



## OPEN ACCESS

## EDITED BY

Jorge Hernandez-Vara,  
Hospital Universitari Vall D'Hebron, Spain

## REVIEWED BY

Yoo Sung Song,  
Seoul National University Bundang Hospital,  
Republic of Korea  
Yuhei Takado,  
National Institutes for Quantum and  
Radiological Science and Technology,  
Japan  
Hitoshi Shimada,  
Niigata University, Japan

## \*CORRESPONDENCE

Takashi Hanakawa  
✉ hanakawa.takashi.2s@kyoto-u.ac.jp

RECEIVED 04 October 2023

ACCEPTED 22 January 2024

PUBLISHED 19 February 2024

## CITATION

Wakasugi N, Takano H, Abe M, Sawamoto N,  
Murai T, Mizuno T, Matsuoka T, Yamakuni R,  
Yabe H, Matsuda H, Hanakawa T and  
Parkinson's and Alzheimer's disease  
Dimensional Neuroimaging Initiative (PADNI)  
(2024) Harmonizing multisite data with the  
ComBat method for enhanced Parkinson's  
disease diagnosis via DAT-SPECT.  
*Front. Neurol.* 15:1306546.  
doi: 10.3389/fneur.2024.1306546

## COPYRIGHT

© 2024 Wakasugi, Takano, Abe, Sawamoto,  
Murai, Mizuno, Matsuoka, Yamakuni, Yabe,  
Matsuda, Hanakawa and Parkinson's and  
Alzheimer's disease Dimensional  
Neuroimaging Initiative (PADNI). This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Harmonizing multisite data with the ComBat method for enhanced Parkinson's disease diagnosis via DAT-SPECT

Noritaka Wakasugi<sup>1</sup>, Harumasa Takano<sup>1</sup>, Mitsunari Abe<sup>1</sup>,  
Nobukatsu Sawamoto<sup>2</sup>, Toshiya Murai<sup>3</sup>, Toshiki Mizuno<sup>4</sup>,  
Teruyuki Matsuoka<sup>5,6</sup>, Ryo Yamakuni<sup>7</sup>, Hirooki Yabe<sup>7</sup>,  
Hiroshi Matsuda<sup>8</sup>, Takashi Hanakawa<sup>1,9\*</sup> and Parkinson's and  
Alzheimer's disease Dimensional Neuroimaging Initiative  
(PADNI)

<sup>1</sup>Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo, Japan, <sup>2</sup>Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan, <sup>3</sup>Department of Psychiatry, Kyoto University Graduate School of Medicine, Kyoto, Japan, <sup>4</sup>Department of Neurology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan, <sup>5</sup>Department of Psychiatry, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan, <sup>6</sup>Department of Psychiatry, NHO Maizuru Medical Center, Kyoto, Japan, <sup>7</sup>Radiation Medical Science Center for the Fukushima Health Management Survey, Fukushima Medical University, Fukushima, Japan, <sup>8</sup>Department of Biofunctional Imaging, Fukushima Medical University, Fukushima, Japan, <sup>9</sup>Department of Integrated Neuroanatomy and Neuroimaging, Kyoto University Graduate School of Medicine, Kyoto, Japan

**Background:** Dopamine transporter single-photon emission computed tomography (DAT-SPECT) is a crucial tool for evaluating patients with Parkinson's disease (PD). However, its implication is limited by inter-site variability in large multisite clinical trials. To overcome the limitation, a conventional prospective correction method employs linear regression with phantom scanning, which is effective yet available only in a prospective manner. An alternative, although relatively underexplored, involves retrospective modeling using a statistical method known as "combatting batch effects when combining batches of gene expression microarray data" (ComBat).

**Methods:** We analyzed DAT-SPECT-specific binding ratios (SBRs) derived from 72 healthy older adults and 81 patients with PD registered in four clinical sites. We applied both the prospective correction and the retrospective ComBat correction to the original SBRs. Next, we compared the performance of the original and two corrected SBRs to differentiate the PD patients from the healthy controls. Diagnostic accuracy was assessed using the area under the receiver operating characteristic curve (AUC-ROC).

**Results:** The original SBRs were  $6.13 \pm 1.54$  (mean  $\pm$  standard deviation) and  $2.03 \pm 1.41$  in the control and PD groups, respectively. After the prospective correction, the mean SBRs were  $6.52 \pm 1.06$  and  $2.40 \pm 0.99$  in the control and PD groups, respectively. After the retrospective ComBat correction, the SBRs were  $5.25 \pm 0.89$  and  $2.01 \pm 0.73$  in the control and PD groups, respectively, resulting in substantial changes in mean values with fewer variances. The original SBRs demonstrated fair performance in differentiating PD from controls (Hedges's  $g = 2.76$ ; AUC-ROC = 0.936). Both correction methods improved discrimination performance. The ComBat-corrected SBR demonstrated

comparable performance ( $g = 3.99$  and  $AUC-ROC = 0.987$ ) to the prospectively corrected SBR ( $g = 4.32$  and  $AUC-ROC = 0.992$ ) for discrimination.

**Conclusion:** Although we confirmed that SBRs fairly discriminated PD from healthy older adults without any correction, the correction methods improved their discrimination performance in a multisite setting. Our results support the utility of harmonization methods with ComBat for consolidating SBR-based diagnosis or stratification of PD in multisite studies. Nonetheless, given the substantial changes in the mean values of ComBat-corrected SBRs, caution is advised when interpreting them.

#### KEYWORDS

dopamine transporter single-photon emission computed tomography, multicenter cohort study, harmonization, Parkinson's disease, combatting batch effects when combining batches of gene expression microarray data

## 1 Introduction

Dopamine transporter (DAT) single-photon emission computed tomography (SPECT) imaging with  $^{123}\text{I}$ -labeled N-(3-fluoropropyl)-2 $\beta$ -carbomethoxy-3 $\beta$ -(4-iodophenyl) nortropine ( $^{123}\text{I}$ -FP-CIT) is a valuable method to detect the loss of dopaminergic neuron terminals in the striatum. DAT-SPECT is widely used for supporting a clinical diagnosis of Parkinsonian syndromes, including Parkinson's disease (PD) and dementia with Lewy bodies, in both clinical and research fields (1–3). In the evaluation of DAT-SPECT, both quantitative and visual assessments play a crucial role. Radiologists often utilize a comprehensive approach for evaluating possible patients with PD by incorporating quantitative metrics such as the specific-to-non-displaceable binding ratio (SBR) alongside visual evaluations. SBR is a scalar value, which is often incorporated into clinical research and serves as a reference in classifying between PD and non-PD.

DAT-SPECT is considered essential in PD research, particularly in the selection and stratification of participants for disease-modifying therapy (DMT) clinical trials (4). Given the large number of patients enrolled in clinical trials for DMT, it is necessary to conduct such trials in a multisite study setting. However, a drawback of multisite studies is the variance in measurements across the participating sites. There are still unknowns about the best effort to handle inter-site differences in the DAT-SPECT methodology.

DAT-SPECT is used to compute a specific-to-non-displaceable binding ratio (SBR), which serves as a semiquantitative measure of  $^{123}\text{I}$ -FP-CIT radionuclide accumulation. While the primary factor influencing SBR decline is the loss of dopaminergic neurons associated with PD (3, 5, 6), it is noteworthy that SBR also decreases with age and exhibits gender-based variations (2). Moreover, differences in SPECT scanners and operations for SBR measurement (e.g., how to place striatal and reference regions) across study sites influence the SBRs (1). Therefore, age (2, 7–10), sex (2, 7–9, 11), and differences in SPECT scanners and procedures for SBR computation across sites (2) have the potential to confound SBR values, suggesting that their correction may improve diagnostic accuracy.

Various methods may contribute to the correction of DAT-SPECT data. A conventional method prospectively corrects for the factors that affect DAT-SPECT results, such as age, sex, and procedures/scanners (2, 8). Although the effectiveness of the prospective correction is well

established (9, 12–14), this method necessitates advanced phantom scanning. Alternatively, a data-driven method called “combatting batch effects when combining batches of gene expression microarray data” (ComBat) (15) may be useful for correcting inter-site differences in DAT-SPECT. ComBat corrects for systematic differences in the mean and variance of data across different sites, as already successfully applied to other multisite data, including genomes and magnetic resonance imaging (MRI) (15, 16). The implementation of ComBat is straightforward and can be seamlessly integrated into an individual researcher's analysis pipeline. Furthermore, ComBat can be retrospectively applied to previously completed research datasets.

While ComBat presents a simple and powerful correction method for multicenter DAT-SPECT data, its full effectiveness has not yet been explored. To assess the efficacy of the ComBat correction, we analyzed DAT-SPECT data performed on healthy controls (HCs) and patients with PD (PDs) from multiple clinical centers. We hypothesized that SBR corrected by the ComBat would demonstrate comparable diagnostic power with prospectively corrected SBR. Additionally, we aimed to elucidate a potential bias in the data distribution introduced by the harmonization methods.

## 2 Materials and methods

### 2.1 Participants

We utilized data from a total of 81 individuals with PDs and 72 HCs who were registered to the Parkinson's and Alzheimer's disease Dimensional Neuroimaging Initiative (PADNI) study<sup>1</sup> in four participating sites.

PADNI is a neuroimaging cohort study comprising patients with dementia or PD and healthy older adults. We determined the number of HCs to exceed the minimum sample size required for the semiquantitative analysis of the striatum, as detailed in a previous report (17). Participants were recruited at four sites: the National Center of Neurology and Psychiatry (NCNP), Kyoto University (KU), Kyoto

1 <https://padni.org/>

Prefectural University of Medicine (KPUM), and Fukushima Medical University (FMU) (Table 1). All the participants exhibited cognitive functioning sufficient for giving informed consent. The inclusion criteria were as follows: age  $\geq 50$  years and having a study partner inform of the participant's activities of daily living (ADL). The exclusion criteria were as follows: use of medications affecting dopamine uptake (e.g., selective serotonin reuptake inhibitors and tricyclic antidepressants), neurological and psychiatric disorders other than PD (e.g., cerebral infarction and major depressive disorder), allergies to alcohol or iodine, and concurrent plans to participate in clinical trials. Notably, FMU did not provide DAT data from the healthy-aged persons.

PDs were consistent with clinically established or probable PD in the International Parkinson and Movement Disorder Society (MDS) criteria (18). The MDS criteria, a widely used criteria for the clinical diagnosis of PD, request clinical examination by movement disorders experts. In our study, we have employed the MDS criteria as the gold standard of PD diagnosis for the teacher data in the subsequent ROC analysis.

HCs maintained independent ADL, showing neither clinical evidence indicating psychiatric or neurological disorders nor abnormal measures in the blood, neurological, and psychological tests required for entry to PADNI (see the "Clinical and Neuropsychological Assessments" section). The participants had no apparent structural abnormalities, which may have affected their cognitive or motor function, on T1- and T2-weighted brain structural MRIs (19).

## 2.2 Acquisition and processing of imaging data

### 2.2.1 Original SBR prior to site-effect mitigation

Five imaging devices across four sites were used for SPECT-computed tomography (SPECT-CT) scans (Table 2). The chosen reconstruction method involved X-ray-CT attenuation correction without scatter correction, based on our previous findings showing comparable SBRs with or without scatter correction (2).

Subsequent to the reconstruction, data were processed by a technician/physician using a routine procedure at each site. DAT-SPECT data were processed using DatView to compute the SBR (2). Striatal volumes of interest (VOIs) were defined on transaxial slices within a 44-mm-thick slab centered on the highest striatal signal (17), with a fixed striatal VOI volume of 11.2 cm<sup>3</sup>. The SBR was calculated as the ratio of the striatal binding count to the non-specific binding count in the background at each site. Age and sex were corrected at each site, yielding the original SBR (2). Striatal VOI and background selection relied on the technician/physician at each site and thus varied across sites, yielding the original SBR, including the site effects.

Additionally, an anthropomorphic striatal phantom filled with <sup>123</sup>I solution was scanned with the SPECT scanner at each site to acquire data for the prospective scanner correction (2, 20).

TABLE 1 The participants' demographic information in each center.

Center	HCs (n)	PDs (n)	Age (years)			Sex (male/female)		
			HCs	PDs	Total	HCs	PDs	Total
NCNP	45	16	67.4 (8.23)	72.3 (7.94)	68.6 (8.38)	25/20	13/3	38/23
KPUM	6	5	68.3 (8.38)	76.4 (6.06)	72.0 (8.22)	3/3	1/4	4/7
KU	21	48	68.6 (9.38)	66.5 (8.64)	68.0 (9.15)	14/7	20/28	34/35
FMU	0	12	n/a	66.5 (5.43)	66.5 (5.43)	n/a	7/5	7/5

Data shown are the mean (SD). NCNP, National Center of Neurology and Psychiatry; KPUM, Kyoto Prefectural University of Medicine; KU, Kyoto University; FMU, Fukushima Medical University; HCs, healthy controls; PDs, patients with Parkinson's disease; n/a, not applicable; SD, standard deviation.

TABLE 2 The single-photon emission computed tomography (SPECT) scanner information in each center.

Center	Interval between infusion and scan (hours)	Dose of intravenous <sup>123</sup> I-FP-CIT (MBq)	SPECT scanner	Reconstruction	Attenuation correction	Scatter correction	Detector system
NCNP	3.0	158–186	Siemens Symbia T6 + LMEGP, GE Discovery NM/CT 670 pro + ELEGP	OSEM (iterations 2, subsets 18), OSEM (iterations 4, subsets 15)	CT	None	Dual-head
KPUM	3.0	158	GE Discovery NM/CT670QSP + ELEGP	OSEM (iterations 5, subsets 6)	None	None	Dual-head
KU	3.0	167	GE NMCT 870DR + ELEGP	FBP	None	None	Dual-head
FMU	3.0	167	Toshiba GCA-9300R + FANHR	OSEM (iterations 6, subsets 10)	None	None	Triple-head

NCNP, National Center of Neurology and Psychiatry; KPUM, Kyoto Prefectural University of Medicine; KU, Kyoto University; FMU, Fukushima Medical University; <sup>123</sup>I-FP-CIT, <sup>123</sup>I-labeled N-(3-fluoropropyl)-2 $\beta$ -carbomethoxy-3 $\beta$ -(4-iodophenyl) nortropane; MBq, mega Becquerel; LMEGP, Low-Medium Energy General Purpose; ELEGP, extended low energy general purpose; FANHR, fanbeam high resolution; OSEM, ordered subset expectation maximization method; FBP, filtered back projection; CT, computed tomography.

## 2.2.2 Prospective correction of SBR using phantom scanning data

We calculated prospectively corrected SBR as a benchmark for site-effect correction. In terms of SBR, the site effects can mainly be broken down into two components: differences in SPECT scanners (the scanner effects) and differences in the striatal VOI and background selection step (the procedure effects). To correct the scanner effects, linear equations were created to convert the SBRs of the standardized striatal phantom scanned by each site's scanner to those by the NCNP scanner (2). This equation was applied to the original SBR to calibrate across the scanners. Furthermore, at the central site (NCNP), two researchers (NW and HT) independently recalculated the SBR for each participant from the reconstructed DAT images using the standardized Southampton method with an iso-contour threshold range of 30–60% (14). The SBRs computed by the two researchers were averaged for each participant. The SBRs following both scanner correction and procedure correction were designated as the fully prospectively corrected SBRs.

## 2.2.3 Harmonizing SBR using ComBat correction method

For the retrospective correction, we employed the ComBat method (15), proposed as an accurate and simple correction technique to remove site effects across various research disciplines.

The ComBat operates as an empirical Bayes-based method designed to minimize differences between data, including inter-site and scanner variations (15). We applied the ComBat method to the original SBR (Supplementary Figure S1). The imaging sites and the scanners were chosen as the variables to be removed from the model, while the clinical diagnosis was retained as a non-site-specific variable.

## 2.3 Clinical and neuropsychological assessments

All participants underwent standardized clinical and neuropsychological assessments across the groups. The cutoff scores for defining HCs were as follows: (i) the MDS-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (18) part III score  $\leq 6$  (excluding postural and action tremors); (ii) the mini-mental state examination (21) score  $\geq 28$ ; (iii) the Clinical Dementia Rating (CDR) (22) score of 0; and (iv) the Japanese edition of the rapid eye movement sleep behavior disorder screening questionnaire (23) score  $\leq 5$ . All participants underwent the Japanese version of the Montreal Cognitive Assessment (24), frontal assessment battery (25), trail-making test (TMT) (26), and Japanese version of the odor stick identification test (OSIT-J) (27) to investigate their cognitive and physical conditions comprehensively.

## 2.4 Statistical analysis

Clinical scores were analyzed across the diagnostic groups and the participating sites, using mean analysis with *t*-test and one-way analysis of variance (ANOVA), respectively. The significance threshold for mean analysis and ANOVA was set at  $p < 0.05$  and  $p < 0.05$  with false discovery rate (FDR) correction for the *post-hoc t*-test after ANOVA.

The primary analysis aimed to assess the discriminative performance of ComBat-corrected SBRs in comparison to the original and prospectively corrected SBRs. We evaluated this by comparing the area under the curve (AUC) of a receiver operating characteristic (ROC) curve (AUC-ROC), which served as the index of diagnostic accuracy for classifying HCs and PDs with the SBR. The Youden index was used to set diagnostic thresholds for HCs and patients with PD in the ROC analysis. Additionally, the effect size of the SBR between HCs and PDs was examined separately for the original and two correction methods, using Hedges's *g*.

We conducted a detailed examination of the relationship between the ComBat-corrected SBR and the scanner-corrected SBR. This analysis was undertaken to assess the behavior of ComBat correction using scanner-corrected SBR as a reference. Our analysis comprised two principal components: (1) the calculation of the correlation coefficient to determine the linear association between ComBat-corrected SBR and scanner-corrected SBR and (2) a plot of residuals from a regression model.

These analyses were conducted using Python scikit-learn 0.24.2<sup>2</sup> and statsmodels 0.12.1.<sup>3</sup> Results visualization was accomplished using matplotlib 3.1.1<sup>4</sup> and seaborn 0.9.0.<sup>5</sup> Behavioral data and SBR for each correction were graphed on "Raincloud plots" (28) using ptitprince 0.2.5.<sup>6</sup>

## 2.5 Ethical approval and participant consent

This study was registered under the University Hospital Medical Information Network Clinical Trials Registry (UMIN000036297) and fulfills the criteria set by the International Committee of Medical Journal Editors. The National Center of Neurology and Psychiatry Ethics Committee, Japan, approved the study protocols (approval number: A2018-086). The principles of the Declaration of Helsinki were strictly adhered to throughout the study.

All participants provided written informed consent to participate and for the publication of images in Figure 1, which displays anonymized SPECT imaging results for individual participants.

## 3 Results

### 3.1 Demographics and clinical assessments of participants

The average age of HCs across all facilities was  $67.2 \pm 8.26$  (SD) years; the average age was  $66.4 \pm 8.23$  at NCNP,  $69.6 \pm 8.65$  at KPUM, and  $68.6 \pm 9.38$  at KU. For PDs (Parkinson's Disease patients), the average age was  $68.5 \pm 8.52$  years (SD), detailed as follows: NCNP ( $72.3 \pm 7.94$ ), KPUM ( $76.4 \pm 6.06$ ), KU ( $66.5 \pm 8.64$ ), and FMU ( $66.5 \pm 5.43$ ).

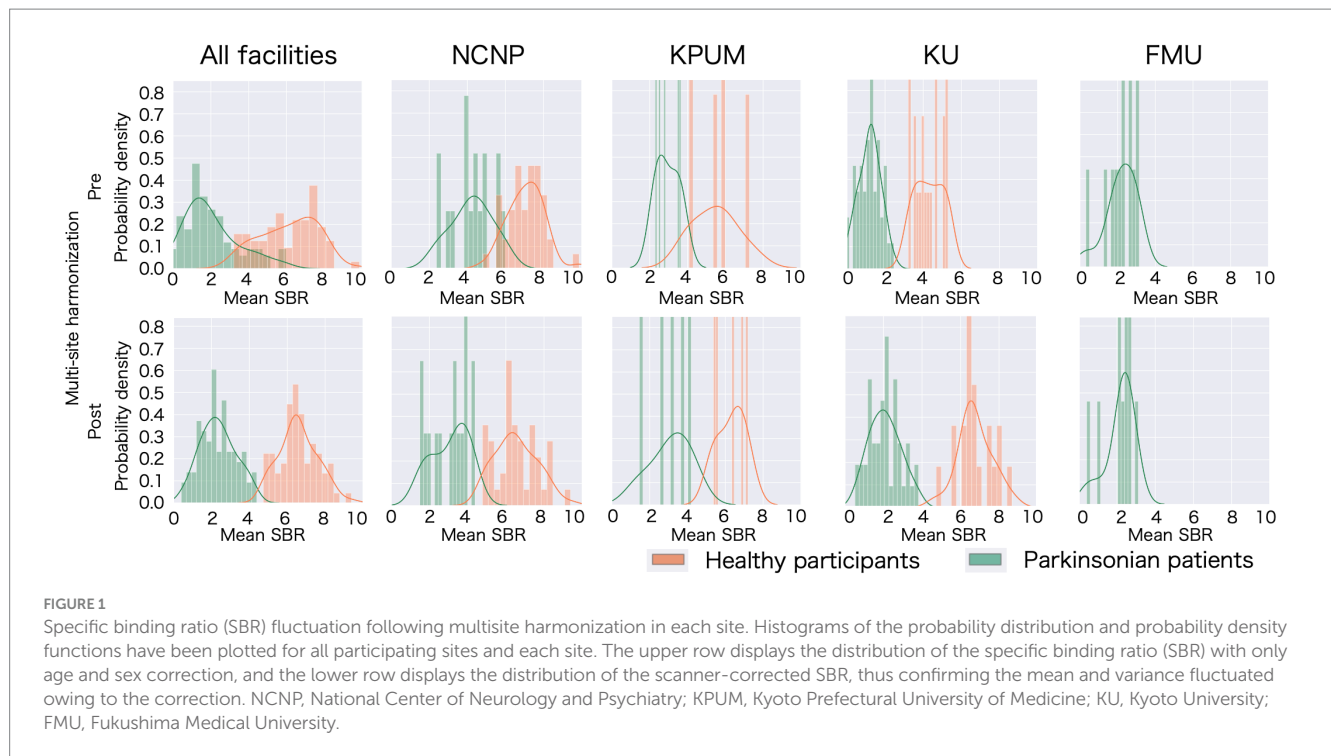
<sup>2</sup> <https://scikit-learn.org/>

<sup>3</sup> <https://www.statsmodels.org>

<sup>4</sup> <https://matplotlib.org/>

<sup>5</sup> <https://seaborn.pydata.org/>

<sup>6</sup> <https://github.com/pog87/PtitPrince>



HCs comprised 42 men and 30 women, with individual site breakdowns of 25 men/20 women at NCNP, 3 men/3 women at KPUM, and 14 men/7 women at KU. PDs consisted of 41 men and 40 women, detailed as 13 males/3 women at NCNP, 1 man/4 women at KPUM, 20 men/28 women at KU, and 7 men/5 women at FMU.

The comprehensive clinical and neuropsychological assessments supported the diagnostic classification of HCs and PDs. Among the PD group, all individuals scored  $\geq 8$  on the MDS-UPDRS part III (excluding postural and action tremors). We noticed significant disparities in the MDS-UPDRS part III scores across the sites (Supplementary Table S1).

In contrast, none of the HCs had apparent cognitive decline or a movement disorder (Table 3); the MDS-UPDRS part III scores (excluding postural and action tremors) were  $< 6$  in all HCs. While the average cognitive function scores were above the cutoff value for dementia in both groups, a comparison between the groups revealed a significant reduction in cognitive functioning among the PD group.

### 3.2 Effects of multisite harmonization

A significant difference was observed in the original SBR across the sites for both HCs ( $F[2, 13.65] = 47.14, p < 0.001$  by Brown–Forsythe-corrected one-way ANOVA) and PDs ( $F[3, 35.69] = 52.19, p < 0.001$ ) (Figure 1; Supplementary Table S2). The *post-hoc* *t*-tests following ANOVA showed that the original SBR was lower in a single site (KU), especially in PDs (Supplementary Table S2). After pooling the data from all sites, the original SBRs were  $6.13 \pm 1.54$  (mean  $\pm$  standard deviation) for HCs and  $2.03 \pm 1.41$  for PDs (Figure 2), yielding a small effect size with the Hedges’s *g* of 2.76 and a fair discrimination performance with the AUC-ROC of 0.936 (Figure 3). The threshold set by ROC analysis was 4.01.

The prospectively corrected SBR, addressing both scanner and procedure, achieved the highest effect size and discrimination

performance (Supplementary Figure S2). Hence, we adopted the fully corrected SBRs as a representative of the prospectively corrected SBRs. The prospective correction effectively removed the site difference of SBRs in HCs ( $F[2, 32.83] = 0.55, p = 0.58$ ), but not entirely in PDs ( $F[3, 18.86] = 7.11, p = 0.005$ ), potentially reflecting significant disparities in the MDS-UPDRS part III scores across the sites. The prospectively corrected SBR was  $6.52 \pm 1.06$  for HCs and  $2.40 \pm 0.99$  for PDs (Figure 2), yielding a medium effect size (Hedges’s *g* = 4.32) and high discrimination performance with the AUC-ROC of 0.992 (Figure 3). The threshold set by ROC analysis was 4.55.

After applying the ComBat correction, the SBR did not differ across the sites in either HCs ( $F[2, 13.16] = 0.01, p = 0.91$ , Brown–Forsythe-corrected one-way ANOVA) or PDs ( $F[3, 37.56] = 0.24, p = 0.62$ , Brown–Forsythe-corrected one-way ANOVA). The ComBat-corrected SBRs were  $5.25 \pm 0.89$  for HCs and  $2.01 \pm 0.73$  for PDs (Figure 2), yielding a medium effect size (Hedges’s *g* = 4.32) and a high discrimination performance with the AUC-ROC of 0.987 (Figure 3). The threshold set by ROC analysis was 3.35. The ComBat-corrected SBR showed diagnostic accuracy comparable to the prospectively corrected SBRs.

The correlation coefficient between the ComBat-corrected SBR and the scanner-corrected SBR was 0.97, indicating a very high correlation (Figure 4A), and the regression model of scanner-corrected SBR using ComBat-corrected SBR converged with small residuals (Figure 4B).

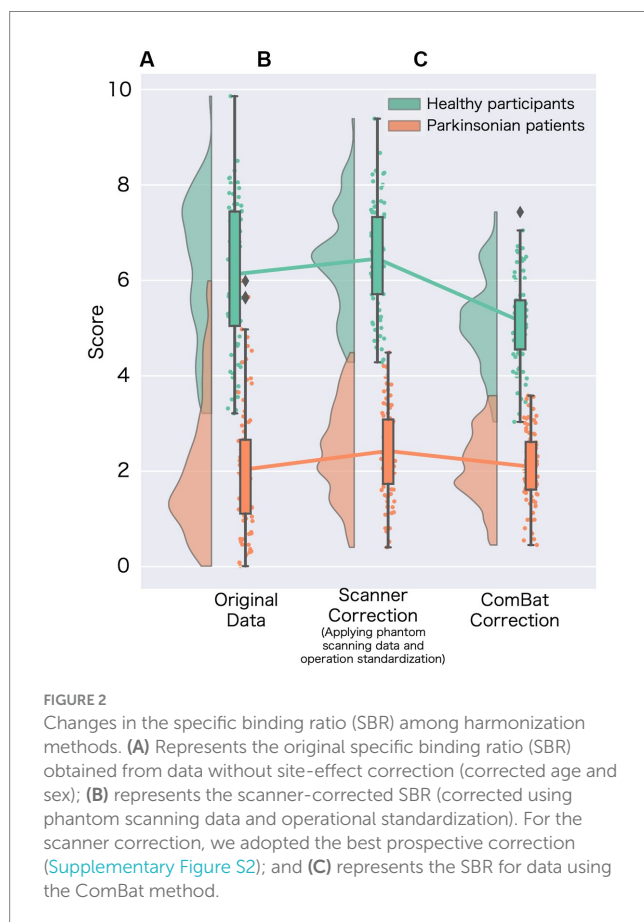
### 3.3 Analyses of data distribution before and after harmonization

The modest separation between PDs and HCs with the original SBR data was (see the section “Effects of multisite harmonization”) likely because the pooled original SBR data across all sites were not normally distributed [ $p = 0.05$  (HC),  $p < 0.001$  (PD)] (Shapiro–Wilk

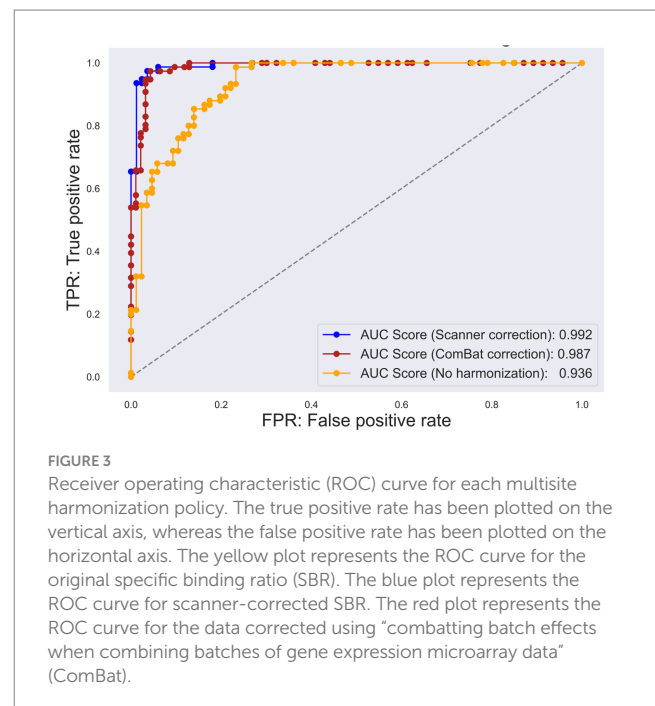
TABLE 3 The results of clinical evaluations.

		HCs	PDs	Mean analysis (PDs vs. HCs, <i>t</i> -test)
MDS-UPDRS	Part I	2.17 (2.61)	10.72 (4.98)	$t(151) = 12.38, p < 0.001$
	Part II	0.40 (1.21)	12.53 (8.57)	$t(151) = 13.84, p < 0.001$
	Part III	1.29 (2.26)	29.77 (18.40)	$t(151) = 14.38, p < 0.001$
	Part IV	n/a <sup>1</sup>	3.13 (5.59)	n/a
MMSE		29.16 (1.69)	27.43 (3.075)	$t(151) = -2.60, p = 0.005$
CDR Sum of boxes		0.08 (0.20)	0.98 (1.89)	$t(151) = 3.68, p < 0.001$
Global CDR		0 (0)	0.22 (0.35)	$t(151) = 3.84, p < 0.001$
MOCA-J <sup>2</sup>		26.2 (2.45)	24.9 (3.81)	$t(151) = -3.19, p = 0.002$
RBDSQ		2.03 (1.97)	4.77 (2.91)	$t(151) = 7.08, p < 0.001$
FAB		16.7 (1.17)	14.9 (3.15)	$t(151) = -3.67, p < 0.001$
TMT-A (seconds)		40.5 (14.1)	70.0 (61.7)	$t(151) = 4.74, p < 0.001$
TMT-B (seconds)		73.0 (42.1)	109.1 (69.2)	$t(151) = 4.04, p < 0.001$
OSIT-J		9.23 (2.67)	(2.86)	$t(151) = -10.91, p < 0.001$

<sup>1</sup>Healthy subjects did not undergo MDS-UPDRS part IV because they are free of parkinsonism and anti-parkinsonian drugs. <sup>2</sup>MoCA-J scores are adjusted for years of education. HCs, healthy controls; PDs, patients with Parkinson's disease; MDS, International Parkinson and Movement Disorder Society; UPDRS, Unified Parkinson's Disease Rating Scale; MMSE, Mini-Mental State Examination; CDR, Clinical Dementia Rating; MOCA-J, Japanese version of the Montreal Cognitive Assessment; RBDSQ, rapid eye movement sleep behavior disorder screening questionnaire; FAB, frontal assessment battery; TMT, trail-making test; OSIT-J, Japanese version of the odor stick identification test; n/a, not applicable.



test)]. The separation between PDs and HCs was more pronounced after ComBat correction (Figure 1), which followed an improved normal distribution [ $p = 0.55$  (HC),  $p = 0.46$  (PD) (Shapiro-Wilk test)], especially in sites with large sample sizes (e.g., Ncnp). The data



distribution for prospective correction also showed the same improvement as ComBat [ $p = 0.35$  (HC),  $p = 0.50$  (PD) (Shapiro-Wilk test)]. These changes in data distribution accounted for the shift of the effect size from small (Hedges's  $g = 2.76$ ) to medium (Hedges's  $g = 3.99$  following the ComBat correction and Hedges's  $g = 4.32$  following the prospective correction).

Comparing the averages among the three types of SBR (original, prospective correction, ComBat correction), there was a significant variation in the average SBR for both HCs and PDs [ $F(2, 175.93) = 21.31, p < 0.001$  by Brown-Forsythe-corrected one-way ANOVA (HCs),  $F(2, 190.00) = 3.25, p = 0.04$  by Brown-Forsythe-corrected one-way ANOVA

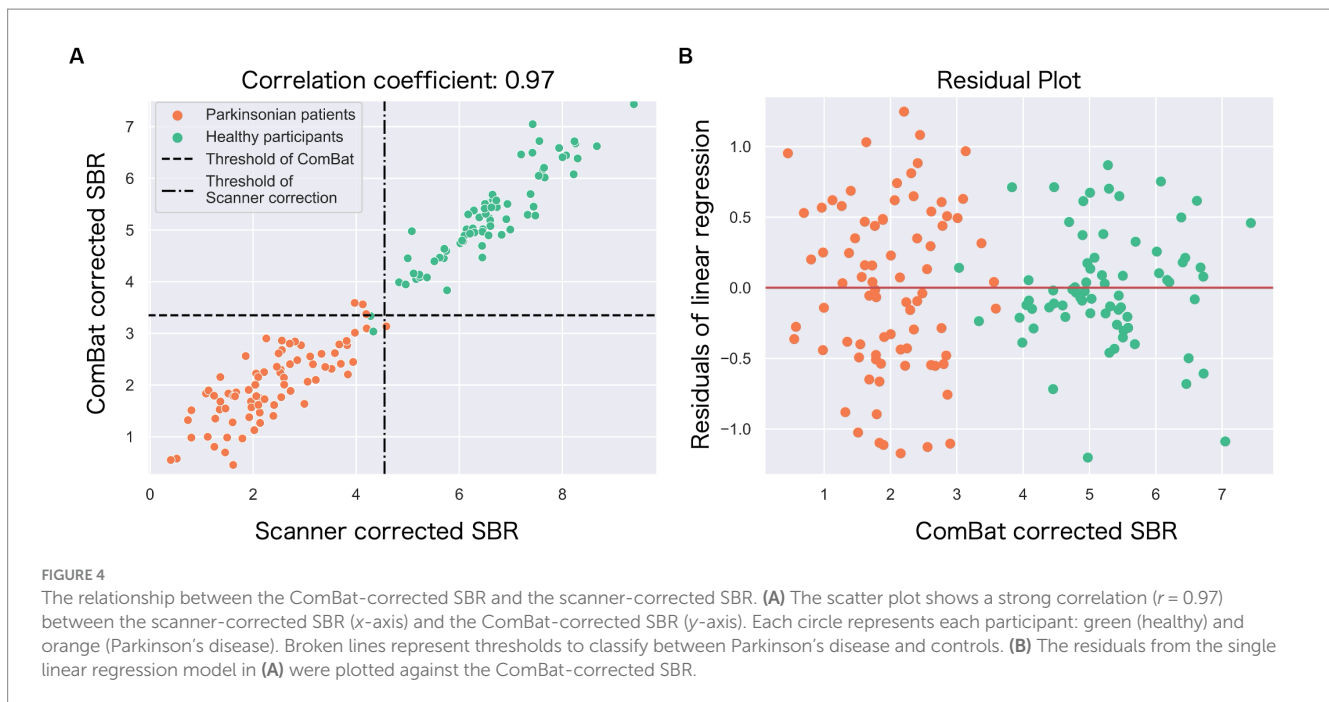


FIGURE 4

The relationship between the ComBat-corrected SBR and the scanner-corrected SBR. (A) The scatter plot shows a strong correlation ( $r = 0.97$ ) between the scanner-corrected SBR (x-axis) and the ComBat-corrected SBR (y-axis). Each circle represents each participant: green (healthy) and orange (Parkinson's disease). Broken lines represent thresholds to classify between Parkinson's disease and controls. (B) The residuals from the single linear regression model in (A) were plotted against the ComBat-corrected SBR.

(PDs)). In the *post-hoc t*-test, ComBat-corrected SBR was significantly lower than the other correction methods in both HCs and PDs ( $p$ -FDR < 0.05).

### 3.4 Assessment of false positives and false negatives from a clinical perspective

After the ComBat, two HCs were categorized as PDs, while three clinical PDs were categorized as HCs.

We investigated the scanner-corrected SBR in individuals who were misclassified with ComBat. In the PD group, the three patients misclassified by ComBat had ComBat-corrected SBRs of 3.38, 3.58, and 3.56, respectively, which were above the cutoff threshold of 3.35. By contrast, their scanner-corrected SBRs were 3.97, 4.12, and 4.19 below the cutoff threshold of 4.55, and thus all were correctly classified as PD. In the HC group, the two patients misclassified as having PD by ComBat had ComBat-corrected SBRs of 3.03 and 3.33. Their scanner-corrected SBRs were 4.33 and 4.28, respectively, below the cutoff threshold. Thus, these two HCs were judged as PD by both ComBat-corrected and scanner-corrected SBR. These results indicate that, for this dataset, the ComBat correction was slightly less sensitive but equally specific compared to the scanner correction.

In reviewing clinical data, none of the two HCs (false positives) displayed increased MDS-UPDRS III scores or global cognitive decline (CDR score = 0). However, one of the HCs completed the TMT-B in 82 s, approaching the cutoff value and suggesting a potential latent decline in executive functioning. The other HCs had an OSIT-J score of 5 points, thus indicating mild olfactory impairment. These participants are currently being followed up in the PADNI cohort to monitor the development of parkinsonism, or cognitive decline. The three PDs (false negatives) displayed MDS-UPDRS part III scores (excluding tremors) of 14, 15, and 8, respectively, indicating relatively mild motor symptoms for patients with PD.

## 4 Discussion

This study demonstrated that the harmonization procedure to remove site-specific effects improved the accuracy of detecting the loss of dopaminergic terminals in the striatum in multicenter DAT-SPECT data.

DAT-SPECT is an established method for the differential diagnosis of PD and related disorders (29). The DAT data without harmonization displayed an acceptable level of accuracy for differentiating PDs and HCs in reference to the MDS PD criteria as the gold standard. Without harmonization, 93.6% of the diagnostic accuracy in a multicenter study appeared considerably high to support the clinical diagnosis of PD. Together with other clinical assessments, including the MDS-UPDRS part III score olfactometry, DAT-SPECT in clinical practice appeared sufficiently accurate, even for the original data. This suggests that the effect size of the dopaminergic terminal loss in PD was substantially larger than the other previously reported factors (2, 7–12, 30, 31). This interpretation appears reasonable because the loss of dopaminergic terminals progresses before the clinical onset of PD (32, 33), and a substantial reduction of the SBR is already present at the onset, even in mild cases (34).

This study demonstrated that correction for differences in the sites improved the diagnostic accuracy of dopaminergic denervation. The prospective correction for site differences comprised two factors as follows: (i) the correction across scanners by applying a linear transformation equation based on data from a phantom filled with  $^{123}\text{I}$  solution for each SPECT scanner; and (ii) the standardization of human operation to set up the VOI in software computing the SBR (i.e., procedure standardization). Both factors exerted significant effects on the SBR (Supplementary Figure S2). The application of scanner correction achieved a high level of agreement with the clinical diagnosis of PD. The prospective correction certainly improved the effect size of the SBR, differentiating PDs and HCs from mild to medium, with an improvement rate of approximately 5% (from the AUC-ROC). The degree of AUC-ROC improvement following the

prospective correction may not appear monumental. However, this level of difference should exert tremendous effects in large-scale studies, such as a randomized control trial for DMT. In clinical trials involving thousands of participants, a 5% difference in diagnostic accuracy will result in over a hundred misdiagnoses (35–37). A cohort based on an accurate diagnostic test should yield a more specific outcome of the intervention and save enormous time and financial costs in these large-scale studies. Therefore, we strongly recommend scanner correction and operation standardization to reduce false findings from DAT-SPECT when managing a large-scale multicenter SPECT study. The correction will be considerably greater in clinical trials in prodromal PD, which comprises marginal differences in the SBRs from HCs.

The scanner correction removed the site effects in HCs; however, the correction only reduced the site effects in PDs (Supplementary Table S2). As DAT-SPECT reflects the severity of PD, this finding can be attributed to the difference in the severity of PD across the sites (Supplementary Table S1) (38). Hence, the scanner correction likely removed the technical differences across sites, leaving the difference in the participants' factors unaffected. This is favorable when we consider analyses using inter-individual differences after harmonization.

The correction with ComBat improved the classification accuracy, which was comparable to the correction after both scanner correction and operation correction. The ComBat harmonization has been principally used in genomic and MRI studies as a simple and robust method for correcting measurement bias across sites. In the present study, we observed a strong correlation between the ComBat-corrected and scanner-corrected SBRs ( $r=0.970$ ) (see Figure 4A). Furthermore, most residuals from the single regression model were within one SD of the mean (Figure 4A). Thus, the ComBat-corrected SBRs closely mirrored the scanner- and operation-corrected SBRs, supporting the potential of ComBat technology to achieve multisite harmonization of DAT-SPECT. ComBat correction is a powerful method that may replace the laborious method, such as phantom scanning, at each site. In addition, it appears useful during the inability to perform phantom scans, for example, for already completed research projects. Moreover, it should be effective while analyzing a public neuroimaging dataset (39).

However, the ComBat correction appears to have a limitation. The ComBat-corrected SBRs classified three PDs as HCs, corresponding to patients with “scans without evidence of dopaminergic deficit” (SWEDD). With the prospective scanner correction, however, there was no SWEDD in the present dataset. When considering the differences in the definition of PD between these two correction methods, it should be noted that there are problems inherent to multicenter studies to define PD according to SBRs.

The accuracy of the original SBR in this study was 93.6%. The MDS criteria used for inclusion criteria of PDs in the present study include “normal functional neuroimaging of the presynaptic dopaminergic system” as an exclusion item. Ideally, PD diagnostic accuracy should be 100% at the time of the original SBR, and SWEDD should not be included in the PDs. This discrepancy may be due to several factors.

Perhaps one of the biggest factors is the variation in diagnosis between different institutions. PD diagnosis was made clinically at each facility based on the MDS PD diagnostic criteria, including evaluation of dopaminergic denervation.

However, the criteria for defining abnormal or normal DAT-SPECT vary across hospitals, with each site having its own method for computing the SBR and employing adjunctive criteria such as laterality to define normal or abnormal DAT-SPECT. Most hospitals set their own threshold, often suggested by analysis software tailored for each site. Thus, the 93.6% accuracy of the original DAT, at least in part, reflected an under-triage of DAT-SPECT findings in a multicenter setting where only a standard SBR threshold was used without the employment of adjunctive criteria or site-specific optimization.

Importantly, our study demonstrated that the two harmonization methods significantly improved accuracy, almost reaching 100%. This finding underscores the potential of careful *a priori* or *a posteriori* data harmonization in multicenter settings to achieve accuracy comparable to that of tailored optimization at individual sites. It is noteworthy that no SWEDDs remained in the data after the best harmonization efforts, supporting the diagnosis of each site. Notably, those PD patients who were misclassified as HCs by the ComBat (“SWEDDs”) yet correctly labeled by the scanner and operation correction had only mild motor symptoms. Therefore, caution is advised to use the ComBat correction when classifying PD, especially with mild symptoms, according to the multicenter studies.

Another factor would be the inclusion of apparently healthy-aged individuals with abnormal DAT-SPECT, which could be either false positives or indicative of truly prodromal PD. This occurrence is inevitable despite our rigorous participant screening process, which included neurological examinations, cognitive tests, questionnaires, blood tests, and MRI scans. In any event, the harmonization methods effectively reduced the false positives.

The mean and variance of a single site can significantly affect the corrected data, thus compromising its generalizability to third parties or in meta-analyses. While this limitation is a problem in the ComBat method, where individual participant data are incorporated into the model for inter-institutional difference correction, the problem can be avoided in prospective correction because the model is constructed from a separate large-scale database with healthy participants. Therefore, we recommend model-based corrections for age, sex, and site effects whenever possible.

In the present study, some sites included sample sizes for each diagnosis that were insufficient to generate a linear model that included ComBat. The small sample size can affect the accuracy of the model because ComBat is a linear model. However, based on the results of the correlation analysis (Figure 4), we showed that the ComBat method obtained performance equivalent to the prospective correction, at least in our dataset.

Because of their Bayesian nature (15), ComBat-corrected SBRs are under the strong influence of measured values at each site, regardless of the appropriateness of their original values. This influence should not be observed with the scanner and operation correction, which corrects raw data using phantom scan data, and thus SBR values at each site do not affect the correction model. This is an advantage of the prospective correction. Therefore, if an accurate linear model with appropriately minimized residuals were constructed, an SBR should be taken as a golden standard. In our study, the ComBat-corrected SBRs showed lower values than the scanner- and operation-corrected data. It appears that the values after the ComBat correction were strongly influenced by the low overall SBRs from a single site (Figure 1; Supplementary Table S2). Moreover, the site with low SBRs contributed a substantially large number of participants to the



multicenter study compared with other sites. The unbalance of data size across the sites is a limitation of the present study. Further research is needed to assess the generalizability of the conclusion from the present study.

This study was conducted using screen visits from the PADNI study. At this initial visit, each participant who was incorrectly diagnosed by SBR with site-effect correction had atypical clinical data for their respective groups. Another limitation of this study was that PD diagnosis depended on clinical symptoms, levodopa responsibility, and olfactory tests alone, without intense tests to exclude atypical parkinsonism, for example, with  $^{123}\text{I}$ -metaiodobenzylguanidine SPECT. The PADNI will follow up with these participants to monitor possible parkinsonism progression and decreased SBR.

We compared the correction methods for evaluating dopaminergic terminal loss in multisite DAT-SPECT data. The site-effect correction improved the diagnostic accuracy and effect size, despite lacking data from healthy participants at one institute. A multisite database with a completely standardized SBR will enable reliable, large-scale multisite research, overcoming the study-wise limitation at each site. Furthermore, the ComBat correction reasonably improved the diagnostic accuracy of PD compared to prospective scanner correction. The ComBat correction is applicable during unavailable phantom scanning, for example, to compare data with publicly available datasets.

We conducted a comparative analysis of correction methods to assess dopaminergic terminal loss in multisite DAT-SPECT data. The site-effect correction led to improvements in diagnostic accuracy and effect size, even in the absence of data from healthy subjects at one institute. The establishment of a multisite database featuring fully standardized SBRs holds the potential to facilitate reliable, large-scale multisite research, effectively mitigating the inherent limitations of individual sites. Furthermore, the ComBat correction method demonstrated a noteworthy enhancement in the diagnostic accuracy of PD diagnoses, approaching the performance of prospective correction. Importantly, ComBat correction offers the advantage of applicability in scenarios where phantom scanning data are unavailable, making it a valuable tool for comparisons with publicly available datasets.

In this study, some sites included sample sizes for each diagnosis that were insufficient to generate a linear model that included ComBat. At least for this dataset, the ComBat method achieved comparable performance to prospective correction, but further research is needed to assess generalizability more accurately.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the National Center of Neurology and Psychiatry Ethics Committee, Japan. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

NW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HT: Investigation, Writing – review & editing. MA: Project administration, Validation, Writing – review & editing. NS: Project administration, Validation, Writing – review & editing. TMu: Project administration, Validation, Writing – review & editing. TMi: Project administration, Validation, Writing – review & editing. TMA: Project administration, Validation, Writing – review & editing. RY: Project administration, Validation, Writing – review & editing. HY: Project administration, Validation, Writing – review & editing. HM: Supervision, Writing – review & editing. TH: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was partially supported by the Japan Agency for Medical Research and Development (AMED, [www.amed.go.jp](http://www.amed.go.jp)) (18dm0207070, 18dm0307003) and the Japan Society for the Promotion of Science (JSPS, [www.jps.go.jp](http://www.jps.go.jp)) KAKENHI (19H05726) to TH.

## Acknowledgments

We thank the Editage group (<https://www.editage.jp/>) for editing the draft of this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2024.1306546/full#supplementary-material>

## References

- Iwabuchi Y, Kameyama M, Matsusaka Y, Narimatsu H, Hashimoto M, Seki M, et al. A diagnostic strategy for parkinsonian syndromes using quantitative indices of DAT SPECT and MIBG scintigraphy: an investigation using the classification and regression tree analysis. *Eur J Nucl Med Mol Imaging*. (2021) 48:1833–41. doi: 10.1007/s00259-020-05168-0
- Matsuda H, Murata M, Mukai Y, Sako K, Ono H, Toyama H, et al. Japanese multicenter database of healthy controls for [(123)I]FP-CIT SPECT. *Eur J Nucl Med Mol Imaging*. (2018) 45:1405–16. doi: 10.1007/s00259-018-3976-5
- Shimizu S, Hirao K, Kanetaka H, Namioka N, Hatanaka H, Hirose D, et al. Utility of the combination of DAT SPECT and MIBG myocardial scintigraphy in differentiating dementia with Lewy bodies from Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. (2016) 43:184–92. doi: 10.1007/s00259-015-3146-y
- Postuma RB, Iranzo A, Hu M, Högl B, Boeve BF, Manni R, et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*. (2019) 142:744–59. doi: 10.1093/brain/awz030
- Nam SB, Kim K, Kim BS, Im HJ, Lee SH, Kim SJ, et al. The effect of obesity on the availabilities of dopamine and serotonin transporters. *Sci Rep*. (2018) 8:4924. doi: 10.1038/s41598-018-22814-8
- Pak K, Kim K, Lee MJ, Lee JM, Kim BS, Kim SJ, et al. Correlation between the availability of dopamine transporter and olfactory function in healthy subjects. *Eur Radiol*. (2018) 28:1756–60. doi: 10.1007/s00330-017-5147-7
- Eusebio A, Azulay JP, Ceccaldi M, Girard N, Mundler O, Guedj E. Voxel-based analysis of whole-brain effects of age and gender on dopamine transporter SPECT imaging in healthy subjects. *Eur J Nucl Med Mol Imaging*. (2012) 39:1778–83. doi: 10.1007/s00259-012-2207-8
- Honkanen EA, Nojonen T, Hirvilampi R, Lindholm K, Parkkola R, Joutsa J, et al. Sex correction improves the accuracy of clinical dopamine transporter imaging. *EJNMMI Res*. (2021) 11:82. doi: 10.1186/s13550-021-00825-3
- Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, et al. European multicentre database of healthy controls for [(123)I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. *Eur J Nucl Med Mol Imaging*. (2013) 40:213–27. doi: 10.1007/s00259-012-2276-8
- Werner RA, Lapa C, Sheikhbahei S, Marcus C, Solnes LB, Du Y, et al. Impact of aging on semiquantitative uptake parameters in normal rated clinical baseline [(123)I]Ioflupane single photon emission computed tomography/computed tomography. *Nucl Med Commun*. (2019) 40:1001–4. doi: 10.1097/MNM.0000000000001061Supplementa ry Material
- Nobili F, Naseri M, De Carli F, Asenbaum S, Booij J, Darcourt J, et al. Automatic semi-quantification of [(123)I]FP-CIT SPECT scans in healthy volunteers using bas Gan version 2: results from the ENC-DAT database. *Eur J Nucl Med Mol Imaging*. (2013) 40:565–73. doi: 10.1007/s00259-012-2304-8
- Buchert R, Kluge A, Tossici-Bolt L, Dickson J, Bronzel M, Lange C, et al. Reduction in camera-specific variability in [(123)I]FP-CIT SPECT outcome measures by image reconstruction optimized for multisite settings: impact on age-dependence of the specific binding ratio in the ENC-DAT database of healthy controls. *Eur J Nucl Med Mol Imaging*. (2016) 43:1323–36. doi: 10.1007/s00259-016-3309-5
- Buchert R, Lange C, Spehl TS, Apostolova I, Frings L, Jonsson C, et al. Diagnostic performance of the specific uptake size index for semi-quantitative analysis of I-123-FP-CIT SPECT: harmonized multi-center research setting versus typical clinical single-camera setting. *EJNMMI Res*. (2019) 9:37. doi: 10.1186/s13550-019-0506-9
- Tossici-Bolt L, Hoffmann SM, Kemp PM, Mehta RL, Fleming JS. Quantification of [(123)I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio. *Eur J Nucl Med Mol Imaging*. (2006) 33:1491–9. doi: 10.1007/s00259-006-0155-x
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. (2007) 8:118–27. doi: 10.1093/biostatistics/kxj037
- Maikusa N, Zhu Y, Uematsu A, Yamashita A, Saotome K, Okada N, et al. Comparison of traveling-subject and com bat harmonization methods for assessing structural brain characteristics. *Hum Brain Mapp*. (2021) 42:5278–87. doi: 10.1002/hbm.25615
- Schmitz-Steinkrüger H, Lange C, Apostolova I, Amthauer H, Lehnert W, Klutmann S, et al. Impact of the size of the normal database on the performance of the specific binding ratio in dopamine transporter SPECT. *EJNMMI Phys*. (2020) 7:34. doi: 10.1186/s40658-020-00304-z
- Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. (2015) 30:1591–601. doi: 10.1002/mds.26424
- Koike S, Tanaka SC, Okada T, Aso T, Yamashita A, Yamashita O, et al. Brain/MINDS beyond human brain MRI project: a protocol for multi-level harmonization across brain disorders throughout the lifespan. *Neuroimage Clin*. (2021) 30:102600. doi: 10.1016/j.nicl.2021.102600
- Tossici-Bolt L, Dickson JC, Sera T, de Nijs R, Bagnara MC, Jonsson C, et al. Calibration of gamma camera systems for a multicentre European 123I-FP-CIT SPECT normal database. *Eur J Nucl Med Mol Imaging*. (2011) 38:1529–40. doi: 10.1007/s00259-011-1801-5
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. (1975) 12:189–98. doi: 10.1016/0022-3956(75)90026-6
- Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. (1993) 43:2412–4. doi: 10.1212/wnl.43.11.2412-a
- Miyamoto T, Miyamoto M, Iwanami M, Kobayashi M, Nakamura M, Inoue Y, et al. The REM sleep behavior disorder screening questionnaire: validation study of a Japanese version. *Sleep Med*. (2009) 10:1151–4. doi: 10.1016/j.sleep.2009.05.007
- Fujiwara Y, Suzuki H, Yasunaga M, Sugiyama M, Ijuin M, Sakuma N, et al. Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal cognitive assessment. *Geriatr Gerontol Int*. (2010) 10:225–32. doi: 10.1111/j.1447-0594.2010.00585.x
- Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a frontal assessment battery at bedside. *Neurology*. (2000) 55:1621–6. doi: 10.1212/WNL.55.11.1621
- Tombaugh TN. Trail making test a and B: normative data stratified by age and education. *Arch Clin Neuropsychol*. (2004) 19:203–14. doi: 10.1016/S0887-6177(03)00039-8
- Kobayashi M, Nishida K, Nakamura S, Oishi M, Shiozaki T, Majima Y, et al. Suitability of the odor stick identification test for the Japanese in patients suffering from olfactory disturbance. *Acta Otolaryngol Suppl*. (2004) 124:74–9. doi: 10.1080/03655230410017715
- Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization [version 1; peer review: 2 approved]. *Wellcome Open Res*. (2019) 4:1–46. doi: 10.12688/wellcomeopenres.15191.1
- Suwijn SR, van Boheemen CJ, de Haan RJ, Tissingh G, Booij J, de Bie RM. The diagnostic accuracy of dopamine transporter SPECT imaging to detect nigrostriatal cell loss in patients with Parkinson's disease or clinically uncertain parkinsonism: a systematic review. *EJNMMI Res*. (2015) 5:1–8. doi: 10.1186/s13550-015-0087-1
- Schmitz-Steinkrüger H, Lange C, Apostolova I, Mathies FL, Frings L, Klutmann S, et al. Impact of age and sex correction on the diagnostic performance of dopamine transporter SPECT. *Eur J Nucl Med Mol Imaging*. (2021) 48:1445–59. doi: 10.1007/s00259-020-05085-2
- Tossici-Bolt L, Dickson JC, Sera T, Booij J, Asenbaum S, Bagnara MC, et al. [(123)I]FP-CIT ENC-DAT normal database: the impact of the reconstruction and quantification methods. *EJNMMI Phys*. (2017) 4:8. doi: 10.1186/s40658-017-0175-6
- Hustad E, Aasly JO. Clinical and imaging markers of prodromal Parkinson's disease. *Front Neurol*. (2020) 11:395. doi: 10.3389/fneur.2020.00395
- Iranzo A, Santamaría J, Valldeoriola F, Serradell M, Salamero M, Gaig C, et al. Dopamine transporter imaging deficit predicts early transition to synucleinopathy in idiopathic rapid eye movement sleep behavior disorder. *Ann Neurol*. (2017) 82:419–28. doi: 10.1002/ana.25026
- Fearnley JM, Lees AJ. Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain*. (1991) 114:2283–301. doi: 10.1093/brain/114.5.2283
- Vijjaratnam N, Simuni T, Bandmann O, Morris HR, Foltynie T. Progress towards therapies for disease modification in Parkinson's disease. *Lancet Neurol*. (2021) 20:559–72. doi: 10.1016/S1474-4422(21)00061-2
- LeWitt PA, Aradi SD, Hauser RA, Rascol O. The challenge of developing adenosine A2A antagonists for Parkinson disease: Istradefylline, praladetan, and tozadenant. *Parkinsonism Relat Disord*. (2020) 80:S54–63. doi: 10.1016/j.parkreldis.2020.10.027
- Stocchi F, Rascol O, Hauser RA, Huyck S, Tzontcheva A, Capece R, et al. Randomized trial of praladetan, given as monotherapy, in patients with early Parkinson disease. *Neurology*. (2017) 88:2198–206. doi: 10.1212/WNL.0000000000004003
- Ceravolo R, Frosini D, Poletti M, Kiferle L, Pagni C, Mazzucchi S, et al. Mild affective symptoms in de novo Parkinson's disease patients: relationship with dopaminergic dysfunction. *Eur J Neurol*. (2013) 20:480–5. doi: 10.1111/j.1468-1331.2012.03878.x
- Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson progression marker initiative (PPMI). *Prog Neurobiol*. (2011) 95:629–35. doi: 10.1016/j.pneurobio.2011.09.005