

Visual versus automated analysis of [I-123]FP-CIT SPECT scans in parkinsonism

Elina Mäkinen BM^{1, 2, 4}, Juho Joutsa MD^{1, 2, 3}, Jarkko Johansson MSc³, Maija Mäki MD⁴, Marko Seppänen MD⁴, Valtteri Kaasinen MD^{1, 2, 3}

1. Department of Neurology, University of Turku, Turku, Finland
2. Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland
3. Turku PET Centre, University of Turku, Turku, Finland
4. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

Corresponding author:

Dr. Elina Mäkinen

Department of Neurology,

Turku University Hospital

POB 52, FIN-20521

Turku, Finland

Tel. +358-2-3130000

Email: elmama@utu.fi

Supported in part by the Finnish Parkinson Foundation, Turku University Foundation and Turku University Hospital (ERVA-funds). This was an independent academic study. The authors declare no conflicts of interest.

ABSTRACT

Purpose The clinical evaluation of dopamine transporter (DAT) SPECT scans typically relies on visual analysis in combination with an automated semi-quantitative method. The interpretation of the results may be difficult in cases that show disagreement between the two methods on the borderline of abnormality. The frequency and clinical characteristics of such cases are unclear.

Materials and Methods Automated semi-quantitative analyses and independent visual analyses by two experienced nuclear medicine physicians and four inexperienced raters were performed for 120 patients with clinically uncertain parkinsonism scanned with brain [I-123]FP-CIT SPECT. Agreement was evaluated with kappa statistics. The clinical characteristics of patients who had discrepant findings between the two analysis methods were investigated.

Results The expert raters outperformed non-experts in terms of agreement between visual and automated analyses ($\kappa=0.66, 0.72$ vs. $0.23-0.54$) and between raters ($\kappa=0.81$ vs. $0.44-0.63$). Twelve patients showed discrepant findings between the visual and automated analyses. These patients were older compared to other patients ($p=0.023$), had 17.6% lower mean striatal tracer binding compared to normal scans ($p=0.003$) and 62.7% higher compared to abnormal scans ($p<0.0001$). After a minimum of 4.5 years of clinical follow-up, none of these patients developed neurodegenerative parkinsonism.

Conclusions Clinical DAT SPECT scans show discrepancies between visual and automated analyses in 10% of cases. The patients with discrepant findings are older, show normal to slightly abnormal tracer binding, and importantly, do not develop neurodegenerative parkinsonism syndromes. Visual analyses by experienced raters are

reliable, but the diagnostic accuracy in discrepant cases can be improved by an automated method.

Keywords: SPECT, dopamine transporter, parkinsonism, clinical, analysis

INTRODUCTION

Brain dopamine transporter (DAT) imaging with [I-123]FP-CIT SPECT is widely used in the differential diagnosis of parkinsonian syndromes (Kägi et al. 2010; Brooks 2012; Kupsch et al. 2012; Varrone et al. 2013). [I-123]FP-CIT is a tracer that binds to DAT located on the presynaptic terminal cell membranes of dopaminergic neurons, and thereby, the tracer binding may identify the loss of dopamine, DAT expression or the number of functional dopaminergic neurons (Kupsch et al. 2012; Ba and Martin 2015). The method can be used to differentiate essential tremor (ET) and Alzheimer's disease from neurodegenerative dopaminergic parkinsonian syndromes, including Parkinson's disease (PD), progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and dementia with Lewy bodies (LBD) (Kägi et al. 2010; Brooks 2012; Ba and Martin 2015). In addition, it appears that secondary parkinsonism such as drug-induced parkinsonism (DIP) and vascular parkinsonism which do not generally affect nigrostriatal projections, can be differentiated from neurodegenerative parkinsonism using DAT imaging (Brooks 2012; Ba and Martin 2015).

Commonly, an expert rater, often a nuclear medicine physician, performs the initial visual interpretation of DAT SPECT cans. The visual approach is generally accepted and is often the preferred method of analyzing DAT scans while there are also several

automated semi-quantitative methods available (Badiavas et al. 2011). Most of these methods take advantage of binding ratios that are calculated by comparing tracer uptake in regions of interests (ROIs), most commonly the whole striatum, the caudate nucleus or the anterior/posterior putamen, that show specific binding to reference areas free of specific binding (occipital or cerebellar cortex) (Joutsa et al. 2015). A commonly used quantitative index is the specific binding ratio (SBR), the ratio of specific to nonspecific uptake, which is presumably related to the density of the presynaptic terminals or DAT (Badiavas et al. 2011; Varrone et al. 2013). The advantage of the automated method is the lack of subjectivity or rater-induced bias (Koch et al. 2005; Pencharz et al. 2014) compared to the visual analysis that is subjective and possibly training-dependent (Scherfler and Nocker 2009; Kahraman et al. 2012). On the other hand, the disadvantage of the automated method could be the simplistic mathematical approach that can be vulnerable to systematic artefacts or errors (Scherfler and Nocker 2009; Buchert et al. 2015), and does not allow room for clinical interpretation in complex or atypical cases.

Some current protocols for the clinical evaluation of DAT SPECT scans emphasize the parallel and equal roles of visual and automated analyses. However, the extent to which automated semi-quantitative analysis should be used (not at all, equal or supplementary to the visual analysis, or even as a primary method of analysis) remains unclear (Ottaviani et al. 2006; Filippi et al. 2008; Suárez-Piñera et al. 2011; Buchert et al. 2015). In a recent study by Albert et al., a combination of visual and semi-quantitative analyses seemed to lead to the best diagnostic accuracy in clinical follow-up (Albert et al. 2016). In another study by Söderlund et al., the visual analyses between experts were very consistent, but a combined approach of visual and automated analyses of tracer binding

created better reproducibility. The study also suggested that observers with less experience tended to over-report scans as abnormal (Söderlund et al. 2013). Buchert et al. reported that visual evaluation of a standardized slab view display of DAT scans showed agreement with semi-quantitative analysis in as many as 90% of cases. However, in their study, readers without any experience showed similar performance compared to experienced readers in the visual analysis (Buchert et al. 2015). On the basis of these studies, it remains unclear whether training or experience in visual analysis provides any benefit in the diagnostic accuracy. Furthermore, the clinical characteristics of cases that exhibit disagreement between visual and automated analyses have not been studied.

If tracer binding appears visually abnormal but the interpretation of the automated analysis is normal – or *vice versa* – how should these cases be interpreted? The issue is of high importance because false interpretations can lead to false diagnoses and unjustified treatments (Suárez-Piñera et al. 2011; Hauser and Grosset 2012; Seibyl et al. 2014). It would also be cost-effective if accurate clinical analyses could be performed by non-experts. This study aimed to investigate the frequency and the characteristics of DAT SPECT imaging cases with discrepant findings in clinical visual and automated analyses. We further studied if visual scan analysis could be performed by non-experts by comparing nuclear medicine specialists to raters without prior experience.

METHODS

Subjects

The study sample included 120 patients who had undergone a brain [I-123]FP-CIT SPECT due to clinically uncertain parkinsonism. We initially identified 489 patients with appropriate clinical data and [I-123]FP-CIT SPECT images from our database (scanned in 2007-2012). Age at scan, sex, scan date, scanner, duration of motor symptoms before scanning, predominant side of the motor symptom, presence of tremor and the post-scan clinical diagnosis in follow-up were available for every patient (Kaasinen et al. 2014). The 489 scans were visually categorized into one of four categories of striatal tracer binding by one of the investigators (1=Normal, 2=Slightly abnormal, 3=Abnormal and 4=Clearly abnormal) (Benamer et al. 2000; Staff et al. 2009; Suárez-Piñera et al. 2011; Kupsch et al. 2012). The uptake was categorized as normal (grade 1) when it was symmetrical and observed in all striatal nuclei in both hemispheres. The uptake was categorized as slightly abnormal (grade 2) if there was a visually detectable reduction in putaminal uptake in one or both hemisphere(s) in association with normal or almost normal caudate uptake in both hemispheres. The uptake was categorized as abnormal (grade 3) when there was clear bilateral reduction (even if asymmetric) in putaminal uptake, with mostly preserved caudate nuclei in both hemispheres. The uptake was categorized as clearly abnormal (grade 4) with clear bilateral reduction in putaminal uptake in association with a binding loss in one or both caudate nuclei. Next, the category distributions were calculated and the data was sorted up randomly. Different samples of 120 patients were extracted, and the sample with the most representative distribution of scans for each of the four categories was selected as the final study sample without clinical information. The sample of 120 scans was considered to be sufficiently small but statistically acceptable for reliable and consistent

individual visual ratings. The most common final diagnoses were PD (n=49), essential tremor (n=16) and medication-induced parkinsonism (n=8). The demographics of the study sample are presented in Table 1. The study was approved by the ethical committee of the local hospital district and was conducted according to the principles of Helsinki.

Scanning and image preprocessing

Patients were imaged on either a GE Infinia II Hawkeye SPECT/CT scanner (GE Medical Systems, Milwaukee, WI) or a Picker Irix gamma camera (Picker International, Uniontown, OH). The data were acquired using LEHR collimators, the images were reconstructed using the ordered subsets expectation algorithm of Hybrid Recon Neurology software (version 1.0.15, Hermes Medical Solutions AB, Stockholm, Sweden), and a 3-D Gaussian post-filter with 0.7 cm full-width at half-maximum, as previously described (Kaasinen et al. 2014).

Automated semi-quantitative analysis

All scans were analyzed using BRASS automated analysis software (ROI-method) (version 3.6; Hermes Medical Solutions, Stockholm, Sweden). Scanner-specific corrections were used for the BRASS analyses (Kaasinen et al. 2014). SBRs for six regions; right and left caudate, right and left anterior putamen, and right and left posterior putamen, were then calculated using the occipital cortex as the reference region: $SBR = (ROI_{caudate \text{ or } putamen} - ROI_{occipital})/ROI_{occipital}$. Semi-quantitative uptake was defined as abnormal if it was more than two standard deviations below the reference mean in any of the six analyzed regions (Varrone et al. 2013). Out of the 120 patients, there were four abnormal cases that had caudate binding defects without parallel losses in putaminal binding.

Image layouts for visual analysis

In the initial phase of image selection, Vinci software (version 3.x, Max Planck Institute, Cologne, Germany) was used to generate horizontal views of the scans from 489 patients. Twelve striatal slices with a thickness of 4.7 mm were used in the visualizations using the Speckle rainbow color scale (range = SBR 0.0-5.5). Each image included four larger and eight smaller scan slices using the neurological convention (right side of the image = right side of the patient). After the selection of the final study sample of 120 patients, four representative images of each category were selected from the original data of patients who were not included in these 120 scans. The layout of the images and examples of each category are shown in Fig. 1.

Visual analyses

The 120 images were individually analyzed by six independent raters. Two of the raters were experienced nuclear medicine physicians with extensive experience in analyzing [I-123]FP-CIT scans (>10 years). The other raters were non-experts, including two registered nuclear medicine nurses with some knowledge of [I-123]FP-CIT SPECT imaging but no knowledge of SPECT image analysis and two non-medical laymen with no previous knowledge of brain imaging. The raters were asked to categorize the 120 patients to one of the four categories of striatal tracer binding (Fig. 1) and specify possible inter-hemispheric differences. The visual analysis of the images was performed by visual interpretation of the images with the help of examples (Fig. 1), category definitions and the following clinical patient information: age at scan, sex, duration of motor symptoms at scanning and the predominant side of motor symptoms.

All raters received both written and oral instructions on how to analyze the images, and they were blinded to the results of the automated analysis and each other's ratings.

Statistical analyses

The agreement between dichotomous (normal/abnormal) visual analyses and the automated analysis, as well the inter-rater agreement were calculated by using kappa statistics (Cohen's unweighted κ) (Cohen 1960). The strength of agreement was defined based on κ values: less than 0.20 was considered to be poor agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 good agreement and 0.81-1.00 excellent agreement (Altman 1991). The differences in the agreement between experts and non-experts were compared using Z-tests. Patient characteristics between discrepant and non-discrepant cases were compared with one-way ANOVA with *post hoc* Bonferroni correction and chi-square tests. The difference between cameras in discrepant cases was compared with chi-square test. The level of statistical significance was set at $p < 0.05$.

RESULTS

Visual vs. automated analysis

Of the total 120 scans, 26 (21.7%) were categorized as normal by automated analysis and both experts, and eight scans (6.7%) were categorized as normal by automated analysis and either one of the experts (non-discrepant normal scans, n=34, 28.3%). Similarly, 73 scans (60.8%) were categorized as abnormal by automated analysis and both experts, and one scan (0.8%) was categorized as abnormal by automated analysis and one of the experts (non-discrepant abnormal scans, n=74, 61.7%). Twelve (10%) of the cases showed discrepancy between expert visual and automated analyses (discrepant scans). Of these 12 scans, nine (7.5%) were categorized as normal by automated analysis and abnormal by both experts, and three (2.5%) were categorized as abnormal by automated analysis and normal by both experts. The main demographic characteristics of the patients are presented in Table 1.

Discrepant scans had 17.6% lower mean striatal SBRs compared to normal scans ($p=0.003$) and 62.7% higher SBRs compared to abnormal scans ($p<0.0001$). Patients with discrepant scans were older compared to patients with normal scans (72.6 vs. 62.4 years, $p=0.023$) without significant age differences compared to patients with abnormal scans ($p=0.33$) (Table 1, Fig. 2). Discrepant cases did not differ from other cases in terms of gender distribution or in the predominant side of motor symptoms (Table 1). Of the 74 patients scanned with the GE Infinia scanner, seven (9.5%) had discrepant results; respectively, discrepant results were observed in 5/46 (10.9%) patients scanned with the Picker Irix scanner. There were no significant differences in the prevalences of discrepant cases between cameras ($p=0.80$).

The clinical characteristics of the 12 patients with discrepant results are presented in Table 2, and their SPECT scans are displayed in Fig. 3. Out of the nine patients who were visually categorized as abnormal while automated analysis categorized them as normal, eight (88.9%) were aged 70 or older at the time of imaging, six (66.7%) had cognitive defect at the time of imaging or developed dementia within five years after imaging, and four (44.4%) received a final diagnosis of drug-induced parkinsonism (DIP). The three discrepant cases that were categorized as abnormal by the automated method were 2.0-2.5 standard deviations below the reference SBR mean in at least one striatal region. For patient number 10 the reduced region was the left anterior putamen (-2.26), for patient number 11 the right anterior putamen (-2.36), the left anterior putamen (-2.34), the left posterior putamen (-2.46) and the left caudatus (-2.21), and for patient number 12 the right anterior putamen (-2.01), the left anterior putamen (-2.31), the left posterior putamen (-2.08), the right caudatus (-2.03) and the left caudatus (-2.13). Patients 11 and 12, who showed uniform slight reductions in binding over the whole striatum uni- or bilaterally, were later diagnosed as having essential tremor.

Effect of expertise in visual analysis

Expert raters showed good agreement between visual and automated analyses [$\kappa=0.66$ (95% CI 0.51-0.80) and $\kappa=0.72$ (95% CI 0.58-0.85)] whereas the agreement was fair to moderate in non-experts. Experts also showed clearly superior inter-rater agreement in dichotomous visual analysis [$\kappa=0.81$ (95% CI 0.70-0.93)] compared to non-experts. (Table 3). Expert raters spent more time on the analyses compared to non-experts (60-90 s vs. 30-45 s per scan, respectively).

DISCUSSION

One out of every ten clinical DAT SPECT scans in our data showed discrepancy between visual and automated analyses. Most of these cases were visually interpreted to show slightly abnormal striatal tracer binding whereas the automated analysis indicated normal binding. Interestingly, none of these patients developed degenerative dopaminergic diseases, such as PD, after a minimum of 4.5 years of clinical follow-up. From a practical clinical point of view, this suggests that scans on the borderline of abnormality with discrepant imaging findings should probably be interpreted as normal. One should also note that the symptom duration in discrepant cases tended to be longer than in other patients. This supports the benign nature of discrepancies as longer motor symptom durations at the time of imaging appear to be associated with a higher probability of a normal scan (Jaakkola et al. 2016). In addition, our results suggest that the accuracy of visual interpretation of DAT SPECT scans is highly dependent on the level of training as the performance of experienced nuclear medicine physicians in the visual analysis was clearly superior to that of non-experts, if we consider the automated analysis as golden standard.

Patients with discrepant imaging findings were older than patients with normal findings in striatal tracer binding. This implies that the visual interpretation was affected by age-related dopaminergic changes. Particularly patients who were interpreted as abnormal in the visual analysis but normal by the automated method were older. Many of these patients also had, or later developed, cognitive problems or dementia, which could be another factor misleading the interpretation. Furthermore, several patients received a final diagnosis of DIP. Therefore, combinations of older age, cognitive defect and anti-psychotic medication would appear to cause interpretation difficulties in DAT imaging.

Possible slight age-related decreases in tracer binding (Varrone et al. 2013; Kaasinen et al. 2015; Albert et al. 2016), in combination with developing dementia, could possibly visually mimic early PD and affect striatal visual shape recognition. While possible age-related defects seem to be difficult to detect in the visual analysis, the semi-quantitative method can be more helpful as it employs age-corrected reference values (Varrone et al. 2013). Furthermore, DIP occurs commonly in patients receiving neuroleptics (Ba and Martin 2015), and it has been reported to be more common among female and elderly patients (Kägi et al. 2010). DIP is generally thought to be associated with normal DAT imaging findings (Kägi et al. 2010; Brooks 2012; Ba and Martin 2015), although some patients might show slight degeneration of the nigrostriatal system (Lorberboym et al. 2006). In the present study, several DIP cases appeared to visually show slight reductions of tracer binding in the putamen (cases 3-6 in Fig. 3). Because dementia and DIP are known to be indications where radionuclide imaging is deemed to be useful, the discrepancies in the interpretations of these cases are particularly alarming.

Three discrepant patients were (falsely) categorized as abnormal with the automated semi-quantitative method. It is of interest to note that two of these patients received a final diagnosis of essential tremor. These patients appeared to have uniform slight reductions in tracer binding over the whole striatum in the automated analysis. While the official indication of DAT SPECT is the differentiation between PD and ET (Scherfler and Nocker 2009; Brooks 2012; Kupsch et al. 2012), it has also been suggested that ET, as a heterogeneous disorder, could be a risk factor for PD (Kägi et al. 2010; Jiménez-Jiménez et al. 2012; Thenganatt and Jankovic 2016). Consistent with the present results, semi-quantitative analysis has previously been reported to show

slightly reduced tracer binding in ET compared to healthy controls (Gerasimou et al. 2012; Waln et al. 2015). Moreover, it seems that ET could affect all parts of the striatum, unlike PD which induces asymmetrical reductions that are greatest in the putamen (Gerasimou et al. 2012).

The accuracy of visual analysis can indeed be improved by training, as the experts performed better than non-experts both in the agreement with the automated analysis and in the inter-rater agreement. Non-experts showed wide divergence in the visual evaluations, and they particularly tended to categorize normal cases incorrectly as abnormal, similarly to the less experienced readers in the study by Söderlund et al. (Söderlund et al. 2013). Only 7-21% of scans were categorized as normal by non-experts whereas both experts categorized 28% and the automated analysis categorized 36% of scans as normal. It thus seems evident that the visual interpretation of DAT SPECT scans should be performed by trained readers, ideally with years of experience. It is noteworthy that even without the support of an automated method, the visual evaluations by experts were consistent and reliable. The great majority of cases can thus be correctly diagnosed by visual analysis alone, whereas the value of the semi-quantitative analysis is emphasized in borderline cases that appear to show mild uptake defects. It should also be noted that the prevalence of discrepant cases is dependent on the proportion of borderline cases in the sample. In the recent study by Seibyl et al., excellent accuracy and inter-rater agreement were reported for expert image readers using only the visual analysis of [I-123]FP-CIT SPECT scans (Seibyl et al. 2014). Very good inter-rater agreement between experts in the visual analyses was also reported in an earlier study by Hauser et al. (Hauser and Grosset 2012). Nevertheless, the combined use of semi-quantification with the visual analysis is recommended by the European

Association of Nuclear Medicine (EANM) (Scherfler and Nocker 2009; Buchert et al. 2015), and this combined approach was superior in the studies by Söderlund et al. and Albert et al. (Söderlund et al. 2013; Albert et al. 2016). The findings by Albert et al. are consistent with the results of the present study when suggesting that the use of the semi-quantitative method is important especially in scans that are visually inconclusive or show dopaminergic striatal reductions (Albert et al. 2016). Furthermore, the vital role of the semi-quantitative methods was further underlined by Filippi et al. on PD diagnostics (Filippi et al. 2008), and in the study by Waln et al., in which semi-automated quantification showed sensitivity even in the differentiation between healthy controls and patients with ET (Waln et al. 2015).

In conclusion, our results suggest that although visual assessment of CIT SPECT images is challenging, it is generally reliable when performed by experienced observers. However, there is discrepancy between visual and automated analyses in 10% of cases. Patients with discrepant interpretations do not seem to develop neurodegenerative parkinsonism syndromes, highlighting the need for cautious interpretation in such cases. Diagnostic accuracy in these cases can be improved with the assessment using an automated semi-quantitative analysis method.

Acknowledgements

We thank the personnel of the Department of Nuclear Medicine in Turku University Hospital and Dr. Tero Vahlberg for assistance with the statistical analyses.

Conflicts of Interest and Sources of Funding

The study was supported in part by the Finnish Parkinson Foundation, Turku University Foundation and Turku University Hospital (ERVA-funds). This was an independent academic study. The authors declare that they have no conflicts of interest.

REFERENCES

- Albert NL, Unterrainer M, Diemling M, Xiong G, Bartenstein P, Koch W, Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, Kapucu L, Kluge A, Ziebell M, Darcourt J, Nobili F, Pagani M, Sabri O, Hesse S, Borghet TV, Van Laere K, Tatsch K, la Fougère C (2016) Implementation of the European multicentre database of healthy controls for [(123)I]FP-CIT SPECT increases diagnostic accuracy in patients with clinically uncertain parkinsonian syndromes. *Eur J Nucl Med Mol Imaging*. doi:10.1007/s00259-015-3304-2
- Altman DG (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London
- Ba F, Martin WR (2015) Dopamine transporter imaging as a diagnostic tool for parkinsonism and related disorders in clinical practice. *Parkinsonism Relat Disord* 21:87-94. doi:10.1016/j.parkreldis.2014.11.007
- Badiavas K, Molyvda E, Iakovou I, Tsolaki M, Psarrakos K, Karatzas N (2011) SPECT imaging evaluation in movement disorders: far beyond visual assessment. *Eur J Nucl Med Mol Imaging* 38:764-773. doi:10.1007/s00259-010-1664-1
- Benamer TS, Patterson J, Grosset DG, Booij J, de Bruin K, van Royen E, Speelman JD, Horstink MH, Sips HJ, Dierckx RA, Versijpt J, Decoo D, Van Der Linden C, Hadley DM, Doder M, Lees AJ, Costa DC, Gacinovic S, Oertel WH, Pogarell O, Hoeffken H, Joseph K, Tatsch K, Schwarz J, Ries V (2000) Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. *Mov Disord* 15:503-510. doi: 10.1002/1531-8257(200005)15:3<503::AID-MDS1013>3.0.CO;2-V

Brooks DJ (2012) Parkinson's disease: diagnosis. *Parkinsonism Relat Disord* 18:31-33.

doi:10.1016/S1353-8020(11)70012-8

Buchert R, Hutton C, Lange C, Hoppe P, Makowski M, Bamoussa T, Platsch G, Brenner

W, Declerck J (2015) Semiquantitative slab view display for visual evaluation of 123I-FP-CIT SPECT. *Nucl Med Commun.*

doi:10.1097/MNM.0000000000000467

Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37-46.

Filippi L, Bruni C, Padovano F, Schillaci O, Simonetti G (2008) The Value of Semi-

Quantitative Analysis of 123I-FP-CIT SPECT in Evaluating Patients with Parkinson's Disease. *Neuroradiol J* 21:505-509. doi:

10.1177/197140090802100406

Gerasimou G, Costa DC, Papanastasiou E, Bostanjiopoulou S, Arnaoutoglou M,

Moralidis E, Aggelopoulou T, Gotzamani-Psarrakou A (2012) SPECT study with I-123-Ioflupane (DaTSCAN) in patients with essential tremor. Is there any correlation with Parkinson's disease? *Ann Nucl Med* 26:337-344.

doi:10.1007/s12149-012-0577-4

Hauser RA, Grosset DG (2012) [123I]FP-CIT (DaTscan) SPECT brain imaging in patients with suspected parkinsonian syndromes. *J Neuroimaging* 22:225-230.

doi:10.1111/j.1552-6569.2011.00583.x

Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37-46.

Jaakkola E, Joutsa J, Kaasinen V (2016) Predictors of normal and abnormal outcome in clinical brain dopamine transporter imaging. *J Neural Transm (Vienna)*

123:205-209. doi: 10.1007/s00702-015-1495-0

- Jiménez-Jiménez FJ, Alonso-Navarro H, García-Martín E, Agúndez JA (2012) The relationship between Parkinson's disease and essential tremor: review of clinical, epidemiologic, genetic, neuroimaging and neuropathological data, and data on the presence of cardinal signs of parkinsonism in essential tremor. *Tremor Other Hyperkinet Mov (N Y)* 2:tre-02-75-409-3.
- Joutsa J, Johansson J, Kaasinen V (2015) Is Occipital Cortex a Valid Reference Region in 123I-FP-CIT SPECT Imaging? *Clin Nucl Med* 40:615-616. doi:10.1097/RLU.0000000000000794
- Kaasinen V, Joutsa J, Nojonen T, Johansson J, Seppänen M (2015) Effects of aging and gender on striatal and extrastriatal [(123)I]FP-CIT binding in Parkinson's disease. *Neurobiol Aging* 36:1757-1763. doi:10.1016/j.neurobiolaging.2015.01.016
- Kaasinen V, Kinoshita M, Joutsa J, Seppänen M, Nojonen T (2014) Differences in striatal dopamine transporter density between tremor dominant and non-tremor Parkinson's disease. *Eur J Nucl Med Mol Imaging* 41:1931-1937. doi:10.1007/s00259-014-2796-5
- Kahraman D, Eggers C, Schicha H, Timmermann L, Schmidt M (2012) Visual assessment of dopaminergic degeneration pattern in 123I-FP-CIT SPECT differentiates patients with atypical parkinsonian syndromes and idiopathic Parkinson's disease. *J Neurol* 259:251-260. doi:10.1007/s00415-011-6163-1
- Koch W, Radau PE, Hamann C, Tatsch K (2005) Clinical testing of an optimized software solution for an automated, observer-independent evaluation of dopamine transporter SPECT studies. *J Nucl Med* 46:1109-1118
- Kupsch AR, Bajaj N, Weiland F, Tartaglione A, Klutmann S, Buitendyk M, Sherwin P, Tate A, Grachev ID (2012) Impact of DaTscan SPECT imaging on clinical

management, diagnosis, confidence of diagnosis, quality of life, health resource use and safety in patients with clinically uncertain parkinsonian syndromes: a prospective 1-year follow-up of an open-label controlled study. *J Neurol Neurosurg Psychiatry* 83:620-628. doi:10.1136/jnnp-2011-301695

Kägi G, Bhatia KP, Tolosa E (2010) The role of DAT-SPECT in movement disorders. *J Neurol Neurosurg Psychiatry* 81:5-12. doi:10.1136/jnnp.2008.157370

Lorberboym M, Treves TA, Melamed E, Lampl Y, Hellmann M, Djaldetti R (2006) [123I]-FP/CIT SPECT imaging for distinguishing drug-induced parkinsonism from Parkinson's disease. *Mov Disord* 21:510-514. doi:10.1002/mds.20748

Ottaviani S, Tinazzi M, Pasquin I, Nothdurfter W, Tomelleri G, Fincati E, Nordera G, Moretto G, Fiaschi A, Smania N, Giorgetti P, Antonini A (2006) Comparative analysis of visual and semi-quantitative assessment of striatal [123I]FP-CIT-SPET binding in Parkinson's disease. *Neurol Sci* 27:397-401. doi:10.1007/s10072-006-0718-4

Pencharz DR, Hanlon P, Chakravartty R, Navalkisoor S, Quigley AM, Wagner T (2014) Automated quantification with BRASS reduces equivocal reporting of DaTSCAN (123I-FP-CIT) SPECT studies. *Nucl Med Rev Cent East Eur* 17:65-69. doi:10.5603/NMR.2014.0019

Scherfler C, Nocker M (2009) Dopamine transporter SPECT: how to remove subjectivity? *Mov Disord* 24:721-724. doi:10.1002/mds.22590

Seibyl JP, Kupsch A, Booij J, Grosset DG, Costa DC, Hauser RA, Darcourt J, Bajaj N, Walker Z, Marek K, McKeith I, O'Brien JT, Tatsch K, Tolosa E, Dierckx RA, Grachev ID (2014) Individual-reader diagnostic performance and between-reader agreement in assessment of subjects with Parkinsonian syndrome or

dementia using 123I-ioflupane injection (DaTscan) imaging. *J Nucl Med* 55:1288-1296. doi:10.2967/jnumed.114.140228

Staff RT, Ahearn TS, Wilson K, Counsell CE, Taylor K, Caslake R, Davidson JE, Gemmell HG, Murray AD (2009) Shape analysis of 123I-N-omega-fluoropropyl-2-beta-carbomethoxy-3beta-(4-iodophenyl) nortropane single-photon emission computed tomography images in the assessment of patients with parkinsonian syndromes. *Nucl Med Commun* 30:194-201. doi:10.1097/MNM.0b013e328314b863

Suárez-Piñera M, Prat ML, Mestre-Fusco A, Fuertes J, Mojal S, Balaguer E (2011) [Interobserver agreement in the visual and semi-quantitative analysis of the 123I-FP-CIT SPECT images in the diagnosis of Parkinsonian syndrome]. *Rev Esp Med Nucl* 30:229-235. doi:10.1016/j.rem.2011.02.009

Söderlund TA, Dickson JC, Prvulovich E, Ben-Haim S, Kemp P, Booij J, Nobili F, Thomsen G, Sabri O, Koulibaly PM, Akdemir OU, Pagani M, van Laere K, Asenbaum-Nan S, George J, Sera T, Tatsch K, Bomanji J (2013) Value of semiquantitative analysis for clinical reporting of 123I-2- β -carbomethoxy-3 β -(4-iodophenyl)-N-(3-fluoropropyl)nortropane SPECT studies. *J Nucl Med* 54:714-722. doi:10.2967/jnumed.112.110106

Thenganatt MA, Jankovic J (2016) The relationship between essential tremor and Parkinson's disease. *Parkinsonism Relat Disord* 22:162-165. doi:10.1016/j.parkreldis.2015.09.032

Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, Kapucu OL, Kluge A, Knudsen GM, Koulibaly PM, Nobili F, Pagani M, Sabri O, Vander Borgh T, Van Laere K, Tatsch K (2013) European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects,

gender differences and evaluation of different methods of analysis. *Eur J Nucl Med Mol Imaging* 40:213-227. doi:10.1007/s00259-012-2276-8

Waln O, Wu Y, Perlman R, Wendt J, Van AK, Jankovic J (2015) Dopamine transporter imaging in essential tremor with and without parkinsonian features. *J Neural Transm (Vienna)* 122:1515-1521. doi:10.1007/s00702-015-1419-z

Figure captions

Fig. 1 Four categories of striatal tracer binding used in the visual analysis of images. Each rater was instructed to categorize images of 120 patients using this sample as reference. The raters were also instructed to indicate the hemisphere with the most prominent reduction in tracer binding. In the images, the right hemisphere is on the right side of the image and the left hemisphere is on the left side. The color scale bar indicates specific binding ratios (SBRs).

Fig. 2 Group-differences in (A) the mean striatal specific binding ratio (SBR) and (B) age. The normal group denotes the 26 patients who were categorized as normal by both visual and automated ROI analysis, the abnormal group denotes the 73 patients who were categorized as abnormal by both visual and automated ROI analysis, and the discrepant group denotes the 12 patients who had different results in the two analysis methods. Statistical significances denote post hoc Bonferroni-corrected p-values after one-way ANOVA. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, NS=non-significant.

Fig. 3 Cases that were categorized abnormal in the visual analysis but normal in the automated analysis (cases 1-9) and normal in the visual analysis but abnormal in the automated analysis (cases 10-12). The clinical characteristics of the cases are presented in Table 2 with corresponding case numbers. Four representative striatal slices of each case are presented. The color scale bar indicates specific binding ratios (SBRs).

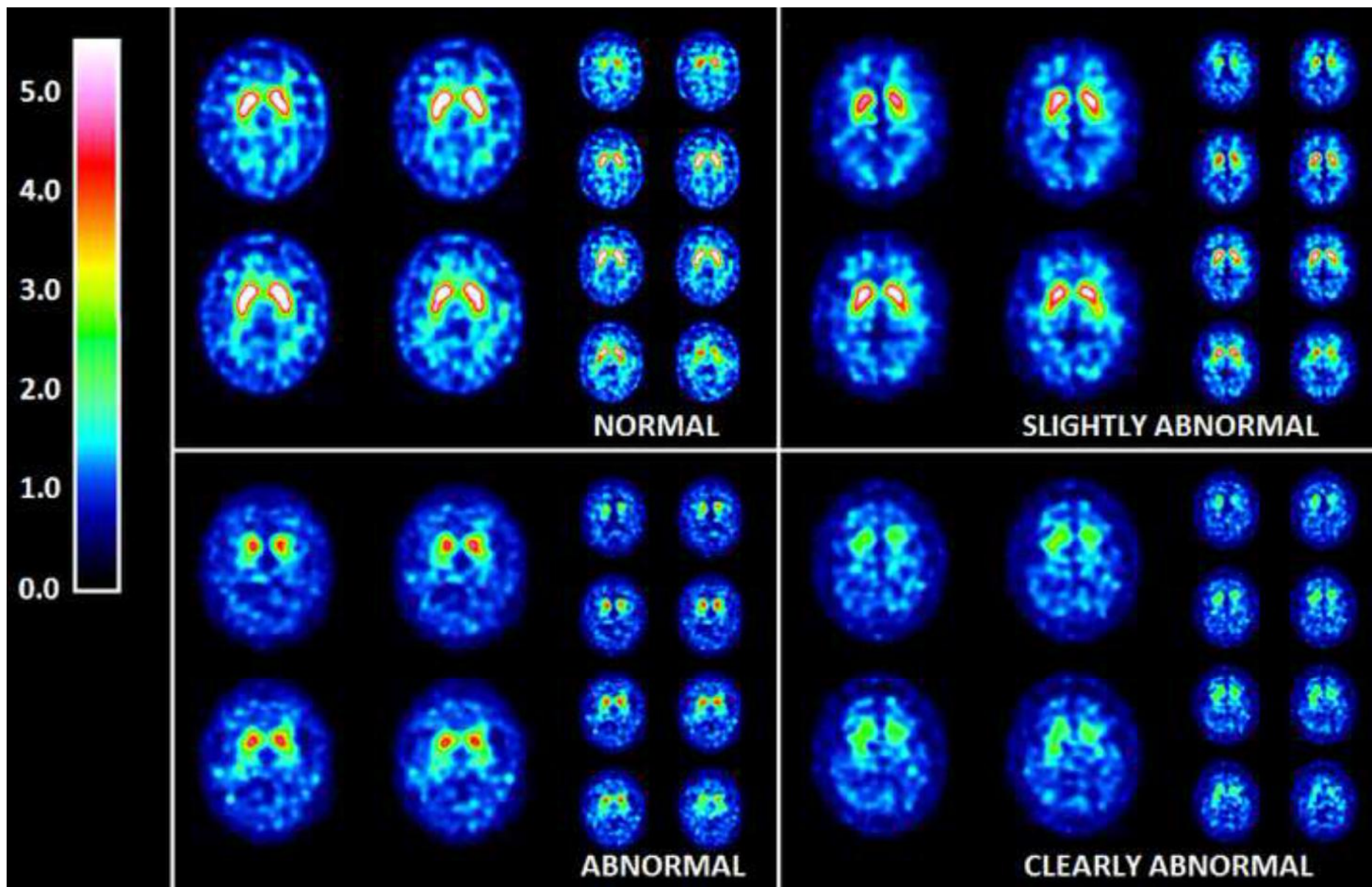


Figure 2

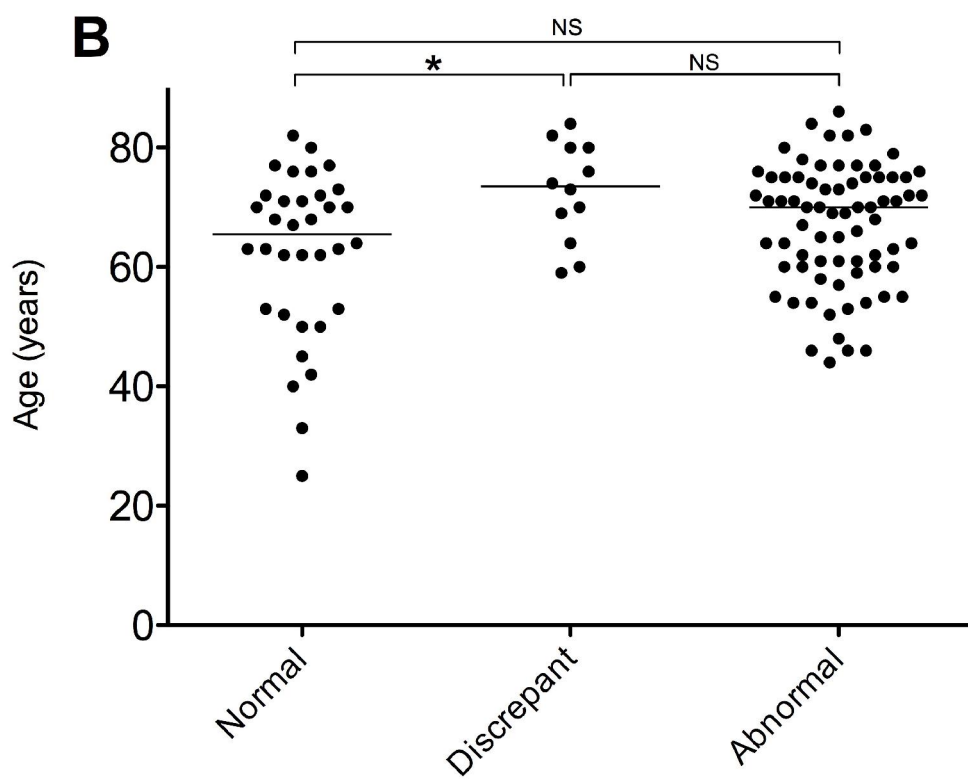
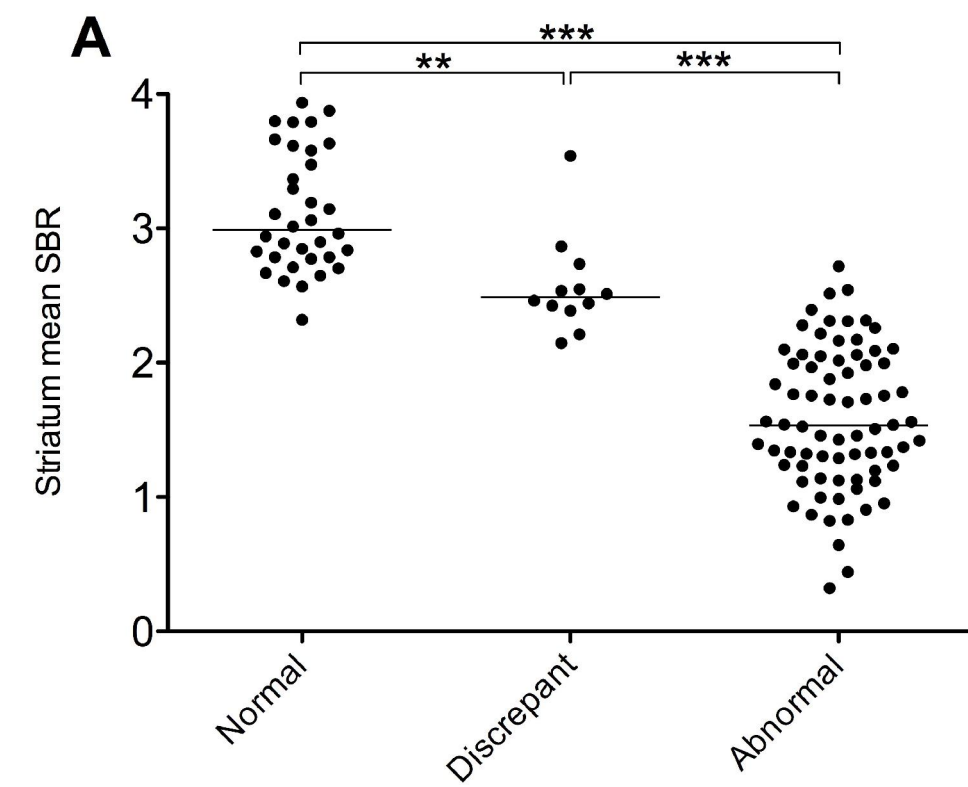


Figure 3

[Click here to download Figure Fig. 3.tiff](#) ■

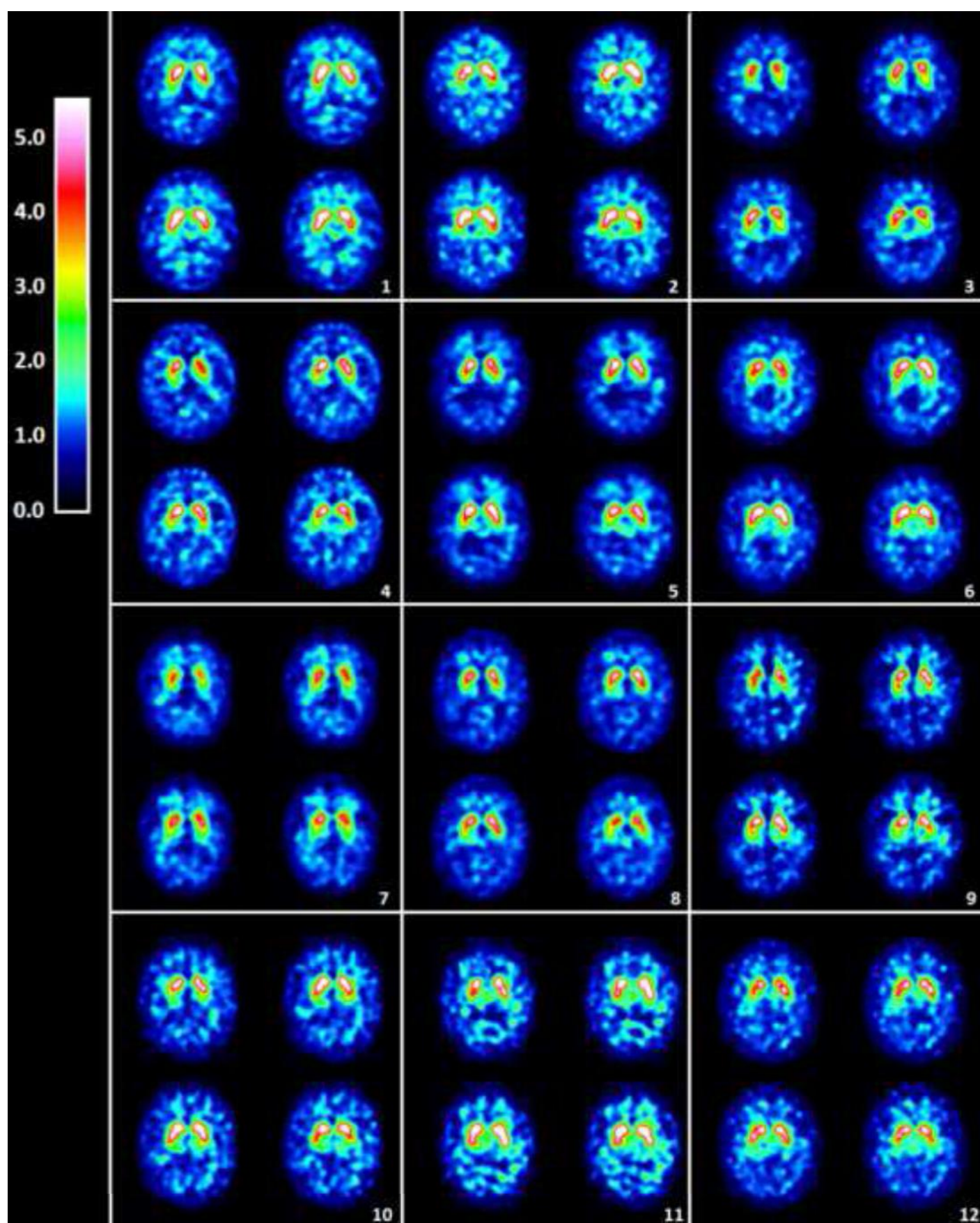


Table 1. Main demographic characteristics of the studied sample categorized according to discrepancies between visual and automated analyses. Values are n or mean (SD). *P*-values denote one-way ANOVA or chi-square tests. Post hoc Bonferroni-corrected statistical subgroup differences are presented in Fig. 2. SBR = specific binding ratio, L/R/S = left/right/symmetrical.

	Non-discrepant normal	Discrepant	Non-discrepant abnormal	F-value	P-value
n	34	12	74	-	-
Age (years)	62.4 (13.8)	72.6 (8.49)	67.0 (10.1)	4.1	0.019
Sex (f/m)	19/15	5/7	31/43	-	0.38
Predominant side of motor symptoms (L/R/S)	8/15/11	1/7/4	28/31/15	-	0.19
Symptom duration (years)	4.3 (7.2)	6.0 (8.3)	2.6 (3.7)	2.6	0.082
Striatum mean (SBR)	3.12 (0.45)	2.57 (0.36)	1.58 (0.57)	120.4	<0.0001

Table 2. Individual clinical characteristics of the cases that showed discrepant findings between visual and automated analyses.
 PD = Parkinson's disease, DIP = drug-induced parkinsonism, ET = essential tremor.

Discrepancy	No.	Age (yrs)	Sex (M/F)	Clinical reason for scanning (from referral)	Symptom duration (yrs)	Predominant side of motor symptoms (Left/Right /Symmetric)	Current diagnosis ¹	Cognitive defect ²
Brass normal, visual abnormal	1	59	M	Suspected parkinsonism plus sdr	2	L	Alzheimer's disease	Yes
	2	70	F	Re-evaluation of PD diagnosis	11	R	Undetermined non-progressive parkinsonism ³	No
	3	73	F	Differential diagnosis between PD and DIP	0,5	R	DIP (Risperidone 2 mg/day)	Yes
	4	80	F	Differential diagnosis between PD and DIP	5	R	DIP (Prochlorperazine, dose unknown)	Yes
	5	84	F	Suspected PD	0,5	R	DIP (Perfenazine 8 mg/day) + Alzheimer's disease	Yes
	6	82	F	Differential diagnosis between PD and DIP	1,5	S	DIP (Risperidone, dose unknown) + Alzheimer's disease	Yes
	7	74	M	Unclear parkinsonian sdr	5	S	ET	No
	8	76	F	Re-evaluation of PD diagnosis	5	S	Unknown	No
	9	80	M	Unclear parkinsonian sdr	1	R	Vascular parkinsonism + vascular dementia	Yes
Brass abnormal, visual normal	10	60	M	Unclear parkinsonian sdr	1	S	Cervical degenerative disease	No
	11	64	M	Unclear parkinsonian sdr	30	R	ET	No
	12	69	M	Differential diagnosis between PD and ET	10	R	ET	No

¹Minimum of 4,5 years of clinical follow-up.

²Cognitive defect at the time of imaging or within 5 years after imaging.

³Parkinsonism for 20 years prior to imaging, no progression, no levodopa response.

Table 3. Effect of expertise. Inter-rater agreement and the agreement between visual analysis and automated analysis in experts and non-experts. κ -values with 95% confidence intervals on standard error (SE) are shown. Agreement with automated analysis was superior in experts compared to all the non-experts. Differences between experts and non-experts 2 and 3 were statistically significant (*= $p < 0.01$).

Rater	Agreement with automated analysis	Comparison to Expert 1	Comparison to Expert 2	Inter-rater agreement 1	Inter-rater agreement 2	Inter-rater agreement 3
Expert 1	0.66 (0.51-0.80)	-	-	0.81 (0.70-0.92)	0.75 (0.65-0.84)	0.60 (0.43-0.69)
Expert 2	0.72 (0.58-0.85)	-	-			
Non-expert 1	0.54 (0.38-0.69)	$p=0.27263$	$p=0.08949$	0.44 (0.23-0.66)	0.61 (0.50-0.72)	0.39 (0.26-0.52)
Non-expert 2	0.23 (0.09-0.36)	$p=0.00002^*$	$p=0.00000^*$			
Non-expert 3	0.47 (0.30-0.64)	$p=0.00513^*$	$p=0.00054^*$	0.63 (0.45-0.82)	0.79 (0.71-0.88)	0.42 (0.28-0.546)
Non-expert 4	0.52 (0.36-0.68)	$p=0.21569$	$p=0.06660$			

Inter-rater agreement 1=Inter-rater agreement for two categories (normal/abnormal).

Inter-rater agreement 2=Inter-rater agreement for four categories.

Inter-rater agreement 3=Inter-rater agreement for categorizing inter-hemispheric differences.