Application of the Anatomical Fiducials Framework to a Clinical Dataset of Patients with Parkinson's Disease

Mohamad Abbass[a], Greydon Gilmore[a,b], Alaa Taha[c], Ryan Chevalier[c], Magdalena Jach[c], Terry M. Peters[b,d,e,f,g], Ali R. Khan[b,d,e,f,g,h], Jonathan C. Lau[a,b,i]

a. Department of Clinical Neurological Sciences, London Health Sciences Centre, Western University, London, ON, Canada.
b. School of Biomedical Engineering, Western University, London, Ontario, Canada.
c. Department of Physiology, Western University, London, Ontario, Canada.
d. Imaging Research Laboratories, Robarts Research Institute, Western University, London, Ontario, Canada
e. Centre for Functional and Metabolic Mapping, Robarts Research Institute, Western University, London, Ontario, Canada
f. Department of Medical Biophysics, Western University, London, Ontario, Canada
g. Brain and Mind Institute, Western University, London, Ontario, Canada
h. Graduate Program in Neuroscience, Western University, London, Ontario, Canada
i. Department of Neurosurgery, Emory University, Atlanta, Georgia, USA

Corresponding Author:

Jonathan C. Lau, MD, PhD, FRCSC

Department of Neurosurgery, Emory University, Atlanta, GA, USA

jonathan.c.lau@gmail.com, jonathan.c.lau@emory.edu

Declarations of interest: none

Abstract

Establishing spatial correspondence between subject and template images is necessary in neuroimaging research and clinical applications such as brain mapping and stereotactic neurosurgery. In the absence of other quantitative approaches, a point-based set of anatomical fiducials (AFIDs) was recently developed and validated to serve as a quantitative measure of image registration based on salient anatomical features. In this study, we sought to apply the AFIDs protocol to the clinic, specifically focussing on structural magnetic resonance images (MRI) obtained from patients with Parkinson's Disease (PD). We first confirmed that AFIDs could be placed to millimetric accuracy in the PD dataset with results comparable to those in normal control subjects. With localization error established, we evaluated subject-to-template registration using this framework by aligning the clinical scans to standard template space using a robust open preprocessing workflow for MRI scans. We found that registration errors from this workflow as measured using AFIDs were higher than previously reported suggesting the need for optimization of image processing pipelines for clinical grade datasets. Finally, we examined the utility of using point-to-point distances between AFID points as a morphometric biomarker of PD, finding evidence of reduced distances between AFIDs around the left temporal horn, brainstem and pineal gland in the clinical group, structures that circumscribe regions known to be affected in PD including the substantia nigra. Overall, we provide evidence that AFIDs can be successfully applied in a clinical setting and utilized to provide localized and quantitative measures of registration error. AFIDs provide clinicians and researchers with a common, open framework for quality control and validation of spatial correspondence and the location of anatomical structures, facilitating accurate aggregation of imaging datasets and comparisons between various neurological conditions.

Keywords: Registration; Accuracy; Fiducials; Deep brain stimulation; Parkinson's Disease; Biomarker.

1. Introduction

Non-invasive imaging techniques such as magnetic resonance imaging (MRI) have allowed for insights into the anatomy and function of the central nervous system. A critical aspect in neuroimaging research and clinical application is to establish accurate spatial correspondence between images [1], allowing for the combination and comparison of multimodal data across subjects and populations. Establishing spatial correspondence requires the specification of a common stereotactic 3D coordinate reference frame, and the registration of 3D images to that reference frame [2]. Researchers have established numerous common reference frames, including those based on both individuals and populations[3].

Establishing correspondence between various brain images has historically relied on linear transformations [3,4]. Over the last few decades, various non-linear transformations have been implemented allowing for more accurate registration between brain images [3]. These transformations have most commonly been evaluated using measures of overlap between regions of interest (ROIs). ROIs that have been used include subcortical structures such as the thalamus or areas of the basal ganglia [5,6]. Measures of spatial correspondence within these relatively large ROIs are known to be quite coarse and fail to capture subtle misregistration between images [5]. Inspired by classical stereotactic methods [7], a set of anatomical fiducials (AFIDs) were validated using an open framework and proposed as an intuitive way to quantify alignment using point-based distance measures between brain structures. This method was validated in individual subject and template scans and was found to be more sensitive to registration errors than ROI-based voxel overlap measures [5].

With the increasing use of MRI in research and clinical settings, accurate assessment of registration between image sequences is necessary. Since clinical outcomes in stereotactic neurosurgery depend on accuracy at the millimeter scale [8], a robust framework for assessing correspondence between brain images is required for optimal neurosurgical planning. In this work, we sought to evaluate the reproducibility and utility of the AFIDs framework in a clinical population with Parkinson's Disease (PD).

[1]AC, anterior commissure; AFID, anatomical fiducial; AFLE, anatomical fiducial localization error; AFRE, anatomical fiducial registration error; AM, anteromedial; AL, anterolateral; ICC, intraclass correlation; IPF, interpeduncular fossa; LMS, lateral mesencephalic sulcus; MCP, mid-commissural point. LV, lateral ventricle; PC, posterior commissure; PD, Parkinson's disease; PMJ, pontomesenphalic junction.

2. Material and Methods

2.1 Subject demographics and MRI acquisition

Subject scans used in this study were obtained from 39 individuals diagnosed with PD (age: 60.30 ± 6.70, gender: 33.3% female). For all subjects, the MRI sequence used was a post gadolinium enhanced volumetric T1-weighted (T1w) image (echo time = 1.5 ms, inversion time = 300 ms, flip angle = 20°, receiver bandwidth = 22.73 kHz, field of view = 26 cm x 26 cm, matrix size = 256 × 256, slice thickness = 1.4 mm, resolution = 1.25 × 1.25 × 1.50 mm) (Signa, 1.5 T, General Electric, Milwaukee, Wisconsin, USA). The subject data were collected at University Hospital in London, ON, Canada. The study was approved by the Human Subject Research Ethics Board (HSREB) office at the University of Western Ontario (REB# 109045).

2.2 AFID placement

The individual scans were imported into 3D Slicer version 4.10.0 [9]. The subject scans were first transformed into anterior commissure (AC)-posterior commissure (PC) space (AC-PC space), and the raters were required to initially place 4 of the AFIDs, which included: AC (AFID01), PC (AFID02) and two additional on the midline. The built-in "AC-PC Transform" function in 3D Slicer was used to align the AC-PC horizontally in-line in the anteroposterior plane. Adequate alignment was subjectively judged by each rater, who then placed the remaining AFIDs as previously outlined [5].

Five raters were initially trained to place AFIDs using publicly available brain images: MNI2009bNonlinAsym[3], deepbrain7t[10] and PD25-T1MPRAGE[11] . Quality assurance was performed to ensure each rater was placing the AFIDs on the templates below a minimum threshold of error (Euclidean error < 2.00 mm when compared with ground truth placements). Once the raters had received adequate feedback about their initial ratings during the training phase, they then independently performed the AFIDs protocol in the subject scans. Two raters (MA and GG) had prior neuroanatomy experience and were deemed "expert", while three (AT, MJ and RC) had no prior neuroanatomy experience and were deemed "novice". The novice raters had no experience with medical imaging so additional training was provided on navigating an MRI

sequence in 3D Slicer (i.e. left/right, axial/coronal/sagittal views etc.). A total of 6240 AFIDs were placed.

2.3 Analysis in subject space

The 3D coordinates of each AFID were exported and subsequently analyzed in MATLAB (vR2018b). The anatomical fiducial localization error (AFLE) was calculated as the Euclidean distance between each individually placed AFID and the group mean, in each of the 32 AFIDs in each scan. Therefore, 6240 AFLE measurements were made for each manually placed AFID. Outliers were determined as having an AFLE of greater than 10.0 mm and are reported in the results.

To determine each rater's deviation from the group mean, the mean rater AFLE across all 39 subjects was calculated for each AFID. AFLE was then dichotomized between expert and novice raters by calculating the mean AFLE among these two groups. Wilcoxon rank-sum tests were used to determine significance in AFLEs between expert and novice raters. Bonferroni correction was used to account for multiple comparisons with an adjusted p-value of 0.05/32 as a threshold for significance. The overall AFLE for each AFID was then calculated as the mean AFLE across all raters.

Rater reliability was assessed using intraclass correlation (ICC), which was calculated in each dimension. A two-way random effects model with single measurement type was used, ICC(2,1) as determined by Shrout and Fliess [12]. ICC among all raters, expert raters and novice raters was calculated.

2.4 Analysis in MNI Space

To assess and quantify registration error, the subject scans were non-linearly transformed to MNI2009bAsym brain template space using fMRIPrep 1.5.4 ([13]; RRID:SCR_016216), which is based on Nipype 1.3.1 ([14]; RRID:SCR_002502). Specifically, the T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection[15], distributed with ANTs 2.2.0 ([16]; RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as the target template. Brain

tissue segmentation of cerebrospinal fluid, white-matter and gray-matter was performed on the brain-extracted T1w using the fast algorithm from FSL 5.0.9 ([17]; RRID:SCR_002823). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed using a symmetric diffeomorphic image registration method (antsRegistration; ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c ([18]; RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym). Many internal operations of fMRIPrep use Nilearn 0.6.0 ([19]; RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

We calculate the Euclidean distance between each individually placed AFID transformed to MNI space and the group mean for each AFID placed in MNI space. We term this the Anatomical Fiducial Registration Error (AFRE). The mean AFRE across all subjects and raters was then calculated in the same manner as for the AFLE. We calculated the mean AFRE for linearly and non-linearly registered images. Wilcoxon rank-sum tests were used to determine significance between AFREs obtained following both linear and non-linear registration. Bonferroni correction was used to account for multiple comparisons with an adjusted p-value of 0.05/32 as a threshold for significance.

2.5 Distance between AFIDs as a biomarker of disease

We sought to utilize the AFIDs protocol to investigate unique morphometric features in our PD patient population. As such, we computed all pairwise Euclidean distances between AFIDs, generating 496 distance measures (32*31/2). We compared these values to distances obtained from a control group of 30 subjects from the OASIS-1 database with AFIDs previously placed [5]. Wilcoxon rank-sum tests were used to determine significant differences in pairwise distances between the two groups, and Bonferroni correction was used to account for multiple comparisons with an adjusted p-value of 0.05/496 being used as a threshold for significance.

All raw and processed data along with the processing scripts that were used in this manuscript are available at: https://github.com/greydongilmore/afids-clinical

3. Results

3.1 AFID Placement

Out of all 6240 AFIDs placed, 21 were deemed outliers by using a threshold AFLE of greater than 10 mm (0.33%). None of the outliers were placed by expert raters, therefore 0.55% fiducials placed by novice raters were outliers. All outliers involved placements at some component of the lateral ventricles, and were as follows (number of outliers for this structure in brackets): right lateral ventricle at AC (5), left lateral ventricle at AC (6), right lateral ventricle at PC (2), right anterolateral temporal horn (1), right superior anteromedial horn (1), left superior anteromedial horn (1), right inferior anteromedial horn (2), left inferior anteromedial horn (2) and right ventral occipital horn (1).

3.2 Analysis in Subject Space

Figure 1 depicts mean distance from the mid-commissural point by each rater for the 32 AFIDs. The mean overall AFLE across all AFIDs was 1.57 mm ± 1.16 mm. The mean AFLE across all raters for each AFID can be seen in Figure 2. AFID 25 and 26 (left and right lateral ventricle at AC respectively) had the highest AFLE at 2.63 mm ± 1.75 mm and 2.79 mm ± 1.95 mm respectively. The AFIDs with the lowest overall AFLE were AFID 01-02 (anterior commissure and posterior commissure respectively), with AFLEs of 0.70 mm ± 0.78 mm and 0.55 mm ± 0.34 mm respectively.

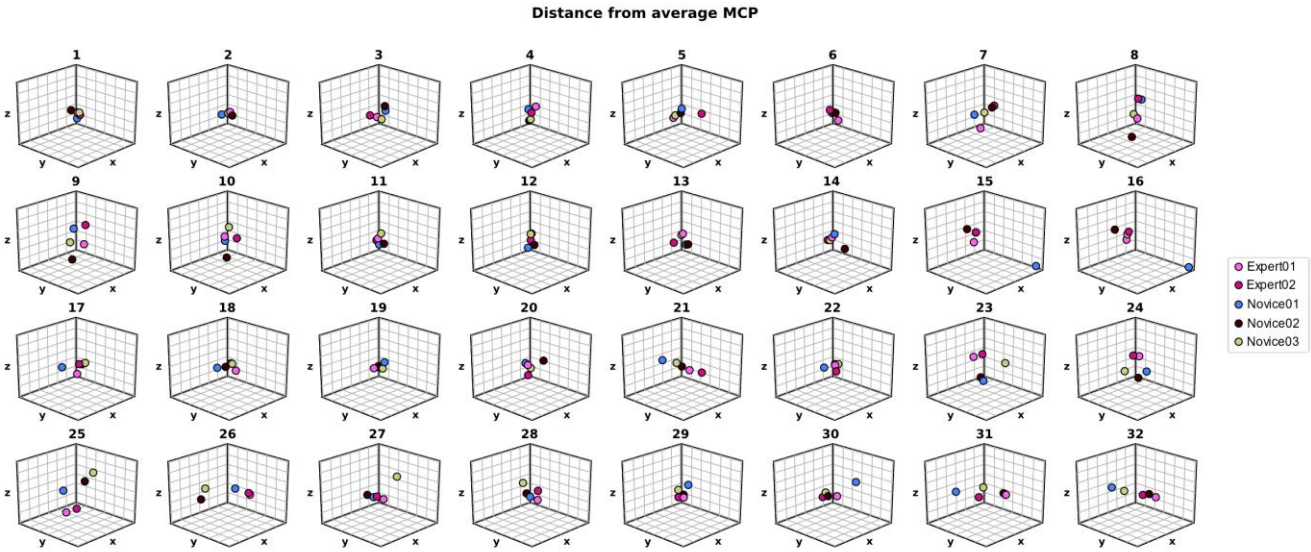

Figure 1 – Mean individual rater Euclidean distance from mid-commissural point (MCP) for all anatomical fiducials in subject space.

Table 1 represents the mean AFLE obtained by expert and novice raters. Expert raters overall had a lower mean AFLE (1.33 mm ± 0.79 mm), compared to novice raters (1.73 mm ± 1.30 mm). Wilcoxon rank-sum tests for AFLE between expert and novice raters with Bonferroni correction for multiple comparisons are shown in Table 1. Expert raters had a lower AFLE in 29 of the 32 AFIDs. 6 AFIDs had significantly different AFLEs between raters, 5 of which were higher in the novice raters. The superior interpeduncular fossa (AFID05) however had a greater AFLE obtained by expert raters compared to novice raters.

Fiducial	**Fiducial Name**	**Expert AFLE (mm)**	**Novice AFLE (mm)**
1	AC	0.54 ± 0.36	0.81 ± 0.45
2*	PC	0.41 ± 0.75	0.65 ± 1.09
3	Infracollicular sulcus	0.92 ± 1	1.15 ± 0.71
4	PMJ	0.86 ± 1.28	1.03 ± 1.92
5*	Superior interpeduncular fossa	1.60 ± 1.48	1.16 ± 1.23
6	R superior LMS	1.35 ± 1.75	1.44 ± 1.35
7	L superior LMS	1.55 ± 1.5	1.43 ± 1.72
8	R inferior LMS	1.61 ± 1.45	1.99 ± 1.9
9	L inferior LMS	1.68 ± 1.27	1.94 ± 1.44
10	Culmen	1.35 ± 0.6	1.86 ± 0.68
11	Intermammillary sulcus	0.64 ± 0.71	0.78 ± 0.85
12	R MB	0.78 ± 0.78	0.87 ± 0.92
13	L MB	0.85 ± 1.47	0.82 ± 1.34
14	Pineal gland	1.41 ± 1.35	1.53 ± 1.29
15*	R LV at AC	1.32 ± 1.41	2.74 ± 1.45
16*	L LV at AC	1.43 ± 1.39	2.93 ± 1.2
17	R LV at PC	1.30 ± 1.31	2.02 ± 1.02
18	L LV at PC	1.16 ± 1	1.65 ± 0.93
19	Genu of CC	0.96 ± 0.89	1.22 ± 1.07
20*	Splenium	0.98 ± 1.11	1.47 ± 1.67
21	R AL temporal horn	1.39 ± 1.35	1.64 ± 1.6
22	L AL temporal horn	1.48 ± 1.51	2.00 ± 1.38
23	R superior AM temporal horn	1.45 ± 1.56	1.78 ± 1.55
24	L superior AM temporal horn	1.56 ± 2.19	1.94 ± 2.4
25	R inferior AM temporal horn	2.29 ± 2.57	2.85 ± 2.34
26	L inferior AM temporal horn	2.46 ± 1.55	3.01 ± 1.48
27	R indusium griseum origin	1.51 ± 1.87	2.10 ± 1.64
28	L indusium griseum origin	1.75 ± 1.42	2.04 ± 1.26
29	R ventral occipital horn	1.34 ± 1.37	1.91 ± 1.58
30	L ventral occipital horn	1.48 ± 2.17	2.24 ± 1.3
31	R olfactory sulcal fundus	1.73 ± 1.85	2.23 ± 1.25
32*	L olfactory sulcal fundus	1.55 ± 1.34	2.29 ± 1.33
Mean		**1.33 ± 0.79**	**1.73 ± 1.30**

Table 1 - Mean anatomical fiducial localization error (AFLE) with standard deviation calculated for expert raters (MA and GG) and novice raters (AT, MJ and RC). Wilcoxon rank-sum tests were obtained for each anatomical fiducial (AFID) between expert and novice raters, with a significance threshold of 0.05/32. 6 AFIDs had significantly different AFLEs obtained by novice and expert raters. All but one AFID (05) had higher AFLEs obtained by novice raters. AC, anterior commissure; AL, anterolateral; AM, anteromedial; CC, corpus callosum; IPF, interpeduncular fossa; MB, mammillary body; LMS, lateral mesencephalic sulcus; LV, lateral ventricle; PC, posterior commissure; PMJ, pontomesenphalic junction.

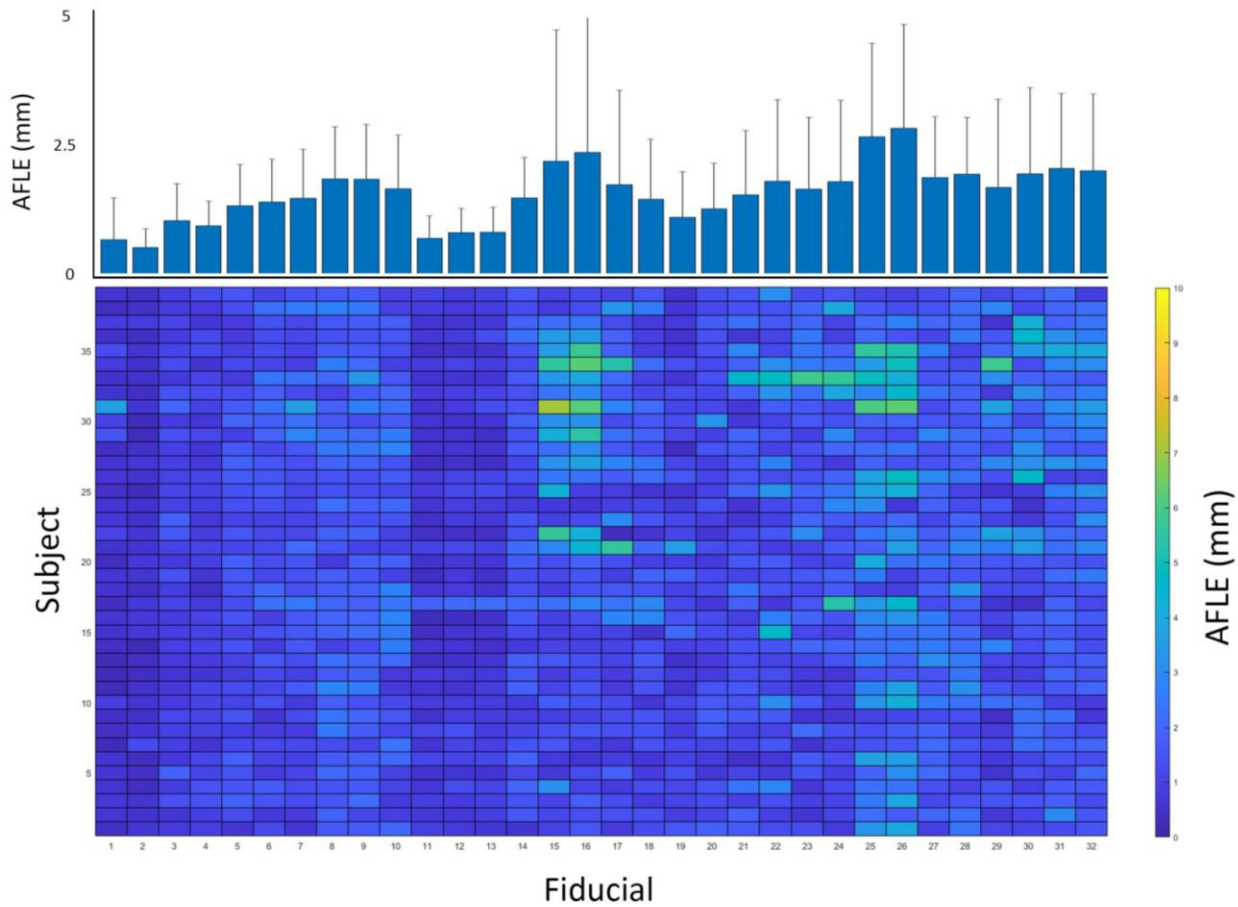

Figure 2 – Mean anatomical fiducial localization error (AFLE) for each anatomical fiducial (AFID) and subject. Bottom colormap represents mean AFLEs across all raters for each AFID and subject, illustrating the distribution of AFLEs across all subjects and AFIDs. Top bar graph represents the mean AFLEs for each AFID across all 39 subjects + standard deviation. AFIDs 1, 2 had the lowest AFLEs, while AFIDs 25 and 26 had the greatest AFLEs.

To illustrate the differences in AFLE obtained across all 39 subjects and 32 AFIDs, the mean AFLE across the 5 raters was obtained. This produced a 39 by 32 matrix which is represented as a colormap in figure 2. Each cell in the matrix represents the mean AFLE across the 5 raters for that subject and AFID. This figure illustrates the distribution of errors across the 39 subjects. Some fiducials with a high AFLE, such as AFIDs 25 and 26 demonstrate a consistently higher error across most subjects. However, other fiducials such as AFIDs 15 and 16 only demonstrate a higher AFLE in a subset of subjects.

Intraclass correlation coefficient (ICC) was calculated for each AFID between all raters, expert raters and novice raters, summarized in Table 2. The mean ICC across all AFIDs was 0.814 between all raters, 0.912 between expert raters and 0.777 between novice raters. The superior

interpeduncular fossa (AFID 05) had the lowest ICC among both expert and novice raters (0.708 and 0.544 respectively). Otherwise, novice raters also had a lower inter-rater agreement when placing AFIDs associated with the temporal horns (AFIDs 23 to 26). The left anteromedial temporal horn (AFID26) had the second lowest ICC calculated at 0.567 between novice raters, but had an ICC of 0.963 between expert raters.

Fiducial	**Fiducial Name**	**Novice ICC**	**Expert ICC**	**Total ICC**
1	AC	0.674	0.958	0.771
2	PC	0.855	0.964	0.895
3	Infracollicular sulcus	0.877	0.974	0.911
4	PMJ	0.805	0.917	0.841
5	Superior interpeduncular fossa	0.544	0.708	0.568
6	R superior LMS	0.726	0.822	0.747
7	L superior LMS	0.739	0.831	0.748
8	R inferior LMS	0.796	0.885	0.814
9	L inferior LMS	0.818	0.890	0.801
10	Culmen	0.877	0.936	0.903
11	Intermammillary sulcus	0.798	0.826	0.816
12	R MB	0.765	0.849	0.798
13	L MB	0.770	0.812	0.782
14	Pineal gland	0.756	0.835	0.757
15	R LV at AC	0.778	0.972	0.846
16	L LV at AC	0.764	0.970	0.841
17	R LV at PC	0.762	0.967	0.830
18	L LV at PC	0.872	0.971	0.908
19	Genu of CC	0.937	0.975	0.952
20	Splenium	0.886	0.979	0.922
21	R AL temporal horn	0.873	0.961	0.904
22	L AL temporal horn	0.723	0.953	0.777
23	R superior AM temporal horn	0.706	0.876	0.755
24	L superior AM temporal horn	0.637	0.914	0.661
25	R inferior AM temporal horn	0.625	0.943	0.704
26	L inferior AM temporal horn	0.567	0.963	0.649
27	R indusium griseum origin	0.829	0.931	0.866
28	L indusium griseum origin	0.836	0.866	0.853
29	R ventral occipital horn	0.924	0.990	0.947
30	L ventral occipital horn	0.926	0.991	0.946
31	R olfactory sulcal fundus	0.748	0.884	0.780
32	L olfactory sulcal fundus	0.673	0.867	0.737
Mean		**0.777**	**0.912**	**0.814**

Table 2 - Intraclass correlation coefficient (ICC) calculated for each anatomical fiducial (AFID) across 39 subjects, across all raters, expert raters (MA and GG) and novice raters (AT, MJ and RC). ICC was calculated using a two-way random effects model with a single measurement type. The mean ICC in these three groups was obtained across all AFIDs. AC, anterior commissure; AL, anterolateral; AM, anteromedial; CC, corpus callosum; IPF, interpeduncular fossa; MB, mammillary body; LMS, lateral mesencephalic sulcus; LV, lateral ventricle; PC, posterior commissure; PMJ, pontomesenphalic junction.

3.3 Analysis in MNI Space

To demonstrate the use of AFIDs in determining registration error, subject scans were linearly and non-linearly transformed to the MNI2009bAsym brain template. AFRE, the mean Euclidean distance from each transformed AFID to the mean location of that AFID placed in template space, was calculated. Linear and non-linear AFRE for each AFID is presented in Table S1. The mean non-linear AFRE is 3.34 mm ± 1.94 mm, and the linear AFRE is 6.41mm ± 2.90 mm. Wilcoxon rank-sum tests for AFRE between linear and non-linear registration with Bonferroni correction for multiple comparisons are shown in Table S1. 29 of the 32 AFIDs had a significantly greater AFRE when using linear registration compared to non-linear registration.

Figure 3 demonstrates the mean non-linear AFRE across all subjects and raters for each AFIDs. The anterior commissure (AFID01) had the smallest AFRE calculated at 1.11 mm ± 1.06 mm. The right and left ventral occipital horns (AFID29+30) had the largest AFRE at 6.81 mm ± 2.94 mm and 7.36 mm ± 3.41 mm respectively. A colormap of non-linear AFRE across the 5 raters for each subject and AFID is illustrated in figure 3. This figure demonstrates that the AFIDs with the smallest registration error (AFIDs 1, 2, 11, 12, 13, 31 and 32) were robustly decreased across most subjects. Alternatively, AFIDs 29 and 30 had large registration errors across multiple subjects.

3.4 Distance between AFIDs as a biomarker of disease

496 unique Euclidean pairwise distances were calculated between AFIDs for our PD patients (n = 39) and OASIS-1 subjects (n = 30). Figure 4 represents the differences between the mean of each pairwise distance, calculated by subtracting the mean distances in the OASIS-1 dataset from the mean distances in the PD subject dataset (therefore a positive value indicates a greater pairwise distance in the OASIS-1 subjects). Wilcoxon rank-sum tests were used, and statistically significant differences are indicated in figure 4. Significance was determined after Bonferroni correction (i.e. by obtaining a p-value $< 0.05/496$).

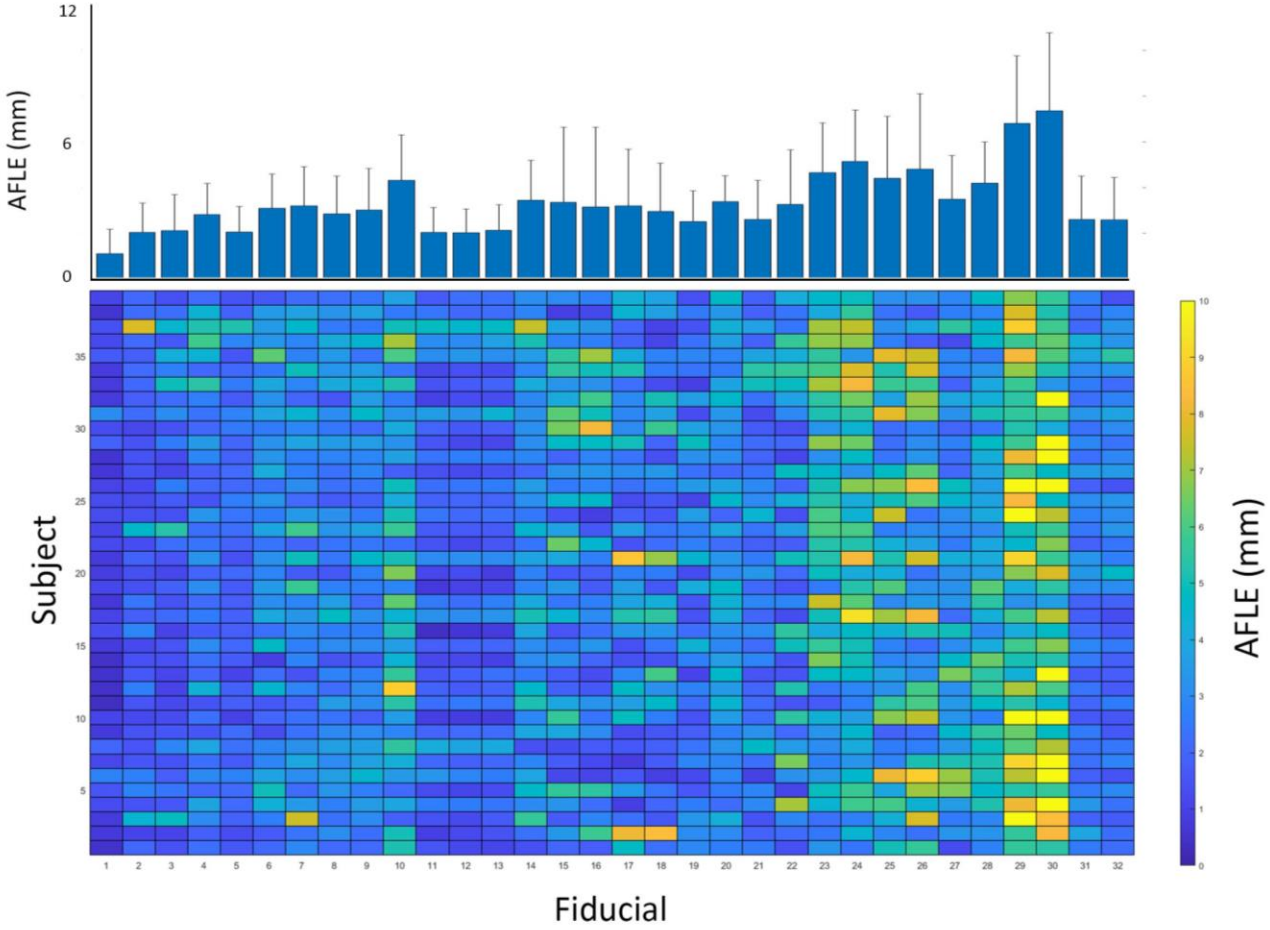

Figure 3 – Mean anatomical registration error (AFRE) for each anatomical fiducial (AFID) and subject. Bottom colormap represents mean non-linear AFREs across all raters for each AFID and subject, illustrating the distribution of non-linear AFREs across all subjects and AFIDs. Top bar graph represents the mean non-linear AFRE for each AFID across all 39 subjects + standard deviation. AFIDs 1, 2, 11, 12, 13, 31 and 32 had decreased AFREs across most subjects. AFIDs 29 and 30 had large AFREs across most subjects.

Between our PD and OASIS-1 datasets, 40 pairwise distances were statistically significantly different (Table S2). The largest distances as a percentage of the distance in the PD dataset were in the left anterolateral to left inferior anteromedial temporal horn (AFID 22 and 26; 2.82mm, 29.6%), the pontomesencephalic junction to the superior interpeduncular fossa (AFID 4 and 5; 2.47mm, 26.5%) and the infracollicular sulcus to the pineal gland (AFIDs 3 and 14; 1.90mm, 18.7%).

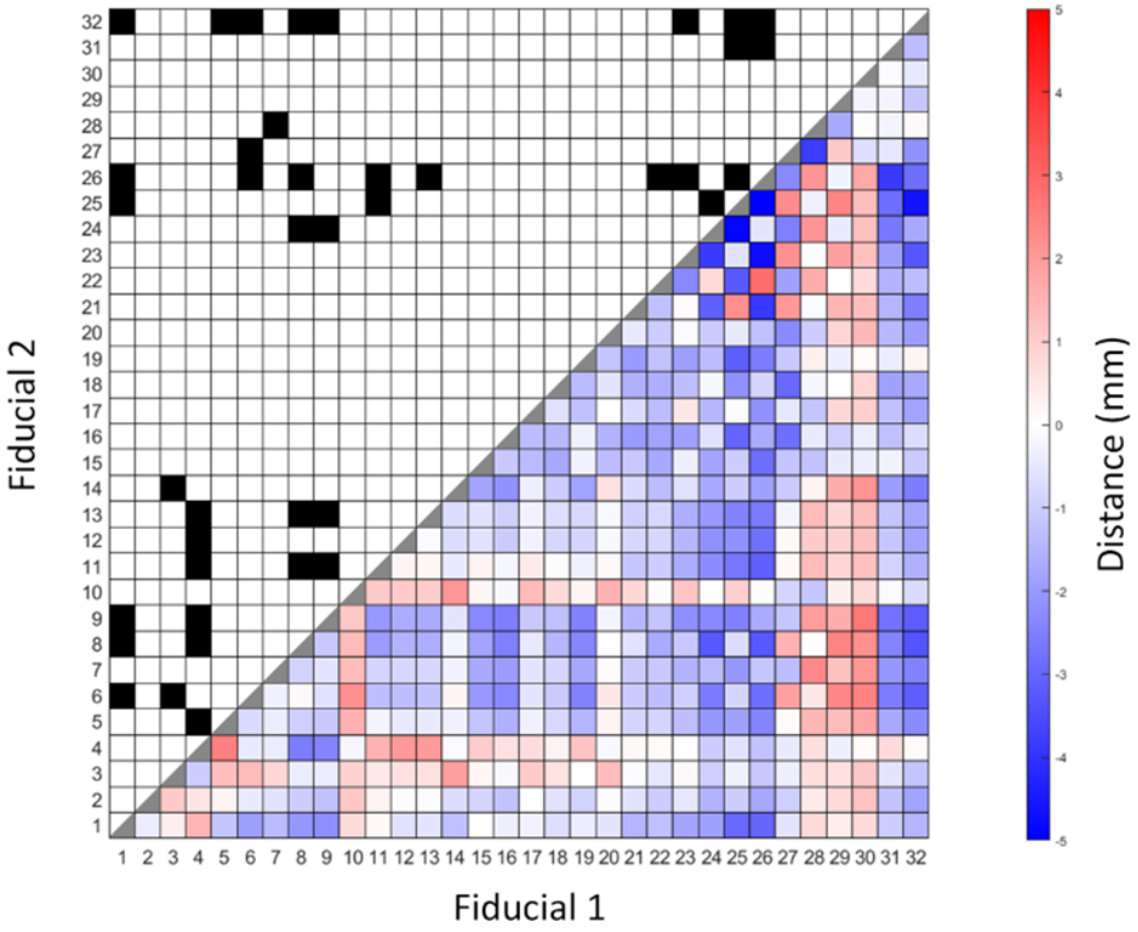

Figure 4 – Summary of mean pairwise distances between each anatomical fiducial (AFID) with significant differences. Bottom right shows heatmap representing the difference between mean pairwise distances between each AFID for OASIS-1 subjects and Parkinson's disease (PD) patients. Positive differences represent a greater pairwise distance in the OASIS-1 subjects relative to PD patients. Significant differences illustrated in top left of figure, designated by a black box. Significance is determined by Wilcoxon rank-sum tests with Bonferroni correction, with a significance threshold of 0.05/496. 40 pairwise distances reached thresholds of statistical significance between PD vs controls (see Table S2 for details).

4. Discussion

4.1 Summary

AFIDs were developed as a method to provide a point-based distance to evaluate brain image correspondence [5]. Using a set of standard templates and individual subject datasets, we previously found that AFID placement protocol was reproducible and more sensitive to local registration error compared to commonly applied voxel overlap measures. The current study sought to apply the AFIDs framework to a clinical dataset, using a set of MRI images obtained in a population of patients with PD. We first demonstrate that AFIDs can be placed with low error

by both novices and experts. We then demonstrate the use of AFIDs to evaluate the transformation of our clinical images to a standard MNI brain template. We obtained point-based measures to evaluate local registration error for each subject. Finally, by comparing patients with controls, we provide evidence that the distance between AFIDs could be a biomarker of PD that does not rely on any special imaging scans other than a volumetric structural T1 weighted MRI scan.

4.2 Accuracy of AFID placement

To investigate the accuracy of AFID placement in a clinical setting, we obtained AFLE measurements among novice and expert raters. A mean AFLE of 1.57 mm ± 1.16 mm was obtained across all raters and clinical images. Expert raters generally placed AFIDs with greater accuracy than novice raters, as evidenced by a lower mean AFLE (1.33mm ± 0.79 mm compared to 1.73 mm ± 1.30 mm) and a greater inter-rater reliability. This suggests that prior knowledge of neuroanatomy does aid in the placement of AFIDs, and that expert raters are more accurate in their use of the AFIDs framework in clinical applications reliant on accurate MRI registration.

The points with lowest AFLE across all raters were the anterior and posterior commissures (AFID01-02; 0.70 mm ± 0.78 mm and 0.55 mm ± 0.34 mm respectively), which has also been found previously [5,20]. Points associated with the ventricular system had the largest AFLE. However, we find that errors associated with these AFIDs are not always homogeneously distributed across subjects. For instance, while the right and left inferior anteromedial temporal horns (AFID25-26) had a high AFLE in most subject scans, the right and lateral ventricles at the anterior commissure (AFID15-16) only had a high AFLE in select subject scans (Figure 2). This may be a consequence of the decreased quality clinical images may be subject to, perhaps making some of these structures more difficult to resolve. Anatomical variability across subjects likely also contributes to an increased AFLE.

Overall, our findings suggest that both novice and expert raters are able to place AFIDs within a margin of error in the millimeter range. In fact, our overall AFLE is comparable with errors obtained in our previous work [5], with a mean AFLE of 1.27 mm in high resolution template scans and 1.58 mm in individual scans. We had also found that fiducials around the ventricular system had an increased error, with AFLEs in the 2-3 mm range. Therefore, despite the

heterogeneous nature of clinically obtained imaging, we provide evidence that AFIDs can be placed with millimetric accuracy.

4.3 Registration error

We evaluated the use of AFIDs to provide a point-based quantitative metric for registration of our clinical images to a standard MNI template. We used fMRIPrep to perform linear and non-linear techniques for our registration. We found a mean non-linear AFRE of 3.34 mm ± 1.94 mm and mean linear AFRE of 6.41mm ± 2.90 mm. 29 of the 32 AFIDs had significantly lower AFREs with non-linear registration compared to linear registration. Non-linear registration has evidence for improved registration accuracy; however, accuracy for these registration methods were previously assessed by calculating voxel overlap in specific ROIs [21–23]. We demonstrated that point-based accuracy measures are able to provide a more localized quantification of registration error [5]. In this paper, we provide evidence that the AFIDs protocol can be used in a clinical set of images to provide localized quantification of registration error. Furthermore, we demonstrate a decreased registration errors obtained through non-linear registration compared to linear registration.

Figure 3 demonstrates the distribution of non-linear registration errors across each AFID and subject. These findings highlight the utility of performing a point-based measure of registration error, as we are able to quantify local areas of registration error for each patient. A schematic such as this may have utility in clinical and research settings where brain image registration is required for a set of subjects, allowing for focal areas of misregistration to be quickly identified. In our particular case, we can see that AFIDs 29 and 30 had a large AFRE across most patients.

The non-linear AFRE we obtained was higher than we previously reported using MRI images from the OASIS database (1.80 mm ± 2.09 mm). Registration may have been affected by the variable quality of clinical images, baseline structural differences in PD patients, and the use of gadolinium-enhanced images for which fMRIPrep is not optimized. We elected to use fMRIPrep due to its focus on robustness rather than accuracy, and because it has been demonstrated to achieve accurate registration in the use of traditional voxel overlap measures [20]. We used fMRIPrep in our previous work to define a baseline for future refinement [5] and elected to use it in this study to aid in directly comparing our results.

4.4 Distance between AFIDs as a biomarker of disease

The AFIDs framework provided us with a novel method to investigate for biomarkers of PD. We compared pairwise distances between AFIDs in our clinical population to control subjects in the OASIS database. A difference in pairwise distance may represent relative morphometric changes (atrophy or hypertrophy) in the cerebral tissue between AFIDs. In our clinical population, we observed the largest differences between the left anterolateral to left inferior anteromedial temporal horn, the pontomesencephalic junction to the superior interpeduncular fossa and the infracollicular sulcus to the pineal gland. These distances were all smaller in our PD patient population.

Voxel-based morphometric studies in PD patients have resulted in inconsistent findings, with conflicting reports of volumetric changes in the substantia nigra (SN) and various cortical areas [24,25]. The SN has reportedly been associated with smaller volumes in patients with PD [26,27]; although other studies have either reported no difference [28] or increased volumes in PD patients [29]. Widespread cortical atrophy has been reported in PD patients with no cognitive impairment [30,31], and volumetric decrease in the hippocampus and temporoparietal cortex has been associated with cognitive decline in PD patients [32]. Our results may be in keeping with some of these findings. In particular, a decreased distance between the pontomesencephalic junction to the superior interpeduncular fossa and the infracollicular sulcus to the pineal gland may be a manifestation of a decrease in SN volume. Additionally, hippocampal atrophy may result in a decreased distance between the left anteromedial and anterolateral temporal horn. In fact, Camicioli et al. (2003) demonstrated a decrease in hippocampal volumes in patients with PD, with an association between decreased left hippocampal volumes and cognitive decline in PD [33]. Finally, these significant differences in local point-wise distance highlight the need to exercise caution when projecting findings in normal controls to patient groups as there can be differences in local brain shape.

4.5 Limitations + future directions

This study has a number of limitations. Although we demonstrate that on average expert raters had a lower AFLE than novice raters, investigating AFLE in 32 AFIDs introduces multiple comparisons which required statistical correction. However, the sample size of this study may not

provide sufficient power in demonstrating significant differences in AFLE for each AFID. Additionally, clinical imaging may be subject to variable image quality which can add subjectivity in placing AFIDs. This may result in higher AFLEs and AFREs, although these results may be more representative of the AFIDs framework applied in a clinical setting. Given that PD is a degenerative disease with minimal imaging findings, we are unable to assess the AFIDs framework in a clinical setting with patients who have mass lesions such as brain tumours.

Further work is required to automate the placement of fiducials, providing clinicians with an efficient method to characterize image registration without the subjectivity of manual AFID placement. Although we demonstrate the use of AFIDs to investigate biomarkers for patients with PD, further work is required to further investigate the robustness of our findings and to provide more data that can be used to investigate for more subtle biomarkers of neurological diseases.

4.6 Conclusion

In summary, we demonstrate that the AFIDs framework can be applied to a clinical population of PD patients with millimetric accuracy. Successful utilization of AFIDs in the context of neurosurgical planning for stereotactic procedures can provide accurate and quantitative measures of image registration, potentially improving outcomes from such procedures. Additionally, we demonstrate how distances between AFIDs could be used as a biomarker to investigate morphological differences in neurodegenerative diseases. AFIDs provide researchers with the benefit of a common, open framework that can be applied across different studies, allowing for an aggregation of clinical datasets and comparisons between various neurological conditions.

References

1. Chakravarty MM, Sadikot AF, Germann J, Bertrand G, Collins DL. Towards a validation of atlas warping techniques. Med Image Anal. 2008;12: 713–726.
2. Pirker W, Katzenschlager R. Gait disorders in adults and the elderly : A clinical guide. Wien Klin Wochenschr. 2017;129: 81–95.
3. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, et al. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011;54: 313–327.
4. Evans AC, Marrett S, Neelin P, Collins L, Worsley K, Dai W, et al. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage. 1992;1: 43–53.
5. Lau JC, Parrent AG, Demarco J, Gupta G, Kai J, Stanley OW, et al. A framework for evaluating correspondence between brain images using anatomical fiducials. Hum Brain Mapp. 2019;40: 4163–4179.
6. Poldrack RA. Region of interest analysis for fMRI. Social Cognitive and Affective Neuroscience. 2007. pp. 67–70. doi:10.1093/scan/nsm006
7. Atlas D’anatomie Stéréotaxique: Réperage Radiologique Indirect Des Noyaux Gris Centraux Des Régions Mésencéphalo-Sous-Optique Et Hypothalamique De L'hommeAtlas D'anatomie Stéréotaxique: Réperage Radiologique Indirect Des Noyaux Gris Centraux Des Régions Mésencéphalo-Sous-Optique Et Hypothalamique De L'homme. By TalairachJ., DavidM., TournouxP., CorredorH. and KvasinaT.. A volume of 294 pages, with 99 figures in addition to 91 plates. Published by Masson & Cie, 120 Boulevard Saint-Germain, Paris 6e, France, 1957. Price, paper bound, 9,000 fr.; cloth bound, 10,000 fro. Radiology. 1958. pp. 272–272. doi:10.1148/71.2.272b
8. Li Z, Zhang J-G, Ye Y, Li X. Review on Factors Affecting Targeting Accuracy of Deep Brain Stimulation Electrode Implantation between 2001 and 2015. Stereotactic and Functional Neurosurgery. 2016. pp. 351–362. doi:10.1159/000449206
9. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging. 2012. pp. 1323–1341. doi:10.1016/j.mri.2012.05.001
10. Lau JC, MacDougall KW, Arango MF, Peters TM, Parrent AG, Khan AR. Ultra-High Field Template-Assisted Target Selection for Deep Brain Stimulation Surgery. World Neurosurg. 2017;103: 531–537.

11. Xiao Y, Fonov V, Mallar Chakravarty M, Beriault S, Al Subaie F, Sadikot A, et al. A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson′s disease cohort. Data in Brief. 2017. pp. 370–379. doi:10.1016/j.dib.2017.04.013
12. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979. pp. 420–428. doi:10.1037//0033-2909.86.2.420
13. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods. 2019;16: 111–116.
14. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform. 2011;5: 13.
15. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29: 1310–1320.
16. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal. 2008;12: 26–41.
17. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging. 2001;20: 45–57.
18. Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009. p. S102. doi:10.1016/s1053-8119(09)70884-5
19. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. Front Neuroinform. 2014;8: 14.
20. Liu Y, Dawant BM. Automatic Localization of the Anterior Commissure, Posterior Commissure, and Midsagittal Plane in MRI Scans using Regression Forests. IEEE Journal of Biomedical and Health Informatics. 2015. pp. 1362–1374. doi:10.1109/jbhi.2015.2428672
21. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009;46: 786–802.

22. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, et al. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010;98: 278–284.
23. Chakravarty MM, Sadikot AF, Germann J, Hellier P, Bertrand G, Collins DL. Comparison of piece-wise linear, linear, and nonlinear atlas-to-patient warping techniques: analysis of the labeling of subcortical nuclei for functional neurosurgical applications. Hum Brain Mapp. 2009;30: 3574–3595.
24. Heim B, Krismer F, De Marzi R, Seppi K. Magnetic resonance imaging for the diagnosis of Parkinson's disease. Journal of Neural Transmission. 2017. pp. 915–964. doi:10.1007/s00702-017-1717-8
25. Pyatigorskaya N, Gallea C, Garcia-Lorenzo D, Vidailhet M, Lehericy S. A review of the use of magnetic resonance imaging in Parkinson's disease. Therapeutic Advances in Neurological Disorders. 2014. pp. 206–220. doi:10.1177/1756285613511507
26. Menke RA, Scholz J, Miller KL, Deoni S, Jbabdi S, Matthews PM, et al. MRI characteristics of the substantia nigra in Parkinson's disease: A combined quantitative T1 and DTI study. NeuroImage. 2009. pp. 435–441. doi:10.1016/j.neuroimage.2009.05.017
27. Minati L, Grisoli M, Carella F, De Simone T, Bruzzone MG, Savoiardo M. Imaging degeneration of the substantia nigra in Parkinson disease with inversion-recovery MR imaging. AJNR Am J Neuroradiol. 2007;28: 309–313.
28. Péran P, Cherubini A, Assogna F, Piras F, Quattrocchi C, Peppe A, et al. Magnetic resonance imaging markers of Parkinson's disease nigrostriatal signature. Brain. 2010. pp. 3423–3433. doi:10.1093/brain/awq212
29. Cho Z-H, Oh S-H, Kim J-M, Park S-Y, Kwon D-H, Jeong H-J, et al. Direct visualization of Parkinson's disease by in vivo human brain imaging using 7.0T magnetic resonance imaging. Movement Disorders. 2011. pp. 713–718. doi:10.1002/mds.23465
30. Jubault T, Gagnon J-F, Karama S, Ptito A, Lafontaine A-L, Evans AC, et al. Patterns of cortical thickness and surface area in early Parkinson's disease. NeuroImage. 2011. pp. 462–467. doi:10.1016/j.neuroimage.2010.12.043
31. Lyoo CH, Ryu YH, Lee MS. Topographical distribution of cerebral cortical thinning in patients with mild Parkinson's disease without dementia. Movement Disorders. 2010. pp. 496–499. doi:10.1002/mds.22975

32. Weintraub D, Dietz N, Duda JE, Wolk DA, Doshi J, Xie SX, et al. Alzheimer's disease pattern of brain atrophy predicts cognitive decline in Parkinson's disease. Brain. 2012. pp. 170–180. doi:10.1093/brain/awr277
33. Camicioli R, Milar Moore M, Kinney A, Corbridge E, Glassberg K, Kaye JA. Parkinson's disease is associated with hippocampal atrophy. Movement Disorders. 2003. pp. 784–790. doi:10.1002/mds.10444

Supplementary

Fiducial	Fiducial Name	Linear AFRE (mm)	Non-linear AFRE (mm)
1*	AC	4.42 ± 2.50	1.11 ± 1.06
2*	PC	4.17 ± 1.97	2.05 ± 1.27
3*	Infracollicular sulcus	4.21 ± 2.23	2.12 ± 1.56
4*	PMJ	6.71 ± 2.45	2.82 ± 1.36
5*	Superior interpeduncular fossa	4.03 ± 1.77	2.06 ± 1.12
6*	R superior LMS	4.39 ± 1.94	3.10 ± 1.50
7*	L superior LMS	4.70 ± 2.23	3.21 ± 1.70
8*	R inferior LMS	4.29 ± 1.89	2.86 ± 1.65
9*	L inferior LMS	4.67 ± 1.87	3.01 ± 1.82
10*	Culmen	6.52 ± 2.91	4.32 ± 1.98
11*	Intermammillary sulcus	4.72 ± 2.43	2.05 ± 1.09
12*	R MB	4.89 ± 2.45	2.02 ± 1.05
13*	L MB	5.02 ± 2.39	2.14 ± 1.11
14	Pineal gland	4.16 ± 2.08	3.45 ± 1.74
15*	R LV at AC	8.66 ± 4.47	3.36 ± 3.28
16*	L LV at AC	8.99 ± 4.62	3.16 ± 3.49
17*	R LV at PC	7.70 ± 4.07	3.21 ± 2.47
18*	L LV at PC	8.08 ± 4.27	2.96 ± 2.10
19*	Genu of CC	6.97 ± 3.17	2.51 ± 1.35
20*	Splenium	6.34 ± 2.79	3.39 ± 1.13
21*	R AL temporal horn	7.07 ± 2.45	2.60 ± 1.72
22*	L AL temporal horn	7.84 ± 2.92	3.26 ± 2.38
23*	R superior AM temporal horn	6.23 ± 2.59	4.66 ± 2.17
24*	L superior AM temporal horn	7.53 ± 2.90	5.14 ± 2.24
25*	R inferior AM temporal horn	8.7 ± 3.58	4.41 ± 2.71
26*	L inferior AM temporal horn	9.06 ± 3.86	4.81 ± 3.30
27*	R indusium griseum origin	6.57 ± 2.87	3.5 ± 1.91
28*	L indusium griseum origin	7.58 ± 3.13	4.2 ± 1.80
29	R ventral occipital horn	9.37 ± 3.84	6.81 ± 2.94
30	L ventral occipital horn	9.02 ± 4.32	7.36 ± 3.41
31*	R olfactory sulcal fundus	6.02 ± 2.85	2.6 ± 1.89
32*	L olfactory sulcal fundus	6.44 ± 2.93	2.58 ± 1.85
Mean		**6.41 ± 2.90**	**3.34 ± 1.94**

Table S1 – Mean anatomical fiducial registration error (AFRE) with standard deviation obtained with linear and non-linear registration of clinical images to MNI space using fMRIPrep. Wilcoxon rank-sum tests were obtained for each anatomical fiducial (AFID) between linear and non-linear AFRE, with a significance threshold of 0.05/32. AFRE was significantly greater in all AFIDs except for AFIDs 14, 29 and 30. AC, anterior commissure; AL, anterolateral; AM, anteromedial; CC, corpus callosum; IPF, interpeduncular fossa; MB, mammillary body; LMS, lateral mesencephalic sulcus; LV, lateral ventricle; PC, posterior commissure; PMJ, pontomesenphalic junction.

Fiducial 1	Fiducial 2	PD Distance	Oasis-1 Distance	Pairwise Difference	Percent Difference
L AL temporal horn	L inferior AM temporal horn	9.6 ± 2.5	12.4 ± 1.7	2.82	29.56
PMJ	Superior interpeduncular fossa	9.3 ± 1.2	11.8 ± 1.3	2.48	26.52
Infracollicular sulcus	Pineal gland	10.2 ± 1.1	12.1 ± 1.2	1.90	18.70
PMJ	R inferior LMS	14.8 ± 1.1	12.2 ± 1.0	-2.56	-17.30
PMJ	L inferior LMS	14.6 ± 1.0	12.2 ± 0.8	-2.41	-16.44
PMJ	R MB	13.4 ± 1.2	15.4 ± 1.4	2.01	15.07
PMJ	L MB	13.3 ± 1.2	15.2 ± 1.3	1.98	14.96
L superior LMS	L indusium griseum origin	17.2 ± 1.9	19.5 ± 2.3	2.33	13.52
R inferior AM temporal horn	L inferior AM temporal horn	46.7 ± 3.7	40.7 ± 4.3	-6.02	-12.88
PMJ	Intermammillary sulcus	13.0 ± 1.2	14.6 ± 1.4	1.51	11.56
Intermammillary sulcus	L inferior AM temporal horn	27.2 ± 2.2	24.0 ± 2.7	-3.12	-11.48
R superior LMS	R indusium griseum origin	16.6 ± 1.9	18.5 ± 1.7	1.86	11.16
L superior AM temporal horn	R inferior AM temporal horn	45.2 ± 3.2	40.3 ± 3.6	-4.90	-10.84
R superior AM temporal horn	L inferior AM temporal horn	44.6 ± 3.1	39.8 ± 4.0	-4.74	-10.63
L inferior LMS	L superior AM temporal horn	22.3 ± 1.8	19.9 ± 2.0	-2.32	-10.42
R inferior AM temporal horn	L olfactory sulcal fundus	44.7 ± 2.5	40.1 ± 2.6	-4.62	-10.35
L MB	L inferior AM temporal horn	25.2 ± 2.1	22.6 ± 2.4	-2.60	-10.31
Intermammillary sulcus	R inferior AM temporal horn	27.0 ± 2.4	24.3 ± 2.5	-2.69	-9.98
L inferior AM temporal horn	L olfactory sulcal fundus	28.5 ± 2.1	25.6 ± 2.3	-2.84	-9.98
R inferior AM temporal horn	R olfactory sulcal fundus	29.1 ± 1.8	26.3 ± 2.3	-2.83	-9.70
AC	L inferior AM temporal horn	32.7 ± 2.1	29.7 ± 2.4	-3.01	-9.20
Infracollicular sulcus	R superior LMS	14.5 ± 0.9	15.9 ± 0.9	1.34	9.20
AC	R inferior AM temporal horn	32.5 ± 2.3	29.6 ± 2.3	-2.94	-9.05
R inferior LMS	L superior AM temporal horn	37.1 ± 2.3	33.8 ± 2.5	-3.27	-8.82
L inferior AM temporal horn	R olfactory sulcal fundus	44.5 ± 2.7	40.7 ± 3.1	-3.83	-8.60
L inferior LMS	Intermammillary sulcus	25.2 ± 1.5	23.2 ± 1.4	-2.03	-8.03
R superior AM temporal horn	L olfactory sulcal fundus	42.6 ± 2.4	39.3 ± 2.7	-3.30	-7.73
R inferior LMS	Intermammillary sulcus	25.3 ± 1.6	23.3 ± 1.7	-1.95	-7.69
R inferior LMS	L inferior AM temporal horn	42.7 ± 2.6	39.4 ± 2.5	-3.28	-7.67
AC	L olfactory sulcal fundus	20.8 ± 1.2	19.3 ± 1.3	-1.47	-7.10
L inferior LMS	L MB	24.3 ± 1.5	22.7 ± 1.3	-1.65	-6.79
Superior interpeduncular fossa	L olfactory sulcal fundus	34.1 ± 1.8	31.8 ± 2.1	-2.30	-6.76
L inferior LMS	L olfactory sulcal fundus	47.6 ± 2.6	44.4 ± 2.8	-3.17	-6.66
R superior LMS	L inferior AM temporal horn	44.1 ± 2.8	41.3 ± 2.5	-2.84	-6.44
R inferior LMS	L olfactory sulcal fundus	53.2 ± 2.6	49.9 ± 2.5	-3.33	-6.26
R superior LMS	L olfactory sulcal fundus	50.2 ± 2.5	47.1 ± 2.3	-3.13	-6.23
R inferior LMS	L MB	26.4 ± 1.6	24.8 ± 1.6	-1.59	-6.03
AC	R superior LMS	32.0 ± 1.9	30.1 ± 1.2	-1.90	-5.95
AC	L inferior LMS	37.3 ± 2.0	35.1 ± 1.8	-2.20	-5.89
AC	R inferior LMS	37.4 ± 2.1	35.3 ± 1.9	-2.05	-5.49

Table S2 - List of all mean pairwise distances (mm) ± standard deviation that are significantly different between OASIS-1 subjects and Parkinson's disease patients. Significance is determined by Wilcoxon rank-sum tests with Bonferroni correction, significance threshold of 0.05/496. AC, anterior commissure; AL, anterolateral; AM, anteromedial; CC, corpus callosum; IPF, interpeduncular fossa; MB, mammillary body; LMS, lateral mesencephalic sulcus; LV, lateral ventricle; PC, posterior commissure; PMJ, pontomesenphalic junction.