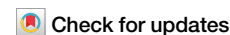


<https://doi.org/10.1038/s41531-025-01232-4>

# Brain age gap as predictor of disease progression in Parkinson's disease



Tom Hähnel<sup>1,2</sup>✉, Shammi More<sup>2</sup>, Felix Hoffstaedter<sup>3,4</sup>, Kaustubh R. Patil<sup>3,4</sup>, Holger Fröhlich<sup>2,5,6,8</sup> & Björn H. Falkenburger<sup>1,7,8</sup>

Parkinson's disease (PD) exhibits high heterogeneity in disease progression, complicating management and increasing required sample sizes for clinical trials. This study evaluates Brain Age Gap (BAG)—the difference between brain age and chronological age—for predicting disease progression in PD. Structural MRI-derived gray matter volumes were analyzed for 451 early-stage PD patients and 172 healthy controls from the PPMI cohort. PD patients had a baseline BAG of 1.1 years, with fast-progressing patients exhibiting a BAG of 3.0 years, whereas slow-progressing patients resembled the BAG of healthy controls. Higher BAG was associated with more severe baseline symptoms, faster cognitive decline in several domains, increased hazard of developing mild cognitive impairment, and faster progression of dopaminergic neuron loss in longitudinal DaTSCANs. BAG-based patient stratification could reduce sample sizes of randomized clinical trials by 23–58%. These findings suggest BAG as a prognostic biomarker of disease progression, which may accelerate the development of disease-modifying treatments.

Parkinson's disease (PD) exhibits high heterogeneity in disease progression, posing challenges for clinical management and limiting the statistical power of clinical trials that investigate new, potentially disease-modifying drugs<sup>1</sup>. PD heterogeneity has been explained by distinct progression subtypes, with individuals within each subtype showing more homogeneous progression patterns<sup>2–6</sup>.

Recently, a randomized clinical trial (RCT) investigating the effect of prasinezumab, a monoclonal antibody developed as a disease-modifying drug to slow PD progression, failed to reach its primary outcome<sup>7</sup>. A secondary analysis revealed that the RCT could have demonstrated a significant treatment effect if patient inclusion had been restricted to a rapid-progressing PD subtype, which would have resulted in a more homogeneous cohort and, on average, faster disease progression<sup>8</sup>. Applied to multiple PD cohorts, similar stratification strategies have consistently improved statistical power<sup>3</sup>. However, these approaches required longitudinal follow-up prior to study enrollment to generate reliable subtype predictions, limiting their feasibility.

Brain aging follows a temporal trajectory of structural changes that are reflected by brain atrophy patterns. Brain age can be estimated from structural magnetic resonance imaging (MRI) using gray matter volumes

(GMV) or white matter volumes (WMV), but GMV has the most significant effect in brain age estimation<sup>9</sup>. Brain age gap (BAG) is the difference between the brain age estimated from structural MRI and the chronological age. BAG has been suggested as a biomarker of overall brain integrity and is increased in neurodegenerative disorders<sup>9</sup>. A positive BAG indicates that the estimated brain age exceeds the chronological age, i.e., that the brain structure resembles that of an older individual. Conversely, a negative BAG indicates that the brain structure resembles that of a younger individual. In people with Alzheimer's disease (AD), BAG is increased by 3 to 8 years<sup>10</sup>. Furthermore, BAG in AD predicts disease progression and conversion from mild cognitive impairment to dementia<sup>11–13</sup>. Studies in PD have reported smaller BAG values, ranging from 0.7 years to 4.4 years<sup>14–24</sup>. The potential for predicting disease progression and patient stratification in RCTs using BAG has not been explored.

We recently identified two PD progression subtypes using a data-driven approach and longitudinal data: a fast-progressing and a slow-progressing subtype<sup>3</sup>. The fast-progressing subtype exhibited a more rapid progression in several cognitive domains, non-motor symptoms, specific gait patterns, and axial symptoms. Moreover, it was associated with worse response to dopaminergic treatment and increased mortality. In contrast,

<sup>1</sup>Department of Neurology, Medical Faculty and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany. <sup>2</sup>Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany. <sup>3</sup>Institute of Neuroscience and Medicine (INM-7), Research Centre Jülich, Jülich, Germany. <sup>4</sup>Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

<sup>5</sup>Bonn-Aachen International Center for IT, University of Bonn, Bonn, Germany. <sup>6</sup>Institute for Digital Medicine, University Medical Center Bonn, Bonn, Germany. <sup>7</sup>German Center for Neurodegenerative Diseases (DZNE), Dresden, Germany. <sup>8</sup>These authors contributed equally: Holger

Fröhlich, Björn H. Falkenburger. ✉e-mail: [tom.haehnel@ukdd.de](mailto:tom.haehnel@ukdd.de)



the slow-progressing subtype exhibited mild disease progression across most symptom domains and a more asymmetrical DaTSCAN uptake ratio. Notably, the existence of both subtypes was validated in multiple independent cohorts. However, longitudinal data was required for subtype predictions.

The objective of this exploratory study is to investigate the potential relationship between BAG and disease progression in PD, encompassing potential associations of BAG with PD progression subtypes, the prediction of progression across multiple symptom domains, and the use of BAG for prognostic patient stratification.

Therefore, we utilized the two PD progression subtypes that were identified within the same dataset, and first compared the atrophy patterns between the fast-progressing and slow-progressing PD subtypes directly using GMV derived from structural T1-weighted (T1w) MRI scans. To estimate BAG in our cohort, we applied multiple brain age estimation workflows from the literature and selected the one that provided the highest accuracy for our dataset. BAG estimates were then compared between PD progression subtypes and correlated with clinical progression rates, features extracted from DaTSCANs, and biomarkers of Alzheimer’s co-pathology. Finally, we analyzed whether baseline BAG can be used to stratify patients and optimize clinical trial designs, leading to reduced sample sizes. To minimize the risk of bias due to people with PD being at different disease

stages, we corrected all analyses using a latent disease time approach published before<sup>3,25</sup>.

Results

Demographics and clinical information

Overall, 451 people with PD and 172 healthy controls (HC) from the Parkinson’s Progression Markers Initiative (PPMI) cohort were included in our analysis (Table 1). HC and people with PD exhibited similar age and sex distributions. People with PD were at early disease stages as indicated by Hoehn & Yahr (H&Y) stages I and II observed in 443 of 451 people with PD, low MDS-Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) III scores (median: 20 points), and high Montreal Cognitive Assessment (MoCA) scores (median: 27 points). A minority of HC (39.0%, *n* = 67) and people with PD (37.7%, *n* = 170) had longitudinal MRI data available, whereas most participants had only a single baseline MRI. PD progression subtype assignments for people with PD were available from a previous analysis<sup>3</sup>. Most people with PD were classified as slow-progressing (65.4%, *n* = 295) and a smaller fraction as fast-progressing (14.6%, *n* = 66) subtype. Because subtype identification required a longitudinal follow-up of specific clinical scores, no subtype information was available for a minority of people with PD (20.0%, *n* = 90). Note that PPMI enrolled participants in two recruitment phases: the first in 2010 and the second in 2020, so participants from the second phase have shorter follow-up periods. This is reflected by the high interquartile range of follow-up durations and visits, as indicated in Table 1.

Table 1 | Demographics and baseline clinical characteristics

|                               | Healthy controls ( <i>n</i> = 172) | Parkinson's disease ( <i>n</i> = 451) |
|-------------------------------|------------------------------------|---------------------------------------|
| Age (years)                   | 61.4 [54.9–68.0]                   | 62.8 [55.3–69.1]                      |
| Sex                           |                                    |                                       |
| Male                          | 64.5% [111]                        | 62.5% [282]                           |
| Female                        | 35.5% [61]                         | 37.5% [169]                           |
| H&Y                           |                                    |                                       |
| H&Y 1                         |                                    | 38.6% [174]                           |
| H&Y 2                         |                                    | 59.6% [269]                           |
| H&Y 3                         |                                    | 1.8% [8]                              |
| H&Y 4                         |                                    | 0.0% [0]                              |
| H&Y 5                         |                                    | 0.0% [0]                              |
| Disease duration (years)      |                                    | 0.4 [0.2–1.0]                         |
| Number of visits              | 14 [5–19]                          | 6 [4–14]                              |
| Follow-up (years)             | 5.0 [1.5–11.0]                     | 2.0 [0.9–7.0]                         |
| MDS-UPDRS I                   |                                    | 5 [3–8]                               |
| MDS-UPDRS II                  |                                    | 5 [3–8]                               |
| MDS-UPDRS III                 |                                    | 20 [15–26]                            |
| MoCA                          |                                    | 27 [26–29]                            |
| SCOPA                         |                                    | 12 [7–20]                             |
| Number of MRIs per individual |                                    |                                       |
| Baseline only                 | 61.0% [105]                        | 62.3% [281]                           |
| 2 MRI                         | 27.9% [48]                         | 4.9% [22]                             |
| >2 MRI                        | 11.0% [19]                         | 32.8% [148]                           |
| Subtype                       |                                    |                                       |
| Fast                          |                                    | 14.6% [66]                            |
| Slow                          |                                    | 65.4% [295]                           |
| Latent disease time (years)   |                                    | 0.2 [–0.9 to 1.2]                     |

Demographic and clinical characteristics of healthy controls and people with PD. Values are reported as median [interquartile range] or as relative frequency [absolute frequency]. Progression subtype assignments were available for 361 of 451 people with PD.  
H&Y Hoehn & Yahr, MDS-UPDRS MDS-Unified Parkinson’s Disease Rating Scale, MoCA Montreal Cognitive Assessment, SCOPA Scales for Outcomes in Parkinson’s disease.

Correcting for different disease stages

A previously published latent disease time approach was used to correct subsequent analyses for differences in disease stages between people with PD<sup>3,25</sup>. Therefore, we aligned people with PD on a latent disease timescale by comparing the individual progression of multiple motor and non-motor symptoms with the cohort’s average, yielding an estimated latent disease time for each person with PD that reflects the actual disease stage. Thereby, a latent disease time of 0 corresponds to the cohorts’ average disease stage at the time of diagnosis. Accordingly, positive or negative latent disease time values correspond to a disease stage that is later or earlier than the average disease stage at the time of diagnosis.

Applying this method, we estimated a mean latent disease time of 0.2 years at baseline visit, with an interquartile range between –0.9 and 1.2 years. (Table 1). Estimated latent disease times exhibited a stronger correlation with Hoehn & Yahr (H&Y) stages ( $\tau$  = 0.23, *p* < 0.0001) compared to the correlation of time since diagnosis with H&Y ( $\tau$  = 0.12, *p* = 0.0027), and this difference in correlations was significant (*p* = 0.004), confirming the validity of our latent disease time approach.

Voxelwise and parcelwise gray matter volume comparison of PD progression subtypes

First, we assessed whether a direct comparison of gray matter volumes (GMV) derived from structural MRI images at baseline can differentiate between fast-progressing and slow-progressing people with PD. GMV in voxelwise space resulted in a visual pattern with a tendency towards more pronounced cortical atrophy for the fast-progressing subtype and more pronounced subcortical atrophy for the slow-progressing subtype (Fig. 1). However, no voxels were significantly different between the two progression subtypes after correction for multiple testing.

In a subsequent approach, we applied an additional 4 mm resampling and 4 mm smoothing to voxelwise gray matter images to reduce noise and the number of statistical tests. This approach produced a similar visual pattern with no voxels reaching statistical significance between the two PD progression subtypes after correcting for multiple testing (Fig. S1). We then performed comparisons between PD progression subtypes using GMV in parcelwise space to further reduce the number of statistical tests. Again, no region exhibited significant differences after correcting for multiple testing, although several cortical regions showed a visual pattern towards more atrophy in the fast-progressing subtype (Figs. S2–5).

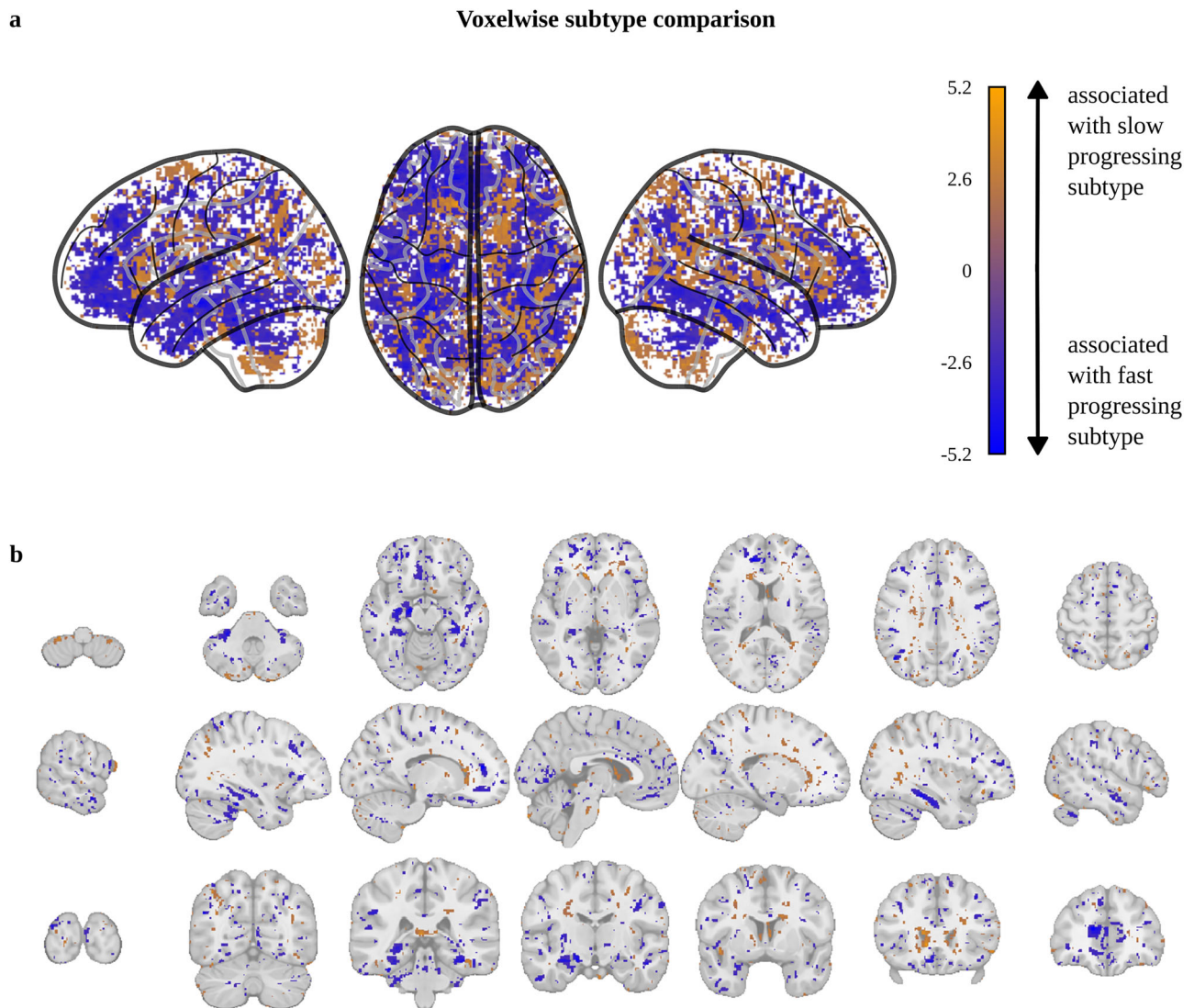


**Fig. 1 | Voxelwise gray matter volume comparison of PD progression subtypes.** Association of atrophy with the fast-progressing (blue) and slow-progressing (orange) subtypes displayed as **a** glass brain visualization and **b** sagittal, coronal, and transverse planes. Voxels with significant PD progression subtype differences before correction for multiple testing are colored depending on their t-value (blue:

associated with fast-progressing subtype, orange: associated with slow-progressing subtype). Note that no voxels remained significant after Benjamini–Hochberg correction for multiple testing. No additional smoothing or resampling was performed for this analysis.

Overall, direct GMV comparison of structural MRI revealed visual patterns associated with PD progression subtypes but did not yield statistically significant differences.

### Brain age model selection and validation in HC

Motivated by these visual GMV patterns, we investigated whether brain age estimations using GMV as features at baseline could capture these patterns and provide information about clinical disease progression in PD.

Although several brain age models have been published, their accuracy varies depending on the datasets on which they were trained and the dataset on which they are applied<sup>26</sup>. Moreover, brain age is typically overestimated in elderly individuals and underestimated in younger individuals. To remove this age-related bias, several bias correction methods have been proposed<sup>27</sup>.

Therefore, we systematically evaluated three brain age models from a recent systematic comparison in combination with two different bias correction methods in our dataset. Once the brain age was estimated by applying the three brain age models, bias correction was

performed in HC, where a BAG of zero is expected, to prevent overfitting (Table S1). The brain age model using lasso regression with no smoothing and 4-voxel resampling (S0\_R4 + LR), combined with bias correction via BAG and chronological age (Beheshti method<sup>27</sup>) achieved the best accuracy. This workflow yielded a mean absolute error of 4.51 years (Fig. S6), consistent with the accuracy of brain age models in the literature<sup>26,28</sup>. Therefore, the S0\_R4 + LR brain age model was used for all further analysis, with the Beheshti bias correction method being retrained on all HC.

Applying this workflow to the complete HC group, brain age correlated strongly with chronological age ( $\rho = 0.90$ ,  $p < 0.0001$ ), with a mean absolute error of 4.34 years (Fig. 2A). As expected, there was no significant difference between mean brain age and mean chronological age in HC (BAG = 0.00 years,  $p = 1$ , Fig. 2B) and BAG was not correlated with chronological age ( $\rho = 0.00$ ,  $p = 1$ , Fig. S6), indicating a successful bias correction. The concordance correlation coefficient for BAG in participants with short-term follow-up MRI was 0.86, demonstrating good retest reliability of our brain age estimations. Longitudinal consistency of brain age was confirmed by a significant correlation between the increase in brain age between two

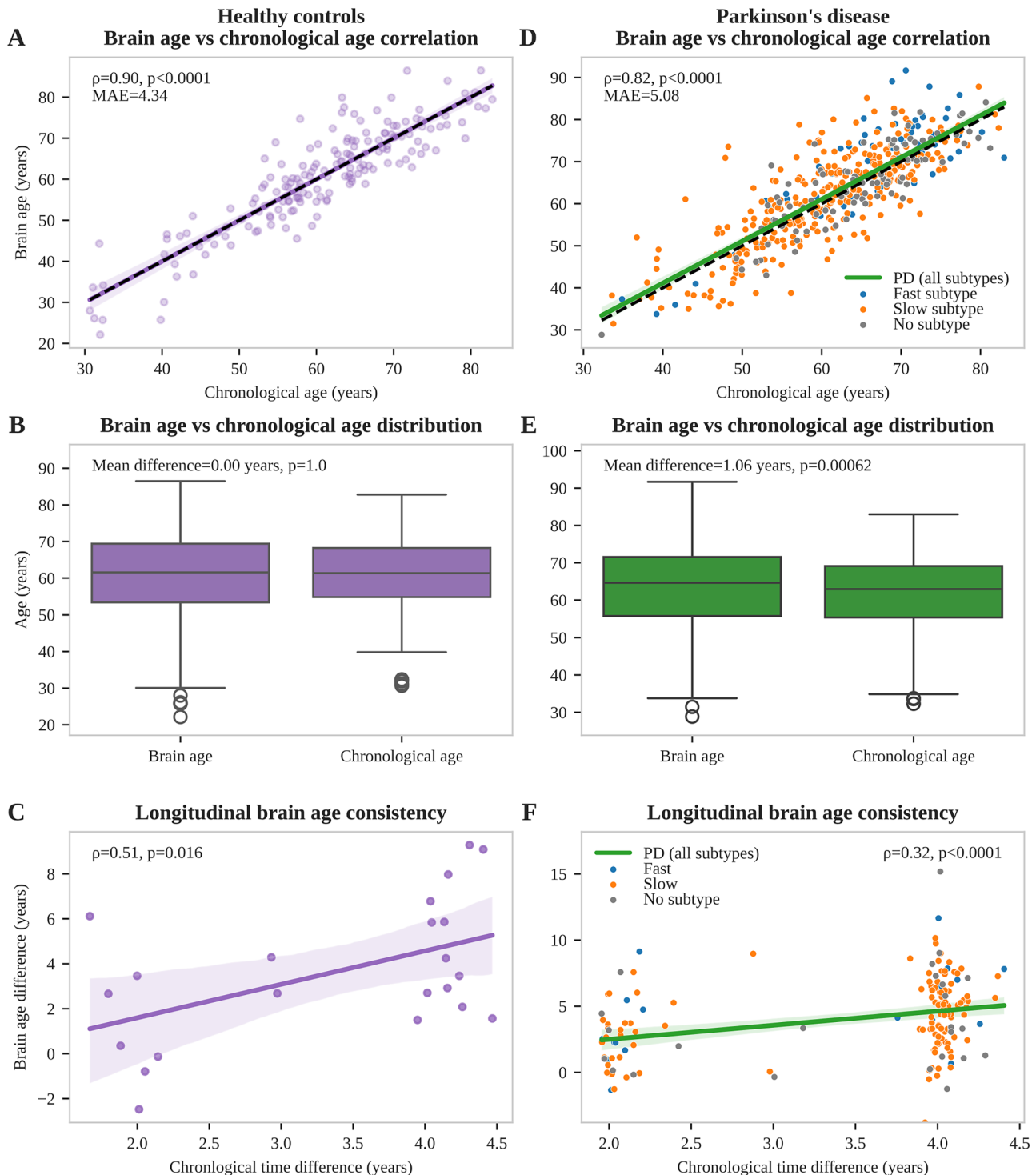


**Fig. 2 | Brain age calculation in healthy controls and Parkinson's disease.** Brain age estimations for healthy controls (purple) are shown in the left column, brain age estimations for the PD cohort (green) in the right column. Individual people with PD are colored depending on their progression subtype (blue: fast-progressing, orange: slow-progressing, gray: no subtype available). **A, D** Correlations of brain age with chronological age. Mean absolute error, Pearson correlation coefficient, and corresponding  $p$ -value are shown. **B, E** Box plots comparing distributions of chronological age and brain age.  $P$ -values of corresponding paired  $t$ -tests are shown

together with the mean age differences. **C, F** Longitudinal consistency analyses showing correlations of brain age difference and chronological age difference between subsequent MRIs. Pearson correlation coefficient and corresponding  $p$ -value are reported. All brain age estimations were based on a bias correction trained on the whole HC group to allow comparison between HC and PD. Cross-validation results for HC are presented in Fig. S6 and were used for model selection. PD Parkinson's disease, MAE mean absolute error.

subsequent MRI scans with the actual time elapsed between them ( $\rho = 0.51$ ,  $p = 0.016$ , Fig. 2C).

Overall, the selected brain age estimation workflow demonstrated state-of-the-art performance compared to recent literature<sup>26,28</sup>.

#### Brain age estimation in PD

When applied to people with PD, brain age estimations correlated strongly with chronological age ( $\rho = 0.82$ ,  $p < 0.0001$ ), with a mean absolute error of 5.08 years (Fig. 2D), slightly higher than that observed in the HC group.

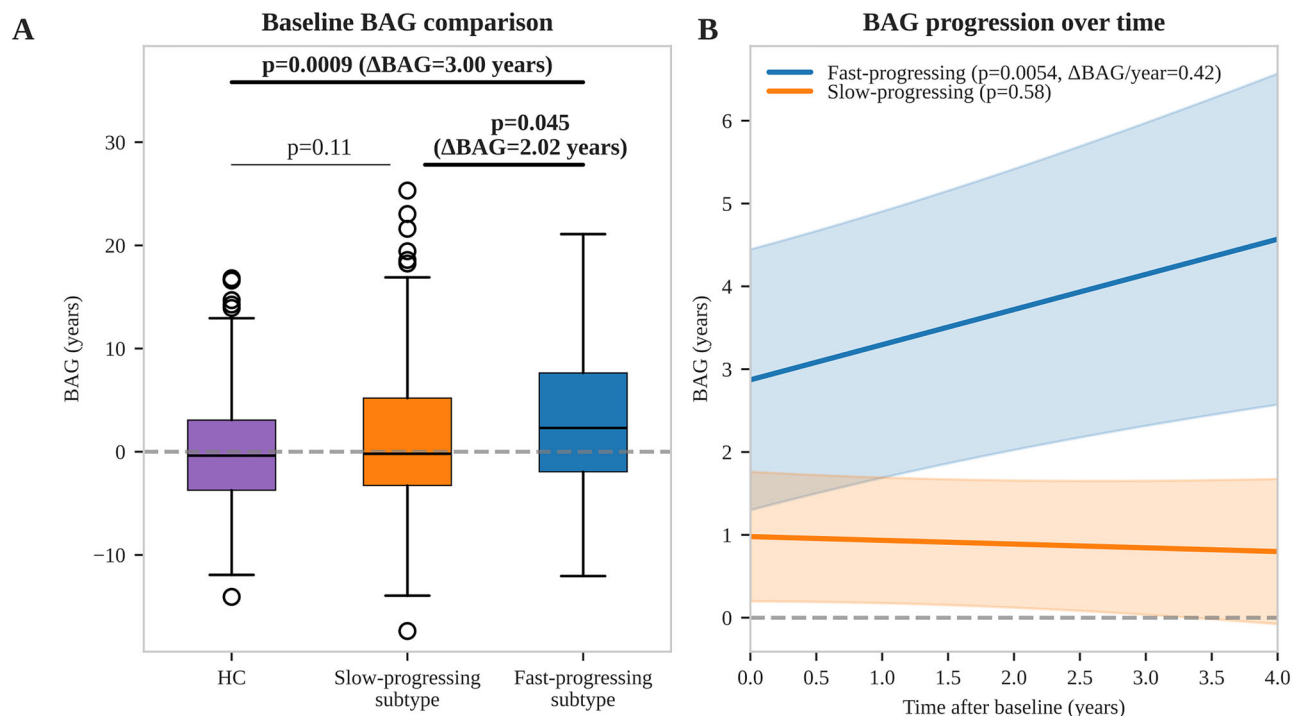


**Fig. 3 | Brain age gap differences between PD progression subtypes and HC.**

**A** BAG comparison between PD progression subtypes and HC at baseline. *P*-values for group differences, corrected for age and sex, are reported alongside the BAG difference for significant differences. The BAG difference between the slow and fast-progressing subtypes was corrected for latent disease time as well. **B** Longitudinal

BAG increase estimated using a linear mixed-effects model for each PD progression subtype. *P*-values for BAG progression over time are shown together with the mean BAG increase per year for significant results. Shaded areas indicate 95% confidence intervals. The horizontal dashed lines correspond to BAG = 0. BAG brain age gap. HC healthy controls.

Estimated brain ages were significantly higher than chronological ages ( $p = 0.00062$ , Fig. 2E), resulting in a mean BAG of 1.06 years. As in HC, BAG did not correlate with chronological age ( $\rho = -0.01$ ,  $p = 0.91$ ), indicating successful bias correction also in PD. The concordance correlation coefficient was 0.89, indicating good retest reliability of BAG estimations. Longitudinal consistency of brain age estimations was confirmed by a significant correlation between the increase in brain age between two subsequent MRI scans with the actual time elapsed between them ( $\rho = 0.32$ ,  $p < 0.0001$ , Fig. 2F).

Overall, the selected brain age estimation workflow provided reliable BAG estimations, with a mean BAG of 1.06 years in PD being consistent with literature reports<sup>14–20</sup>.

### Brain age gap in fast-progressing and slow-progressing PD

We then assessed whether BAG serves as a marker of disease progression by comparing BAG between PD progression subtypes.

The fast-progressing PD subtype exhibited a significantly higher BAG at baseline than HC ( $\Delta\text{BAG} = 3.00$  years,  $p = 0.0009$ , Fig. 3A) and the slow-progressing PD subtype ( $\Delta\text{BAG} = 2.02$  years,  $p = 0.045$ ). No significant difference was observed between the slow-progressing and fast-progressing PD subtypes ( $p = 0.11$ ).

Longitudinal analysis of BAG trajectories revealed a significant annual increase of BAG in the fast-progressing PD subtype of 0.42 years per year ( $p = 0.0054$ , Fig. 3B), meaning a 42% faster brain aging than would be expected in HC. Conversely, the slow-progressing PD subtype showed no significant change in BAG over time ( $p = 0.58$ ).

### Brain age gap correlation with clinical scores

We now sought to distinguish for which clinical domains BAG represents a marker of disease severity and for which clinical domains a predictor of disease progression. Therefore, we evaluated correlations of baseline BAG with (I) baseline outcomes (reflecting disease severity) and (II) longitudinal progression of outcomes (reflecting disease progression).

Baseline BAG was correlated with cognitive scores, motor scores, and scores capturing overall disease severity (Fig. 4). Interestingly, BAG correlated with motor scores and overall disease severity scores at baseline, while there was no association of BAG with progression of these scores. Conversely, BAG showed more significant correlations with cognitive decline over time than with baseline cognitive scores.

No correlation of baseline BAG was evident for clinical scores reporting autonomic dysfunction, non-motor symptoms, anxiety, depression, sleep disturbances, and motor impairment in daily life (Table S2).

Analyzing baseline BAG correlations with biomarkers, we found that higher BAG was correlated with faster degeneration of dopaminergic axon terminals in the caudate nucleus and overall striatum, as observed in longitudinal DaTSCANs. Contrarily, BAG was not correlated with DaTSCAN uptake ratios in these regions at baseline (Table S3). Furthermore, no significant correlations of BAG were found regarding baseline values and longitudinal changes of cerebrospinal fluid markers of Alzheimer's disease pathology and neurofilament light chain (Table S4).

Using a Cox proportional hazards model as a complementary time-to-event approach, we identified higher BAG as a significant predictor for overall cognitive decline and progression to mild cognitive impairment (Fig. 5). In detail, each year increase in baseline BAG was associated with 2% increased hazard for cognitive decline ( $p = 0.04$ ) and 4% increased hazard for mild cognitive impairment ( $p = 0.032$ ). Visual and statistical checks indicated no violation of the proportional hazards assumption (Fig. S7).

Overall, these findings suggest that BAG is predictive of overall cognitive decline, decline in several cognitive domains, and faster neurodegeneration observed in longitudinal DaTSCANs.

### Clinical trial optimization using the brain age gap

Motivated by the finding that BAG correlates with cognitive decline, we assessed how baseline BAG can be used to optimize clinical trial designs by patient stratification.

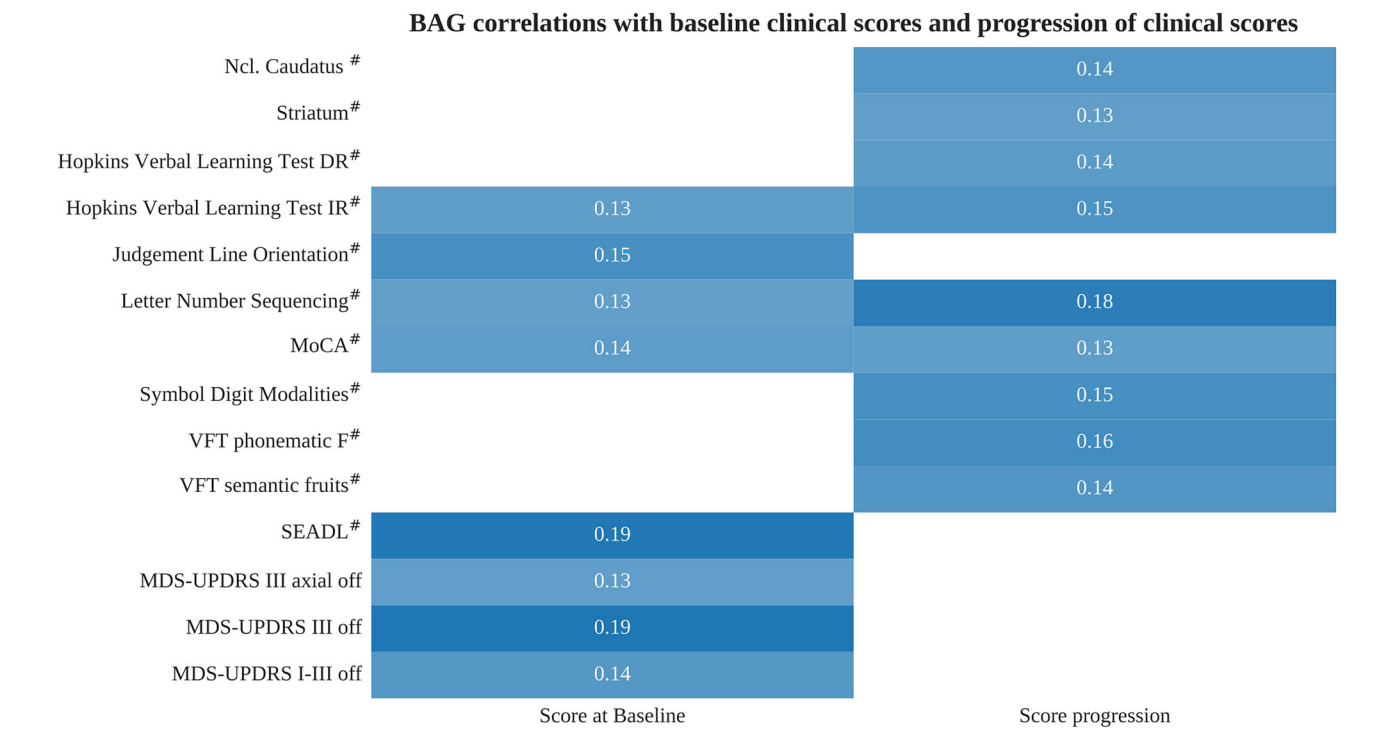


**Fig. 4 | Brain age gap correlations with clinical scores.** Correlations of BAG with baseline outcomes (left column) and progression of outcomes (right column) are depicted with the color indicating the strength of the partial correlation (light blue: low correlation, dark blue: high correlation). Partial correlation coefficients are shown and were corrected for age. Correlations with baseline outcomes were corrected for latent disease time as well. Only correlations with statistical significance after correction for multiple testing using the Benjamini–Hochberg method are shown. Outcomes marked with # were inverted such that positive correlation coefficients always indicate higher BAG being associated with greater baseline severity or faster progression. A complete list, including also outcomes without significant correlations, is provided in Tables S2–S4. BAG brain age gap, DR delayed recall, IR intermediate recall, MDS-UPDRS MDS-Unified Parkinson’s Disease Rating Scale, MoCA Montreal Cognitive Assessment, SEADL Schwab and England Activities of Daily Living scale, VFT verbal fluency task.

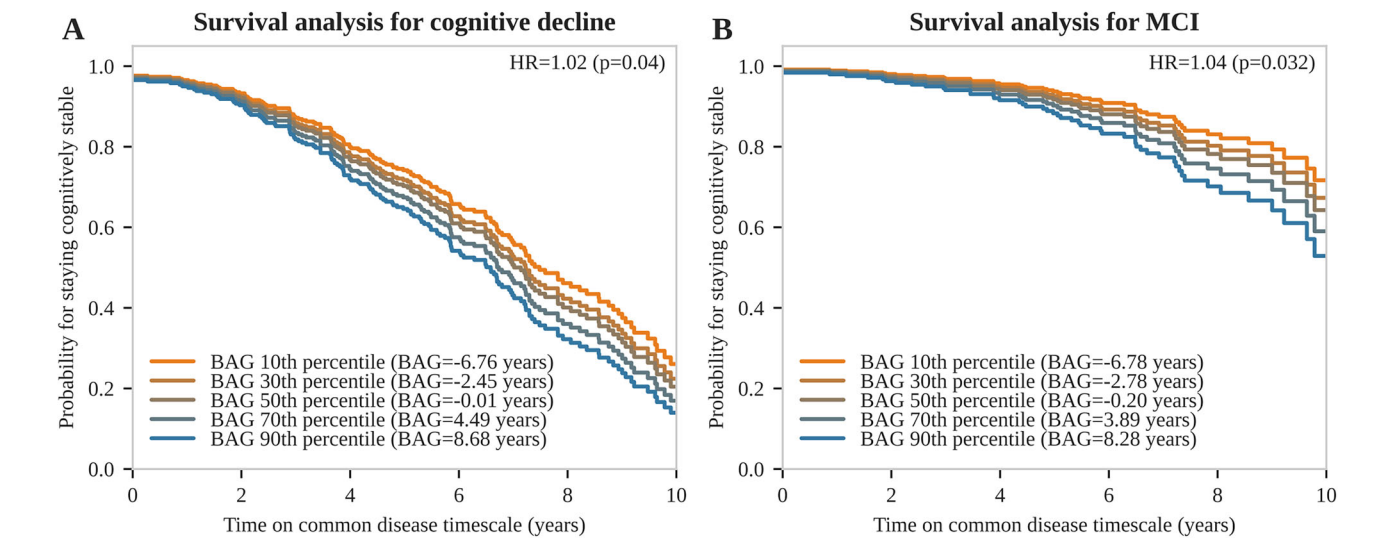


**Fig. 5 | Event-free survival analysis for cognitive decline and mild cognitive impairment.** Cox regression plots for **A** cognitive decline and **B** mild cognitive impairment, depending on BAG percentiles at baseline and including age and sex as additional covariates. Hazard ratios per BAG increase of 1 year are shown with corresponding *p*-values. Cognitive decline was defined as a progression towards the next level on an ordinal scale containing the levels normal cognition, subjective cognitive impairment, mild cognitive impairment, and dementia. BAG brain age gap, HR hazard ratio, MCI mild cognitive impairment.

We simulated a randomized clinical trial (RCT) assessing the effect of a disease-modifying drug on cognitive decline in people with PD. We assumed a 30% treatment-induced reduction of cognitive decline, a two-year follow-up period, and cognitive decline being measured using a cognitive compound score. Without stratification, i.e., including all people with PD, 1218 people with PD were required to achieve 80% statistical power (Fig. 6). Applying a BAG-based stratification—including only people with PD with a BAG above the 50th, 70th, and 90th percentile—reduced the required sample sizes by 23% (*n* = 930), 28% (*n* = 874), and 58% (*n* = 507), respectively.

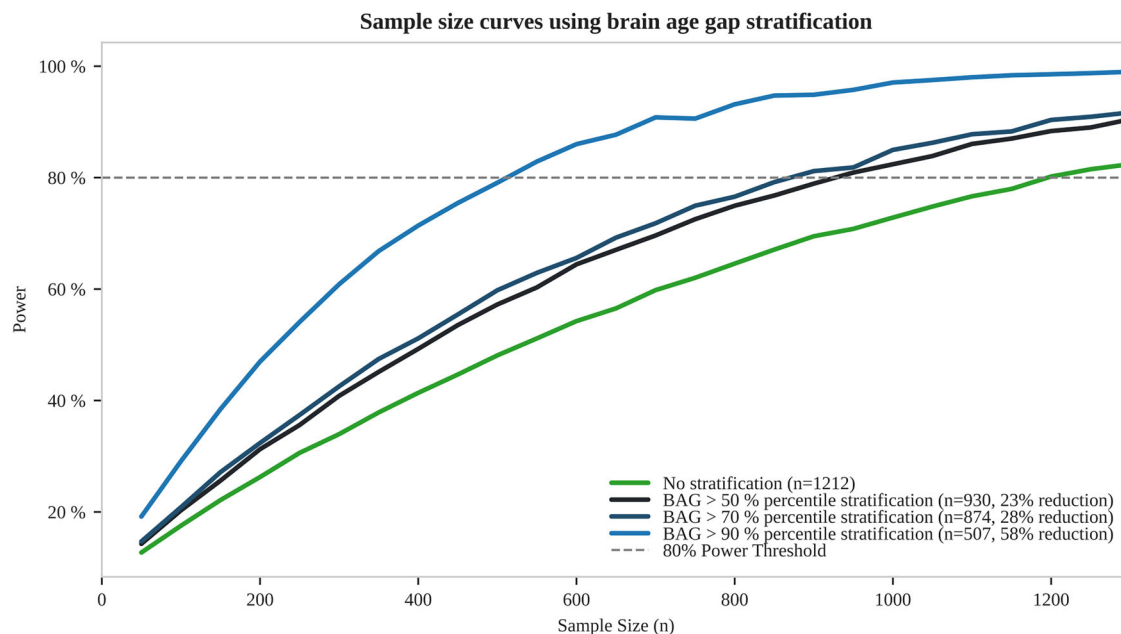


**Fig. 6 | Sample sizes required for simulated clinical trials using brain age stratification.** Estimated power and sample size curves for a clinical trial depending on the BAG threshold used as an inclusion criterion: no stratification applied to people with PD (green), inclusion of people with PD with a baseline BAG greater than the 50th percentile (black), 70th percentile (dark blue), and 90th percentile (light blue).

A treatment effect of 30% on the progression rate of the cognitive composite score, a two-year observation period, and a significance level  $\alpha = 0.1$  were assumed. Sample sizes and relative reductions compared to the unstratified scenario are reported for 80% power (dashed line). BAG brain age gap.

Repeating this analysis with single cognitive scores instead of the cognitive composite score, sample size reductions were found for all cognitive outcomes, with median reductions being even higher: 44% (50th BAG percentile), 54% (70th BAG percentile), and 70% (90th BAG percentile), indicating that our approach is not limited to a specific cognitive score. However, absolute sample sizes were significantly higher for single cognitive scores compared to the cognitive compound score (Table S5).

## Discussion

Our exploratory study demonstrates that GMV from structural MRI scans conveys information about disease progression in PD. Traditional voxelwise or parcelwise GMV analysis using linear models and adjusting for important covariates cannot detect robust atrophy patterns. However, BAG models provide a new, machine learning-based approach for capturing disease-related patterns in MRI scans. While a mean BAG of about 1 year in PD aligns with recent literature, our work provides additional insight that increased BAG is driven by a small fraction of fast-progressing people with PD, whereas BAG in most people with PD resembles that of HC. Additionally, BAG increased by 0.42 years per year in fast-progressing people with PD, leading to even larger differences between progression subtypes over time. Higher BAG was associated with more severe symptoms at baseline and predicted faster cognitive decline and faster degeneration of dopaminergic axon terminals from the substantia nigra. PD patients can be stratified based on their baseline BAG to enrich clinical trials with fast-cognitive-declining patients. Thereby, sample sizes of clinical trials could be reduced by 23–58% in our simulations, depending on the fraction of patients that can still be included.

Our analysis showed a mean BAG difference of 1.06 years in PD. This value is within the lower range of BAG values reported for PD in the literature (0.7–4.4 years)<sup>14–24,29</sup> and notably lower than BAG in Alzheimer's disease (about 3–8 years)<sup>10</sup>. The relatively low BAG value compared to other PD publications may be due to PPMI inclusion criteria, which restricted participants to the early disease stage with a maximum of 2 years since diagnosis. Our results and those of other authors indicate that BAG increases with disease duration for at least some patients, potentially leading to higher BAG values in advanced disease stages<sup>19,21,23</sup>.

Furthermore, the use of mean absolute error as a model selection criterion has been the subject of controversy, given that complex machine learning methods used in brain age models are more likely to overfit the training data<sup>26,30</sup>. Thus, these models may not generalize well to neurodegenerative disorders, contrary to providing lower BAG in the training cohort. As we (1) performed model selection on a dataset that was not used to train the brain age model and (2) selected the bias correction method using cross-validation, an overfit is relatively unlikely. Additionally, BAG in people with PD estimated by the other brain age workflows exhibited only minimal differences compared to the approach that provided the minimal mean absolute error (Table S6).

Our findings are in line with previous publications of BAG being associated with scores reporting motor symptoms<sup>18,21,23,29,31</sup>, cognitive impairment<sup>15,18,21–23,29,31</sup>, and overall disease severity<sup>19,24</sup>. However, we did not observe correlations between BAG and cerebrospinal fluid markers of Alzheimer's dementia, neurofilament light chain, or uptake ratios extracted from DaTSCAN analysis at baseline, indicating that BAG provides an independent biomarker. In line with this, Teipel et al. found only minimal BAG differences in LRRK2 and GBA PD subtypes, indicating that BAG values may also be independent of genetic causes of PD<sup>18</sup>.

Literature on BAG as a predictor of disease progression in PD is sparse. It has been shown that BAG is higher in people with PD who converted to MCI than in people with PD who remained cognitively unimpaired<sup>22</sup>. While this could have reflected the association of higher BAG being found in advanced disease stages, our study demonstrates that BAG is truly predictive of a faster rate of cognitive decline, with the hazard of MCI being increased by 4% for each BAG year. While significant, the observed effect size of BAG on cognitive decline is relatively small, indicating that other factors impacting cognitive decline may be clinically more relevant. Further research should investigate the interactions of multiple predictors, including genetics, lifestyle factors, non-motor symptoms, and Alzheimer's co-pathology.

Higher BAG values at baseline have been reported as predictors of faster decline in visuo-executive functions (Symbol Digit Modalities Test), verbal learning and memory (Hopkins Verbal Learning Test), and overall cognition (Montreal Cognitive Assessment, MoCA)<sup>18</sup>. While our analysis

confirms these findings, we provide additional evidence that high BAG values also predict faster cognitive decline in executive function/language (semantic and phonematic Verbal Fluency Tasks), and attention/concentration (Letter Number Sequencing test) domains. Thus, BAG should be interpreted as a predictor of overall cognitive decline in PD rather than being restricted to a specific cognitive domain.

Importantly, there was no association between A $\beta$ 42 or phospho-tau-levels and BAG, indicating that BAG reports cognitive decline independent of the co-occurrence of Alzheimer pathology. Moreover, BAG was not correlated with neurofilament light chain, indicating that brain atrophy patterns leading to increased BAG are not primarily caused by an overall faster neurodegeneration. Conversely, BAG at baseline was associated with faster decline of uptake ratios observed in longitudinal DaTSCAN measurements.

While higher BAG was correlated with faster progression of several cognitive scores, we found no correlation with progression of non-cognitive domains, including motor symptoms and other non-motor symptoms. Therefore, BAG should be considered one of several factors contributing to the fast-progressing subtype, primarily explaining its cognitive characteristics.

We compared BAG between two PD progression subtypes that were validated in multiple external cohorts before<sup>3</sup>. BAG was significantly elevated only in the fast-progressing subtype, which comprises a small fraction of people with PD (18%) and exhibited 42% faster brain aging than HC. Surprisingly, in the slow-progressing subtype, which represents the majority (82%) of people with PD, BAG was not elevated and did not increase over time. These findings support the notion that PD should be viewed as distinct subtypes with potentially different biological origins, rather than a continuum, and that BAG serves as a biomarker of the fast-progressing PD subtype.

Our voxelwise GMV comparison between PD progression subtypes indicated a visually pronounced cortical atrophy in the fast-progressing subtype and subcortical atrophy in the slow-progressing subtype. These visual atrophy patterns are in accordance with the higher cognitive impairment and faster cognitive decline observed in the fast-progressing PD subtype<sup>3</sup>. A potential biological explanation behind these visual patterns may be provided by the  $\alpha$ -synuclein Origin and Connectome Model (SOC Model), which expects that the body-first subtype exhibits already more severe and bilateral cortical  $\alpha$ -synuclein pathology at the time of diagnosis than the brain-first subtype<sup>2,32</sup>. Notably, the body-first subtype has been discussed as a potential biological explanation behind the fast-progressing subtype, as it exhibits similar clinical characteristics<sup>3</sup>. However, other hypotheses, like distinct  $\alpha$ -synuclein strain types, have also been discussed as biological drivers of PD progression subtypes<sup>33</sup>.

The Federal Drug Agency (FDA) acknowledges prognostic enrichment strategies as they may increase the efficiency of drug development and support precision medicine (see “Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products - Guidance for Industry”, [www.fda.gov/media/121320/download](http://www.fda.gov/media/121320/download), published in 2019). By enriching cohorts with patients who are likely to experience the outcome of interest (i.e., clinical progression), this strategy allows for a reduction in the sample size required for RCTs. Here, we showed that BAG-based prognostic enrichment lowers the required sample size in RCTs by 23–58%. Our RCT simulations using cognitive primary endpoints are in line with the ongoing PreCoDe trial that investigates the effect of prasinezumab on a cognitive composite score in PD<sup>34</sup>. However, most previous RCTs have used a motor or combined endpoint<sup>7,35</sup>. The expected sample size reduction needs to be weighed against an increased screening failure rate, which was set to 50–90% in our simulations. Thus, the benefit of prognostic patient stratification strategies depends strongly on weighing the costs of additional screening visits against the lower number of follow-up visits required. Furthermore, prognostic patient stratification may limit the approval area of the evaluated drug, as only a minority of PD patients will be included in the trial. Our study presents concrete sample size and power calculations that will allow

researchers to make an informed decision on the potential benefits and disadvantages of prognostic patient stratification.

Our presented enrichment strategy allows patient stratification based on a single and objective parameter obtained at baseline. This provides a complementary approach to clinical trial optimization strategies we previously published: (I) using longitudinal follow-up of multiple clinical scores to enrich RCTs with fast-progressing PD patients, and (II) using primary outcomes derived from digital biomarkers assessing motor functions<sup>3,36</sup>. Further research should systematically investigate combinations of different clinical trial optimization strategies. Importantly, BAG-based enrichment strategies could be applied to other neurological diseases like Alzheimer's disease, where an association of BAG with disease progression has been observed<sup>12,13,37</sup>.

Our results are based on the PPMI cohort, which restricts PD patients to the early disease stage at inclusion. While this demonstrates the high prognostic potential of BAG in early disease stages, it offers only limited insight into PD in advanced disease stages. For example, it remains unclear whether BAG remains consistently low in slow-progressing people with PD or whether it will increase in later disease stages. Furthermore, our study is limited by the small number of PD patients with longitudinal MRI scans. As such, this study should be considered exploratory and requires replication in independent cohorts and broader populations to prove the robustness and generalizability of our results.

We limited our systematic comparison of brain age models to workflows based on GMV. Recent studies have shown that brain age estimations might be slightly more accurate if a combination of GMV, WMV, and cerebrospinal fluid volumes is used<sup>26</sup>. Furthermore, regional brain age models have been proposed, potentially providing better interpretability by restricting features to specific regions of interest<sup>16,38</sup>.

Although BAG shows promise for patient stratification in clinical trials, its utility for predicting outcomes in individual people with PD is limited, as most BAG estimations and observed group differences fall within the model's mean absolute error of 4.34 years. Integrating multimodal data has been shown to increase accuracy in predicting individual clinical outcomes in PD<sup>39</sup>. Further research should therefore explore whether combining BAG and other data modalities (e.g., digital biomarkers) could improve predictions of disease progression at the individual level.

Our exploratory study demonstrates that BAG may serve as a prognostic biomarker for faster neurodegeneration and cognitive decline in PD, particularly distinguishing fast-progressing people with PD. BAG offers potential to optimize clinical trial designs, thereby accelerating the development of new disease-modifying drugs.

## Methods

### PPMI cohort

We analyzed HC and people with PD from the Parkinson's Progression Markers Initiative (PPMI, NCT04477785, registered 2020-07-15) with visits between 2010 and 2024. PPMI included HC and untreated, de-novo PD patients whose clinical diagnosis was confirmed by a pathological dopamine transporter SPECT (DaTSCAN). Furthermore, PD inclusion was limited to individuals with a clinical diagnosis within two years before the baseline visit. Written informed consent for data collection and sharing was obtained from all participants by PPMI. Ethical guidelines on human data collection were adhered to. The PPMI project was approved by the Institutional Review Board or Independent Ethics Committee of all participating sites in Europe, including Attikon University Hospital (Greece), Hospital Clinic de Barcelona and Hospital Universitario Donostia (Spain), Innsbruck University (Austria), Paracelsus-Elena-Klinik Kassel/University of Marburg (Germany), Imperial College London (UK), Pitié-Salpêtrière Hospital (France), University of Salerno (Italy), and in the USA, including Emory University, Johns Hopkins University, University of Alabama at Birmingham, PD and Movement Disorders Center of Boca Raton, Boston University, Northwestern University, University of Cincinnati, Cleveland Clinic Foundation, Baylor College of Medicine, Institute for Neurodegenerative Disorders, Columbia University Medical Center, Beth Israel Medical

Center, University of Pennsylvania, Oregon Health and Science University, University of Rochester, University of California at San Diego, and University of California, San Francisco.

### Parkinson's disease progression subtypes

PD progression subtypes were identified in a recent publication using a latent time joint mixed-effects model (LTJMM) and variational deep embedding with recurrence (VADER) in the PPMI cohort<sup>3</sup>. PD subtype assignments were available for PD patients with longitudinal assessments of non-motor and motor scores (i.e., at least two visits with MDS-UPDRS I-IV, MoCA, PIGD score, and SCOPA-AUT).

### Latent disease time calculation

Disease progression was modeled as a multivariate linear process using a latent time joint mixed-effects model (LTJMM) as described in previous publications<sup>3,25</sup>.

$$y_{ijk} = x_i\beta_k + \gamma_k(t_{ijk} + \delta_i) + \alpha_{0ik} + \alpha_{1ik}t_{ijk} + \epsilon_{ijk} \quad (1)$$

We denoted  $y_{ijk}$  as outcome  $k$  measured at visit  $j$  for individual  $i$ . We adjusted for age at diagnosis and sex using covariates  $x_i$  with coefficients  $\beta_k$  shared across individuals. Cohort-level progression  $\gamma_k$  was represented by a fixed mean slope for each outcome  $k$ , common to all individuals. Time was measured as time since diagnosis  $t_{ijk}$ . To align trajectories across people with PD, we introduced an individual-specific time shift  $\delta_i$  shared across outcomes, so trajectories were defined on a common latent disease time ( $t_{ijk} + \delta_i$ ). Individual heterogeneity was modeled with random intercepts  $\alpha_{0ik}$  and random slopes  $\alpha_{1ik}$ . Measurement errors  $\epsilon_{ijk}$  and time shifts  $\delta_i$  were assumed to be normally distributed with mean zero, and random effects ( $\alpha_{0ik}$ ,  $\alpha_{1ik}$ ) followed a multivariate normal distribution with mean of zero.

Models were fitted via Markov chain Monte Carlo (MCMC) with 4 chains, 25,000 iterations, and 12,500 warm-up steps using the R packages `ltjmm`<sup>40</sup> and `rstan`<sup>41</sup>. All outcomes were min–max normalized to their theoretical ranges; MoCA scores were inverted to ensure positive slopes across outcomes. Convergence was assessed visually and using  $\hat{R}$  statistics, ensuring that it was below 1.05.

The following outcomes were used in LTJMM to capture progression of a wide variety of motor and non-motor scores: MDS-UPDRS I-IV, Postural Instability and Gait Dysfunction score (PIGD), MoCA, and Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction (SCOPA).

To validate our LTJMM approach, we calculated Kendall tau-b rank correlation coefficients for disease stages reported by H&Y stages with (I) time since diagnosis and (II) latent disease times. The difference between the two correlations was assessed statistically using a bootstrapping approach.

### MRI acquisition

For training the brain age estimation model, four large neuroimaging datasets were used: Cambridge Centre for Ageing and Neuroscience (CamCAN,  $N = 651$ )<sup>42</sup>, Information eXtraction from Images (IXI,  $N = 562$ ) (<https://brain-development.org/ixi-dataset/>), the enhanced Nathan Kline Institute-Rockland Sample (eNKI,  $N = 597$ )<sup>43</sup>, and the 1000 brains study (1000BRAINS;  $N = 1143$ )<sup>44</sup> as described in a recent publication<sup>26</sup>. MRI inclusion was restricted to HC between the ages of 18 and 90 years with no diagnosis of psychiatric, neurological, or major medical conditions.

PPMI T1-weighted (T1w) MRIs were acquired for HC and people with PD using 3D volumetric sequence (e.g., MP-RAGE, IR-FSPGR) in the sagittal plane, with a slice thickness of 1.2 mm or less, and details being reported online (<https://www.ppmi-info.org/study-design/research-documents-and-sops>).

### MRI preprocessing

T1w images were preprocessed using the Computational Anatomy Toolbox (CAT) 12.8<sup>45</sup>, replicating the workflow of a recent publication<sup>26</sup>. For precise normalization and segmentation, T1w images first underwent affine

registration with increased accuracy ( $\text{accstr} = 0.8$ ). Following bias field correction and tissue class segmentation, normalization was performed using an optimized Geodesic Shooting approach with  $\text{regstr} = 1$ <sup>46</sup>. We utilized 1 mm Geodesic Shooting templates and generated output images at 1 mm isotropic resolution. The normalized GM segments were then modulated for both linear and non-linear transformations and used for further analyses.

### Voxelwise and parcelwise MRI analysis

PD progression subtypes were compared using GMV in voxelwise and parcelwise space. Therefore, we used a linear model with the voxelwise or parcelwise GMV as the dependent variable and the subtype as the independent variable, integrating age, sex, total intracranial volume, and latent disease time as covariates. Visualization was done by coloring voxels/parcels based on their  $t$ -value when the difference between progression subtypes was significant. Visualizations were created based on test results prior to multiple testing correction, as indicated in the figure legends. Additionally, test results after correcting for multiple testing using the Benjamini–Hochberg method are reported<sup>47</sup>. For voxelwise GMV comparison, analysis was restricted to voxels inside the CAT 12.8 brain mask.

In a second approach, we applied an additional 4-voxel smoothing using a full-width at half maximum Gaussian kernel and 4-voxel resampling via linear interpolation to reduce the number of voxels and improve signal-to-noise ratio.

For parcelwise GMV comparison, we combined variants of Schaefer's cortical atlas (100, 400, 800, or 1200 parcels)<sup>48</sup>, Fan's subcortical atlas (36 parcels)<sup>49</sup>, and Buckner's cerebellar atlas (37 parcels)<sup>50</sup>. We calculated the mean GMV of all voxels within each parcel (173, 473, 873, and 1273 total parcels).

Parcelwise and voxelwise analyses were executed using Python packages `nibabel` 5.2.1<sup>51</sup> and `nilearn` 0.10.4<sup>52</sup>.

### Brain age models

Brain age models were selected from a recent publication that systematically compared 128 different brain age estimation workflows that combined different feature representations and machine learning algorithms<sup>26</sup>. All brain age models were trained on the four HC datasets described above.

We selected three models from this publication, each performing best on a specific task:

1. Within-dataset brain age estimation (lowest mean absolute error): voxelwise feature representation using 4 mm smoothing and 4 mm resampling with Gaussian process regression (S4\_R4 + GPR)
2. Cross-dataset brain age estimation (lowest mean absolute error): voxelwise feature representation using 4 mm smoothing, 4 mm resampling, and subsequent principal component analysis with Gaussian process regression (S4\_R4 + PCA + GPR)
3. Correlation with clinical scores (highest Pearson's correlation): voxelwise feature representation using no smoothing and 4 mm resampling with lasso regression (S0\_R4 + LR)

Several studies have shown that brain age is often overestimated in elderly individuals and underestimated in younger individuals<sup>53</sup>. To remove this age-related bias, we compared two bias correction methods applied to the brain age estimations:

1. Cole et al. method<sup>54</sup>: linear model that estimates bias correction using estimated brain age
2. Beheshti et al. method<sup>27</sup>: linear model that estimates bias correction using BAG and chronological age

To select the optimal brain age estimation workflow for our PPMI dataset, we applied all three brain age models to estimate brain age on the PPMI HC cohort. Subsequently, brain age was corrected using the two bias correction methods described above in a 5-fold cross-validation fashion. We selected the workflow with the lowest mean absolute error on the HC test sets. For subsequent use in the PD cohort, we

retrained the best bias correction method on the entire PPMI HC cohort.

The final selected brain age estimation workflow was further validated for retest reliability and longitudinal consistency. We calculated the concordance correlation coefficient for BAG, assuming that BAG should remain relatively constant over short intervals. Therefore, we used longitudinal MRI scans with intervals of less than 1.5 years. Longitudinal consistency was assessed by comparing the increase in brain age between two subsequent MRI scans and the actual time between both MRI scans, using longitudinal MRI with more than 1.5 years in between. If more than two MRI scans were available for an individual, the first and last MRI scans were used.

### BAG group comparisons

BAG differences between HC, the fast-progressing subtype, and the slow-progressing subtype were compared pairwise using a linear model with BAG as the dependent variable and group as the independent variable. Age and sex were included as covariates. For comparison of the fast-progressing and slow-progressing subtypes, latent disease time was included additionally as a covariate.

BAG progression was analyzed by fitting a linear mixed-effects model for each group, with BAG as the dependent variable and time since first MRI as the independent variable. Fixed and patient-specific random intercepts and slopes were modeled.

### BAG correlations with clinical scores

BAG was correlated with (1) baseline outcomes and (2) progression of outcomes. For both analyses, some clinical scores were inverted so that higher scores always indicated more severe impairment or faster progression. Partial correlation coefficients were calculated with age as a covariate, using the Python package pingouin 0.5.4<sup>55</sup>. For correlations with baseline outcomes, latent disease time was included as an additional covariate. Progression of outcomes was modeled on the latent disease timescale by fitting a linear mixed-effects model. Fixed and patient-specific random intercepts and slopes were included.

Correlation coefficients and respective *p*-values were calculated for the following outcomes: (1) clinical scores, (2) DaTSCAN uptake ratios, and (3) cerebrospinal fluid parameters. Benjamini–Hochberg correction for multiple testing was performed within each outcome group<sup>47</sup>. DaTSCAN uptake ratios were analyzed separately for caudate nucleus, putamen, and striatum with (1) the mean of the left and right sides and (2) the asymmetry index as parameters. Cerebrospinal fluid parameters included amyloid beta 1–42, neurofilament light chain, phospho-tau, and phospho-tau/amyloid beta 1–42 ratio.

### Event-free survival analysis

Event-free survival analysis was performed using a Cox proportional hazards model for the events (1) occurrence of mild cognitive impairment (MCI), and (2) cognitive decline indicated by progression on an ordinal scale that included the items normal cognition, subjective cognitive impairment, mild cognitive impairment, and dementia. Modeling was performed using age at baseline visit, sex, and BAG at baseline as covariates with time modeled on the latent disease timescale. The proportional hazards assumption was checked by (1) performing statistical tests for any time-varying coefficients with *p*-values being corrected for multiple testing using the Benjamini–Hochberg method, and (2) visual inspection of scaled Schoenfeld residuals. The analysis was implemented using the Python package lifelines 0.30.0<sup>56</sup>.

### Sample size calculation

We assessed the effect of enriching people with PD with high baseline BAG in clinical cohorts by simulating an RCT with a disease-modifying treatment. Single cognitive scores or a cognitive compound score were used as the primary outcome. The cognitive compound score was calculated by (1) computing z-scores for cognitive outcomes based on the baseline mean and

baseline standard deviation, (2) inverting z-scores for some outcomes so that higher scores always indicate more severe cognitive impairment, and (3) averaging the z-scores across cognitive tests. Therefore, we selected the four cognitive tests with the lowest sample size in the scenario without BAG stratification that were available at baseline (i.e., Hopkins Verbal Learning Test-Intermediate Recall, Letter Number Sequencing, Montreal Cognitive Assessment, Symbol Digit Modalities). Measurements were assumed every six months over two years. A power of 80% and a significance level of 0.1 were chosen. We assumed a treatment effect corresponding to a 30% reduction in the rate of disease progression (as measured by the cognitive composite score). For simplicity, we used equally sized control and treatment groups, excluding different treatment dosage arms. Power and sample size were calculated using the PPMI PD cohort and a linear mixed-effects model, following Edland et al.<sup>57</sup>, implemented in R 4.3.2 with the longpower package 1.0.24<sup>58</sup>.

### Statistical analyses

All statistical tests were performed two-sided with a significance level of 0.05 using Python 3.12 with packages statsmodels 0.14.2<sup>59</sup> and scikit-learn 1.5.0<sup>60</sup>. Box plots display the median, interquartile range (IQR), whiskers extending to 1.5 times the IQR, and outliers as circles beyond the whiskers.

### Data availability

Clinical data is available from the PPMI study group ([www.ppmi-info.org](http://www.ppmi-info.org)).

### Code availability

Source code for statistical analyses will be published at <https://github.com/t-haehnel/Brainage-PD-Progression> under the MIT license upon manuscript acceptance. Source code for brain age estimation workflows is available at [https://github.com/juaml/brainage\\_estimation](https://github.com/juaml/brainage_estimation).

Received: 15 October 2025; Accepted: 25 November 2025;

Published online: 08 December 2025

### References

- Greenland, J. C., Williams-Gray, C. H. & Barker, R. A. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *Eur. J. Neurosci.* **49**, 328–338 (2019).
- Borghammer, P. The  $\alpha$ -Synuclein Origin and Connectome Model (SOC Model) of Parkinson's disease: explaining motor asymmetry, non-motor phenotypes, and cognitive decline. *J. Parkinsons Dis.* **11**, 455–474 (2021).
- Hähnel, T. et al. Progression subtypes in Parkinson's disease identified by a data-driven multi cohort analysis. *npj Parkinsons Dis.* **10**, 95 (2024).
- Birkenbihl, C. et al. Artificial intelligence-based clustering and characterization of Parkinson's disease trajectories. *Sci. Rep.* **13**, 2897 (2023).
- Fereshtehnejad, S.-M. et al. New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes. *JAMA Neurol.* **72**, 863–873 (2015).
- Dadu, A. et al. Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *npj Parkinsons Dis.* **8**, 172 (2022).
- Pagano, G. et al. Trial of prasinezumab in early-stage Parkinson's disease. *N. Engl. J. Med.* **387**, 421–432 (2022).
- Pagano, G. et al. Prasinezumab slows motor progression in rapidly progressing early-stage Parkinson's disease. *Nat. Med.* **30**, 1096–1103 (2024).
- Sajedi, H. & Pardakhti, N. Age prediction based on brain MRI image: a survey. *J. Med. Syst.* **43**, 279 (2019).
- Liu, S. et al. A systematic review: brain age gap as a promising early diagnostic biomarker for Alzheimer's disease. *J. Neurol. Sci.* **475**, 123563 (2025).

11. Yin, C. et al. Deep learning to quantify the pace of brain aging in relation to neurocognitive changes. *Proc. Natl. Acad. Sci. USA* **122**, e2413442122 (2025).
12. Gaser, C. et al. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PLoS ONE* **8**, e67346 (2013).
13. Liu, W. et al. Risk prediction of Alzheimer's disease conversion in mild cognitive impaired population based on brain age estimation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 2468–2476 (2023).
14. Dular, L., Špiclin, Ž. & For The Alzheimer's Disease Neuroimaging Initiative. The Australian Imaging Biomarkers And Lifestyle Flagship Study Of Ageing Analysis of brain age gap across subject cohorts and prediction model architectures. *Biomedicine* **12**, 2139 (2024).
15. Beheshti, I., Mishra, S., Sone, D., Khanna, P. & Matsuda, H. T1-weighted MRI-driven brain age estimation in Alzheimer's disease and Parkinson's disease. *Aging Dis.* **11**, 618 (2020).
16. Chen, Y.-S. et al. Multiscale brain age prediction reveals region-specific accelerated brain aging in Parkinson's disease. *Neurobiol. Aging* **140**, 122–129 (2024).
17. Nguyen, H.-D., Clément, M., Mansencal, B. & Coupé, P. Brain structure ages—a new biomarker for multi-disease classification. *Hum. Brain Mapp.* **45**, e26558 (2024).
18. Teipel, S. J. et al. Brain age in genetic and idiopathic Parkinson's disease. *Brain Commun.* **6**, fcae382 (2024).
19. Yan, S. et al. The influence of accelerated brain aging on coactivation pattern dynamics in Parkinson's disease. *J. Neurosci. Res.* **102**, e25357 (2024).
20. Yoshinaga, K. et al. Age-disproportionate atrophy in Alzheimer's disease and Parkinson's disease spectra. *Alzheimers Dement.* **17**, e70048 (2025).
21. Eickhoff, C. R. et al. Advanced brain ageing in Parkinson's disease is related to disease duration and individual impairment. *Brain Commun.* **3**, fcab191 (2021).
22. Charissé, D. et al. Brain age and Alzheimer's-like atrophy are domain-specific predictors of cognitive impairment in Parkinson's disease. *Neurobiol. Aging* **109**, 31–42 (2022).
23. Beheshti, I., Booth, S. & Ko, J. H. Differences in brain aging between sexes in Parkinson's disease. *NPJ Parkinsons Dis.* **10**, 35 (2024).
24. Chai, L. et al. Estimated Brain Age in Healthy Aging and Across Multiple Neurological Disorders. *J. Magn. Reson. Imaging* <https://doi.org/10.1002/jmri.29667> (2024).
25. Hähnel, T. et al. Predictive modeling to uncover Parkinson's disease characteristics that delay diagnosis. *NPJ Parkinsons Dis.* **11**, 64 (2025).
26. More, S. et al. Brain-age prediction: a systematic comparison of machine learning workflows. *NeuroImage* **270**, 119947 (2023).
27. Beheshti, I., Nugent, S., Potvin, O. & Duchesne, S. Bias-adjustment in neuroimaging-based brain age frameworks: a robust scheme. *NeuroImage Clin.* **24**, 102063 (2019).
28. Beheshti, I. et al. Predicting brain age using machine learning algorithms: a comprehensive evaluation. *IEEE J. Biomed. Health Inform.* **26**, 1432–1440 (2022).
29. Beheshti, I., Perron, J. & Ko, J. H. Neuroanatomical signature of the transition from normal cognition to MCI in Parkinson's disease. *Aging Dis.* <https://doi.org/10.14336/AD.2024.0323> (2024).
30. Bashyam, V. M. et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).
31. Chen, C.-L. et al. Advanced brain aging in Parkinson's disease with cognitive impairment. *NPJ Parkinsons Dis.* **10**, 62 (2024).
32. Horsager, J. et al. Brain-first versus body-first Parkinson's disease: a multimodal imaging case-control study. *Brain* **143**, 3077–3088 (2020).
33. Bräuer, S., Schniewind, I., Dinter, E. & Falkenburger, B. H. Recursive seed amplification detects distinct  $\alpha$ -synuclein strains in cerebrospinal fluid of patients with Parkinson's disease. *Acta Neuropathol. Commun.* **13**, 13 (2025).
34. University Hospital Tuebingen. A randomized, double-blind, placebo-controlled, 104-week proof-of-concept study to evaluate the efficacy of intravenous prasinezumab in participants with Parkinson's disease carrying a severe mutation in the GBA gene. <https://clinicaltrials.gov/study/NCT07055087> (2025).
35. Vijiaratnam, N. et al. Exenatide once a week versus placebo as a potential disease-modifying treatment for people with Parkinson's disease in the UK: a phase 3, multicentre, double-blind, parallel-group, randomised, placebo-controlled trial. *Lancet* **405**, 627–636 (2025).
36. Raschka, T. et al. Objective monitoring of motor symptom severity and their progression in Parkinson's disease using a digital gait device. *Sci. Rep.* **15**, 25541 (2025).
37. Condado, J. G. et al. BrainAge moderates associations between Alzheimer's disease biomarkers and cognitive decline: a meta-analysis across A4/LEARN, HABS and ADNI cohorts. Preprint at medRxiv <https://doi.org/10.1101/2025.07.07.25331026> (2025).
38. Ran, C., Yang, Y., Ye, C., Lv, H. & Ma, T. Brain age vector: A measure of brain aging with enhanced neurodegenerative disorder specificity. *Hum. Brain Mapp.* **43**, 5017–5031 (2022).
39. Brzenczek, C., Klopfenstein, Q., Hähnel, T., Holger, F. & Glaab, E. Integrating digital gait data with metabolomics and clinical data to predict outcomes in Parkinson's disease. *NPJ Dig. Med.* **7**, 235 (2024).
40. Donohue, M. mdonohue / Itjmm — Bitbucket. <https://bitbucket.org/mdonohue/Itjmm/src/master/> (2017).
41. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Soft.* **76**, 1–32 (2017).
42. Taylor, J. R. et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage* **144**, 262–269 (2017).
43. Nooner, K. B. et al. The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front. Neurosci.* **6**, 152 (2012).
44. Caspers, S. et al. Studying variability in human brain aging in a population-based German cohort—rationale and design of 1000BRAINS. *Front. Aging Neurosci.* **6**, 149 (2014).
45. Gaser, C. et al. CAT: a computational anatomy toolbox for the analysis of structural MRI data. *GigaScience* **13**, giae049 (2024).
46. Ashburner, J. & Friston, K. J. Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage* **55**, 954–967 (2011).
47. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
48. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
49. Fan, L. et al. The Human Brainnetome Atlas: a new brain atlas based on connectonal architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
50. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).
51. Brett, M. et al. nipy/nibabel: 5.2.1. *Zenodo* <https://doi.org/10.5281/zenodo.10714563> (2024).
52. Chamma, A. et al. nilearn. *Zenodo* <https://doi.org/10.5281/zenodo.10948303> (2024).
53. De Lange, A.-M. G. & Cole, J. H. Commentary: Correction procedures in brain-age prediction. *NeuroImage: Clin.* **26**, 102229 (2020).
54. Cole, J. H. et al. Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
55. Vallat, R. Pingouin: statistics in Python. *J. Open Source Soft.* **3**, 1026 (2018).

56. Davidson-Pilon, C. lifelines: survival analysis in Python. *J. Open Source Soft.* **4**, 1317 (2019).
57. Ard, M. C. & Edland, S. D. Power calculations for clinical trials in Alzheimer's disease. *J. Alzheimers Dis.* **26**, 369–377 (2011).
58. Iddi, S. & Donohue, M. C. Power and sample size for longitudinal models in R – the longpower package and Shiny app. *R. J.* **14**, 264–282 (2022).
59. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* 92–96 (SciPy, 2010).
60. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

## Acknowledgements

The study was conducted as part of AIPD, which was funded by the European Union under the grant agreement number 101168379. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

## Author contributions

T.H.: conceptualization, methodology, formal analysis, writing – original draft, visualization. F.H.: methodology, software, writing – review & editing. B.H.: supervision, writing – review & editing, H.F.: supervision, writing – review & editing, K.R.: writing – review & editing. S.M.: methodology, writing – review & editing

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declares no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41531-025-01232-4>.

**Correspondence** and requests for materials should be addressed to Tom Hähnel.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025