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

John P. Seibyl*1, Andreas Kupsch2, Jan Booji3, Donald G. Grosset4, Durval C. Costa5, Robert A. Hauser6, Jacques Darcourt7, Nin Bajaj8, Zuzana Walker9,10, Kenneth Marek1, Ian McKeith11, John T. O’Brien12, Klaus Tatsch13, Eduardo Tola14, 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 Neuroradiological 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, 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 Clinico 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 $^{123}$I Injection (DaTscan, $^{123}$I-Ioflupane) is an important adjuent to support the clinical diagnosis. Understanding individual-reader diagnostic performance of $^{123}$I-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 $^{123}$I-Ioflupane in subjects with either movement disorders or 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 $^{123}$I-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; Ioflupane $^{123}$I Injection; DaTscan

DOI: 10.2967/jnumed.114.140228

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 $^{123}$I-Ioflupane SPECT Studies Included in This Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Study phase and design</th>
<th>Study period and population</th>
<th>No. of study centers</th>
<th>Subjects enrolled/evaluable for efficacy</th>
<th>No. of image readers and no/timing of reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (4)</td>
<td>• Phase 3&lt;br&gt;• Multicenter, open-label, nonrandomized&lt;br&gt;• Single-dose&lt;br&gt;• Expert clinical diagnosis at baseline according to published consensus criteria as the RCD</td>
<td>August 1997 to February 1998; healthy volunteers and subjects with clinical diagnosis of Parkinson disease, multiple-system atrophy, progressive supranuclear palsy, or essential tremor</td>
<td>6</td>
<td>250/220</td>
<td>• On-site reader blinded to subjects' clinical information&lt;br&gt;• Panel of 5 readers blinded to clinical information&lt;br&gt;• Once, at baseline</td>
</tr>
<tr>
<td>B (5,6)</td>
<td>• Phase 3&lt;br&gt;• Multicenter, open-label, nonrandomized&lt;br&gt;• Single-dose&lt;br&gt;• Expert clinical diagnosis at 12 mo as the RCD</td>
<td>December 2003 to June 2006; subjects with dementia (features of possible DLB, Alzheimer disease, or vascular dementia)</td>
<td>40</td>
<td>351/326</td>
<td>• On-site reader&lt;br&gt;• Panel of 3 readers blinded to clinical information&lt;br&gt;• Once, at baseline</td>
</tr>
<tr>
<td>C (1)</td>
<td>• Phase 3&lt;br&gt;• Multicenter, open-label, nonrandomized&lt;br&gt;• Repeat-dose (maximum of 3)&lt;br&gt;• Expert clinical diagnosis at 36 mo as the RCD</td>
<td>January 1999 to June 2005; healthy volunteers and subjects with clinical features of early Parkinson disease or tremor (mainly essential tremor)</td>
<td>10</td>
<td>202/102</td>
<td>• On-site reader blinded to subjects' clinical information&lt;br&gt;• Panel of 3 readers blinded to clinical information&lt;br&gt;• Three: baseline, month 18, month 36</td>
</tr>
<tr>
<td>D (3,8)</td>
<td>• Phase 4&lt;br&gt;• Multicenter, open-label, nonrandomized&lt;br&gt;• Single-dose&lt;br&gt;• Expert clinical diagnosis at 24 mo as the RCD</td>
<td>November 2000 to November 2003; subjects with movement disorder (uncertain as to PS or non-PS)</td>
<td>15</td>
<td>125/78</td>
<td>On-site reader with access to subjects' clinical information</td>
</tr>
<tr>
<td>E (2,7)</td>
<td>• Phase 4&lt;br&gt;• Multicenter, open-label, randomized&lt;br&gt;• Single-dose&lt;br&gt;• Expert clinical diagnosis at 12 mo as the RCD</td>
<td>October 2006 to January 2011; subjects with clinically uncertain Parkinsonian syndromes</td>
<td>19</td>
<td>273 (135 in imaging group)/92 (imaging group)</td>
<td>• On-site reader blinded to subjects' clinical information&lt;br&gt;• Once, at baseline</td>
</tr>
</tbody>
</table>

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 $^{123}$I Injection (DaTscan or DaTScan or $^{123}$I-Ioflupane [GE Healthcare] or $^{123}$I-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 $^{123}$I-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
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 rela-
tive 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 consis-
tent acquisition, reconstruction, and processing methods across the
studies. SPECT images were acquired using a variety of devices, in-
cluding both multi- and single-head γ cameras and multidetector sin-
gle-slice systems. Each γ camera system was capable of SPECT ac-
quisition 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 ra-
diotrace 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 experi-
ence and had been trained in person by an expert nuclear medicine
physician. 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

### Table 1
Demographic Characteristics and Clinical Diagnosis (per Reference Standard) by Study

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

*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.
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)) or 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 (J)). Disease state, specifically movement disorders (study A (4), study C (J), study D (3,8), and study E (2,7)) versus dementia (study B (5,6)), was evaluated. We also assessed early/uncertain (study C (J), 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 (J)), month 24 (study D (3,8)), or month 36 (study C (J)). Timing of 123I-ioflupane SPECT imaging was also evaluated, with imaging performed at baseline (all 5 studies), month 18 (study C (J)), and month 36 (study C (J)). 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 (J)). 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 \( \kappa \) statistic. The Fleiss \( \kappa \) 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
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 (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 specificity were observed. Figure 2 shows the sensitivity and specificity using the clinical diagnosis of PD 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 ($P < 0.0001$ for ITD and PP) and specificity in the ITD population ($P = 0.0491$ 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-iофлюпан 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 $\kappa$ for most comparisons were consistently very good (almost perfect using Landis and Koch categorizations) (16) whether comparing between

---

**FIGURE 2.** Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing $^{123}$I-iофлюпан 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 (1)). M = month; N = number of subjects.

**FIGURE 3.** Sensitivity (blue) and specificity (gold) for individual BIE (solid bars) and on-site (open bars) readers comparing $^{123}$I-iофлюпан 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 (1)). 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 123I-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 123I-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 clinical information in these cases of 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 123I-ioflupane image results (17). Marshall et al. reported a similar tendency in PS (7). 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...
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 123I-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.

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 123I-ioflupane SPECT imaging.
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-year 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.

ACKNOWLEDGMENT

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

REFERENCES


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

John P. Seibyl, Andreas Kupsch, Jan Booij, Donald G. Grosset, Durval C. Costa, Robert A. Hauser, Jacques Darcourt, Nin Bajaj, Zuzana Walker, Kenneth Marek, Ian McKeith, John T. O’Brien, Klaus Tatsch, Eduardo Tolosa, Rudi A. Dierckx and Igor D. Grachev

*J Nucl Med.*
Published online: June 12, 2014.
Doi: 10.2967/jnumed.114.140228

This article and updated information are available at:
http://jnm.snmjournals.org/content/early/2014/06/11/jnumed.114.140228

Information about reproducing figures, tables, or other portions of this article can be found online at:
http://jnm.snmjournals.org/site/misc/permission.xhtml

Information about subscriptions to JNM can be found at:
http://jnm.snmjournals.org/site/subscriptions/online.xhtml

*JNM* ahead of print articles have been peer reviewed and accepted for publication in *JNM*. They have not been copyedited, nor have they appeared in a print or online issue of the journal. Once the accepted manuscripts appear in the *JNM* ahead of print area, they will be prepared for print and online publication, which includes copyediting, typesetting, proofreading, and author review. This process may lead to differences between the accepted version of the manuscript and the final, published version.