, the PDD subgroup was more accurately classified compared to PD-MCI (balanced accuracy 91% vs. balanced accuracy 65%). These results highlight that whilst cognitive impairment in PD is a heterogeneous entity, there is a distinct subgroup that will progress to dementia early on, with no differences in disease duration, average time to MCI or baseline cognitive profile, but with profound abnormalities in brain structure during the pre-symptomatic phase.

Predicting Cognitive Decline: Combined Clinical-Neuroimaging Features

Finally, we combined the clinical and imaging parameters. Using a grid search, the optimal model had a clinical weighting of 0.53 and imaging weighting of 0.01 (Fig. 5e and 5f). The average classification metrics (Supplementary Table 2) using the combined clinical imaging data were: balanced prediction accuracy 74%, AUC 0.85, sensitivity 62%, specificity 87%, positive predictive value 75% and negative predictive value 76% (permutation testing $P < 0.001$). Applying the final model to the independent test data resulted in a balanced accuracy of 78%, AUC 0.86, sensitivity 63%, specificity 92%, positive predictive value 80% and

413 negative predictive value 82%. These results show that the combined model is the most
 414 accurate, and generalises to unseen data. Again, positive predictions were more likely in PDD
 415 (sensitivity 100%), compared to PD-MCI (sensitivity 53%). Combination of clinical and
 416 neuroimaging features in the model provides positive prediction probabilities for developing
 417 PDD, PD-MCI and PD-NON of 0.61, 0.32 and 0.08 respectively. Whilst the clinical data is the
 418 major contributor to this (Fig. 5e) the additional MRI data improves the specificity by better
 419 discriminating the PD-NON cohort.

420

421 **DISCUSSION**

422 This work demonstrates that extensive grey and white matter loss is present at least three years
 423 before cognitive impairment clinically manifests in Parkinson's disease. Moreover, the
 424 structural abnormalities in newly diagnosed Parkinson's disease who later develop PD-MCI
 425 overlap with those encountered in the group developing PDD suggesting an intermediate stage.
 426 Those who develop PD-MCI or PDD are significantly older with more severe motor disability,
 427 hyposmia, depression and higher frequency of RBD symptoms at baseline. Those destined for
 428 early PDD compared to PD-MCI have more severe hyposmia. Finally, we show that future
 429 cognitive impairment in Parkinson's disease can be predicted using clinical variables (AUC =
 430 0.81), structural imaging (AUC = 0.72) or a combination of these two modalities (AUC = 0.85),
 431 and that these models accurately predict the emergence of early PDD (sensitivity 100%). These
 432 results suggest there is a distinct subgroup that experience earlier cognitive decline and are
 433 associated with a characteristic pattern of brain atrophy.

434

435 Parkinson's disease is diagnosed once the motor symptoms become clinically detectable. This
 436 transition occurs once 60-70% of the dopaminergic neurons in the substantia nigra have been
 437 lost, and is the hallmark of stage III disease (Braak *et al.*, 2003). In our study, there was no
 438 difference between motor symptom onset to diagnosis and baseline assessment in the three
 439 groups, suggesting that they passed this threshold at a similar time. Following a diagnosis of
 440 Parkinson's disease, there are three main pathological entities that can interact or occur in
 441 isolation which may contribute to the development of dementia: Lewy-body disease,
 442 Alzheimer's pathology and cerebrovascular damage (Hely *et al.*, 2008; Melzer *et al.*, 2012;
 443 Shimada *et al.*, 2013; Gomperts *et al.*, 2016; Irwin *et al.*, 2017). Here, we have shown that a
 444 spectrum of atrophy is present at diagnosis and can predict future cognitive decline, particularly
 445 in the cerebellum, temporo-parietal cortex and primary olfactory structures. The grey matter
 446 loss detected in the group destined for PDD corresponded well with regions of atrophy in
 447 established PDD (Melzer *et al.*, 2012) but with much less involvement of the frontal regions
 448 and the striatum. Additionally, regions of reduced grey matter in the cerebellum, precuneus
 449 and posterior cingulate overlapped with abnormalities found in patients with clinical dementia
 450 with Lewy bodies or PDD who also have evidence of A β Alzheimer's pathology on PET
 451 amyloid imaging (Shimada *et al.*, 2013). It may be that this particular sub-group, who rapidly
 452 decline cognitively and are known to progress more rapidly to death (Hely *et al.*, 2008), contain

a significant proportion of individuals with co-existing Alzheimer's pathology, a hypothesis which is consistent with the histological data (Irwin *et al.*, 2017).

The PD-MCI group displayed an attenuated but similar pattern of grey matter loss in the temporal lobe, lingual gyrus and cerebellum. This may represent an intermediate form between PD-NON and PDD. The SVM feature map represents regions where damage can jointly contribute to the emergence of PD-CI irrespective of the cause, and involved areas known to be affected by dementia with Lewy bodies, Alzheimer's disease and PDD (Colloby *et al.*, 2014) (Fig. 6). This may account for variation seen in previous studies where the precise results will depend upon the mixture of pathologies sampled by that particular cohort.

Recently, Schrag and colleagues used the PPMI dataset to develop a model to predict future cognitive impairment in recently diagnosed *de-novo* Parkinson's disease (Schrag *et al.*, 2017) followed up for 2 years. They found the five most strongly associated variables with cognitive impairment were age, UPSIT score, RBDQ score, CSF A β 42, and mean caudate DAT uptake. Combining these with logistic regression gave an AUC of 0.80 for predicting future cognitive impairment. The PPMI cohort analysed in our study have a longer follow-up (up to 5 years), which allowed us to include a higher proportion of patients with cognitive impairment. We confirmed that baseline age, UPSIT score, GDS and RBDQ scores are significantly different in those destined for cognitive decline, but also show differences between SCOPA-AUT in agreement with previous work (Anang *et al.*, 2014). Additionally, we have achieved a model that can more accurately predict future cognitive decline (AUC 0.85) and correctly classifies 100% destined for PDD. Whilst the differences in prediction accuracies may partly be due to greater numbers of PD-MCI or PDD in this work, the major advantage of our approach is it only requires non-invasive data. Despite the MRI component only making a minor contribution to the final model in this work, it did make a significant difference. Additionally, this work uses an extremely simple imaging modality and methodological approach, and it is likely that multimodal, quantitative MRI measures may further improve these by better detecting microstructural brain abnormalities at an individual subject level (Lambert *et al.*, 2013). However, future work is required to acquire these high quality quantitative datasets for this patient cohort.

The small group size of the Parkinson's disease dementia cohort is a limitation to further subgroup analysis in this work. Additionally, the cognitive profiles that could only provide level one evidence of PD-MCI or PDD, which would contribute to the clinical heterogeneity and may well account for the lower specificity in predicting PD-MCI. When this analysis was undertaken, the genetic cohort arm of the PPMI was only 54% complete, and none had been followed up for more than three years, which will account for the low proportions of SNCA or GBA mutations (Schapira 2015) associated with cognitive impairment in this work. The MRI data collected via the PPMI combines a variety of imaging resolutions, scanner manufacturers and MRI field strengths, hence we chose to analyse a more homogenous 3T cohort to avoid a major source of non-experimental error (Tardif *et al.*, 2010). Finally, this work used a relatively simple classifier and possibly a combination of quantitative multimodal imaging, detailed clinical phenotyping and more advanced classification methods may improve classification parameters further.

CONCLUSION

In summary, this work demonstrates significant abnormalities in cortical structure at least three years before dementia clinically manifests in Parkinson's disease and demonstrates that future cognitive impairment in Parkinson's disease can be accurately predicted using non-invasive baseline data. This provides a quick and simple technique to stratify individuals with PD early in the disease course.

505 ***Acknowledgements***

506 Data used in the preparation of this article were obtained from the Parkinson’s Progression
 507 Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on
 508 the study, visit www.ppmi-info.org. PPMI – a public-private partnership – is funded by the
 509 Michael J. Fox Foundation for Parkinson’s Research and funding partners, including AbbVie,
 510 Avid, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline,
 511 Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Servier, Teva, and
 512 UCB.

514 ***Funding***

515 CL is funded by a Medical Research Council grant (MR/R006504/1), and was also supported
 516 by the Guarantors of Brain Post Doctoral Fellowship for this work. ME receives research
 517 support from a Medical Research Council Grant (MR/M02363X/1)

REFERENCES

- Aarsland D, Andersen K, Larsen JP, Lolk A, Kragh-Sorensen P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch.Neurol* 2003; 60: 387-92.
- Aarsland D, Kurz MW. The epidemiology of dementia associated with Parkinson disease. *J Neurol Sci* 2010; 289: 18-22.
- Anang JB, Gagnon JF, Bertrand JA, Romenets SR, Latreille V, Panisset M, *et al.* Predictors of dementia in parkinson disease A prospective cohort study. *Neurology* 2014;10-1212.
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage* 2005, Jul 1;26(3):839-51.
- Ashburner J, Friston KJ. Diffeomorphic registration using geodesic shooting and gauss-newton optimisation. *Neuroimage* 2011; 55(3): 954-67.
- Braak H, Tredici KD, Rüb U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic parkinsons disease. *Neurobiology of Aging* 2003; 24(2): 197-211.
- Colloby SJ, O'Brien JT, Taylor JP. Patterns of cerebellar volume loss in dementia with Lewy bodies and Alzheimers disease: A VBM-DARTEL study, *Psychiatry Res* 2014; 223: 187-91.
- Compta Y, Pereira JB, Ríos J, Ibarretxe-Bilbao N, Junqué C, Bargalló N, *et al.*, Combined dementia-risk biomarkers in parkinson's disease: A prospective longitudinal study. *Parkinsonism Relat Disord* 2013; 19 (8): 717-24.
- Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiol Behav* 1984; 32: 489-502
- Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RS, Helms G, *et al.*, Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *Neuroimage* 2011; 55(4):1423-34.
- Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, *et al.*, Clinical diagnostic criteria for dementia associated with parkinson's disease. *Mov Disord* 2007; 22(12): 1689-707.
- Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. *Hum Brain Mapp* 1995; 3: 165-89.
- Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, *et al.*, 2008. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 23: 2129-70.
- Gomperts SN. Lewy Body Dementias: Dementia With Lewy Bodies and Parkinson Disease Dementia. *Continuum (Minneap Minn)*, 2016; 22: 435-63.

551 Hanganu A, Bedetti C, Jubault T, Gagnon JF, Mejia-Constain B, Degroot C, *et al.*, Mild
552 cognitive impairment in patients with parkinson's disease is associated with increased cortical
553 degeneration. *Mov Disord* 2013; 28(10): 1360-9.

554 Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of
555 Parkinson's disease: The inevitability of dementia at 20 years. *Mov Disord* 2008; 23(6): 837-
556 44.

557 Hu M, Szewczyk-Królikowski K, Tomlinson P, Nithi K, Rolinski M, Murray C, *et al.*,
558 Predictors of cognitive impairment in an early stage Parkinson's disease cohort. *Mov Disord*
559 2014; 29(3): 351-9.

560 Hutton C, De Vita E, Ashburner J, Deichmann R, Turner R. Voxel-based cortical thickness
561 measurements in MRI. *Neuroimage* 2008; 40(4): 1701-10.

562 Hutton C, Draganski B, Ashburner J, Weiskopf N. A comparison between voxel-based cortical
563 thickness and voxel-based morphometry in normal aging. *Neuroimage* 2009, Nov 1; 48(2):371-
564 80.

565 Irwin DJ, Grossman M, Weintraub D, Hurtig HI, Duda JE, Xie SX, *et al.*, Neuropathological
566 and genetic correlates of survival and dementia onset in synucleinopathies: A retrospective
567 analysis. *Lancet Neurol* 2017; 16(1): 55-65.

568 Lambert C, Lutti A, Helms G, Frackowiak R, Ashburner J. Multiparametric brainstem
569 segmentation using a modified multivariate mixture of gaussians. *NeuroImage: Clinical* 2013;
570 2: 684-94.

571 Lee JE, Cho KH, Song SK, Kim HJ, Lee HS, Sohn YH, *et al.*, Exploratory analysis of
572 neuropsychological and neuroanatomical correlates of progressive mild cognitive impairment
573 in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 2013;jnnp-2013.

574 Litvan I, Goldman JG, Tröster AI, Schmand BA, Weintraub D, Petersen RC, *et al.*, Diagnostic
575 criteria for mild cognitive impairment in parkinson's disease: Movement disorder society task
576 force guidelines. *Mov Disord* 2012; 27(3): 349-56.

577 Mak E, Su L, Williams GB, Firbank MJ, Lawson RA, Yarnall AJ, *et al.*, Baseline and
578 longitudinal grey matter changes in newly diagnosed Parkinson's disease: ICICLE-PD study.
579 *Brain*, 2015; 138: 2974-86.

580 Malone IB, Leung KK, Clegg S, Barnes J, Whitwell JL, Ashburner J, *et al.*, Accurate automatic
581 estimation of total intracranial volume: A nuisance variable with less nuisance. *Neuroimage*
582 2015; 104: 366-72.

583 Melzer TR, Watts R, MacAskill MR, Pitcher TL, Livingston L, Keenan RJ, *et al.*, Grey matter
584 atrophy in cognitively impaired parkinson's disease. *Journal of Neurology, Neurosurgery &*
585 *Psychiatry* 2012; 83(2): 188-94.

586 Parkinson Progression Marker Initiative. The parkinson progression marker initiative (PPMI).
587 *Prog Neurobiol* 2011; 95(4): 629-35.

588 Pereira JB, Svenningsson P, Weintraub D, Brønnick K, Lebedev A, Westman E, *et al.*, Initial
589 cognitive decline is associated with cortical thinning in early parkinson disease. *Neurology*
590 2014; 82(22): 2017-25.

591 Schapira AH. Glucocerebrosidase and Parkinson disease: recent advances. *Molecular and*
592 *Cellular Neuroscience* 2015; 66: 37-42.

593 Schrag A, Siddiqui UF, Anastasiou Z, Weintraub D, Schott JM. Clinical variables and
594 biomarkers in prediction of cognitive impairment in patients with newly diagnosed parkinson's
595 disease: A cohort study. *Lancet Neurol* 2017; 16(1): 66-75.

596 Shimada H, Shinotoh H, Hirano S, Miyoshi M, Sato K, Tanaka N, *et al.*, Beta-Amyloid in
597 Lewy body disease is related to Alzheimer's disease-like atrophy. *Mov Disord* 2013; 28: 169-
598 75.

599 Soh SE, Morris ME, McGinley JL. Determinants of health-related quality of life in Parkinson's
600 disease: a systematic review. *Parkinsonism Relat Disord* 2011; 17: 1-9.

601 Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor
602 dominant and postural instability/gait difficulty groups with the movement disorder society
603 unified Parkinson's disease rating scale: comparison with the unified Parkinson's disease rating
604 scale. *Mov Disord* 2013; 28: 668-70.

605 Stiasny-Kolster K, Mayer G, Schafer S, Moller JC, Heinzel-Gutenbrunner M, Oertel WH. The
606 REM sleep behavior disorder screening questionnaire--a new diagnostic instrument, *Mov*
607 *Disord*, 2007; 22: 2386-93.

608 Tardif CL, Collins DL, Pike GB. Regional impact of field strength on voxel-based
609 morphometry results. *Hum Brain Mapp* 2010; 31(7): 943-57.

610 Yarnall AJ, Breen DP, Duncan GW, Khoo TK, Coleman SY, Firbank MJ, *et al.*, Characterizing
611 mild cognitive impairment in incident Parkinson disease the ICICLE-PD study. *Neurology*
612 2014; 82(4): 308-16.

613

LEGENDS

Table 1: Summary of baseline demographic, clinical and psychometric parameters.

Results that survive Bonferonni corrected $P < 0.001$ are highlighted in bold. NS – Not significant on Mann-Whitney U test ($P < 0.05$). Dash indicates not significant on Krushkal-Wallis test (or chi squared test for categorical descriptors).

Table 2: Full cognitive profile data at baseline showing the percentage of individuals with abnormal scores.

Figure 1: Cohort selection from the PPMI data download. Confidence level refers to PPMI documented confidence in the assigned cognitive category. *Abbreviations: HC = Healthy Control; PD = Parkinson's disease; PRODROME = Prodromal cohort; GENPD = Genetic cohort with PD; GENUA = Genetic cohort unaffected; REGPD = Genetic Registry with PD; REGUA = Genetic registry unaffected; PD-NON = PD without cognitive impairment; PD-MCI = PD mild cognitive impairment; PDD = Parkinson's disease dementia.*

Figure 2: Longitudinal progression in motor and non-motor test scores over five years of follow-up. Note the average time to MCI and PDD was 1.8y and 3.45y respectively. Solid line = PD-NON, dash line = PD-MCI, dotted line = PDD. *Abbreviations: UPDRS – Unified Parkinsons Disease Rating Scale (MDS version); H&Y = Hoehn and Yahr; MoCA – Montreal Cognitive Assessment; SCOPA-AUT - Scales for Outcomes in Parkinson's disease – Autonomic; RBD – Rapid eye movement behavioural disorder questionnaire; STAI - State-Trait Anxiety Inventory; Depression - Geriatric Depression Scale; Epworth – Epworth Sleepiness Scale; QUIP - Questionnaire for Impulsive-Compulsive Disorder in Parkinson's Disease.*

Figure 3: Grey matter voxel-based morphometry. Significant regions of reduced grey matter density in PD subjects with preclinical cognitive impairment (MCI or dementia).

Figure 4: White matter voxel-based morphometry. Significant regions of reduced white matter density in PD subjects with preclinical cognitive impairment (MCI or dementia).

Figure 5: SVM results. A - Clinical optimisation, with 11 features providing the maximum average AUC. B - Receiver Operator Curve (ROC) for clinical predictions. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data. C - Imaging optimisation, with 269 features providing the maximum average AUC (see Figure 6 for the feature map). D - Receiver Operator Curve for imaging predictions. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data. E - Combined clinical-imaging grid search to define weighting parameters to produce weighted-average kernel for the analysis of clinical and neuroimaging data. F - Receiver Operator Curve for the combination of clinical and neuroimaging data. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data.

Figure 6: Spatial distribution of the final 269 imaging features contributing to the optimal SVM predictions. The ordering of the features has been preserved, corresponding to the largest (1) to smallest (269) SVM weights. Whilst these correspond with the voxel-based morphometry results, there is greater representation of the cerebellum and dorso-lateral temporal lobe, and the addition of regions such as the bilateral striatum, orbitofrontal cortex and entorhinal cortex.

	VARIABLE	Group Average (Standard Deviation)			Statistical Comparison <i>P</i> Value		
		PD-NON	PD-MCI	PDD	PD-NON vs PD-MCI	PD-NON vs PDD	PD-MCI vs PDD
DEMOGRAPHICS	N	207	102	9	-	-	-
	Baseline age (Years)	59.20 (9.61)	65.36 (7.69)	69.34 (4.68)	<0.0000001	0.001	NS
	Gender (Male:Female %)	58:42	72:28	78:22	-	-	-
	Education (years)	15.64 (3.12)	15.24 (3.18)	14.89 (4.29)	-	-	-
	Age at diagnosis (years)	58.22 (9.75)	64.56 (7.55)	68.53 (5.38)	<0.0000001	0.001	NS
	Average length of PPMI follow-up (years)	3.25 (1.35)	3.62 (1.26)	4.57 (0.59)	0.01	0.002	0.02
	Time from symptom onset to PD diagnosis (years)	1.84 (2.74)	1.47 (1.77)	1.07 (0.78)	-	-	-
	Time from PD diagnosis to baseline assessment and MRI	0.97 (1.36)	0.77 (1.15)	0.81 (0.81)	-	-	-
	Time from baseline to MCI	-	1.88 (1.19)	1.72 (1.05)	N/A	-	-
	Time from baseline to dementia	-	-	3.45 (1.51)	N/A	N/A	N/A
GENE	LRRK2 (%)	15.00	5.00	0.00	-	-	-
	GBA (%)	5.00	3.00	0.00	-	-	-
	SNCA (%)	0.00	0.00	0.00	-	-	-
MOTOR	Left symptoms dominant (%)	42.51	36.27	11.11	-	-	-
	Right symptoms dominant (%)	52.66	55.88	88.89	-	-	-
	Symmetrical symptoms (%)	2.42	6.86	0.00	-	-	-
	UPDRS Part I	4.22 (3.79)	5.86 (4.35)	7.00 (5.00)	0.0004	NS	NS
	UPDRS Part II	5.29 (3.93)	6.73 (4.97)	9.11 (5.30)	0.01	0.02	NS
	UPDRS Part III	18.72 (9.10)	22.55 (9.49)	28.00 (12.85)	0.0007	0.02	NS
	Total UPDRS Score	28.23 (12.44)	35.14 (13.19)	44.11 (14.79)	0.00002	0.002	NS
	H&Y	1.58 (0.51)	1.72 (0.47)	1.78 (0.44)	-	-	-
	Tremor dominant phenotype (%)	68.12	71.57	55.56	-	-	-
	PIGD dominant phenotype (%)	23.19	22.55	33.33	-	-	-
NON-MOTOR	Indeterminant phenotype (%)	8.69	5.88	11.11	-	-	-
	Total SCOPA-AUS	10.36 (6.85)	13.39 (7.59)	15.22 (8.54)	0.0001	0.05	NS
	RBD Questionnaire total	3.89 (2.45)	5.46 (3.00)	7.44 (3.78)	0.000005	0.005	NS
	Percentage with RBD (RBDQ≥5)	31.88	55.88	77.77	0.00005	0.03	NS
	Olfactory: UPSIT	24.48 (8.19)	21.33 (8.45)	11.56 (4.98)	0.002	0.00003	0.0009
	Epworth Sleepiness Scale	6.06 (3.52)	7.28 (4.47)	7.44 (4.82)	-	-	-
	Geriatric Depression Score	2.10 (2.69)	3.54 (3.35)	3.00 (2.40)	0.000006	NS	NS
	STAI: State Score	47.30 (5.23)	47.26 (5.32)	46.44 (5.10)	-	-	-
	STAI: Trait Score	49.52 (4.43)	48.36 (4.71)	49.56 (4.33)	-	-	-
	STAI: State-Trait Score	93.26 (7.85)	91.99 (8.48)	92.89 (7.42)	-	-	-
COGNITIVE	QUIP	0.32 (0.60)	0.42 (0.85)	0.11 (0.33)	-	-	-
	MoCA	27.11 (2.63)	27.18 (2.15)	25.67 (2.55)	-	-	-
	HVLT-R free recall (T-score)	47.83 (10.86)	45.34 (11.55)	37.78 (9.34)	0.01	0.005	NS
	HVLT-R delayed recall (T-score)	47.47 (11.77)	46.84 (13.14)	39.00 (10.23)	NS	0.01	0.04
	HVLT-R retention (T-score)	49.32 (11.97)	49.27 (13.12)	45.00 (10.27)	-	-	-
	HVLT-R recognition	51.60 (13.10)	46.14 (12.32)	44.00 (15.55)	0.03	NS	NS
	Benton Judgement of Line	12.07 (2.67)	11.52 (3.03)	9.36 (2.50)	NS	0.002	0.02
	Letter-Number Sequencing	11.58 (3.00)	10.50 (2.66)	9.33 (4.03)	0.001	NS	NS
	Symbol Digit Modalities Test (T-score)	44.22 (9.24)	40.91 (8.57)	38.53 (10.20)	0.004	0.03	NS
	Semantic Fluency (T-score)	50.78 (10.56)	47.62 (10.73)	46.67 (10.98)	0.01	NS	NS

Table 1: Summary of baseline demographic, clinical and psychometric parameters. Results that survive Bonferonni corrected $P < 0.001$ are highlighted in bold. NS – Not significant on Mann-Whitney U test ($P < 0.05$). Dash indicates not significant on Krushkal-Wallis test (or chi squared test for categorical descriptors).

	Percentage of cohort with abnormal* baseline scores (%)		
Final cognitive diagnosis	PD-NON	PD-MCI	PDD
MoCA	20.3	18.6	30.0
HVLT-R free recall	21.7	29.4	66.7
HVLT-R delayed recall	20.8	27.5	55.6
HVLT-R retention	16.4	18.6	30.0
HVLT-R recognition discrimination	16.4	34.3	30.0
Benton Judgement of Line Orientation Total	7.72	15.69	44.4
Letter-Number Sequencing	4.3	7.8	30.0
Symbol Digit Modalities Test	22.7	23.5	55.6
Semantic Fluency	9.1	18.6	11.1

Table 2: Full cognitive profile data at baseline showing the percentage of individuals with abnormal scores.

Figure 1: Cohort selection from the PPMI data download. Confidence level refers to PPMI documented confidence in the assigned cognitive category. *Abbreviations: HC = Healthy Control; PD = Parkinson's disease; PRODROME = Prodromal cohort; GENPD = Genetic cohort with PD; GENUA = Genetic cohort unaffected; REGPD = Genetic Registry with PD; REGUA = Genetic registry unaffected; PD-NON = PD without cognitive impairment; PD-MCI = PD mild cognitive impairment; PDD = Parkinson's disease dementia.*

Figure 2: Longitudinal progression in motor and non-motor test scores over five years of follow-up. Note the average time to MCI and PDD was 1.8y and 3.45y respectively. Solid line = PD-NON, dash line = PD-MCI, dotted line = PDD. *Abbreviations: UPDRS – Unified Parkinsons Disease Rating Scale (MDS version); H&Y = Hoehn and Yahr; MoCA – Montreal Cognitive Assessment; SCOPA-AUT - Scales for Outcomes in Parkinson's disease – Autonomic; RBD – Rapid eye movement behavioural disorder questionnaire; STAI - State-Trait Anxiety Inventory; Depression - Geriatric Depression Scale; Epworth – Epworth Sleepiness Scale; QUIP - Questionnaire for Impulsive-Compulsive Disorder in Parkinson's Disease.*

Figure 3: Grey matter voxel-based morphometry. Significant regions of reduced grey matter density in PD subjects with preclinical cognitive impairment (MCI or dementia).

Figure 4: White matter voxel-based morphometry. Significant regions of reduced white matter density in PD subjects with preclinical cognitive impairment (MCI or dementia).

Figure 5: SVM results. A - Clinical optimisation, with 11 features providing the maximum average AUC. B - Receiver Operator Curve (ROC) for clinical predictions. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data. C - Imaging optimisation, with 269 features providing the maximum average AUC (see Figure 6 for the feature map). D - Receiver Operator Curve for imaging predictions. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data. E - Combined clinical-imaging grid search to define weighting parameters to produce weighted-average kernel for the analysis of clinical and neuroimaging data. F - Receiver Operator Curve for the combination of clinical and neuroimaging data. The dashed line indicates the ROC for the model (from training data) while the solid line indicates the ROC for the independent test data.

Figure 6: Spatial distribution of the final 269 imaging features contributing to the optimal SVM predictions. The ordering of the features has been preserved, corresponding to the largest (1) to smallest (269) SVM weights. Whilst these correspond with the voxel-based morphometry results, there is greater representation of the cerebellum and dorso-lateral temporal lobe, and the addition of regions such as the bilateral striatum, orbitofrontal cortex and entorhinal cortex.

SUPPLEMENTARY MATERIAL:

Supplementary Figure 1: Cortical thickness. Regions of reduced cortical thickness ($P < 0.001$) in PD subjects with preclinical cognitive impairment (MCI or dementia).

Supplementary Figure 2: Summary of Support Vector Machine training and feature selection pipeline. The full dataset, X , is divided into training, X_t , and independent, X_i , test datasets. At each point $X_{t'}$ in steps 1-3, a random number generator is seeded using the timestamp, and the training-test splits re-allocated. The following is performed on the training data, X_t : 1) Using an iterative process, ten-fold cross validation is used to obtain average absolute values for the feature weights, W . The feature with the highest weight, W_{max} , is identified and removed from X_t and the position (i.e. the relative importance) of this feature is stored. This process is repeated until no features remain. Step 1) is then repeated to construct a matrix of feature positions from which the average feature location vector, F , is calculated. 2) The identified features are then sequentially added to create matrix $X_{t'}$ according to their average position from lowest (i.e. most important) to highest (i.e. least important). At each iteration, ten-fold cross validation is used to calculate the average area under the curve which is stored in vector A . 3) The features that correspond to the maximum value in A (i.e. the best model) are then applied in a ten-fold cross validation of the test dataset, X_i , to calculate the overall model accuracy metrics for the training data. 4) The same features are applied to the independent test data, X_i , and the accuracy metrics extracted. Note at each point, $X_{t'}$, in steps 1-3, a random number is used to assign subjects within each 10-fold cross validation.

Supplementary Table 1: Percentage of individuals at baseline with abnormal test scores affecting single or multiple domains, using the definitions provided by the MDS-Task Force on PD-MCI.

Supplementary Table 2: SVM accuracy metrics, top is the average over 10-fold cross-validation, bottom is the independent training set. Standard deviations provided in brackets. *Significantly higher $P < 0.05$ compared to imaging alone. **Significantly higher compared to clinical alone.

Supplementary Material 1: Baseline clinical variables used for building the SVM model

Supplementary Figure 1: Cortical thickness. Regions of reduced cortical thickness ($P < 0.001$) in PD subjects with preclinical cognitive impairment (MCI or dementia).

Supplementary Figure 2: Summary of Support Vector Machine training and feature selection pipeline. The full dataset, X , is divided into training, X_t , and independent, X_i , test datasets. At each point X_t in steps 1-3, a random number generator is seeded using the timestamp, and the training-test splits re-allocated. The following is performed on the training data, X_t : 1) Using an iterative process, ten-fold cross validation is used to obtain average absolute values for the feature weights, W . The feature with the highest weight, W_{max} , is identified and removed from X_t and the position (i.e. the relative importance) of this feature is stored. This process is repeated until no features remain. Step 1) is then repeated to construct a matrix of feature positions from which the average feature location vector, F , is calculated. 2) The identified features are then sequentially added to create matrix $X_{t'}$ according to their average position from lowest (i.e. most important) to highest (i.e. least important). At each iteration, ten-fold cross validation is used to calculate the average area under the curve which is stored in vector A . 3) The features that correspond to the maximum value in A (i.e. the best model) are then applied in a ten-fold cross validation of the test dataset, X_i , to calculate the overall model accuracy metrics for the training data. 4) The same features are applied to the independent test data, X_i , and the accuracy metrics extracted. Note at each point, X_t , in steps 1-3, a random number is used to assign subjects within each 10-fold cross validation.

		BASELINE COGNITIVE CLASSIFICATION (Percentage of Cohort)		
		NORMAL	SINGLE DOMAIN DEFICIT	MULTIPLE DOMAIN DEFICIT
COHORT	PD-NON	59.91	14.00	26.09
	PD-MCI	39.22	16.67	44.11
	PDD	11.11	22.22	66.67

Supplementary Table 1: Percentage of individuals at baseline with abnormal test scores affecting single or multiple domains, using the definitions provided by the MDS-Task Force on PD-MCI.

10-FOLD CROSS VALIDATION AVERAGE			
	CLINICAL	IMAGING	JOINT
AUC	0.81 (0.11)	0.72 (0.13)	0.85 (0.09)*
Balanced Accuracy (%)	73 (0.09)*	69 (0.08)	74 (0.07)*
Sensitivity (%)	62 (13)*	46 (13)	62 (12)*
Specificity (%)	85 (12)	93 (6)**	87 (6)
PPV (%)	77 (16)	82 (14)	75 (9)
NPV (%)	76 (7)	71 (6)	76 (6)
INDEPENDENT TEST DATA			
	CLINICAL	IMAGING	JOINT
AUC	0.89	0.66	0.86
Balanced Accuracy (%)	76	61	78
Sensitivity (%)	63	37	63
Specificity (%)	88	86	92
PPV (%)	75	58	80
NPV (%)	82	71	82

Supplementary Table 2: SVM accuracy metrics, top is the average over 10-fold cross-validation, bottom is the independent training set. Standard deviations provided in brackets. *Significantly higher $P < 0.05$ compared to imaging alone. **Significantly higher compared to clinical alone.

SUPPLEMENTARY MATERIAL 1:

Baseline clinical variables used for building the SVM model

1. Baseline age
2. Gender
3. Years of education
4. UPDRS part 1 score
5. UPDRS part 2 score
6. UPDRS part 3 score
7. UPDRS total score
8. Hoehn and Yahr scale
9. Tremor score
10. PIGD score
11. TREM/PIGD ratio (normalised to a continuous scale as described in the methods)
12. SCOPA-AUT score
13. RBD questionnaire score
14. UPSIT score
15. Epworth score
16. Geriatric depression scale
17. State-trait anxiety inventory - State component
18. State-trait anxiety inventory - Trait component
19. State-trait anxiety inventory - Total score
20. QUIP total score
21. MoCA total score
22. Hopkins verbal learning test revised - Total recall (T-score)
23. Hopkins verbal learning test revised - Delayed recall (T-score)
24. Hopkins verbal learning test revised - Retention (T-score)
25. Hopkins verbal learning test revised - Recognition discrimination index (T-score)
26. Benton judgment of line orientation (age and education scaled score)
27. Letter number sequence (scaled score)
28. Symbol digit modalities test (T-score)
29. Semantic fluency (scaled score)