Standardized Semi-Quantitative Evaluation of [123I]FP-CIT SPECT in a Multicenter Study

Andrea Skanjeti1,2, Giancarlo Castellano3, Baldassarre Elia4, Francesca Dazzara1, Michela Zotta3, Gabriela Mariam Delgado Yabar1, Matteo Manfredi3, Marco Iudicello5, Maurizio Zibetti6, Patrizia Ferrero7, Paola Filippi8, Riccardo Emanuele Pellerito4, Valerio Podio1

1Nuclear Medicine Unit, San Luigi Gonzaga Hospital, University of Turin, Orbassano, Italy
2Medical Science Department, University A. Avogadro, Novara, Italy
3Nuclear Medicine Unit, San Giovanni Battista Hospital, University of Turin, Turin, Italy
4Nuclear Medicine Unit, Umberto I Hospital, Turin, Italy
5Neurology 1, San Luigi Gonzaga Hospital, Orbassano, Italy
6Neurology 4, San Giovanni Battista Hospital, University of Turin, Turin, Italy
7Neurology 2, San Giovanni Battista Hospital, University of Turin, Turin, Italy
8Neurology, Maria Vittoria Hospital, Turin, Italy
Email: askanjeti@yahoo.it

Received August 1, 2013; revised September 1, 2013; accepted September 8, 2013

Copyright © 2013 Andrea Skanjeti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background and Purpose: To the best of our knowledge no multicenter studies have been published using standardized semi-quantitative evaluation of [123I]FP-CIT scan (DAT-SPECT). The aims of this study were: 1) to cross-compare semi-quantitative software-assisted evaluations of DAT-SPECTs performed in three centers with different equipments; 2) to assess the accuracy of semi-quantitative evaluations of DAT-SPECT and 3) to identify the threshold with the best accuracy, sensitivity and specificity in a patient population with suspected parkinsonian syndrome.

Materials and Methods: Two hundred twenty patients (mean age at the time of SPECT acquisition, 67.4 ± 9.5 yy) acquired in three centers (Ospedale San Luigi Gonzaga; Ospedale San Giovanni Battista Molinette; Ospedale Mauriziano Umberto I) were included. All of them underwent DAT-SPECT from January 2006 to July 2010. All exams were analyzed with the freely available software BASGAN and semi-quantitative data were used to predict disease. In particular, analyses were based on the values from the most deteriorated putamen and caudate, normalized for age and corrected for equipment. ROC analysis was performed and area under the curve (AUC) was estimated.

Results: Analysis showed high AUCs (0.898, 0.864, 0.900 and 0.891 for each center and for the multicenter setting, respectively) confirming the very good accuracies reached. The best cut-off were 0.72 and 0.82 for putamen and caudate respectively. These thresholds allowed sensitivities and specificities in each center and in the multicenter setting of 76% and 96%, 91% and 82%, 93% and 90%, 86% and 89% respectively. No significant differences were observed between sensibility and specificity in each center.

Conclusion: A unique threshold useful for all centers with high and similar sensitivities and specificities is possible after correction for age and equipments. The high accuracy reached in this multicenter trial by the semi-quantitative analysis seems similar to accuracies from qualitative analysis in other multicenter studies.

Keywords: Movement Disorder; Parkinson Disease; [123I]FP-CIT-SPECT; Semi-Quantitative Evaluation; Multicenter Study

1. Introduction

Molecular imaging of dopamine transporters, in particular with [123I]FP-CIT (a dopamine trasporter ligand, commercial name: DaTSCAN®), plays a well established role in the diagnosis and follow-up of patients with parkinsonian syndrome since the 90 s for the contribution in diagnostic accuracy, patients’ quality of life and cost-effectiveness evaluation [1-4]. The assessment of [123I]FP-CIT scan (DAT-SPECT) is usually qualitative, however questions about the reproducibility of the technique have been raised recently [5-7]. Several authors, as well as the European Association of Nuclear Medicine (EANM), encourage both collegial discussion and the adoption of standardized interpretation criteria based on automated, semi-quantitative evaluations [8,9] with the view to en-
sure high and uniform diagnostic value among different institutions. This, in turn, is a requirement for multicenter studies.

It is well known that multicenter trials have acquired significant importance thanks to the larger data pools and higher statistical power reached; moreover they led to standardization with improvement of patients management. However, in nuclear medicine field, where different equipments, acquisition protocols, data analysis and interpretation/presentation of clinical results have been observed [10], comparability of semi-quantified results is still to be assured [11]. Consequently, among few multicenter studies that evaluated [123I]FP-CIT in patients suspected for neurodegenerative parkinsonian syndrome (NPS) [3] or suspected Lewy Body Dementia [6] all of them used a qualitative approach to assess the exam. To the best of our knowledge, no multicenter trial evaluated the accuracy of standardized and semi-quantified [123I]FP-CIT exam, although a large database with data from normal subjects acquired in 13 different centers has been recently published [12].

Various methods have been explored in order to assess DAT-SPECT semi-quantitatively [8]. Italian Association of Nuclear Medicine (AIMN) has proposed a free software, BASGAN v2, that semi-quantifies the uptake of striata and compare each patient’s data with data from a healthy subjects population database [13]. Recently our group evaluated accuracy and reproducibility of software-based analyses [9]. High accuracy was obtained and this type of evaluation showed to be useful to assist the physician in doubtful cases. The intra- and inter-operator reproducibility was very high, also. Consequently a larger trial among three centers with different equipments was undertaken.

The aims of multicenter study from three centers with different equipments were: 1) to cross-compare data from semi-quantitative software-assisted evaluations of DAT-SPECT; 2) to assess the accuracy of semi-quantitative evaluations of DAT-SPECT and 3) to identify the threshold with the best accuracy, sensitivity and specificity in a patient population with suspected parkinsonian syndrome.

2. Materials and Methods

2.1. Patient Selection

Two hundred twenty patients (mean age at the time of SPECT acquisition, 67.4 ± 9.5 years) from three centers (Ospedale San Luigi Gonzaga, Orbassano “center A”; Ospedale San Giovanni Battista Molinette, Torino “center B”; Ospedale Mauriziano Umberto I, Torino “center C”) were included in this trial. Patients underwent DAT-SPECT from January 2006 to July 2010, all having been referred by neurologists with clinical expertise in movement disorders. All patients showed symptoms suggesting initial onset of the disease. Drugs interfering with DAT-SPECT, as previously reported [14], were discontinued. These included psychostimulants, antidepressants, muscarinic receptor antagonists and anorexic drugs.

In NPS the diagnosis was defined by the clinician blinded to DAT-SPECT after at least 2 - 4 years from the scan according to clinical criteria. In case of Parkinson Disease (PD) the fulfillment of Step 1 of UK Brain Bank criteria was used. In NPS-free patients the diagnosis was established similarly, based on clinical criteria independently from DAT-SPECT report, particularly Findley & Koller criteria were used to identify Essential Tremor (ET). One hundred thirty-eight patients (63%) were diagnosed with NPS. Eighty-two (37%) were found NPS-free and were diagnosed as described in Table 1. Seventy-eight patients were scanned in the first center (mean age 68 years), and 71 in each of the other two centers (mean age respectively 65 and 69 years). The study was carried out in accordance with the ethical guidelines of the local Ethics Committee for Clinical Investigation. All patients gave their informed consent prior to DAT-SPECT.

Table 1. Patients NPS (Neurodegenerative Parkinsonian Syndrome) affected and NPS-free for each center and in the multicenter setting. For each center is shown the number of patients, their mean age and sex.

<table>
<thead>
<tr>
<th></th>
<th>Center A</th>
<th></th>
<th>Center B</th>
<th></th>
<th>Center C</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N°</td>
<td>Mean age</td>
<td>Males</td>
<td>N°</td>
<td>Mean age</td>
<td>Males</td>
<td>N°</td>
</tr>
<tr>
<td>NPS</td>
<td>54</td>
<td>68.7 y</td>
<td>32</td>
<td>44</td>
<td>63.6 y</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>NPS free</td>
<td>24</td>
<td>66.8 y</td>
<td>12</td>
<td>27</td>
<td>67.9 y</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>68.1 y</td>
<td>44</td>
<td>71</td>
<td>65.2 y</td>
<td>43</td>
<td>71</td>
</tr>
</tbody>
</table>

Copyright © 2013 SciRes.
era was used equipped with fan-beam collimator, circular orbit with radius variable according to the patients. One hundred twenty views (30 sec/view) were acquired using a step-and-shoot protocol at 3° interval (matrix 128 × 128; peak: 159 ± 10% keV). In this case the pixel size was variable due to the collimator geometry, however in all cases a voxel size of roughly 3 mm matrix was reconstructed.

In center C Siemens E.CAM 2002 dual head gamma-camera was used equipped with low energy, high resolution parallel hole collimators. One-hundred-twenty views (40 sec/view) were acquired using a step-and-shoot protocol at 3° interval (matrix 128 × 128; zoom 1.45, circular orbit, peak: 159 ± 10% keV); pixel size was 3.3 mm.

All exams, both acquired views and reconstructed transaxial images, were visually assessed by nuclear physicians and total counts ranged 1.5 - 2.5 millions for each exam. No motion artifacts were identified and quality resulted suitable for diagnostic purposes. In order to meet software requirements, all sets were reconstructed by filtered back-projection (Butterworth, order = 7.0, cut-off = 0.45). Chang algorithm was used for attenuation correction (μ = 0.10 cm⁻¹). Trans-axial images were reconstructed with voxel size in the range required by BASGAN (2.5 - 3.5 mm) and reoriented on the orbito-meatal line. In all cases, BASGAN evaluation was performed and semi-quantitative data from each ganglium were normalized twice. First normalization was done for age, according to data from normal subjects of the same age included in BASGAN software.

After age correction, data from normal subjects showed to be significantly different among centers. Therefore, a second correction was mandatory: mean data from putamen and caudate of NPS-free subjects in each center were used to normalize all patients’ data of that center (normalization for center) as follows:

\[
\text{Put value}_{\text{patient}}^{\text{Mean Put value}_{\text{NPS-free}}} \quad (1)
\]

and

\[
\text{Caud value}_{\text{patient}}^{\text{Mean Caud value}_{\text{NPS-free}}} \quad (2)
\]

for putamen and caudate data respectively. Therefore, all data furtherly used in this work were corrected for age and equipment.

Successive analyses were performed based on data of the most deteriorated putamen and caudate ganglium.

2.3. Statistical Analysis

Receiver operating characteristic (ROC) curve analysis was applied to assess the accuracy of the software-assisted evaluation of DAT-SPECT by estimating area under the curve (AUC); furthermore, the best cut-off (cutoff that allowed to reach the higher accuracy) was defined. Chi Square test, Student’s t test and ANOVA were also used as appropriate, p-value < 0.05 was considered statistically significant. SPSS vers. 19 (IBM Corporation, Armonk, NY, USA) and MedCalc vers. 12 (MedCalc Software, Mariakerke, Belgium) were used for all statistical purposes.

3. Results

Mean values from both NPS patients and NPS-free subjects clustered for each center are illustrated in Table 2. Results did not show any significant difference among the centers, while significant differences were observed between NPS and NPS-free.

### Table 2. Mean putamen and caudate value for each center in NPS affected and NPS-free patients.

<table>
<thead>
<tr>
<th></th>
<th>NPS</th>
<th>NPS-free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Centre A</td>
<td>Centre B</td>
</tr>
<tr>
<td>Putamen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.481</td>
<td>0.464</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.40 - 0.56</td>
<td>0.40 - 0.53</td>
</tr>
<tr>
<td>p-value ANOVA</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>p-value t-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.644</td>
<td>0.635</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.58 - 0.71</td>
<td>0.56 - 0.71</td>
</tr>
<tr>
<td>p-value ANOVA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value t-test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Test characteristics of [123I]FP-CIT analyzed by BASGAN in a multicenter setting.

<table>
<thead>
<tr>
<th></th>
<th>Center A</th>
<th>Center B</th>
<th>Center C</th>
<th>Multicenter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Putamen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.898</td>
<td>0.864</td>
<td>0.900</td>
<td>0.891</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cut-off</td>
<td></td>
<td></td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Sensitivity(%) and 95% C.I.</td>
<td>76 (62 - 87)</td>
<td>91 (78 - 97)</td>
<td>93 (80 - 98)</td>
<td>86 (79 - 91)</td>
</tr>
<tr>
<td>Specificity(%) and 95% C.I.</td>
<td>96 (79 - 99)</td>
<td>82 (62 - 94)</td>
<td>90 (74 - 98)</td>
<td>89 (80 - 95)</td>
</tr>
<tr>
<td><strong>Caudate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.836</td>
<td>0.837</td>
<td>0.864</td>
<td>0.846</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cut-off</td>
<td></td>
<td></td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Sensitivity(%) and 95% C.I.</td>
<td>74 (60 - 85)</td>
<td>82 (70 - 93)</td>
<td>80 (64 - 91)</td>
<td>78 (70 - 84)</td>
</tr>
<tr>
<td>Specificity(%) and 95% C.I.</td>
<td>79 (58 - 93)</td>
<td>89 (71 - 98)</td>
<td>90 (74 - 98)</td>
<td>87 (77 - 93)</td>
</tr>
</tbody>
</table>

All AUCs were large, statistically significant and similar among centers. No significant differences were observed among centers, nor for putamen, neither for caudate data; furthermore, no significant differences were observed among both sensitivities and specificities from each center and in the entire population. Each center and global accuracies according to putamen semi-quantitative data were respectively 82%, 87%, 92% and 87% while in the case of caudate were respectively 76%, 85%, 85% and 82%.

4. Discussion

Standardization in Nuclear Medicine is a trend followed with success for years. This is required primarily in order to ensure diagnostic quality for patients and secondly to permit an adequate sharing of information between colleagues. Furthermore, standardization is useful to correct errors associated with various instrumentation as well.

However, in the case of dopaminergic system imaging, Morton et al. observed that all exams should be acquired in the same camera type to have homogenous data and phantom data should be used to normalize the database accordingly [15]. On the other hand, both Meyer and coll. and Koch et al. demonstrated that homogenous data could be obtained from different equipments, although corrections factors were mandatory [11,16]. However, to the best of our knowledge, no multicenter studies evaluated the accuracy of semi-quantitative data to predict neurodegenerative parkinsonian syndrome.

This study, the first that performs a standardized semi-quantitative evaluation of DAT-SPECT in a multicenter setting, showed the feasibility of this type of evaluation. Despite similar: patients age, size population, [123I]FP-CIT administered activity, time from injection to scan, patients’ preparation, duration of acquisition, reconstruction and semi-quantitative evaluation, still significant differences persisted among results from NPS-free patients of different centers (data not shown). Indeed, it seems that different equipments yields different semi-quantitative values of the binding potential in basal ganglia. However, correlations between these values were possible in a similar way as published by above-mentioned authors [11,16].

In this study, [123I]FP-CIT was analyzed by BASGAN, age-corrected for normal subjects according to BASGAN and corrected for acquisition instrumentation. Therefore, this test resulted with a comparable and good both sensitivity and specificity assured by a unique threshold (Table 3).

Few multicenter trials about [123I]FP-CIT are available, and among them none followed a semi-quantitative approach [3,6]. In fact, these papers showed a good performance of visual assessment. Therefore, questions can raise whether a quantitative approach is useful, given that can be expensive and time consuming. However, recently Tondeur and coll. raise doubts about the reproducibility of visual assessment. They observed significant difference between observers sensitivity in a three point scale (normal, abnormal and equivocal) qualitative evaluation of DAT-SPECT. The authors further advise a collegial discussion among nuclear physicians in order to avoid heterogeneous assessment [5]. Another paper indirectly showed differences among nuclear physician decisions on reporting [123I]FP-CIT, although in few cases [7]. Finally, for junior nuclear medicine physicians a quantitative approach could be useful until large experience is acquired [9]. Therefore, we strongly believe that this study, although limited, demonstrates the inherent ability of DAT-SPECT to be reliably evaluated by semi-quantitative analysis, although exams could be acquired in different centers.

The retrospective nature of the study in addition with normalization with data from NPS free patients are main limits. Moreover, although ideally post-mortem pathology should confirm diagnosis, we had to use clinical diagnosis. These limits are shared with a majority of clini-
cal trials. Furthermore, an ideal normalization for each center requires exams from healthy subjects: this could be ethically debatable. In our experience the normalization was possible with roughly 20 - 30 subjects for each center. In comparison with other multicenter trials based on qualitative evaluation, the present study reports similar accuracy. In fact the first published study [2] showed an accuracy between 94% and 98%, while more recent studies reported faintly lower accuracies. Marshall et al. [3] in a study with repetition of DAT-SPECT after three years observed an overall accuracy of 84% and McKeith et al. [6] reported an overall accuracy of 86%. The overall accuracy found in our study was roughly similar (87% and 82% for putamen and caudate values respectively). Therefore, in addition to the good reproducibility and good accuracy yielded by the use of BASGAN in a single center setting [9], now high accuracy is observed in this multicenter setting, providing a good standardization of the procedure.

The greater accuracy offered by putamen values as has been previously reported is likely explained by earlier and greater affection of pigmented neurons from ventrolateral substantia nigra pars compacta. These dopaminergic neurons project axons in the dorsal putamen. Successively are affected dorso-medial nigral neurons with projections in the head of caudate [17,18].

A last issue derived from this study was the high specificity. Some authors reported not-high-enough specificity attributed to incorrect qualitative assessment of negligible uptake deficits [7]. In fact, some images could be visually categorized as abnormal due to faint uptake deficits, while when BASGAN assessment is performed, a database of semi-quantitative data from normal subjects is available. This was observed also in a previous study when applied to doubtful exams [9].

In conclusion, this multicenter study identified a unique threshold accurate for all centers, with high and similar sensitivities and specificities. Consequently, the test (standardized DAT-SPECT assessed with BASGAN, age normalized and corrected for instrumentation) shows good accuracy and gives similar results regardless of the center in which it was performed. The software used to this purpose is freely available online and his reproducibility has been already demonstrated to be very good in trained technologists hands [9].

Finally, the evaluation of DAT-SPECT cannot be other than an integral one, obviously! Now a semi-quantitative approach seems possible, it remains to be studied the integration between a visual and semi-quantitative assessment in order to ensure high quality and reproducible DAT-SPECT report.

REFERENCES


http://dx.doi.org/10.1007/s00259-011-2042-3

http://dx.doi.org/10.1007/BF00839063

http://dx.doi.org/10.1007/s00259-005-0036-8

http://dx.doi.org/10.1007/s00259-012-2276-8

http://dx.doi.org/10.1007/s00259-006-0357-2

http://dx.doi.org/10.1007/s00259-007-0621-0

http://dx.doi.org/10.1097/00006231-200512000-00015


http://dx.doi.org/10.1602/neurorx.1.2.243

http://dx.doi.org/10.1093/brain/114.5.2283