
Individual-Reader Diagnostic Performance and Between-Reader Agreement in Assessment of Subjects with Parkinsonian Syndrome or Dementia Using ¹²³I-Ioflupane Injection (DaTscan) Imaging

John P. Seibyl*¹, Andreas Kupsch², Jan Booij³, Donald G. Grosset⁴, Durval C. Costa⁵, Robert A. Hauser⁶, Jacques Darcourt⁷, Nin Bajaj⁸, Zuzana Walker^{9,10}, Kenneth Marek¹, Ian McKeith¹¹, John T. O'Brien¹², Klaus Tatsch¹³, Eduardo Tolosa¹⁴, Rudi A. Dierckx¹⁵, and Igor D. Grachev*¹⁶

¹Institute for Neurodegenerative Disorders, New Haven, Connecticut; ²Department of Neurology and Stereotactic Neurosurgery, Otto-von-Guericke-University, Magdeburg, Germany; ³Department of Nuclear Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; ⁴Institute of Neurological Sciences, Southern General Hospital and University of Glasgow, Glasgow, Scotland; ⁵Nuclear Medicine-Radiopharmacology, Champalimaud Centre for the Unknown, Champalimaud Foundation, Lisbon, Portugal; ⁶Departments of Neurology, Molecular Pharmacology and Physiology, USF Health Byrd Parkinson's Disease and Movement Disorders Center, National Parkinson Foundation Center of Excellence, Tampa, Florida; ⁷Department of Nuclear Medicine, Centre Antoine Lacassagne, TIRO Research Group, University of Nice Sophia-Antipolis, Nice, France; ⁸Department of Clinical Neurology, National Parkinson's Foundation International Centre of Excellence, Nottingham University Hospitals NHS Trust and University of Nottingham, Queen's Medical Centre Campus, Nottingham, United Kingdom; ⁹Division of Psychiatry, University College London, London, United Kingdom; ¹⁰North Essex Partnership University NHS Foundation Trust, Essex, United Kingdom; ¹¹Campus for Ageing and Vitality, Newcastle University, Newcastle upon Tyne, United Kingdom; ¹²Department of Psychiatry, University of Cambridge School of Clinical Medicine and Cambridgeshire and Peterborough Foundation NHS Trust, Cambridge, United Kingdom; ¹³Department of Nuclear Medicine, Municipal Hospital Karlsruhe, Inc., Karlsruhe, Germany; ¹⁴Neurology Service, Hospital Clínic de Barcelona, Universitat de Barcelona, IDIBAPS, Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Catalonia, Spain; ¹⁵Departments of Radiology, Nuclear Medicine, and Molecular Imaging, University Medical Centre Groningen, Groningen, The Netherlands; and ¹⁶Medical Affairs, Life Sciences, GE Healthcare, Princeton, New Jersey

Establishing an early, accurate diagnosis is fundamental for appropriate clinical management of patients with movement disorders or dementia. Ioflupane ¹²³I Injection (DaTscan, ¹²³I-ioflupane) is an important adjunct to support the clinical diagnosis. Understanding individual-reader diagnostic performance of ¹²³I-ioflupane in a variety of clinical scenarios is essential. **Methods:** Sensitivity, specificity, interreader, and intrareader data from 5 multicenter clinical studies were reviewed. The different study designs offered an assortment of variables to assess the effects on the diagnostic performance of ¹²³I-ioflupane: on-site versus 3–5 blinded image readers, number of image evaluations, early/uncertain versus late/confirmed clinical diagnosis as reference standard, and subjects with movement disorders versus dementia. **Results:** Eight hundred eighteen subjects had individual-reader efficacy data available for analysis. In general, sensitivity and specificity were high and comparable between on-site versus blinded independent readers. In subjects with dementia, when the clinical diagnosis was made at month 12 versus baseline, specificity improved from 77.4%–91.2% to 81.6%–95.0%. In subjects with movement disorders, this effect was observed to an even greater extent, when diagnostic performance using month-18 diagnosis as

a reference standard (sensitivity, 67.0%–73.7%; specificity, 75.0%–83.3%) was compared versus month-36 diagnosis (77.5%–80.3% and 90.3%–96.8%, respectively). Diagnostic performance was similar in subjects with dementia (74.4%–89.9% and 77.4%–95.0%, respectively) and subjects with movement disorders (67.0%–97.9% and 71.4%–98.4%, respectively). In most of the comparisons, between-reader agreement was very good (almost perfect), with κ ranging from 0.81 to 1.00. Within-reader agreement, measured in 1 study, was 100% for 3 blinded readers. **Conclusion:** Individual-reader diagnostic performance, as assessed by measuring sensitivity and specificity of ¹²³I-ioflupane to detect the presence or absence of striatal dopaminergic deficit, using the clinical diagnosis as a reference standard, was high in subjects with either movement disorders or dementia and was similar in on-site readers versus blinded analyses. Between- and within-reader agreements were very good (almost perfect). Longer follow-up between imaging and clinical diagnosis improved the diagnostic accuracy, most likely due to improvement in the clinical diagnosis reference standard, rather than changes in reader accuracy.

Key Words: sensitivity; specificity; diagnostic performance; ¹²³I-ioflupane; DaTscan

J Nucl Med 2014; 55:1288–1296

DOI: 10.2967/jnumed.114.140228

Received Mar. 22, 2014; revision accepted Apr. 28, 2014.

For correspondence or reprints contact: John Seibyl, Institute for Neurodegenerative Disorders, 60 Temple St., 8A, New Haven, CT 06510.

E-mail: jseibyl@mnimaging.com

*Contributed equally to this work.

Published online Jun. 12, 2014.

COPYRIGHT © 2014 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

Clinical diagnosis of movement disorders is challenging, particularly to the general neurologist, because of the subtlety and the

TABLE 1
 Details of 3 Phase-3 and 2 Phase-4 ¹²³I-ioflupane SPECT Studies Included in This Analysis

Study	Study phase and design	Study period and population	No. of study centers	Subjects enrolled/evaluable for efficacy	No. of image readers and no./timing of reads
A (4)	<ul style="list-style-type: none"> Phase 3 Multicenter, open-label, nonrandomized Single-dose Expert clinical diagnosis at baseline according to published consensus criteria as the RCD 	August 1997 to February 1998; healthy volunteers and subjects with clinical diagnosis of Parkinson disease, multiple-system atrophy, progressive supranuclear palsy, or essential tremor	6	250/220	<ul style="list-style-type: none"> On-site reader blinded to subjects' clinical information Panel of 5 readers blinded to clinical information Once, at baseline
B (5,6)	<ul style="list-style-type: none"> Phase 3 Multicenter, open-label, nonrandomized Single-dose Expert clinical diagnosis at 12 mo as the RCD 	December 2003 to June 2006; subjects with dementia (features of possible DLB, Alzheimer disease, or vascular dementia)	40	351/326	<ul style="list-style-type: none"> On-site reader Panel of 3 readers blinded to clinical information Once, at baseline
C (1)	<ul style="list-style-type: none"> Phase 3 Multicenter, open-label, nonrandomized Repeat-dose (maximum of 3) Expert clinical diagnosis at 36 mo as the RCD 	January 1999 to June 2005; healthy volunteers and subjects with clinical features of early Parkinson disease or tremor (mainly essential tremor)	10	202/102	<ul style="list-style-type: none"> On-site reader blinded to subjects' clinical information Panel of 3 readers blinded to clinical information Three: baseline, month 18, month 36
D (3,8)	<ul style="list-style-type: none"> Phase 4 Multicenter, open-label, nonrandomized Single-dose Expert clinical diagnosis at 24 mo as the RCD 	November 2000 to November 2003; subjects with movement disorder (uncertain as to PS or non-PS)	15	125/78	On-site reader with access to subjects' clinical information
E (2,7)	<ul style="list-style-type: none"> Phase 4 Multicenter, open-label, randomized Single-dose Expert clinical diagnosis at 12 mo as the RCD 	October 2006 to January 2011; subjects with clinically uncertain Parkinsonian syndromes	19	273 (135 in imaging group)/92 (imaging group)	<ul style="list-style-type: none"> On-site reader blinded to subjects' clinical information Once, at baseline

RCD = Reference clinical diagnosis.

lack of specificity of signs and symptoms in the early stages of the disease (1,2). In some clinical situations, definitive diagnosis is only possible through postmortem neuropathology, and all in-life diagnoses are probable. The advent of in vivo imaging agents that reflect relevant pathophysiology, such as presynaptic nigrostriatal dopaminergic reduction, provides specialists and general practitioners a diagnostic tool that may increase diagnostic confidence and guide treatment. Conceivably, this critically depends on the reliability of individual image readers to accurately identify a normal versus abnormal scan findings.

Numerous clinical trials have been performed to assess the diagnostic accuracy of Ioflupane ¹²³I Injection (DaTscan or DaTSCAN or ¹²³I-ioflupane [GE Healthcare] or ¹²³I-FP-CIT) in detecting the presence or absence of striatal dopaminergic deficit (SDD). These trials varied in their clinical design, providing several variables that can be assessed to determine their impact on diagnostic accuracy. We identified the 5 available phase 3/4 multicenter clinical trials. They

assessed variables that are potentially relevant to individual-reader diagnostic performance, including on-site versus blinded-image-evaluation (BIE) readers; diagnostic cohort, for example, subjects with movement disorders (Parkinsonian syndrome [PS]) versus dementia (dementia with Lewy bodies [DLB]); early/uncertain versus late/confirmed diagnosis; duration of follow-up period after imaging to establish a final clinical diagnosis as a reference standard; and serial imaging over the course of disease progression. Between- and within-reader agreement, when available, were also evaluated to gain better insight regarding the efficacy of ¹²³I-ioflupane imaging in disparate clinical scenarios, to establish best practices, based on evidence.

MATERIALS AND METHODS

The 5 available clinical studies (3 phase 3 and 2 phase 4) (1–8) were used for this pooled analysis (not a meta-analysis of peer-reviewed

TABLE 2
Demographic Characteristics and Clinical Diagnosis (per Reference Standard) by Study

Characteristic and diagnosis	Study					Total (n = 818)
	A (n = 220)	B (n = 326)	C (n = 102)	D (n = 78)	E (n = 92)	
Age (y)						NA
Mean ± SD	62.7 ± 8.87	73.9 ± 7.17	60.4 ± 10.91	64.2 ± 11.99	67.9 ± 12.21	
Minimum	40	54	33	25	19	
Maximum	80	90	79	84	87	
Median	63.5	75.0	61.0	67.0	68.0	
Sex						
Male	136 (62)	187 (57)	57 (56)	41 (53)	50 (54)	471 (58)
Female	84 (38)	139 (43)	45 (44)	37 (47)	42 (46)	347 (42)
Race						
Caucasian	216 (98)	326 (100)	102 (100)	77 (99)	91 (99)	812 (99)
Black	3 (1)	0 (0)	0 (0)	0 (0)	1 (1)	4 (<1)
Asian	1 (<1)	0 (0)	0 (0)	1 (1)	0 (0)	2 (<1)
Other	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
PS (SDD)	158 (72)	0 (0)	71 (70)	48 (62)	50 (54)	327 (40)
Possible PS	158 (72)	0 (0)	5 (5)	48 (62)	0	211 (26)
Probable PS	0 (0)	0 (0)	66 (65)	0 (0)	50 (54)	116 (14)
DLB (SDD)	0 (0)	116 (36)	0 (0)	0 (0)	0 (0)	116 (14)
Possible DLB	0 (0)	27 (8)	0 (0)	0 (0)	0 (0)	27 (3)
Probable DLB	0 (0)	89 (27)	0 (0)	0 (0)	0 (0)	89 (11)
Non-PS/Non-DLB (no SDD)	62 (28)	210 (64)	31 (30)	30 (38)	42 (46)	375 (46)
ET	27 (12)	0 (0)	14 (14)	23 (29)	20 (22)	84 (10)
AD	0 (0)	125 (38)	0 (0)	0 (0)	0 (0)	125 (15)
Other*	35 (16)	85* (26)	17 (17)	7 (9)	22 (24)	166 (20)
SDD present†	158 (72)	116 (36)	71 (70)	48 (62)	50 (54)	443 (54)
SDD absent	62 (28)	210 (64)	31 (30)	30 (38)	42 (46)	375 (46)

*Other includes other (n = 1), no diagnosis (n = 22), and not assessed at 12 mo (n = 62).

†Includes possible and probable PS and possible and probable DLB diagnoses.

NA = not applicable; ET = essential tremor; AD = Alzheimer disease.

Intent-to-diagnose populations used for studies A (4), B (5,6), C (1), and D (3,8). Per-protocol population used for study E (2,7). Data in parentheses are percentages.

publications). All studies were prospective trials with similar designs and objectives—namely, to evaluate the sensitivity and specificity of ¹²³I-ioflupane in detecting the presence or absence of an SDD. Expert clinical diagnosis by a consensus panel or on-site clinical evaluation was used as the reference standard in all studies; however, the timing of when the diagnosis was made relative to duration of illness varied. All of these studies complied with the current version of the Declaration of Helsinki, the International Conference on Harmonization Good Clinical Practice Consolidated Guideline, and applicable laws. Ethics committees or institutional review boards approved each study's protocol and amendments. Subjects or their guardians signed written informed consent forms, which included a provision for subsequent analyses, of which this work is an example.

The 5 studies have been published (1–8), including the inclusion/exclusion criteria. Table 1 provides a summary of the details of the studies. In brief, subjects received 1 dose of 111–287 MBq of ¹²³I-ioflupane, with a small number of subjects receiving up to 3 doses (at

baseline, month 18, and month 36). Supplemental Tables 1, 2, and 3 (supplemental materials are available at <http://jnm.snmjournals.org>) provide details of radioactive dose per study and illustrate the consistent acquisition, reconstruction, and processing methods across the studies. SPECT images were acquired using a variety of devices, including both multi- and single-head γ cameras and multidetector single-slice systems. Each γ camera system was capable of SPECT acquisition and reconstruction to produce transverse slices, including a clear visualization of the striatum (i.e., the head of the caudate nucleus and putamen). Images were acquired within 3–6 h after radiotracer injection (1–8). Images were read on-site or by 3 or 5 BIE readers and classified as normal (SDD absent) or abnormal (SDD present) (9). Image readers had several years' neuroimaging experience and had been trained in person by an expert nuclear medicine physician in the evaluation of ¹²³I-ioflupane images; BIE readers were blinded to patient information with the exception of age, because striatal ¹²³I-ioflupane binding decreases normally with aging (10–12). The content of in-person training was consistent across the

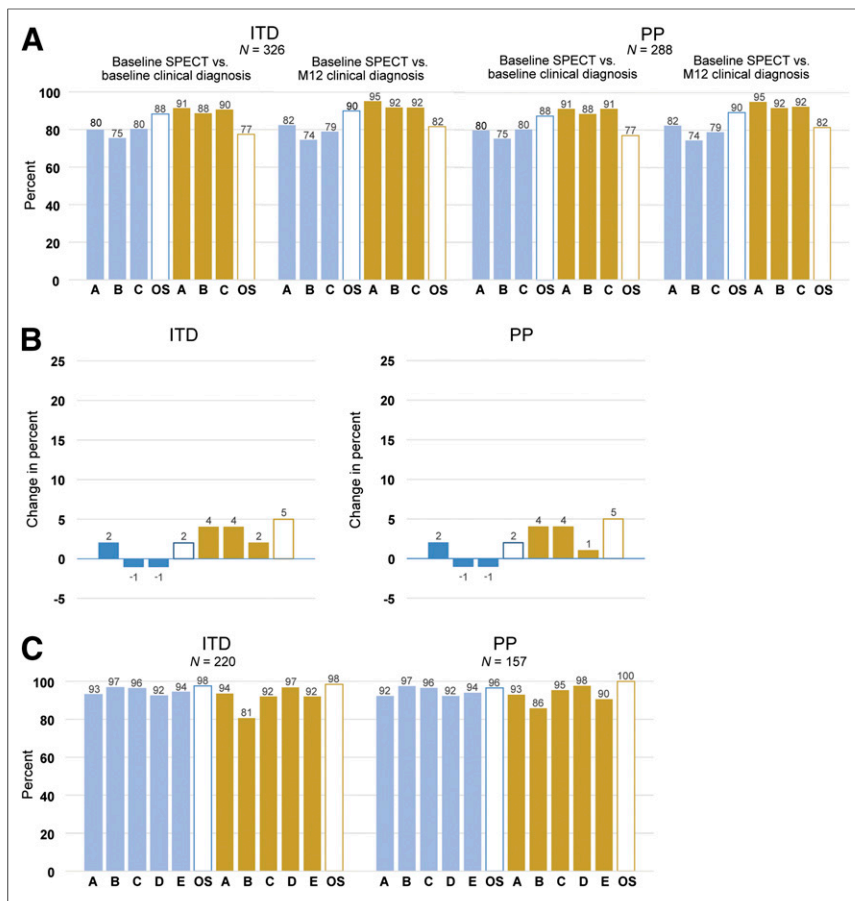


FIGURE 1. (A) Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing ^{123}I -ioflupane SPECT images taken at baseline using clinical diagnosis made at baseline with those at month 12 as reference standard for both ITD and PP populations in subjects with dementia (study B (5,6)). (B) Change in sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers for ^{123}I -ioflupane SPECT images taken at baseline using clinical diagnosis made at baseline vs. at month 12 in subjects with dementia for ITD and PP populations (study B (5,6)). (C) Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing ^{123}I -ioflupane SPECT images taken at baseline using clinical diagnosis made at baseline for ITD and PP populations in subjects with movement disorders (study A (4)). M = month; N = number of subjects; OS = on-site.

5 studies, with only small differences based on the evolving experience obtained from previous trials (13,14).

The studies, although methodologically similar, differed from one another to allow assessment of how common variables encountered in clinical practice might affect the diagnostic performance of the readers. These include comparing blinded (study E (2,7)) with unblinded (study D (3,8)) on-site, institutional readers and comparing on-site with BIE readers (study A (4), study B (5,6), and study C (1)). Disease state, specifically movement disorders (study A (4), study C (1), study D (3,8), and study E (2,7)) versus dementia (study B (5,6)), was evaluated. We also assessed early/uncertain (study C (1), study D (3,8), and study E (2,7)) versus late/confirmed (study A (4)) clinical diagnosis of PS. The timing of when clinical diagnoses (reference standard) were made was evaluated, with all studies including an initial diagnosis at baseline, with others at month 12 (study B (5,6), study E (2,7)), month 18 (study C (1)), month 24 (study D (3,8)), or month 36 (study C (1)). Timing of ^{123}I -ioflupane SPECT imaging was also evaluated, with imaging performed at baseline (all 5 studies), month 18 (study C (1)), and month 36 (study C (1)). Between-reader agreement (BIE

vs. BIE and BIE vs. on-site) was evaluated in 3 of the studies (study A (4), study B (5,6), and study C (1)). Within-reader agreement was assessed in 1 of the studies (study B (5,6)).

Statistical Analysis Software (SAS Institute Inc.) was used to perform the statistical analyses of the studies. Descriptive statistics were used to present demographic data. Populations reported in this paper include intent to diagnose (ITD; all dosed subjects who underwent SPECT imaging and reference clinical diagnosis assessment) and per protocol (PP; all subjects in the ITD populations without a major protocol violation). Sensitivity (equivalent to positive percentage agreement) and specificity (equivalent to negative percentage agreement) were calculated and are reported with 95% confidence intervals (CIs). Individual logistic regression analysis was performed on the effects of duration of follow-up as a continuous variable and reader type (blinded versus on-site). Pairwise between-reader agreements were analyzed using Cohen κ statistic. The Fleiss κ statistic was used as the multiple-summary coefficient for all BIE readers.

RESULTS

Table 2 summarizes the demographic characteristics and clinical diagnoses of the subjects by study. Eight hundred eighteen subjects were included in the ITD efficacy analysis; sex and SDD present/absent were equally represented. The PP population comprised 714 subjects. Reasons for exclusion (subjects may have had more than one) included failure to meet inclusion/exclusion criteria (25), image obtained outside 3- to 6-h window (4), radioactivity dose exceeded 185 MBq (47), lost to follow-up (2), adverse event (1), technical image acquisition issues (11), no final diagnosis available

(28), and protocol violation (3). No images were excluded from reading because of movement artifacts caused by inability to lie still during the scanning procedure.

Diagnostic Accuracy in Dementia and Movement Disorders: On-Site Versus BIE Readers

Sensitivity and specificity were high in subjects with dementia (Fig. 1A) for both the ITD and the PP populations (Supplemental Table 4 provides 95% CIs). On-site reader sensitivity tended to be slightly higher than BIE reader sensitivity in these subjects, whereas specificity tended to be slightly lower (study B (5,6)). However, for subjects with movement disorders (Figs. 1C, 2A and 2B, and 3A and 3B), diagnostic performance for on-site versus BIE readers was similar (Supplemental Table 5 provides 95% CIs). Regression analysis indicated that reader type significantly affected sensitivity ($P = 0.0063$ for ITD and $P = 0.0081$ for PP) and specificity ($P < 0.0001$ for ITD and PP). For phase-4 studies with on-site readers only, sensitivity and specificity were

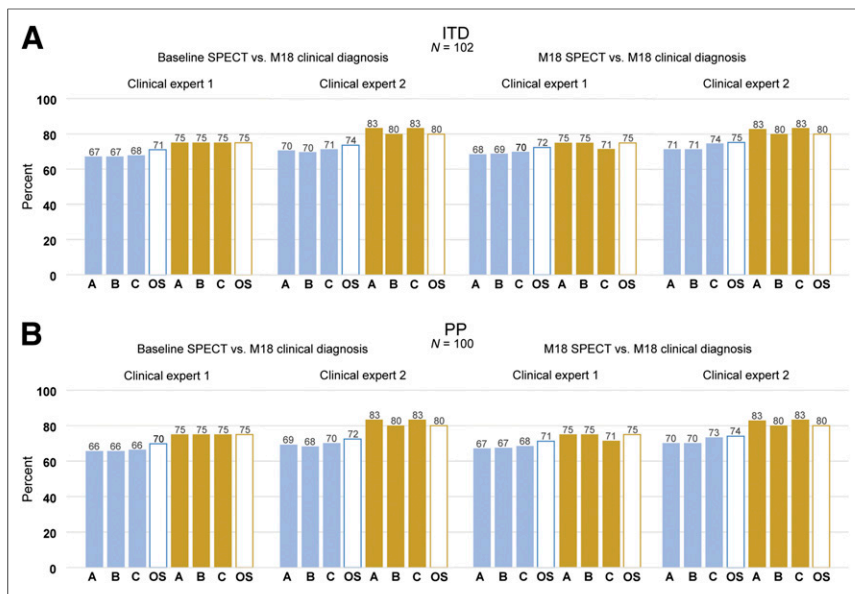


FIGURE 2. Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing ^{123}I -ioflupane SPECT images taken at baseline with those at month 18 using clinical diagnosis made individually by 2 clinical experts at month 18 as reference standard for ITD (A) and PP (B) populations in subjects with movement disorders (study C (7)). M = month; N = number of subjects.

high and similar to other phase-3 studies. One exception was specificity, being lower in study D (nonblinded on-site readers with regard to clinical patient information) (3,8) than in study E

proved when the clinical diagnosis was made later in the disease process. Regression analysis indicated that duration of follow-up significantly affected sensitivity ($P < 0.0001$ for

(blinded on-site readers with regard to clinical patient information) (2,7). Sensitivity and specificity were equally high and comparable in subjects with movement disorders or with dementia.

Timing of Clinical Follow-up

Considering the timing of clinical follow-up and the timing of the clinical diagnosis, there was a slight effect on the diagnostic performance of the reads. Figure 1A shows the sensitivity and specificity across BIE and on-site readers using a clinical diagnosis of DLB made at baseline versus at month 12, and Figure 1B depicts the changes in sensitivity and specificity. Slight improvements in sensitivity and specificity were observed. Figure 2 shows the sensitivity and specificity using the clinical diagnosis of PS made individually by 2 clinical experts at month 18 as the reference standard, whereas Figure 3 depicts results using the month-36 consensus clinical diagnosis. Changes in sensitivity and specificity are displayed in Figure 4. Although the differences are small, reader performance overall improved when the clinical diagnosis was made later in the disease process. Regression analysis indicated that duration of follow-up significantly affected sensitivity ($P = 0.0491$ for ITD and 0.0546 for PP).

Effect of Serial Imaging over Course of Disease on Diagnostic Performance

In 1 of the studies (study C (1)), subjects with movement disorders underwent up to three ^{123}I -ioflupane imaging sessions—that is, at baseline, month 18, and month 36. Figure 3 compares BIE and on-site reader diagnostic performance for each of these imaging sessions using the month 36 clinical diagnosis as the reference standard. Changes in sensitivity and specificity are displayed in Figure 5, showing consistent and stable sensitivity and specificity during the 3-y observation period, independent from the time point of the acquisition of the scans.

Between-Reader Agreement

Figures 6 and 7 depict the between-reader agreement for 3 studies for the ITD and PP populations, respectively, using the Altman method for assessing the strength of agreement (95% CIs available in Supplemental Tables 6 and 7) (15). The κ for most comparisons were consistently very good (almost perfect using Landis and Koch categorizations) (16) whether comparing between

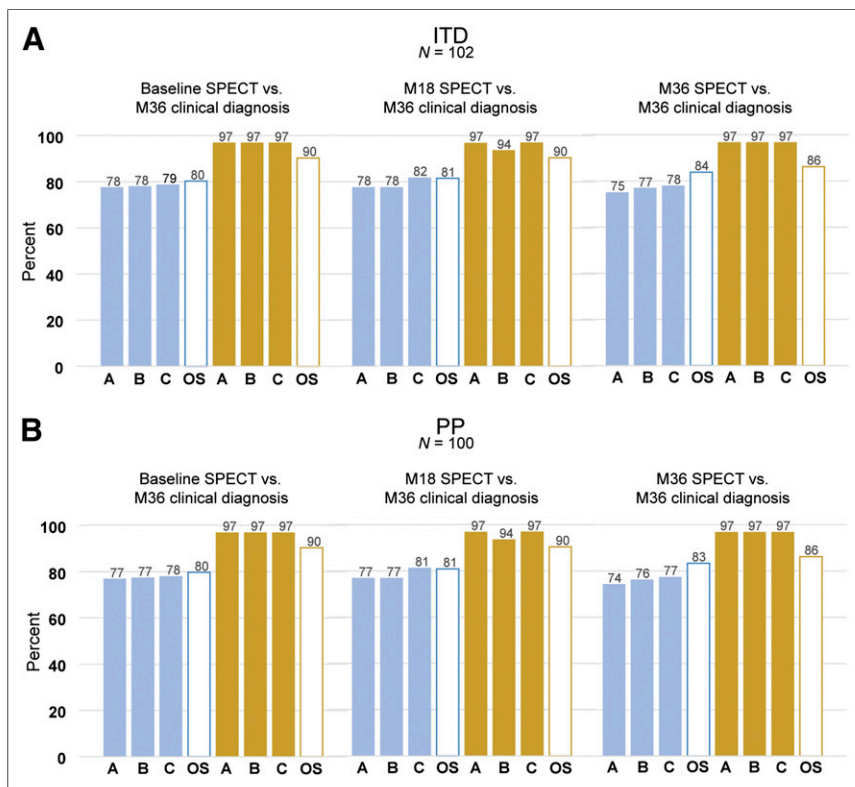


FIGURE 3. Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing ^{123}I -ioflupane SPECT images taken at baseline with those at month 18 versus at month 36 using clinical diagnosis made at month 36 as reference standard for ITD (A) and PP (B) populations in subjects with movement disorders (study C (7)). M = month; N = number of subjects; OS = on-site.

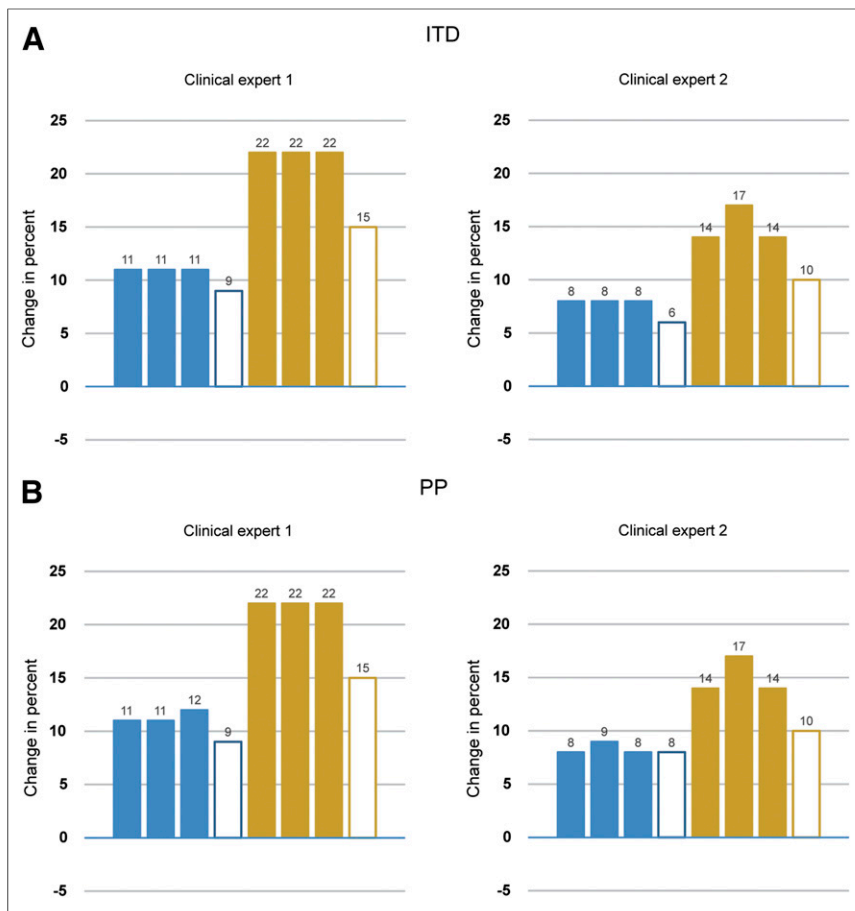


FIGURE 4. Change in sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers for ^{123}I -ioflupane SPECT images taken at baseline using clinical diagnosis made by individual clinical experts at month 18 versus consensus panel at month 36 in subjects with movement disorders for ITD (A) and PP (B) populations (study C (1)). Improvement observed is most likely due to greater accuracy of clinical diagnosis reference standard and not due to changes in interpretation of images.

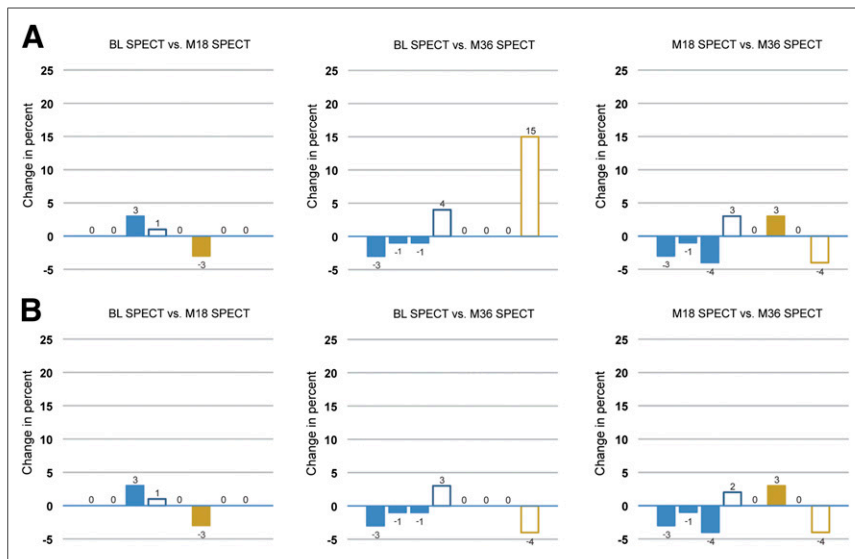


FIGURE 5. Change in sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers for ^{123}I -ioflupane SPECT images taken at baseline versus month 18, baseline versus month 36, and month 18 versus month 36 using clinical diagnosis made at month 36 in subjects with movement disorders for ITD (A) and PP (B) populations (study C (1)). BL = baseline; M = month; N = number of subjects; OS = on-site.

BIE readers or BIE versus on-site readers. Agreement dropped slightly, demonstrating moderate, good, and very good correlation between BIE and on-site readers for reading images from subjects with dementia (Figs. 6C and 7C).

Within-Reader Agreement

In 1 of the studies (study B (5,6)), within-reader agreement was measured. Each of the 3 BIE readers had complete (i.e., 100%) agreement for the 24–26 images read (Supplemental Table 8).

DISCUSSION

This analysis of 5 prospective studies provides a large dataset ($n = 818$) and multiple clinical scenarios in which to evaluate individual ^{123}I -ioflupane SPECT image reader performance. Both on-site readers and BIE readers demonstrated consistently high diagnostic performance. Reader type significantly affected sensitivity and specificity, but the differences did not compromise the diagnostic value of ^{123}I -ioflupane SPECT imaging for either reader types. A slightly higher sensitivity and slightly lower specificity, observed in on-site reading of images from subjects with dementia (study B (5,6)), may be attributed to challenges in maintaining comprehensive blinding. Access to clinical information may have contributed to some loss of objectivity. This trend for lower specificity was also observed in the phase-4 study D (3,8), in which the on-site readers were not blinded. In both study B (5,6) and study D (3,8), the subjects studied had uncertain diagnoses (DLB and clinically uncertain PS, respectively). Conceivably, access to questionable diagnoses may have contributed to reductions in accuracy of the readings. Walker et al. reported that the tendency to overdiagnose DLB contributed to the lack of agreement between the clinical diagnosis and ^{123}I -ioflupane image results (17). Marshall et al. reported a similar tendency in PS (1). We observed that if the clinical diagnosis of DLB was made 12 mo later, with subsequent greater confidence in the diagnosis, slight improvements in specificity were observed. More dramatic increases in sensitivity and specificity were observed when a clinical diagnosis of PS was made at month 36 versus month 18, which again can likely be attributed to greater accuracy when the clinical diagnosis is made later in the disease process. This observed improvement over time was statistically significant by regression analysis. Nonetheless, overall

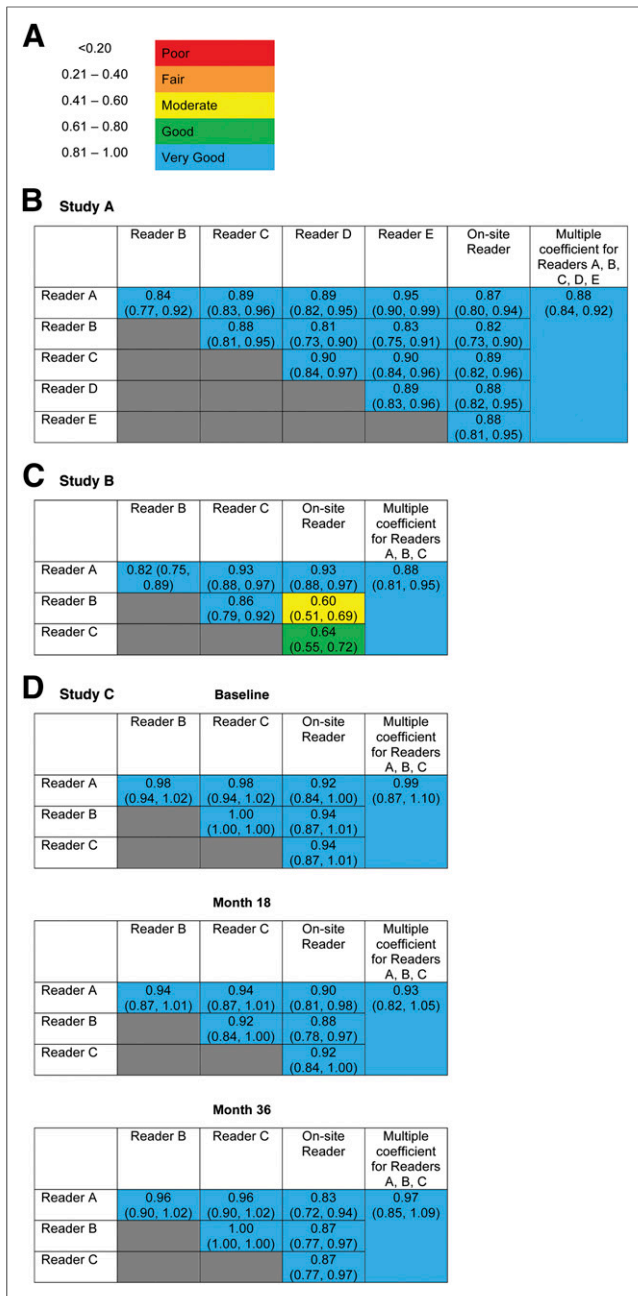


FIGURE 6. Between-reader agreement for 3 studies. κ coefficient (95% CIs)—ITD populations. (A) Strength of agreement for κ . (B) Study A (4)— ^{123}I -ioflupane images taken at baseline in subjects with movement disorders ($n = 220$). (C) Study B (5,6)— ^{123}I -ioflupane images taken at baseline in subjects with dementia ($n = 272$ – 292). (D) Study C (1)— ^{123}I -ioflupane images taken at baseline ($n = 98$ – 102), month 18 ($n = 100$ – 103), and month 36 ($n = 92$ – 97) in subjects with movement disorders. $N =$ number of subjects.

sensitivity and specificity were high in subjects with both movement disorders and dementia.

When patients with movement disorders were serially scanned 3 times over the course of 3 y and the month-36 clinical diagnosis was used as the reference standard, the sensitivity and specificity of the ^{123}I -ioflupane SPECT imaging results were consistently high. These results suggest that significant changes are noted on scans, even early in the course of illness, providing high diagnostic accuracy early in the disease process, and accuracy persists as disease progression occurs.

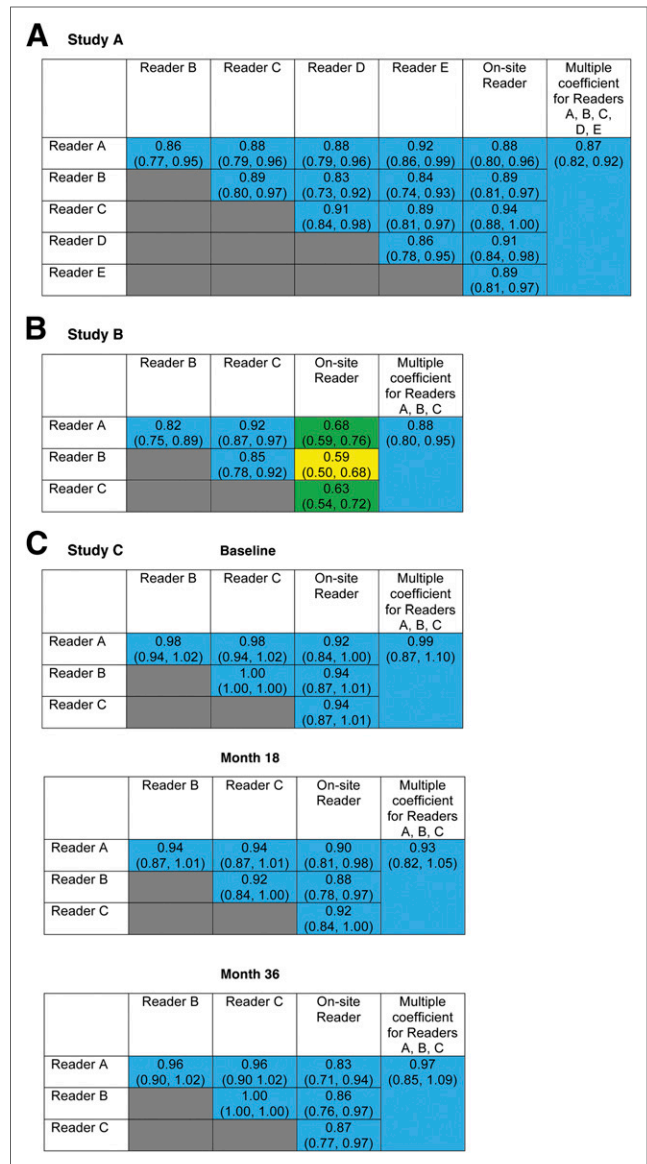


FIGURE 7. Between-reader agreement for 3 studies. κ coefficient (95% CIs)—PP populations. (A) Study A (4)— ^{123}I -ioflupane images taken at baseline in subjects with movement disorders ($n = 157$). (B) Study B (5,6)— ^{123}I -ioflupane images taken at baseline in subjects with dementia ($n = 249$ – 272). (C) Study C (1)— ^{123}I -ioflupane images taken at baseline ($n = 96$ – 100), month 18 ($n = 98$ – 100), and month 36 ($n = 90$ – 95) in subjects with movement disorders. $N =$ number of subjects.

Between-reader agreement was very good (15) (almost perfect) (16), with κ exceeding 0.8 in most comparisons. Agreement dropped to a combination of very good, good, and moderate when comparing BIE readers with on-site readers of images from subjects with DLB, which may be attributable to compromised blinding of the on-site readers or to differences between different disease entities (DLB vs. PS). Overall, the very good between-reader agreement may reflect adequate experience and training of the readers. This interpretation complies with the complete within-reader agreement observed for the 3 readers in the DLB study.

The present analysis has some limitations. Several of the studies summarized in this analysis were performed more than 10 y ago, when there was limited experience with ^{123}I -ioflupane SPECT

imaging. Training methods available today (13,14) had not yet been developed or validated, and nuclear medicine physicians had not yet acquired extensive experience in reading the images (18). However, the strong between-reader agreement observed supports previous conclusions (1) that, on the basis of correct imaging procedures and processing, the visual interpretation of the image is independent of the expert conducting the analysis. Another limitation was that scans were only scored visually; the value of quantitative analysis is evolving and will be incorporated into future studies. Clinical diagnosis was used as the reference standard rather than autopsy. This approach was taken because it is not feasible to wait for neuropathologic confirmation in the clinical research setting with patients in the early stages of disease. Furthermore, expert clinical diagnosis is considered an acceptable reference standard for biomarker validation studies (19). Last, the use of clinical diagnosis as the reference standard may have contributed to minor reductions in specificity, most likely due to potential uncertainties in the clinical diagnosis and possible contribution of subjects who have scans without evidence of dopaminergic deficit. Studies have shown that 11%–21% of patients with an initial diagnosis of Parkinson disease eventually have their diagnoses changed to essential or dystonic tremor (1,20,21). Autopsy confirmation studies have shown ¹²³I-ioflupane SPECT imaging to be very accurate (17,22). The minor reductions observed in specificity in this pooled analysis can be more likely attributed to the uncertainties associated with the clinical diagnoses and less likely to inaccuracies in the diagnostic performance of ¹²³I-ioflupane SPECT imaging. This notion is supported by the improved diagnostic performance of ¹²³I-ioflupane SPECT imaging at longer clinical follow-ups, because initially subtle clinical symptoms may progress to a more characteristic clinical picture (8,23).

CONCLUSION

This combined analysis demonstrates that individual-reader sensitivity and specificity were high for patients with movement disorders and dementia. Slightly lower sensitivities were observed in patients with dementia, most probably due to greater complexity of pathology and less precise clinical diagnosis, the reference standard. These differences were small and do not compromise the value of ¹²³I-ioflupane SPECT imaging. Diagnostic performance was accurate early in the disease process and remained consistently accurate with progression of disease. The sensitivity and specificity were high and relatively stable over the 3-y period of clinical follow-up in PS, regardless of when the scan was taken. Between-reader agreement, for both blinded readers and on-site image readers, was very good (almost perfect) in all populations studied. Within-reader agreement, measured in 1 of the studies, showed complete agreement. This analysis illustrates the utility and robust diagnostic performance of individual readers of ¹²³I-ioflupane SPECT images across a broad variety of clinical scenarios.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734. Financial support to develop this work was provided by GE Healthcare.

Dr. Seibyl is a consultant of GE Healthcare, Piramal, and Navidea and reports personal fees from Molecular Neuroimaging, outside the submitted work. Dr. Kupsch has nothing to disclose.

Dr. Booi reports that he is a consultant of GE Healthcare. Dr. Grosset reports grants and personal fees from Merz Pharma and personal fees from Astellas, Civitas, InVentiv Health, AbbVie, GE Healthcare, Teva, and UCB Pharma, outside the submitted work. Dr. Costa has nothing to disclose. Dr. Hauser reports grants from GE Healthcare, personal fees from GE Healthcare, during the conduct of the study; personal fees from Zanon Company S.p.A, Teva Pharmaceuticals, INC, UCB BIOSCIENCES GmbH, Amdante Steering Committee, AbbVie, Eli Lilly, Cleveland Company, PPMI, Novartis, Biotie, Lundbeck, PharmStrat LLC, GE Healthcare, Allergan, University of Houston, DSMB, Ipsen, Consensus Medical Communications, Pfizer, Azilect Strategic Advisory, Neurocrine Meeting, and Gerson Lehrman; and grants from LS-1, outside the submitted work. Dr. Darcourt has nothing to disclose. Dr. Bajaj reports grants and personal fees from GE Healthcare, outside the submitted work. Dr. Walker reports personal fees from GE Healthcare and Bayer Healthcare, grants from GE Healthcare, grants from Lundbeck, other from GE Healthcare, and personal fees from Novartis, outside the submitted work. Dr. Marek reports personal fees from GE Healthcare and personal fees from Molecular Neuroimaging, outside the submitted work. Dr. McKeith reports grants and personal fees from GE Healthcare, during the conduct of the study; grants and personal fees from GE Healthcare, outside the submitted work. Dr. O’Brien reports grants and other from GE Healthcare, grants and other from Lilly, other from Bayer Healthcare, other from TauRx, other from Cytos, outside the submitted work. Dr. Tatsch reports personal fees from GE Healthcare, during the conduct of the study, and personal fees from GE Healthcare, outside the submitted work. Dr. Tolosa reports personal fees from Novartis, TEVA, Boehringer Ingelheim, UCB, Solvay, Lundbeck, and TEVA; grants from Spaniard Network for research on neurodegenerative Disorders (CIBERNED)—instituto Carlos III (ISCIII), grants from The Michael J. Fox Foundation for Parkinson’s Research (MJFF), and grants from Fondo de Investigaciones Sanitarias de la Seguridad Social (FISS), outside the submitted work. Dr. Dierckx has nothing to disclose. Dr. Grachev reports employment at GE Healthcare, during the conduct of the study. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENT

We acknowledge the writing assistance provided by Stacy Simpson Logan, CMPP, of Winfield Consulting, funded by GE Healthcare.

REFERENCES

1. Marshall VL, Reiningner CB, Marquardt M, et al. Parkinson’s disease is over-diagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [¹²³I]FP-CIT SPECT. *Mov Disord.* 2009;24:500–508.
2. Kupsch AR, Bajaj N, Weiland F, et al. 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.* 2012;83:620–628.
3. Catafau AM, Tolosa E. Impact of dopamine transporter SPECT using ¹²³I-ioflupane on diagnosis and management of patients with clinically uncertain Parkinsonian syndromes. *Mov Disord.* 2004;19:1175–1182.
4. Benamer TS, Patterson J, Grosset DG, et al. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [¹²³I]-FP-CIT SPECT imaging: the [¹²³I]-FP-CIT study group. *Mov Disord.* 2000;15:503–510.
5. McKeith I, O’Brien J, Walker Z, et al. Sensitivity and specificity of dopamine transporter imaging with ¹²³I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study. *Lancet Neurol.* 2007;6:305–313.

6. O'Brien JT, McKeith IG, Walker Z, et al. Diagnostic accuracy of ¹²³I-FP-CIT SPECT in possible dementia with Lewy bodies. *Br J Psychiatry*. 2009;194:34–39.
7. Seibyl J, Grosset DG, Bajaj N, Hauser RA, Marek K, Grachev ID. Accuracy of DaTscan™ [Ioflupane I 123 Injection] in diagnosis of early parkinsonian syndromes (PS) [abstract]. *J Nucl Med*. 2013;54(suppl 2):58.
8. Tolosa E, Borghet TV, Moreno E. Accuracy of DaTSCAN (¹²³I-ioflupane) SPECT in diagnosis of patients with clinically uncertain parkinsonism: 2-year follow-up of an open-label study. *Mov Disord*. 2007;22:2346–2351.
9. Full Prescribing Information for DaTscan (US). GE Healthcare. http://www3.gehealthcare.com/en/Products/Categories/Nuclear_Imaging_Agents/~/_media/Downloads/us/Product/Product-Categories/Nuclear-Imaging-Agents_Non-Gatekeeper/DaTscan/GEHealthcare_DaTscan-Prescribing-Information.pdf. April 2013. Accessed May 18, 2014.
10. Varrone A, Dickson JC, Tossici-Bolt L, et al. 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*. 2013;40:213–227.
11. Lavalaye J, Booij J, Reneman L, Habraken JB, van Royen EA. Effect of age and gender on dopamine transporter imaging with [¹²³I]FP-CIT SPET in healthy volunteers. *Eur J Nucl Med*. 2000;27:867–869.
12. 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;39:1778–1783.
13. Information for Nuclear Imagers. US.datscan.com. <http://us.datscan.com/imagers>. January 2013. Accessed May 18, 2014.
14. Simuni T. Advances in multidisciplinary imaging. Is it essential tremor or a Parkinsonian Syndrome? Medscape.com. <http://www.medscape.org/sites/advances/multidisciplinary-imaging>. June 2013. Accessed June 2, 2014.
15. Altman DG. *Practical Statistics for Medical Research*. London, U.K.: Chapman and Hall; 1991.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
17. Walker Z, Jaros E, Walker RW, et al. Dementia with Lewy bodies: a comparison of clinical diagnosis, FP-CIT single photon emission computed tomography imaging and autopsy. *J Neurol Neurosurg Psychiatry*. 2007;78:1176–1181.
18. Söderlund TA, Dickson JC, Prvulovich E, et al. Value of semiquantitative analysis for clinical reporting of ¹²³I-2-beta-carbomethoxy-3beta-(4-iodophenyl)-N-(3-fluoropropyl)nortropine SPECT studies. *J Nucl Med*. 2013;54:714–722.
19. Gorovets A, Marzella L, Rieves D, Yang L. Efficacy considerations for U.S. Food and Drug Administration approval of diagnostic radiopharmaceuticals. *J Nucl Med*. 2013;54:1479–1484.
20. Whone AL, Watts RL, Stoessl AJ, et al. Slower progression of Parkinson's disease with ropinirole versus levodopa: the REAL-PET study. *Ann Neurol*. 2003;54:93–101.
21. Schneider SA, Edwards MJ, Mir P, et al. Patients with adult-onset dystonic tremor resembling parkinsonian tremor have scans without evidence of dopaminergic deficit (SWEDDs). *Mov Disord*. 2007;22:2210–2215.
22. Colloby SJ, McParland S, O'Brien JT, Attems J. Neuropathological correlates of dopaminergic imaging in Alzheimer's disease and Lewy body dementias. *Brain*. 2012;135:2798–2808.
23. Jennings DL, Seibyl JP, Oakes D, Eberly S, Murphy J, Marek K. (¹²³I)beta-CIT and single-photon emission computed tomographic imaging vs clinical evaluation in Parkinsonian syndrome: unmasking an early diagnosis. *Arch Neurol*. 2004;61:1224–1229.