

RESEARCH ARTICLE

Longitudinal brain structure changes in Parkinson's disease: A replication study

Andrzej Sokołowski¹, Nikhil Bhagwat², Yohan Chatelain¹, Mathieu Dugré¹, Alexandru Hanganu³, Oury Monchi⁴, Brent McPherson², Michelle Wang², Jean-Baptiste Poline², Madeleine Sharp², Tristan Glatard^{1*}

1 Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, **2** Department of Neurology and Neurosurgery, McGill University, Montreal, Canada, **3** Département de Psychologie, Université de Montréal, Montréal, Canada, **4** Département de Radiologie, Radio-Oncologie et Médecine Nucléaire, Université de Montréal, Montréal, Canada

* tristan.glatard@concordia.ca



OPEN ACCESS

Citation: Sokołowski A, Bhagwat N, Chatelain Y, Dugré M, Hanganu A, Monchi O, et al. (2024) Longitudinal brain structure changes in Parkinson's disease: A replication study. PLoS ONE 19(1): e0295069. <https://doi.org/10.1371/journal.pone.0295069>

Editor: Niels Bergsland, University at Buffalo, UNITED STATES

Received: May 3, 2023

Accepted: November 13, 2023

Published: January 31, 2024

Copyright: © 2024 Sokołowski et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The code and results are publicly available at <https://github.com/LivingPark-MRI/hanganu-et-al-2014>. Data used in the notebook were downloaded directly from the PPMI and cannot be shared publicly due to its Data Usage Agreements preventing republishing data. Data was collected by PPMI between 2010 and 2023, we have accessed the database in February 2023. While PPMI provides imaging, clinical, and demographic data, the identifiers are anonymized, hence we don't have access to any personal information such as person's name or address. By

Abstract

Context

An existing major challenge in Parkinson's disease (PD) research is the identification of biomarkers of disease progression. While magnetic resonance imaging is a potential source of PD biomarkers, none of the magnetic resonance imaging measures of PD are robust enough to warrant their adoption in clinical research. This study is part of a project that aims to replicate 11 PD studies reviewed in a recent survey (*JAMA neurology*, 78(10) 2021) to investigate the robustness of PD neuroimaging findings to data and analytical variations.

Objective

This study attempts to replicate the results in Hanganu et al. (*Brain*, 137(4) 2014) using data from the Parkinson's Progression Markers Initiative (PPMI).

Methods

Using 25 PD subjects and 18 healthy controls, we analyzed the rate of change of cortical thickness and of the volume of subcortical structures, and we measured the relationship between structural changes and cognitive decline. We compared our findings to the results in the original study.

Results

(1) Similarly to the original study, PD patients with mild cognitive impairment (MCI) exhibited increased cortical thinning over time compared to patients without MCI in the right middle temporal gyrus, insula, and precuneus. (2) The rate of cortical thinning in the left inferior temporal and precentral gyri in PD patients correlated with the change in cognitive performance. (3) There were no group differences in the change of subcortical volumes. (4) We did not find a relationship between the change in subcortical volumes and the change in cognitive performance.

using PPMI dataset we accepted Data Usage Agreement which in particular prevents us from redistributing the data or sharing patients' identifiers publicly. The PPMI is registered at the National Institutes of Health's clinical trial registry (NCT04477785).

Funding: This work was funded by the Michael J. Fox Foundation for Parkinson's Research (MJFF-021134; <https://www.michaeljfox.org>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

Despite important differences in the dataset used in this replication study, and despite differences in sample size, we were able to partially replicate the original results. We produced a publicly available reproducible notebook allowing researchers to further investigate the reproducibility of the results in Hanganu et al. (2014) when more data is added to PPMI.

Introduction

Parkinson's disease (PD), one of the most common neurodegenerative diseases, is often characterized by akinesia, bradykinesia, tremor, and is commonly associated with mild cognitive impairment which significantly decrease overall quality of life [1]. These symptoms are accompanied by atrophy of the cortical and subcortical brain structures as well as cortical thinning [2]. As a result, there has been interest in determining whether magnetic resonance imaging (MRI) measures of atrophy can be used as a biomarker of cognitive decline. Overall, gray matter atrophy and cortical thinning are present in early PD, while frontal atrophy and temporoparietal thinning are associated with cognitive impairment in PD [3].

Magnetic resonance imaging-derived measures of the structural brain changes occurring in PD have emerged as potential diagnostic and prognostic tools to understand the trajectory of PD. Structural imaging, especially regional cortical thickness and loss of gray matter volume, has been considered helpful in determining a PD diagnosis, progression prognosis, and distinguishing PD from other dementias [2]. However, the need to further investigate the sensitivity, reliability, effect of confounding factors, and overall generalizability of these progression measures has been highlighted [3] and such rigorous validation is likely a factor preventing the adoption of MRI measures as outcome measures in PD clinical research.

The replication of neuroimaging findings has been challenged in multiple ways in recent years. For example, in the study of Botvinik-Nezer et al. [4], 70 independent teams were asked to analyze the same dataset using the methods of their choice to test nine ex-ante hypotheses. The study investigated the variability of the results. Results obtained across research teams did not concur on five out of the nine hypotheses, reaching agreement levels ranging from 21% to 37%. Furthermore, the identification of regional brain atrophy in PD has been of interest as a possible marker of certain symptoms of PD and of the progression of PD [2]. However, studies conducted in non-PD populations have shown that estimates of regional volume [5, 6] and of cortical thickness vary depending on the software toolbox [7, 8]. Overall, a range of factors matter in the replicability of neuroimaging findings, including computational environments [9, 10], analysis tools and versions [11, 12], statistical models [13], and study populations [14].

This study is a part of a reproducibility evaluation project that aims to replicate 11 structural MRI measures of PD reviewed in Mitchel et al. [2]. The goal of the present study is to replicate the work by Hanganu et al. [15] to test whether prior findings regarding structural MRI-derived PD biomarkers replicate in a different dataset using similar analytical methods. Hanganu et al. [15] compared the change of gray matter volume and cortical thinning over time between PD patients with mild cognitive impairment (PD-MCI), PD patients without mild cognitive impairment (PD-non-MCI), and healthy older controls (HC); and also tested the relationship between longitudinal structural changes and cognitive decline in the PD patients. They reported four main findings: (Finding 1) an increased rate of cortical thinning in PD patients with mild cognitive impairment compared to PD patients without MCI (mainly affecting the right temporal regions, insula, and inferior frontal gyrus), and compared to healthy controls (mainly in the right temporal regions and supplementary motor area);

(Finding 2) a correlation between the change in Montreal Cognitive Assessment (MoCA) scores and cortical thinning in the bilateral temporal lobe, right occipital medial lobe, and the left postcentral gyrus in PD patients; (Finding 3) an increased loss of the amygdala and nucleus accumbens volumes as well as overall cortical thickness for the PD-MCI group compared to HC; (Finding 4) a correlation between the change of the right amygdala and thalamus volumes and the change in MoCA scores in PD patients.

The results of Hanganu et al. [15] are of clinical interest because they provide insight into the relationship between structural brain changes and cognitive impairment, thus highlighting possible neural substrates of PD-related cognitive impairment [16]. Our study addresses the issue of MRI measure replicability, investigates crucial elements of reporting the study to make it replicable, and discusses the impact of study design decisions on the replicability. The goal of our study was to attempt to replicate the original findings using a different cohort. We used open data from the Parkinson's Progression Markers Initiative (PPMI; www.ppmi-info.org) in order to construct a similar patient cohort as that used in the original study and we followed the data processing methods and statistical analyses from the original study.

Methods

Participants

The original study included 15 PD-non-MCI, 17 PD-MCI and 18 HC. In order to reconstruct this cohort, PD patients and HC were selected from PPMI to attempt to match the sample size and demographics of the groups in the original study. The following criteria were used to define the PD cohorts: clinical diagnosis of PD, available T1-weighted images at two research visits, Hoehn and Yahr stage I and II (the stage was stable across the two visits for each patient), testing performed at PD OFF state, available MoCA scores, and the absence of any other neurological condition. Data was collected after approval of the local ethics committees of the PPMI's participating sites. All participants provided written informed consent. This study was conducted in accordance with the Declaration of Helsinki and was exempt from the Concordia University's Research Ethics Unit.

Patients were divided into PD-MCI and PD-non-MCI groups. In the original study, MCI was diagnosed on the basis of the presence of subjective complaints of cognitive impairment, objective impairment on two or more neuropsychological tests in one domain of cognitive function and the absence of dementia. In the PPMI dataset, patients are already classified as having MCI or not using a very similar criteria for classification, and thus the existing classification was used. The diagnosis is available in the PPMI's cogstate variable which was used to determine MCI. Diagnosis of MCI in PPMI is administered at the participating sites based on clinician's experience and test outcomes in six domains: attention, memory, orientation, executive abilities, praxis, and language.

Ten PD-MCI (M age = 67.6; SD = 5.8), 15 PD-non-MCI (M age = 63.4; SD = 9.4), and 18 HC participants were selected (M age = 66.9; SD = 6.1). PD-non-MCI and HC group sample sizes match those of the original study but an insufficient number of PD-MCI patients were identified in the PPMI dataset that met all the original inclusion criteria, thus our sample is smaller than the original sample (n = 10 vs n = 17). Initially, 20 PD-MCI patients who had at least two MRI scans were identified in the PPMI. One participant was excluded for not meeting Hoehn and Yahr criterium, one for lacking MoCA score, and additional eight were not tested during PD OFF state. There were 138 PD-non-MCI patients that had at least two MRI scans. Seven subjects were excluded for not meeting Hoehn and Yahr criterion, nine for lacking MoCA score, additional 53 patients were not tested during PD OFF state. The sample that met all the criteria to match the PD-MCI sample was 69. We matched 15 PD-non-MCI

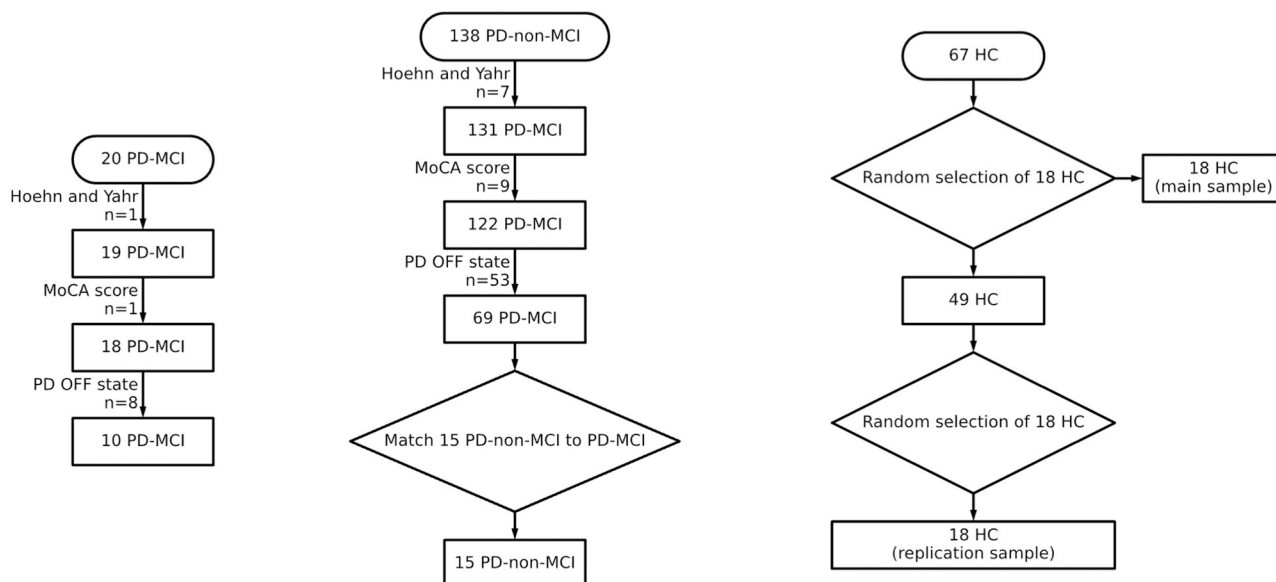


Fig 1. Flow diagram of sample selection.

<https://doi.org/10.1371/journal.pone.0295069.g001>

patients to match the PD-MCI group. Finally, there were 67 healthy control subjects out of whom 18 were randomly selected as a main control cohort. From the remaining 49 participants another set of 18 healthy controls were randomly selected as a replication sample. No participant was excluded due to image preprocessing failure. Flow diagram of sample selection is presented in Fig 1. Descriptive statistics are reported in Table 1.

Image acquisition and preprocessing

Structural MRI images were taken from the PPMI which uses a standardized study protocol and the following parameters: repetition time = 2.3 s, echo time = 2.98 s, inversion time = 0.9 s, slice thickness = 1 mm, number of slices = 192, field of view = 256 mm, and matrix

Table 1. Descriptive statistics.

	Mean ± SD			p value		
	PD-MCI (n = 10)	PD-non-MCI (n = 15)	HC (n = 18)	PD-MCI vs PD-non-MCI	PD-MCI vs HC	PD-non-MCI vs HC
Sex (M/F)	10/0	10/5	15/3	1	1	1
Age (y)	67.6 ± 5.8	63.4 ± 9.4	66.9 ± 6.1	0.21	0.76	0.2
Education (y)	14.6 ± 2.4	14.5 ± 1.9	16.1 ± 2.9	0.92	0.17	0.66
Duration Time 2–1 (m)	23.9 ± 9.0	19.1 ± 5.4	14.5 ± 2.9	0.11	< .001	0.003
Disease duration	4.9 ± 3.3	5.1 ± 3.1	-	0.9	-	-
UPDRS III ON (Time 1)	20.2 ± 14.3	17.7 ± 9.6	-	0.72	-	-
UPDRS III OFF (Time 1)	25.2 ± 13.0	22.3 ± 10.8	-	0.55	-	-
UPDRS III OFF (Time 2)	29.3 ± 13.6	25.7 ± 11.3	-	0.47	-	-
MoCA (Time 1)	24.3 ± 1.9	25.9 ± 1.8	-	0.05	-	-
MoCA (Time 2)	24.5 ± 2.4	27.0 ± 1.7	-	0.006	-	-

Group differences computed with Student's t-test for continuous variables and with χ^2 test for the categorical variable. HC, healthy controls; MoCA, Montreal Cognitive Assessment; UPDRS, PD-non-MCI, Parkinson's disease without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; Unified Parkinson's Disease Rating Scale.

<https://doi.org/10.1371/journal.pone.0295069.t001>

size = 256×256 . However, since PPMI is a multisite project there may be slight differences in the sites' setup. Scans were acquired using different 3T scanners (Philips Achieva $n = 2$; Siemens Prisma fit $n = 8$; Siemens Prisma $n = 4$; Siemens Skyra $n = 2$; Siemens TrioTim $n = 64$; Siemens Verio $n = 6$). There were scans with echo time (TE) that diverged from the standardized protocol: one image with TE = 2.52 s, one image with TE = 1.91 s, three images with TE = 2.91 s, one image with TE = 2.93 s, four images with TE = 2.95 s, four images with TE = 2.96 s, two images with TE = 3.06s. Additionally, two images had TE = 2.94 s, TR = 6.49 s, and TE = 2.91 s, TR 6.26 s.

T1-weighted brain images were processed using FreeSurfer 7.1.1 [17]. The longitudinal pre-processing stream was used to calculate the change in cortical thinning and subcortical volumes [18]. FreeSurfer's recon-all function was used for cortical reconstruction. First, all timepoints were processed cross-sectionally with the default workflow, then an unbiased template from the two timepoints was created for each subject, finally data was processed longitudinally. Specifically an unbiased within-subject template space and image [19] is created using robust, inverse consistent registration [20]. Several processing steps, such as skull stripping, Talairach transforms, atlas registration as well as spherical surface maps and parcellations are then initialized with common information from the within-subject template, significantly increasing reliability and statistical power [18]. The rate of change of cortical thickness between the two timepoints was calculated for each subject. Cortical thickness was smoothed with a 10 mm FWHM kernel. The original study also reported manual correction of misclassified tissue types, which was not performed in our study since the protocol for it was insufficient to replicate. The preprocessed cortical thickness data was used in the vertex-wise analyses and subcortical volumetric data was used in the subcortical analyses.

Statistical analyses

Structural brain images and Montreal Cognitive Assessment (MoCA) scores from the initial and the follow-up visits were analyzed consistently with the 4 main findings reported in the original study. (Finding 1) We tested vertex-wise differences in the change of cortical thickness between HC, PD-MCI, and PD-non-MCI groups with an ANCOVA model. (Finding 2) We tested the correlation between the change of cortical thickness and the change of MoCA scores in PD-MCI, PD-non-MCI, and PD-all (all PD patients) groups. The time between the two visits was added as a covariate in the general linear models. The rate of change of the cortical thickness was calculated with the formula: $(\text{Thickness at Time 1} - \text{Thickness at Time 2}) / (\text{Time 2} - \text{Time 1})$. Subcortical volumes were adjusted for the estimated total intracranial volume as well as the averages of the two time points using regression-based correction, in line with the original study. (Finding 3) We tested the differences in regional volume changes between the three groups using t-tests and (Finding 4) measured the correlations between the change in MoCA scores and change of the subcortical volumes and cortical thickness in each group using Pearson correlation. Whole-brain analyses were corrected using FreeSurfer's `mri_glm-sim` function. Cluster-wise p -value threshold was used at the $p < .05$ level. Volumetric analyses were corrected for multiple comparisons using Bonferroni correction. Group analysis is significant at $p < .007$ while correlational analysis at $p < .003$ which both correspond to $p < .05$ after the correction. Effect sizes in vertex-wise analyses represented by Pearson correlation using FreeSurfer's `mri_segstats` function.

Code availability

We used publicly available software to facilitate reproducing our study. Pandas v. 1.5.2 (doi: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)) was used to define the cohort from PPMI data files. FreeSurfer 7.1.1

was used for image preprocessing and vertex-wise analyses. We used a containerized version of FreeSurfer managed by Boutiques 0.5.25 (doi:[10.5281/zenodo.3839009](https://doi.org/10.5281/zenodo.3839009)). The containerized FreeSurfer analyses were executed through the Slurm batch manager on the Narval cluster (<https://docs.alliancecan.ca/wiki/Narval/en>) hosted at Calcul Québec and part of Digital Research Alliance of Canada.

The code and results are publicly available at <https://github.com/LivingPark-MRI/hanganu-etal-2014>. We developed a Python package (LivingPark utils, available at <https://github.com/LivingPark-MRI/livingpark-utils>) to download and manipulate PPMI data directly from the original PPMI database. As a result, our notebook can be re-executed by anyone with a PPMI account.

Results

Vertex-wise results

We found numerous group differences in the rate of change of cortical thickness. The results are reported in Table 2 and Fig 2.

PD-MCI vs. PD-non-MCI. The two patient groups differed in the rate of cortical thinning. The PD-MCI group, compared to PD-non-MCI patients, had an increased rate of cortical thinning in the right middle temporal and precentral gyri as well as right insula. PD-non-MCI group exhibited an increased thinning in the left precuneus compared to the PD-MCI group.

PD-MCI vs. HC. The HC group had increased cortical thinning in the right precentral and supramarginal gyri, left superior frontal gyrus, and bilateral superior parietal lobule, compared to the PD-MCI group.

PD-non-MCI vs. HC. The HC group had increased cortical thinning in the right precuneus as well as precentral and supramarginal gyri compared to the PD-non-MCI group.

Table 2. Vertex-wise group differences in the rate of change of cortical thickness.

Group and region	Size (mm ²)	MNI X	MNI Y	MNI Z	Max	Effect size	Cluster-wise <i>p</i> -value
PD-MCI vs PD-non-MCI							
L precuneus	901.62	-15.4	-45.2	65.8	2.8711	0.39	.0004
R precentral	808.31	45	-4.9	49.8	-4.3574	0.43	.0026
R insula	689.83	36.1	-16.4	-3.4	-4.285	0.42	.0066
R MTG	636.56	42.1	7.1	-37.1	-4.7243	0.41	.0108
PD-MCI vs HC							
L SPL	1596.42	-12.2	-55.4	59.9	3.4112	0.40	.0002
L SFG	816.25	-8.3	35.1	34	4.2357	0.39	.0006
R precentral	619.04	15	-12.2	61.8	2.4693	0.36	.0144
R supramarginal	578.61	37.3	-29.7	39.5	3.0369	0.38	.0221
R SPL	563.81	29.5	-48.4	42.7	3.9183	0.42	.0256
PD-non-MCI vs HC							
R SMG	1010.34	35.7	-30.6	39.6	4.4171	0.42	.0002
R precentral	580.49	43.5	-9.4	46.2	3.8495	0.41	.0211
R precuneus	537.11	8.6	-71.2	40	3.6102	0.39	.0384

Effect size represented by Pearson correlation. HC, healthy controls; Max, maximum $-\log_{10}(p\text{-value})$ in the cluster; MTG, middle temporal gyrus; PD-non-MCI, Parkinson's disease without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; SFG, superior frontal gyrus; SMG, supramarginal gyrus; SPL, superior parietal lobule.

<https://doi.org/10.1371/journal.pone.0295069.t002>

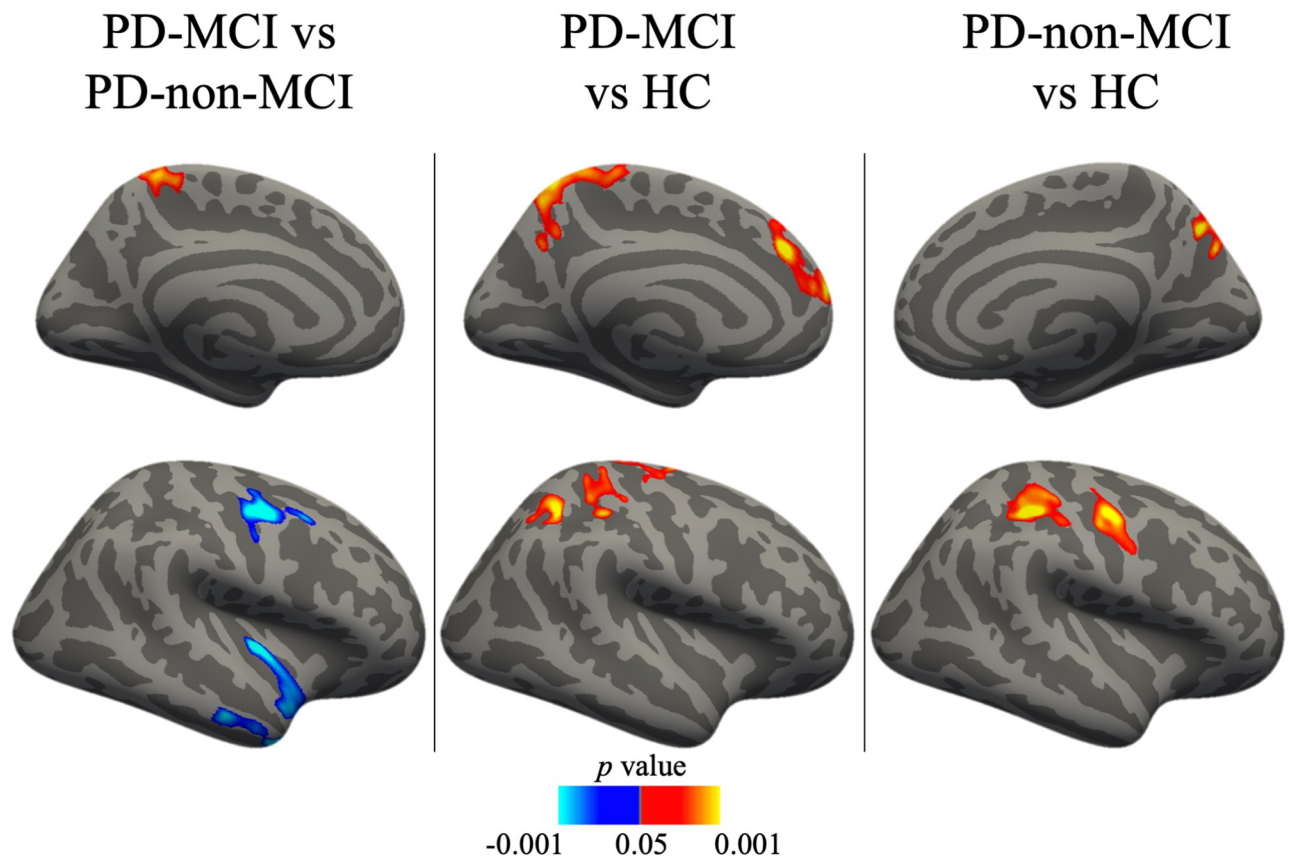


Fig 2. Vertex-wise group differences in the rate of change of cortical thickness. Cold colors represent increased cortical thinning in the first group compared to the second group, warm colors represent increased cortical thinning in the second group compared to the first group. Results corrected with the cluster-wise threshold ($p < .05$). HC, healthy controls; PD-MCI, Parkinson's disease with mild cognitive impairment; PD-non-MCI, Parkinson's disease without mild cognitive impairment.

<https://doi.org/10.1371/journal.pone.0295069.g002>

There was a significant negative correlation between the change in MoCA scores and the rate of change of the right middle frontal gyrus thickness in the PD-MCI group ($p < .05$). There was a positive correlation between the MoCA and the rate of change of the left inferior temporal and precentral gyri across all PD patients ($p < .05$). The correlations were not significant in the PD-non-MCI group. The results are reported in Table 3 and Fig 3.

Table 3. Vertex-wise correlation between the rate of change of cortical thickness and the change in Montreal Cognitive Assessment scores in PD patients.

Group and region	Size (mm ²)	MNI X	MNI Y	MNI Z	Max	Effect size	Cluster-wise p -value
PD-all							
L ITG	592.67	-53.3	-20.1	-33.1	3.4012	0.49	.0177
L precentral	523.77	-24.4	-16.8	64.5	2.5887	0.47	.0406
PD-MCI							
R rostral MTG	749.67	30.1	43.1	16.9	-4.0378	0.77	.0004

Effect size represented by Pearson correlation. ITG, inferior temporal gyrus; L, left; Max, maximum $-\log_{10}(p\text{-value})$ in the cluster; MTG, middle temporal gyrus; PD-all, Parkinson's disease with and without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; R, right.

<https://doi.org/10.1371/journal.pone.0295069.t003>

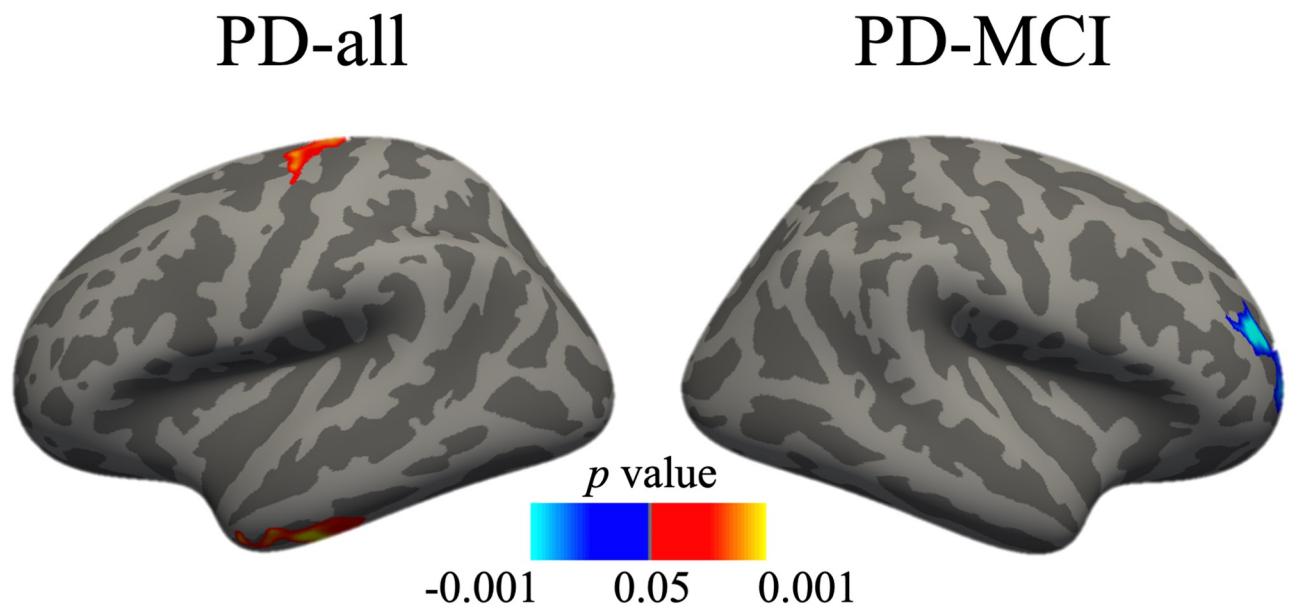


Fig 3. Vertex-wise correlation between the rate of change of cortical thickness and the change in Montreal Cognitive Assessment scores in the PD-all and PD-MCI groups.

<https://doi.org/10.1371/journal.pone.0295069.g003>

Volumetric results

Group comparisons of the volumetric change of subcortical regions revealed no significant differences between the groups ($p > .05$). The results are reported in Table 4.

Correlation analysis did not show any significant correlation between the change of MoCA scores and the change in volume of subcortical structures. Results are reported in Table 5.

Replication results

There were several group differences in the rate of change of cortical thickness in the replication analysis with a different HC sample. The results are reported in Table 6 and Fig 4.

Table 4. Group differences in the volumetric change of the subcortical regions and the overall cortical thickness.

	PD-MCI		PD-non-MCI		HC		PD-MCI vs PD-non-MCI		PD-MCI vs HC		PD-non-MCI vs HC	
	$M \pm SD$	%	$M \pm SD$	%	$M \pm SD$	%	t	p	t	p	t	p
CTh	0.006 ± 0.01	-0.57	0 ± 0.02	-0.11	0.006 ± 0.03	-0.21	0.78	.44	0.002	1	-0.68	.50
Thal	-223.24 ± 335	-2.56	210.34 ± 427	1.25	-51.26 ± 423	-0.68	-2.70	.01	-1.10	.28	1.76	.09
Caud	36.33 ± 229	0.51	-12.99 ± 220	-0.25	-9.36 ± 123	-0.2	0.48	.63	0.59	.56	-0.06	.95
Put	-16.38 ± 166	-0.66	115.21 ± 291	1.03	-86.9 ± 148	-0.92	-1.29	.21	1.16	.26	2.58	.02
Pal	-8.16 ± 223	0.41	31.45 ± 97	1.31	-21.68 ± 99	-0.13	-0.61	.55	0.22	.85	1.54	.13
Hipp	-50.43 ± 138	-1.76	4.56 ± 183	-1.14	24.22 ± 158	-0.49	-0.81	.43	-1.25	.22	-0.33	.74
Amyg	-39.03 ± 141	-4.97	-5.15 ± 148	-2.24	25.98 ± 116	-0.65	-0.57	.57	-1.31	.20	-0.68	.50
Nacc	-9.89 ± 87	-1.23	-3.06 ± 53	-0.82	8.04 ± 59	0.23	-0.24	.81	-0.65	.52	-0.56	.58

Subcortical volumes are presented in mm^3 , cortical thickness in mm. Amyg, amygdala; Caud, caudate; CTh, cortical thickness; HC, healthy controls; Hipp, hippocampus; Nacc, nucleus accumbens; PD-non-MCI, Parkinson's disease without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; Pal, pallidum; Put, putamen; Thal, thalamus. The results are significant at the $p < .007$ corresponding to $p < .05$ after the Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0295069.t004>

Table 5. Correlation between the rate of change of the subcortical volumes and the change in Montreal Cognitive Assessment scores.

Structure	PD-all		PD-MCI		PD-non-MCI	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
R Thalamus	-0.03	.89	0.09	.81	0.07	.80
R Caudate	0.09	.68	-0.18	.61	0.28	.31
R Putamen	0.02	.92	0.14	.7	0.1	.73
R Pallidum	0	1	-0.07	.84	0.33	.22
R Hippocampus	-0.1	.64	-0.04	.92	-0.05	.85
R Amygdala	-0.24	.26	-0.25	.48	-0.27	.33
R Nucleus accumbens	0.04	.87	0.38	.27	-0.12	.66
L Thalamus	0.19	.37	0.14	.71	0.48	.07
L Caudate	0	1	-0.29	.42	0.42	.12
L Putamen	-0.03	.9	-0.17	.63	0.19	.50
L Pallidum	0.05	.81	0.02	.97	0.15	.61
L Hippocampus	-0.01	.96	-0.08	.82	0.07	.81
L Amygdala	0.24	.26	0.32	.36	0.34	.22
L Nucleus accumbens	0.02	.92	-0.07	.85	0.12	.67

L, left; PD-all, Parkinson's disease with and without mild cognitive impairment; PD-non-MCI, Parkinson's disease without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; R, right. The results are significant at the $p < .003$ corresponding to $p < .05$ after the Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0295069.t005>

Table 6. Vertex-wise group differences in the rate of change of cortical thickness in the replication analysis.

Group and region	Size (mm ²)	MNI X	MNI Y	MNI Z	Max	Effect size	Cluster-wise <i>p</i> -value
PD-MCI vs PD-non-MCI							
L postcentral	1099	-20.2	-32	67.1	3.7161	0.41	.0002
L precentral	571.81	-47.7	-0.3	42	-4.5907	0.42	.0225
R precentral	890.47	45	-5	50.5	-4.3883	0.43	.0004
R insula	655.25	36.1	-16.4	-3.4	-4.7014	0.41	.0080
PD-MCI vs HC							
L precentral	2042.8	-35.1	-15.9	65.6	4.3611	0.42	.0002
R supramarginal	927.9	38.9	-29.1	38.8	4.2321	0.42	.0002
PD-non-MCI vs HC							
R supramarginal	1310.89	38.2	-28.9	40.3	5.0518	0.42	.0002
R lateraloccipital	1259.34	28	-88.8	15.4	3.261	0.39	.0002

<https://doi.org/10.1371/journal.pone.0295069.t006>

Group comparisons of the volumetric change of subcortical regions in the replication analysis revealed no significant differences between the groups ($ps > .05$). The mean change of the subcortical volumes in HC group is close to zero which is a results of random selection of the sample. The results are reported in Table 7.

Discussion

This study attempted to replicate the results of Hanganu et al. [15], which focused on the longitudinal changes in the cortical thickness and subcortical gray matter volume in PD patients. Group differences between PD patients with and without MCI were investigated along with the relationship between the structural changes and cognition. We used the analytic methods described in the paper and applied them to a different dataset of PD patients with and without MCI, and HC.

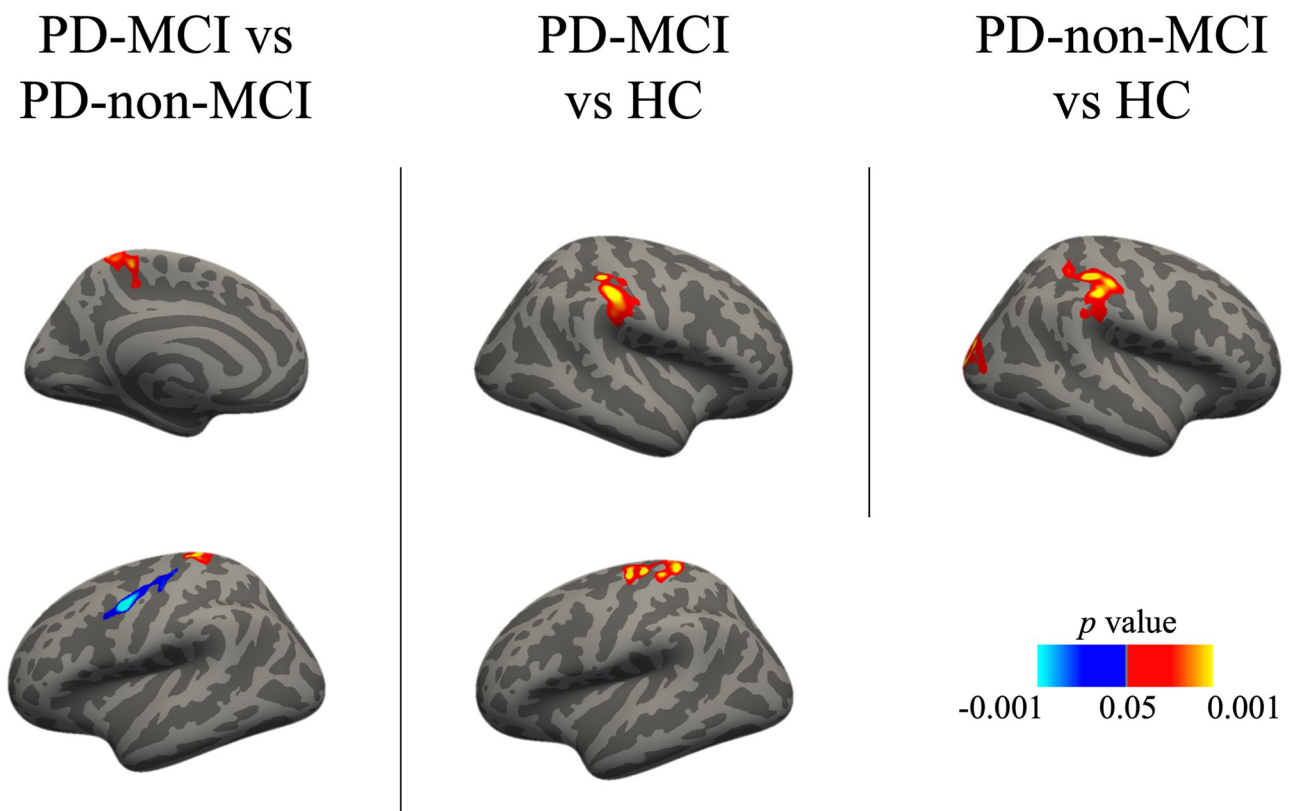


Fig 4. Vertex-wise group differences in the rate of change of cortical thickness in the replication analysis. Cold colors represent increased cortical thinning in the first group compared to the second group, warm colors represent increased cortical thinning in the second group compared to the first group. Results corrected with the cluster-wise threshold ($p < .05$). HC, healthy controls; PD-MCI, Parkinson's disease with mild cognitive impairment; PD-non-MCI, Parkinson's disease without mild cognitive impairment.

<https://doi.org/10.1371/journal.pone.0295069.g004>

Table 7. Group differences in the volumetric change of the subcortical regions and the overall cortical thickness in the replication analysis.

	PD-MCI		PD-non-MCI		HC		PD-MCI vs PD-non-MCI		PD-MCI vs HC		PD-non-MCI vs HC	
	$M \pm SD$	%	$M \pm SD$	%	$M \pm SD$	%	t	p	t	p	t	p
CTh	0.006 ± 0.01	-0.57	0 ± 0.02	-0.11	0.009 ± 0.02	-0.41	0.78	.44	-0.51	.62	-1.30	.20
Thal	-223.24 ± 335	-2.56	210.34 ± 427	1.25	0 ± 227	-1.27	-2.70	.01	-1.90	.07	1.70	.10
Caud	36.33 ± 229	0.51	-12.99 ± 220	-0.25	0 ± 126	-0.47	0.48	.63	0.47	.65	-0.21	.83
Put	-16.38 ± 166	-0.66	115.21 ± 291	1.03	0 ± 113	-0.82	-1.29	.21	-0.31	.76	1.55	.13
Pal	-8.16 ± 223	0.41	31.45 ± 97	1.31	0 ± 136	-0.6	-0.61	.55	-0.12	.90	0.75	.46
Hipp	-50.43 ± 138	-1.76	4.56 ± 183	-1.14	0 ± 212	-0.02	-0.81	.43	-0.67	.51	0.07	.95
Amyg	-39.03 ± 141	-4.97	-5.15 ± 148	-2.24	0 ± 196	0.23	-0.57	.57	-0.55	.58	-0.08	.93
Nacc	-9.89 ± 87	-1.23	-3.06 ± 53	-0.82	0 ± 97	-4.14	-0.24	.81	-0.27	.79	-0.11	.91

Subcortical volumes are presented in mm^3 , cortical thickness in mm. Amyg, amygdala; Caud, caudate; CTh, cortical thickness; HC, healthy controls; Hipp, hippocampus; Nacc, nucleus accumbens; PD-non-MCI, Parkinson's disease without mild cognitive impairment; PD-MCI, Parkinson's disease with mild cognitive impairment; Pal, pallidum; Put, putamen; Thal, thalamus. The results are significant at the $p < .007$ corresponding to $p < .05$ after the Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0295069.t007>

We have replicated the differences in the rate of cortical thinning between the PD-MCI and PD-non-MCI groups (Finding 1 in [15]), which supports the notion that these two patient groups differ in the rate of neurodegeneration. Patients with MCI had increased cortical thinning of the right MTG and insula, which overlaps with the regions reported in the original study. However, we did not replicate the group differences involving the HC group. These differences were not replicated using HC replication sample either. The discrepancies between the results with both HC results further support the notion that this original result is challenging to reproduce.

We found a relationship between the decrease in cognitive performance and cortical thinning of the left precentral gyrus and ITG in PD patients. The ITG cluster partially overlaps with the area reported in the original study. Although the structural changes were not found in the exact same brain voxels, the data supports the role of the thinning of the temporal regions in PD patients' cognition (Finding 2 in [15]). Additionally, we found a relationship between the increase of the cognitive performance and cortical thinning of the right MFG in the PD-MCI group. This result was not reported in the original study and provides additional insight into potential structural changes in PD populations. This is in contrast to previous studies reporting the link between the MFG atrophy and cognitive functioning [21, 22]. However, our result is not entirely surprising since the relationship between the presence of a regional gray matter preservation and the severity of symptoms in dementia has been previously reported [23]. Overall, we have replicated the original vertex-wise results (Findings 1 and 2 in [15]) to a certain extent.

Volumetric results (Findings 3 and 4 in [15]) were not replicated. The original study reported group differences in the change of overall cortical thickness as well as volumes of amygdala and nucleus accumbens. Hanganu et al. [15] reported a correlation between the change in cognitive performance and the change in gray matter volume in the right thalamus and amygdala in PD patients. We found no significant correlations across all subcortical regions. We did not replicate the results with the HC replication sample either but there is consistency between both HC groups. The lack of replication is not only caused by sample selection of HC group but also affected by the sample size, effect may be too small to be detected by such a small sample.

There are several factors beyond the selection of HC group that may explain the differences between the original results and the results of our replication. Most importantly, the two studies differ in the sample sizes for the PD-MCI group. We were unable to select enough PD patients with MCI that met all the inclusion criteria which reduced the chance to replicate the results. This was troublesome despite the fact that we used PPMI, one of the largest publicly available PD datasets. The PPMI is a relatively new initiative and is lacking much longitudinal data. Our Jupyter notebook remains accessible and can be re-run as new data becomes available in PPMI. Importantly, current neuroscience recommends using larger sample sizes to avoid inflation of effect sizes. Sample sizes used in research are increasing and recent data suggest that brain-wide association studies may require as many as thousands of participants to define reliable brain-behavior relationships [24].

Differences between the samples may also affect replicability. There are no established clinical measures to infer disease severity in the brain, but disease duration, UPDRS score (a measure of symptom severity), and medication use are sometimes used as proxy measures. Average disease duration was very similar in our sample compared to the original study suggesting the patients were roughly matched. However, some patients had not yet started PD medications in our sample whereas all patients in the original study were already taking dopaminergic medications, and the UPDRS score was also a few points lower in our sample compared to the original study's sample, both of which suggest that the replication sample we

constructed from the PPMI cohort had slightly milder disease than the sample included in the original study. Our sample is also slightly older than the original cohort. Furthermore, despite using the same inclusion criteria as the original study, it is possible that other differences in sample characteristics may have contributed to the incomplete replication. Conclusions drawn from our data should not be expanded to different clinical populations (e.g., more severe PD patients).

Differences in neuroimaging data acquisition protocols and MRI scanners can also influence replicability. Data used in our study was acquired using different MRI scanners. Although PPMI uses a standardized protocol for data acquisition we cannot rule out the possibility that the reported differences may be related to the variability introduced by using various scanners, even though the vast majority of scans were acquired with a Siemens scanner. Ideally all participants should be scanned with the same machine but the benefits of collecting more data in a multisite project outweighs the advantages of a single-stage acquisition.

Software versions may also have impacted the results. FreeSurfer 5.3. for Centos 4 was used in the original study. This version of FreeSurfer was released 10 years ago and is no longer supported or recommended, hence we used version 7.1.1. for Centos 7 instead. Although the software version shouldn't drastically change the clinical results, it is possible that it introduced variability during preprocessing or during the statistical analysis stage. We used the `mri_glmfit-sim` method to perform the analysis (including cluster correction) while the original study used the QDEC (Query, Design, Estimate, Contrast) method with `mri_surfcluster`. Our method is more stringent but also more reliable. It might have established more reliable borders between the gray and white matters. Filip et al. [25] reported structural group differences in data analyzed with FreeSurfer 5.3. which were not replicated with version 7.1. Previous studies indicated that software version may impact structural brain analyses [7, 9]. We will test the impact of software variability in our future work.

We thoroughly followed the original processing pipeline reported by Hanganu et al. [15]. Nevertheless, it is possible that we have missed some steps which could have influenced the results. We did not perform manual correction of misclassified brain tissue as this procedure cannot be objectively replicated. There might be slight differences in the statistical models that we were not aware of. Any of the aforementioned discrepancies from the original study could have impacted the ability to replicate the results. Once the analyses were conducted, we contacted the authors of the original study to obtain their feedback, which importantly contributed to the discussion section.

Finally, there is a negative bias in replication studies, coming from the fact that researchers conducting replications focus on following original methods rather than getting positive results. Therefore, it is expected that more negative results are reported in replications than in original studies. We encourage scientists to follow all the steps and details from the original studies instead of simply aiming to replicate positive results.

We encountered multiple challenges while attempting to replicate the study by Hanganu et al. [15]. Analysis details are necessarily limited in the methods section of most papers which left us to infer some analytic steps. We also had difficulty constructing a similar cohort from publicly available PD data, which led to some differences in the patient characteristics between the original and replication samples, and to differences in sample size. We have published a Jupyter notebook that the research community can use to replicate our study. While respecting the PPMI data usage agreement, it clearly defines the criteria to define the study population (using the Pandas library), the preprocessing pipeline and software version (through containerized tools), and the statistical model that was used to obtain the results. Our notebook addresses some of the aforementioned challenges encountered during our study and can be re-run over time to update the study as more data gets added to

PPMI. We continue to investigate the replicability of MRI-derived PD biomarkers by replicating other clinical studies.

Author Contributions

Conceptualization: Jean-Baptiste Poline, Madeleine Sharp, Tristan Glatard.

Data curation: Andrzej Sokołowski.

Formal analysis: Andrzej Sokołowski.

Funding acquisition: Jean-Baptiste Poline, Madeleine Sharp, Tristan Glatard.

Investigation: Andrzej Sokołowski, Jean-Baptiste Poline, Madeleine Sharp, Tristan Glatard.

Methodology: Andrzej Sokołowski, Tristan Glatard.

Project administration: Tristan Glatard.

Resources: Tristan Glatard.

Software: Andrzej Sokołowski, Yohan Chatelain, Mathieu Dugré, Tristan Glatard.

Supervision: Tristan Glatard.

Validation: Andrzej Sokołowski, Tristan Glatard.

Visualization: Andrzej Sokołowski.

Writing – original draft: Andrzej Sokołowski.

Writing – review & editing: Andrzej Sokołowski, Nikhil Bhagwat, Yohan Chatelain, Mathieu Dugré, Alexandru Hanganu, Oury Monchi, Brent McPherson, Michelle Wang, Jean-Baptiste Poline, Madeleine Sharp, Tristan Glatard.

References

1. Grafton ST. Contributions of functional imaging to understanding parkinsonian symptoms. *Curr Opin Neurobiol.* 2004; 14(6):715–9. <https://doi.org/10.1016/j.conb.2004.10.010> PMID: 15582373
2. Mitchell T, Lehericy S, Chiu SY, Strafella AP, Stoessl AJ, Vaillancourt DE. Emerging Neuroimaging Biomarkers Across Disease Stage in Parkinson Disease: A Review. *JAMA Neurol.* 2021; 78(10):1262–72. <https://doi.org/10.1001/jamaneurol.2021.1312> PMID: 34459865
3. Yang J, Burciu RG, Vaillancourt DE. Longitudinal Progression Markers of Parkinson's Disease: Current View on Structural Imaging. *Curr Neurol Neurosci Rep.* 2018; 18(12):1–11. <https://doi.org/10.1007/s11910-018-0894-7> PMID: 30280267
4. Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature.* 2020; 582(7810):84–8. <https://doi.org/10.1038/s41586-020-2314-9> PMID: 32483374
5. Gomez-Ramirez J, Quilis-Sancho J, Fernandez-Blazquez MA. A Comparative Analysis of MRI Automated Segmentation of Subcortical Brain Volumes in a Large Dataset of Elderly Subjects. *Neuroinformatics.* 2022; 20(1):63–72. <https://doi.org/10.1007/s12021-021-09520-z> PMID: 33783668
6. Rane S, Plassard A, Landman BA, Claassen DO, Donahue MJ, Initiative for the ADN. Comparison of Cortical and Subcortical Measurements in Normal Older Adults across Databases and Software Packages. *J Alzheimers Dis Rep.* 2017; 1(1):59–70. <https://doi.org/10.3233/ADR-170008> PMID: 29756095
7. Bhagwat N, Barry A, Dickie EW, Brown ST, Devenyi GA, Hatano K, et al. Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses. *Gigascience.* 2021; 10(1):1–13. <https://doi.org/10.1093/gigascience/giaa155> PMID: 33481004
8. Masouleh SK, Eickhoff SB, Zeighami Y, Lewis LB, Dahnke R, Gaser C, et al. Influence of Processing Pipeline on Cortical Thickness Measurement. *Cerebral Cortex.* 2020; 30(9):5014–27. <https://doi.org/10.1093/cercor/bhaa097> PMID: 32377664
9. Gronenschild EHBM, Habets P, Jacobs HIL, Mengelers R, Rozendaal N, van Os J, et al. The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical

- Volume and Cortical Thickness Measurements. PLoS One. 2012; 7(6):e38234. <https://doi.org/10.1371/journal.pone.0038234> PMID: 22675527
10. Glatard T, Lewis LB, da Silva RF, Adalat R, Beck N, Lepage C, et al. Reproducibility of neuroimaging analyses across operating systems. Front Neuroinform. 2015; 9:12. <https://doi.org/10.3389/fninf.2015.00012> PMID: 25964757
 11. Dickie E, Hodge SM, Craddock RC, Poline J, Kennedy DN. Tools Matter: Comparison of Two Surface Analysis Tools Applied to the ABIDE Dataset. Res Ideas Outcomes. 2017; 3:13726.
 12. Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage. 2014; 99:166–79. <https://doi.org/10.1016/j.neuroimage.2014.05.044> PMID: 24879923
 13. Tan A, Ma W, Vira A, Marwha D, Eliot L. The human hippocampus is not sexually-dimorphic: Meta-analysis of structural MRI volumes. Neuroimage. 2016; 124:350–66. <https://doi.org/10.1016/j.neuroimage.2015.08.050> PMID: 26334947
 14. Kennedy DN, Abraham SA, Bates JF, Crowley A, Ghosh S, Gillespie T, et al. Everything matters: The reponim perspective on reproducible neuroimaging. Front Neuroinform. 2019; 13:1. <https://doi.org/10.3389/fninf.2019.00001> PMID: 30792636
 15. Hanganu A, Bedetti C, Degroot C, Mejia-Constain B, Lafontaine AL, Soland V, et al. Mild cognitive impairment is linked with faster rate of cortical thinning in patients with Parkinson's disease longitudinally. Brain. 2014; 137(4):1120–9. <https://doi.org/10.1093/brain/awu036> PMID: 24613932
 16. Bougea A, Maraki MI, Yannakoulia M, Stamelou M, Xiomerisiou G, Kosmidis MH, et al. Higher probability of prodromal Parkinson disease is related to lower cognitive performance. Neurology. 2019; 92(19):e2261–72. <https://doi.org/10.1212/WNL.0000000000007453> PMID: 30944240
 17. Fischl B. FreeSurfer. Neuroimage. 2012; 62(2):774–81. <https://doi.org/10.1016/j.neuroimage.2012.01.021> PMID: 22248573
 18. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage. 2012; 61(4):1402–18. <https://doi.org/10.1016/j.neuroimage.2012.02.084> PMID: 22430496
 19. Reuter M, Fischl B. Avoiding asymmetry-induced bias in longitudinal image processing. Neuroimage. 2011; 57(1):19–21. <https://doi.org/10.1016/j.neuroimage.2011.02.076> PMID: 21376812
 20. Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: A robust approach. Neuroimage. 2010; 53(4):1181–96. <https://doi.org/10.1016/j.neuroimage.2010.07.020> PMID: 20637289
 21. Chung SJ, Yoo HS, Lee YH, Lee HS, Ye BS, Sohn YH, et al. Frontal atrophy as a marker for dementia conversion in Parkinson's disease with mild cognitive impairment. Hum Brain Mapp. 2019; 40(13):3784–94. <https://doi.org/10.1002/hbm.24631> PMID: 31090134
 22. Li L, Ji B, Zhao T, Cui X, Chen J, Wang Z. The structural changes of gray matter in Parkinson disease patients with mild cognitive impairments. PLoS One. 2022; 17(7):e0269787. <https://doi.org/10.1371/journal.pone.0269787> PMID: 35857782
 23. Sokołowski A, Roy ARK, Goh SYM, Hardy EG, Datta S, Cobigo Y, et al. Neuropsychiatric symptoms and imbalance of atrophy in behavioral variant frontotemporal dementia. Hum Brain Mapp. 2023; 44(15):5013–29. <https://doi.org/10.1002/hbm.26428> PMID: 37471695
 24. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. Reproducible brain-wide association studies require thousands of individuals. Nature. 2022; 603(7902):654–60. <https://doi.org/10.1038/s41586-022-04492-9> PMID: 35296861
 25. Filip P, Bednarik P, Eberly LE, Moheet A, Svatkova A, Grohn H, et al. Different FreeSurfer versions might generate different statistical outcomes in case–control comparison studies. Neuroradiology. 2022; 64(4):765–73. <https://doi.org/10.1007/s00234-021-02862-0> PMID: 34988592