A new visual rating scale for Ioflupane imaging in Lewy body disease

Jim J. Lloyd[a,b,*], George Petrides[a], Paul C. Donaghy[b], Sean J. Colloby[b], Johannes Attems[b], John T. O'Brien[c], Gemma Roberts[a,b], Alan J. Thomas[b]

[a] Nuclear Medicine Department, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, United Kingdom
[b] Institute of Neuroscience, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne NE4 5PL, United Kingdom
[c] Department of Psychiatry, University of Cambridge, School of Clinical Medicine, Box 189, Level E4 Cambridge Biomedical Campus, Cambridge CB2 0SP, United Kingdom

ABSTRACT

Background: Dopaminergic loss on ^{123}I-Ioflupane brain imaging is a recognised biomarker for dementia with Lewy bodies. It is usually assessed using a visual rating scale developed for Parkinson's disease, which may not be optimal for dementia with Lewy bodies, as patterns of dopaminergic loss can be different.

Objectives: We aimed to develop a new visual rating scale for ^{123}I-Ioflupane brain images in Lewy body disease that encompasses appearances seen in dementia with Lewy bodies, and validate this against autopsy diagnosis.

Methods: Four experienced observers developed and tested a new scale consisting of two metrics, reflecting overall loss and heterogeneity of loss. 66 subjects were used during development including clinical diagnoses of Alzheimer's disease ($n = 14$), Parkinson's disease ($n = 9$), Parkinson's disease dementia (n = 9), dementia with Lewy bodies ($n = 15$) and normal controls ($n = 19$). The scale was then tested on an independent group of 46 subjects with autopsy confirmed diagnosis: Alzheimer's disease ($n = 11$), Parkinson's disease ($n = 3$), Parkinson's disease dementia ($n = 15$), dementia with Lewy bodies ($n = 12$), normal controls ($n = 4$) and Frontotemporal dementia (n = 1).

Results: In the autopsy validation the sensitivity and specificity of the new scale for Lewy body disease was 97% and 100% respectively, compared with the standard scale which had the same sensitivity (97%), but lower specificity (80%). The new scale had excellent inter rater reliability (intra-class correlation coefficient 0.93).

Conclusion: A new robust and reliable rating scale is described that straightforwardly captures the visual appearance of ^{123}I-Ioflupane brain images. It demonstrated high accuracy in autopsy confirmed cases and offers advantages over the existing visual rating scale.

1. Introduction

Dementia with Lewy bodies (DLB) is the second commonest cause of degenerative dementia after Alzheimer's disease (AD) (Vann Jones and O'Brien, 2014). Although clinical criteria for DLB have high accuracy in specialist centres (sensitivity and specificity both > 80%) (McKeith et al., 2005), it may be low in non-specialist centres. Correct diagnosis of dementia is vital to communicate prognosis to patients and carers and to avoid unnecessary and potentially harmful treatments. Visualisation of nigrostriatal dopaminergic integrity using ^{123}I-Ioflupane (FP-CIT, DaTSCAN, GE Healthcare) brain SPECT imaging has been reported to improve sensitivity in probable DLB (McKeith et al., 2007), to increase certainty of diagnosis in possible DLB (Walker et al., 2015) and is included as an "indicative biomarker" in recent diagnostic criteria (McKeith et al., 2017).

Clinical reporting ^{123}I-Ioflupane scans for DLB is most often by primary visual read which may be supported by semi-quantification. It is helpful to have a systematic robust method to do this visual assessment. A visual rating scale was introduced by Benamer (Benamer et al., 2000) and developed and validated for Parkinson's disease (PD). When applied in DLB (e.g. (McKeith et al., 2007)), this scale has limitations, in particular the scale does not include a pattern of uniform reduction which may be more common in DLB (O'Brien et al., 2004; Walker et al., 2004). Anecdotally, within our centre it was apparent that observers were having to force images into Benamer categories, even though the image did not fit the strict definition.

Kahraman et al. (2012) and Davidsson et al. (2014) employed a system of visual assessment which was similar to Benamer, except that different image categories were defined. Those studies found some success with this system to distinguish PD from atypical PD syndromes, although they did not directly compare with the Benamer method. The application in those papers was different to ours and so unlikely to be

directly applicable, particularly as they did not include a category of moderate generalised loss which may be seen in DLB.

Visual reading is likely to remain a principal mode of assessment due to difficulties in image quantification. Absolute quantification of ^{123}I-FP-CIT SPECT is extremely difficult (Bailey and Willowson, 2013) and therefore semi-quantitative regional analysis is usually employed. Although useful, there are several limitations to this approach. Different methods of semi-quantification exist with no standard approach (Koch et al., 2005; Morton et al., 2005a, 2005b; Poli et al., 2013; Slomka et al., 2001; Tossici-Bolt et al., 2006, 2011, 2017; Varrone et al., 2013). Depending on the method used, specific binding ratios may be sensitive to changes in acquisition and processing (Dickson et al., 2010; Koch et al., 2013, 2014; Tossici-Bolt et al., 2011; Tossici-Bolt et al., 2017). Deriving a suitable local normal range for binding ratios can be problematic (Dickson et al., 2017). Image warping and registration may not fully account for anatomical variations in striatal shape. Quantification may give misleading answers in cases where background is low due to atrophy or artefactually high due to scalp uptake.

Our aim in the current work was to develop a new visual rating scale in Lewy body disease (LBD) that encompasses the pattern and distribution of dopaminergic loss seen in DLB. We wished the new scale to be reproducible, straightforward to implement, capture both the overall uptake and its distribution and be highly accurate in separating LBD from non-LBD cases in a mixed group of cases including a significant proportion with DLB. Our aim was to devise a system that did not rely on assigning images to pre-defined categories, but that would be applicable across the full range of conditions for which 123I-FP-CIT SPECT is used, including DLB. The scale was developed using a group of subjects with robust clinical diagnoses and then tested using independent cases that additionally had proceeded to autopsy confirmation of diagnosis. The new scale was compared to the existing Benamer scale.

2. Materials and methods

Based on our expertise in rating Ioflupane scans and in particular the need to capture balanced loss in DLB we devised a new rating method which we will refer here to as the "Newcastle scale". Data from well characterised subjects (O'Brien et al., 2004) were used for the development and testing of the scale in two phases. Phase one consisted of applying it to cases with clinical diagnoses and setting a threshold for detection of Lewy body disease. In phase two the scale was tested using an independent group of subjects with autopsy diagnosis.

All procedures performed in studies involving human participants were in accordance with NHS and Newcastle Brain Bank ethical approvals and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

2.1. Subjects

Image data was drawn from subjects involved in a previous study (O'Brien et al., 2004) as shown in Tables 1 and 2. Patients were obtained from a community-dwelling population referred to local old age psychiatry and neurology services. Normal controls were recruited from among friends and spouses of patients included in this and other research studies.

Subjects underwent detailed physical, neurological and neuropsychiatric examinations, including the Mini-Mental State Examination (MMSE) (Roth et al., 1986), and the motor subsection of the Unified Parkinson's Disease Rating Scale (UPDRS III) (Fahn, 1987). Diagnosis was made by a consensus panel of experienced dementia clinicians using the NINCDS/ADRDA criteria for AD (McKhann et al., 1984), the consensus criteria for DLB and PDD (Parkinson's disease dementia) (McKeith et al., 1996) and the UK Parkinson's Disease Society Brain Bank criteria for PD (Gibb and Lees, 1988). All AD subjects met criteria for probable AD, 23 DLB subjects fulfilled probable and 4 possible DLB. No subject was on any medication which may affect ^{123}I-Ioflupane uptake.

Forty-six subjects underwent autopsy and neuropathological assessment which was performed blind to clinical diagnoses and ^{123}I-Ioflupane findings. The mean (sd) time between scan and autopsy was 5.74 (3.74) years. Neuropathological findings in these cases have been described previously (Thomas et al., 2017). Six cases fulfilled the neuropathological criteria for both AD and DLB (mixed dementia). In these cases clinical notes were reviewed at baseline and all follow up (blinded to any ^{123}I-Ioflupane results) and the most likely clinical diagnosis at the time of scan was chosen to validate the ^{123}I-Ioflupane results.

In phase one Lewy body disease refers to cases with a clinical diagnosis of PD, PDD or DLB. In phase two this term refers to subjects with Lewy body disease meeting neuropathological criteria (Thomas et al., 2017). This includes subjects with a previous clinical diagnosis of PD, PDD or DLB and who had significant Lewy body pathology at autopsy. Since PD, PDD and DLB may be indistinguishable pathologically, we have classified these subjects according to their final combined clinicopathological diagnoses where LBD is confirmed at autopsy.

2.2. Imaging

Subjects were imaged using a triple-detector rotating gamma camera (Picker 3000XP) fitted with a high resolution fan-beam collimator, 4 h after injection of 150 MBq of ^{123}I-Ioflupane. One hundred and twenty 15 s views over a 360° orbit were acquired on a 128×128 matrix with a square pixel dimension of 3.5 mm. Imaging time was 30 min. Image reconstruction was performed without attenuation correction using filtered back projection with a Butterworth filter (order 13, cut-off 0.3 cycles.cm^{-1}) to produce transverse sections with an axial resolution of 10 mm full width at half maximum (FWHM).

Table 1
Subjects used in phase 1 (clinical diagnoses).

	n	Sex (M:F)	Age	MMSE	UPDRS III
Controls	19	11:8	73.1 ± 6.0	28.4 ± 1.3	0.5 ± 0.8
AD	14	5:9	80.9 ± 5.3	17.1 ± 5.2	6.1 ± 6.3
DLB	15	9:6	75.2 ± 7.1	14.7 ± 5.6	31.1 ± 10.9
PD	9	7:2	75.2 ± 5.2	25.9 ± 2.3	25.3 ± 10.2
PDD	9	7:2	73.0 ± 7.7	19.9 ± 6.2	42.0 ± 14.3
Group tests, statistic, p-value Pair wise tests ns = not significant ($p > 0.05$).		$\chi^2 = 5.8, 0.2$	$F_{4,61} = 3.7, 0.009$ Gabriel's post-hoc tests: AD > con, $p < 0.04$ AD > PDD, $p < 0.04$ Otherwise ns	$F_{4,61} = 26.9, < 0.001$ Gabriel's post-hoc tests: Con, PD > AD, DLB, PDD ($p < 0.04$); Otherwise ns.	$H_4 = 53.0, < 0.001$ Mann-Whitney post-hoc tests: Con, AD < DLB, PD, PDD ($p < 0.02$); Otherwise ns

Table 2
Subjects used for phase 2 (autopsy diagnoses).

	n	Sex (M:F)	Age	MMSE	UPDRS III
Control	4	2: 2	81.3 ± 7.4	28.0 ± 2.2	3.3 ± 3.2
AD	11	7: 4	80.4 ± 4.6	17.2 ± 4.7	8.5 ± 7.7
DLB	12	7: 5	75.6 ± 6.3	18.5 ± 5.2	22.2 ± 14.5
PD	3	3: 0	81.4 ± 3.4	25.7 ± 5.0	39.7 ± 19.9
PDD	15	11: 4	71.1 ± 4.7	21.1 ± 5.6	34.9 ± 11.4
Frontotemporal dementia (FTD)	1	1: 0	78.4	25.0	4.0
Group tests, Statistic, p-value (Performed on all groups except FTD and PD)		$\chi^2 = 1.1, 0.8$	$F_{3,38} = 7.7, < 0.001$	$F_{3,38} = 5.0, 0.005$	$F_{3,38} = 15.7, < 0.001$
Pair-wise tests			Gabriel's Post-Hoc tests	Gabriel's post-hoc tests	Mann-Whitney post-hoc tests
ns = not significant			Con, AD > PDD	Con > AD, DLB	DLB, PDD > Con, AD
			($p < 0.007$)	($p < 0.01$)	($p < 0.03$);
$P > 0.05$			Otherwise ns	Otherwise ns	PDD > DLB ($p = 0.03$)
					Otherwise ns

2.3. Rating and scale development

Four observers were involved at both phases. All observers had experience of research involving ^{123}I-Ioflupane imaging in the context of DLB diagnosis, although not all were imaging specialists. AT is a consultant old age psychiatrist with over 20 years experience of research and clinical practice relating to DLB, but is not involved with clinical reporting. JJL is a consultant medical physicist and expert in image analysis who provides some limited clinical reporting. GP is a consultant radiologist with particular interest in neurological applications and provides a full range of expert independent clinical reports. PD is and an old age psychiatry specialist registrar with 5 years experience in research relating to DLB and imaging who does not provide any image reports.

In compiling the Newcastle scale we aimed to overcome two fundamental assumptions of the standard Benamer scale, devised for PD, which are: a) that when dopaminergic loss occurs it is initially unilateral and, b) that dopaminergic loss invariably occurs with a predictable rostro-caudal gradient, putamen loss always occurring before caudate. This is because previous work has shown that in DLB and PDD dopaminergic loss is much more often both bilateral and uniform than in PD (O'Brien et al., 2004; Walker et al., 2004).

The Newcastle scale considers separately uptake scores in four areas (left and right caudate and putamen) relative to background and then derives two summary metrics, a *total loss score* (sum of all areas, with higher values indicating more severe abnormality) and a *difference score* (maximum difference between any two areas). Individual area uptake scores vary between 0 (normal) and 3 (absent) with a value of 0.5 indicating very mild or equivocal loss. Details are given below and example images and scores are shown in Fig. 1.

In phase one each rater independently blindly rated each subject using the Newcastle scale and the median total and difference scores for all raters were determined. These scores were compared to clinical diagnosis and based on these results, a threshold for abnormality (i.e. the presence of Lewy body disease) was set. In phase two, the process was repeated and using the threshold set in phase one, the results were compared to the final autopsy diagnoses.

In addition for both phases images were rated according to the Benamer scale (Benamer et al., 2000). As noted above, certain images (e.g. Fig. 1e) do not strictly fit within the categories defined by this scale. In these cases raters recorded the pattern of uptake (e.g. significant moderate global reduction) and then allocated a forced "best fit" Benamer category.

2.4. Details of scoring procedure

1. Use a standard colour scale (GE) with the highest uptake at 100% and the bottom of scale at 0%
2. Assess background counts as: normal, very mildly raised, mildly

raised, moderately or highly raised
3. Consider separately each of the following structures: Right Caudate, Right Putamen, Left Caudate and Left Putamen
4. Locate the structure with *highest uptake* and grade this with reference to the background as shown below:

   0-Normal (i.e. normal background)
   0.5-Very mild /equivocal loss (i.e. very mildly raised background)
   1-Mild loss (i.e. mildly raised background)
   2-Moderate loss (i.e. moderately raised background)
   3-Severe loss (i.e. highly raised background)

   Note: To be classed as normal the structure should have a crisp uniform edge and have normal thickness. In cases where the mean uptake appears normal, but the structure is irregular in some way, "ragged", or with a clipped edge, then a score of 0.5 should be used

5. Grade remaining structures taking into account the highest uptake structure and background
6. Derive two metrics as follows:

   Total score = Sum of the grades for all four areas.
   Difference score = Maximum difference in grade between any two areas.

2.5. Statistical analysis

Statistical analysis was carried out using SPSS (v23, IBM). Inter operator agreement for the Newcastle scale was measured used the intra-class correlation coefficient (ICC, two-way mixed effects models) for each striatal region (right caudate, right putamen, left caudate, left putamen) as well as for the total score. The optimum threshold to maximise the combined sensitivity and specificity for Lewy body disease detection was determined in phase one using ROC curve analysis.

3. Results

3.1. Accuracy

Median scores versus clinical diagnosis for phase one are shown in Fig. 2. The total score provided a larger separation between those with and without Lewy Body disease than the difference score. ROC curve analysis (phase one) identified a total score value of 4.5 as providing the highest combined sensitivity and specificity. Defining abnormality in this way as a total median score of > 4 gave a Lewy Body disease sensitivity of 28/33 (85%) and specificity of (31/33) 94%. The overall misclassification rate was 7/66 (11%).

Including the difference score or left/right or anterior/posterior score asymmetry did not improve the accuracy within phase one.
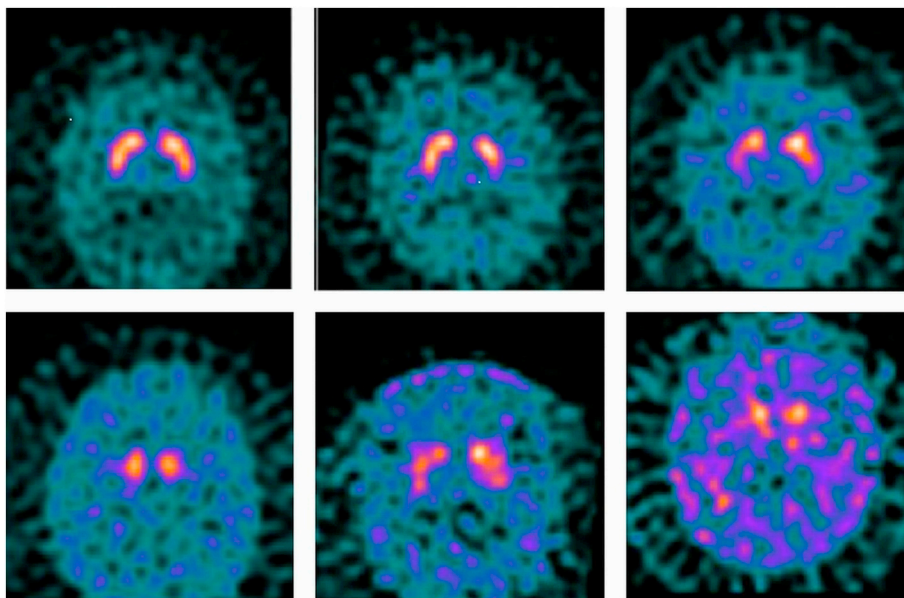
Fig. 1. Example images illustrating application of the Newcastle scale. From left to right, top to bottom:
a) Normal subject Score 0,0,0,0 Total 0 Difference 0
b) Normal subject Score 0.5,0.5,0.5,1 Total 2.5 Difference 0.5
c) DLB patient Score 1,2,1,1 Total 5 Difference 1
d) PD patient Score 1,3,1,3 Total 8 Difference 2
e) DLB patient Score 2,2,1,2 Total 7 Difference 1
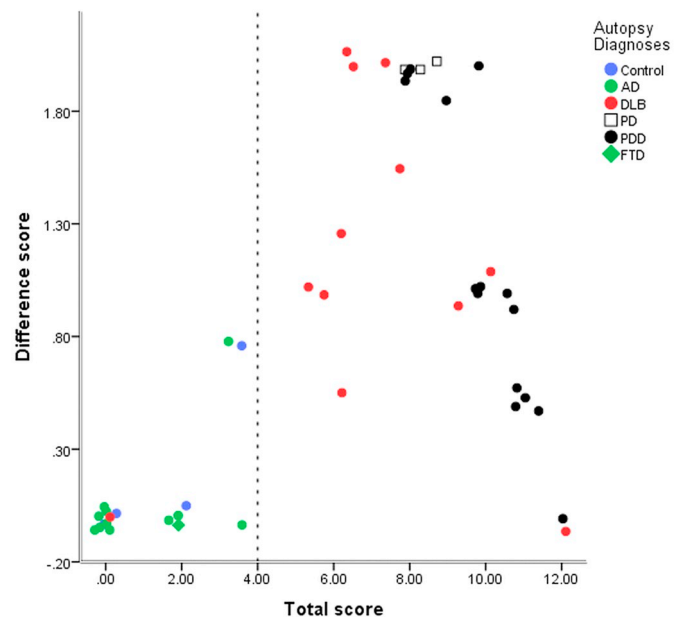f) PDD patient Score 2,3,2,3 Total 10 Difference 1
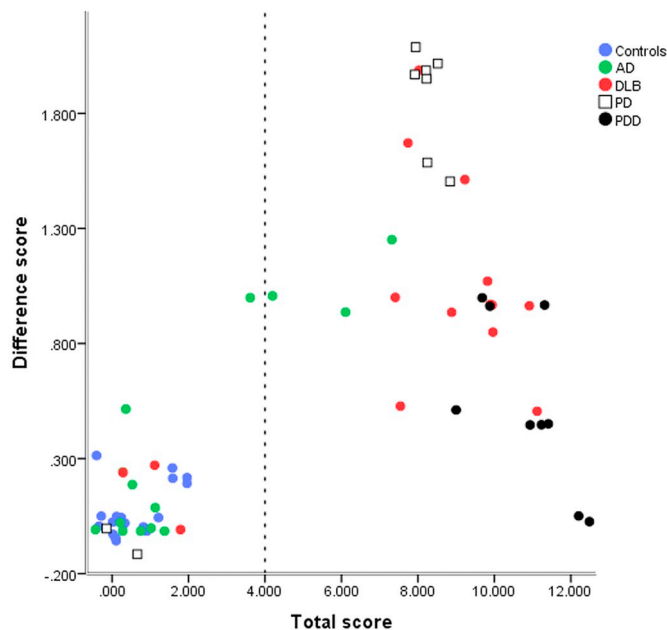
Fig. 2. Median rater total and difference scores for subjects in phase one. (A small random jitter has been applied to separate overlying points).

Fig. 3. Median rater total and difference scores for subjects in phase two. (A small random jitter has been applied to separate overlying points).

Median scores versus autopsy diagnosis for phase two are shown in Fig. 3. Using the Newcastle scale (and threshold as described above) the sensitivity and specificity for Lewy body disease was 30/31 (97%) and 15/15 (100%) respectively (i.e. just one false negative case, with a total and difference score of 0).

In phase one the Benamer scale sensitivity and specificity both had a value of 28/33 (85%). In phase two using the Benamer scale gave sensitivity and specificity of 30/31 (97%) and 12/15 (80%) respectively (i.e. one false negative case and three false positive cases). For phase two, The Newcastle scale gave the same sensitivity (97%) as the Benamer scale and improved specificity (100% vs 80%), although this difference just failed to reach statistical significance ($p = 0.07$).

3.2. Scale applicability and inter-operator reliability

For the Newcastle scale, results from phase one and two were combined ($n = 112$) and the ICC values were: 0.83, 0.94, 0.81, 0.92 and 0.93, for the right caudate, right putamen, left caudate, left putamen and total scores respectively. This represents 'good to excellent' inter-rater reliability (Koo and Li, 2016).

In phase two, there were 11 cases in which the Benamer scale was not strictly applicable, that is the image appearance did not fit with any of the categories of that scale. Seven cases were described as "very mild global loss", of these five were controls, one was AD, and one was DLB; these cases were all allocated to Benamer category 0 as the forced "best-fit". Four cases were described as having "significant moderate global loss" and all had clinical DLB diagnoses; these cases were all allocated to Benamer category 1 as the forced "best-fit".

For phase one and two combined the categories defined in the original Benamer scale (Benamer et al., 2000) were applicable in 90/112 (80%) of cases. Of these cases there was complete rater agreement in 44/90 (49%) and a majority agreement in 39/90 (43%). If PD and PDD cases were excluded then the Benamer scale was applicable in 56/77

(73%) of cases (phase one and two).

4. Discussion

A method of visual rating is presented which in this relatively small study was highly accurate for Lewy body disease diagnosis in autopsy proven cases. The method was straightforward to apply and had excellent observer agreement. Although the group size in each phase was relatively small, the abnormality threshold was determined in one phase and then tested on independent autopsy validated cases. This approach gives a more robust assessment of the scale performance than using a single group of subjects to develop and test the scale.

The aim was to devise a system that would be applicable for use in clinical and research applications relating to DLB. However, DLB is within spectrum of Lewy body diseases involving nigrostriatal dopaminergic dysfunction having similar pathophysiology and clinical management. We therefore did not seek to distinguish between subjects within this spectrum, but aimed to produce a scale that would accommodate all potential appearances within this group and provide a distinction with subjects not having Lewy body disease. Clinically distinguishing between AD and DLB especially in early stages is a far more pertinent question that distinguishing between DLB and PDD. One could devise separate scales to be applicable in different applications, such as a dementia scale and a PD scale, but given the pathophysiological overlap a single scale accommodating the full spectrum of appearances is more desirable. In order to develop and test this scale a wide range of conditions including DLB was employed.

4.1. Accuracy

The important difference between phases one and two is that the latter benefited from autopsy validation and so whilst in phase one, three AD cases were reported as abnormal (Fig. 2) in phase two, no AD cases were reported as abnormal and all abnormal ^{123}I-Ioflupane scans had Lewy body disease. In phase one, there were 7/66 misclassified cases. Two cases initially diagnosed as 'possible' PD had normal ^{123}I-Ioflupane scans. Their baseline UPDRS III scores were 16 and 19 respectively, which didn't progress at one year, while their motor phenotypes were tremor dominant. It seems likely that these patients had essential tremor rather than PD. Three patients with an initial clinical diagnosis of DLB had normal ^{123}I-Ioflupane scans. Review after one year suggested that these patients were likely to have had AD. Two patients initially diagnosed as AD had borderline high scores on the Newcastle Scale; on review with follow up there was a suggestion of AD with developing parkinsonism (UPDRS scores increasing from 13 to 28 and 8 to 19 after one year) and so these may have been mixed AD plus DLB cases.

In phase two, where diagnosis was confirmed at autopsy, all cases were correctly classified using the Newcastle scale apart from one DLB patient who had a normal scan. The neuropathological summary states: "Neuropathological examination revealed AD and DLB (mixed dementia) as well as an old infarct in the right rostral striatum." For this patient, the mean activity ratio for the posterior putamen (as described in (O'Brien et al., 2004) was 2.74, which is closer to the average for controls and AD patients (3.2 +/− 0.54 and 3.01 +/− 0.52 respectively) rather than DLB patients (1.92 +/− 0.68). On the Benamer scale all raters scored this patient as 0. This case, together with two others from a larger study (Thomas et al., 2017), provides evidence that it is possible to have DLB without nigrostriatal dopaminergic loss and therefore false negative cases will arise whatever analysis method is used.

Overall in phase 2 there were 10 (22%) cases where the neuropathological diagnoses differed from the consensus clinical diagnosis at the time of the initial study (O'Brien et al., 2004), although only two cases where the change reversed the presence of any Lewy body disease. Other studies (Thomas et al., 2017; Walker and Walker, 2009)

have indicated that the accuracy of ^{123}I-Ioflupane is higher compared to autopsy confirmed diagnosis than to clinical diagnosis. There is therefore need for caution when developing rating scales or quantification methods without autopsy confirmed diagnosis.

4.2. Distribution of uptake

Part of the rationale for the proposed rating method is to enable a simple visual plot that captures both the total uptake and heterogeneity of uptake in group studies. This could be useful in a research setting for initial exploration of the relationship between disease and image appearance. This is well illustrated in Figs. 2 and 3 and several qualitative points can be made: 1. for separation into groups with and without Lewy body disease, total uptake is more important than distribution of uptake, 2. there is a tendency for uptake to be more uniform in DLB than PD, although there is a large overlap between these two groups, 3. completely uniform reduction was not seen in DLB.

In this study using information about the distribution of uptake did not improve the accuracy of classification into those with and without Lewy body disease. However, the pattern of nigrostriatal loss in PD is known to be asymmetric, particularly in early disease (Djang et al., 2012). Therefore the difference score may potentially be useful in milder and prodromal cases and this warrants further investigation.

4.3. Comparison to established rating scales

The most established visual rating scale (Benamer et al., 2000), was developed for Parkinson's disease and may have limitations for DLB. In particular, it does not include a category for "balanced loss". We found that 20% of cases did not fit into any category on the established scale. Furthermore there was complete agreement among observers on category in only 49% of cases. If the PD and PDD cases were excluded then the number of cases in which the Benamer scale was not applicable rose to 27%. During discussion about this study and previous ones the authors concluded that they often had to make an arbitrary choice of Benamer category to describe an uptake appearance.

Kahraman et al. (2012) devised a scale that was similar to that of Benamer et al. (2000) and found it useful in distinguishing PD from atypical PD syndromes (APD). Effectively they added one additional abnormal category, "eagle wing appearance" with normal caudate uptake and discrete reduction in one or both putamina. The example image they give of this appearance would fall below the threshold for abnormality in our study. All of the subjects in that study had clinical diagnoses of PD or APD and given the relatively low accuracy clinical compared to pathological diagnosis (Hughes et al., 1992), one might expect that some of the 165 subjects did not in fact have any cortical Lewy body disease. However, no subjects were reported as having a normal image appearance and 12% had the mildly abnormal "eagle wing" appearance. A strength of the current study is that we included a wide range of conditions including normal controls and non Lewy Body disease subjects and in phase two we had autopsy confirmation. This enabled us to set a threshold for abnormality such that minor image irregularities can be confidently placed within the normal range. Being able to distinguish between AD subjects with minor decrease in uptake and DLB subjects with significantly abnormal scans is clinically very important. Davidsson et al. (2014) used the same rating scale as that devised by Kahraman et al. (2012), although only 3 of the 121 subjects had an image appearance described as "eagle wing" and therefore the scale applied was very similar to the original Benamer scale in practice. Although visual rating was useful in that study there were no subjects with dementia and therefore it is not possible to assess its applicability in the current context.

4.4. Relevance to clinical practice

In clinical practice it is now common to report based on visual

assessment with the aid of semi-quantification. However several semi-quantification methods are available and the optimum thresholds for abnormality are uncertain. Expert judgement based on visual assessment is still therefore critical. In many cases the decision is clear cut, but in borderline cases the process of applying the Newcastle scale provides a useful method to assist the visual evaluation process. It is not suggested that reporting the total and difference scores explicitly would be helpful, although certain phrases could be associated with particular scores, for example, "mild global loss with more pronounced right sided putamen loss" would be associated with a score of [1,2,1,1]. This can then be correlated with the clinical findings to aid interpretation.

In this study we chose raters with a range of different backgrounds and experience levels and found a high level of inter-rater agreement. This suggests that the proposed scale is likely to be widely applicable in a clinical setting with a range of reporter expertise.

4.5. Applicability in early disease

All patients in this study had confirmed dementia or Parkinson's disease. There is increasing interest in imaging patients at earlier stages (McKeith et al., 2016), where there are some symptoms suggestive of DLB, but patients have mild cognitive impairment rather than dementia. ^{123}I-Ioflupane striatal uptake can be reduced in such prodromal DLB, although to a lesser extent than in probable DLB (Kasanuki et al., 2017). It is likely therefore that the application of any method of image assessment will be more challenging in prodromal disease. In this study a significant number of subjects without Lewy body disease had mild global reduction (but within the normal range). In particular, in phase two there was one control and two AD subjects with median total scores between 3 and 4 that had no evidence of Lewy Body pathology at autopsy. These cases were incorrectly classified as abnormal using the Benamer scale. Caution should therefore be exercised in regarding the appearance of subtle decrease as a reliable sign of early disease. Notwithstanding the challenges of optimum threshold setting, the method described here could be applicable to research into early disease.

4.6. Equipment and processing

The images in this study were acquired on a three headed camera. At the time of the original study (around 2000) this provided image quality in advance of equipment in routine clinical use. Nowadays imaging with a dual headed camera is most common in clinical practice. However, it is likely that the reconstruction method is more relevant in terms of image resolution and visual image appearance. For this study images were processed with filtered back projection and the resolution and appearance was similar to that used in our centre in routine clinical practice using a dual headed camera. Recent advances in image processing using iterative reconstruction can affect image appearance, particularly if resolution recovery, attenuation and scatter correction are employed. These methods can particularly affect the visual appearance of the background. The colour look up table for display can also have an impact on visual assessment. As with any visual rating scale, local application of the scale proposed here will need to take into account the particular equipment and processing used. Depending on the method used, this caveat will also apply to semi-quantitative methods.

4.7. Conclusion

A rating scale is proposed based on autopsy confirmed diagnoses that captures in a straightforward way the visual appearance of ^{123}I-Ioflupane scans. In this relatively small study, it is demonstrated to have high accuracy in differentiating Lewy Body disease from non-Lewy Body disease cases and it offers potential advantages over existing visual rating scales. Further work is warranted to explore the applicability of the rating method in larger groups of subjects and in a range of different contexts.

Disclosure of potential conflicts of interest

Alan Thomas has received support from GE Healthcare, the manufacturer of Ioflupane (123I-FP-CIT, DaTSCAN), for investigator led research. John O'Brien has acted as a consultant for GE Healthcare and received research support for investigator led research. George Petrides and Gemma Roberts have received honoraria from GE Healthcare for educational presentations. All other authors declare that they have no conflict of interest.

References

Bailey, D.L., Willowson, K.P., 2013. An evidence-based review of quantitative spect imaging and potential clinical applications. J. Nucl. Med. 54, 83–89.

Benamer, H.T.S., Patterson, J., Grosset, D.G., Booij, J., Bruin, K.d., Royen, E.v., Speelman, J.D., Horstink, M.H.I.M., Sips, H.J.W.A., Dierckx, R.A., Versijpt, J., Decoo, D., Linden, C.V.D., Hadley, D.M., Doder, M., Lees, A.J., Costa, D.C., Gacinovic, S., Oertel, W.H., Pogarell, O., Hoeffken, H., Joseph, K., Tatsch, K., Schwarz, J., Ries, V., 2000. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. Mov. Disord. 15, 503–510.

Davidsson, A., Georgiopoulos, C., Dizdar, N., Granerus, G., Zachrisson, H., 2014. Comparison between visual assessment of dopaminergic degeneration pattern and semi-quantitative ratio calculations in patients with Parkinson's disease and Atypical Parkinsonian syndromes using DaTSCAN® SPECT. Ann. Nucl. Med. 28, 851–859.

Dickson, J.C., Tossici-Bolt, L., Sera, T., Erlandsson, K., Varrone, A., Tatsch, K., Hutton, B.F., 2010. The impact of reconstruction method on the quantification of DaTSCAN images. Eur. J. Nucl. Med. Mol. Imaging 37, 13.

Dickson, J.C., Tossici-Bolt, L., Sera, T., Booij, J., Ziebell, M., Morbelli, S., Assenbaum-Nan, S., Borght, T., Pagani, M., Kapucu, O.L., Hesse, S., Van Laere, K., Darcourt, J., Varrone, A., Tatsch, K., 2017. The impact of reconstruction and scanner characterisation on the diagnostic capability of a normal database for [(123)I]FP-CIT SPECT imaging. EJNMMI Res. 7, 10.

Djang, D.S., Janssen, M.J., Bohnen, N., Booij, J., Henderson, T.A., Herholz, K., Minoshima, S., Rowe, C.C., Sabri, O., Seibyl, J., Van Berckel, B.N., Wanner, M., 2012. SNM practice guideline for dopamine transporter imaging with 123I-ioflupane SPECT 1.0. J. Nucl. Med. 53, 154–163.

Fahn, S.E.R., 1987. Unified Parkinson's disease rating scale. Recent developments in Parkinson's disease. Mcmillan, New York, pp. 153–163.

Gibb, W.R., Lees, A.J., 1988. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 51, 745–752.

Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J., 1992. Accuracy of clinical-diagnosis of idiopathic Parkinsons-disease – a clinicopathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry 55, 181–184.

Kahraman, D., Eggers, C., Schicha, H., Timmermann, L., Schmidt, M., 2012. Visual assessment of dopaminergic degeneration pattern in 123I-FP-CIT SPECT differentiates patients with atypical parkinsonian syndromes and idiopathic Parkinson's disease. J. Neurol. 259, 251–260.

Kasanuki, K., Iseki, E., Ota, K., Kondo, D., Ichimiya, Y., Sato, K., Arai, H., 2017. 123I-FP-CIT SPECT findings and its clinical relevance in prodromal dementia with Lewy bodies. Eur. J. Nucl. Med. Mol. Imaging 44, 358–365.

Koch, W., Radau, P., Harmann, C., Tatsch, K., 2005. Clinical testing of an optimized software solution for an automated, observer independent evaluation of dopamine transporter SPECT studies. J. Nucl. Med. 46, 10.

Koch, W., Bartenstein, P., la Fougere, C., 2013. 3D-OSEM and FP-CIT SPECT quantification: benefit for studies with a high radius of rotation? Nucl. Med. Commun. 34, 971–977.

Koch, W., Bartenstein, P., la Fougere, C., 2014. Radius dependence of FP-CIT quantification: a Monte Carlo-based simulation study. Ann. Nucl. Med. 28, 103–111.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15, 155–163.

McKeith, I.G., Galasko, D., Kosaka, K., 1996. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. Neurology 47, 1113–1124.

McKeith, I.G., Dickson, D.W., Lowe, J., Emre, M., O'Brien, J.T., Feldman, H., Cummings,

J., Duda, J.E., Lippa, C., Perry, E.K., Aarsland, D., Arai, H., Ballard, C.G., Boeve, B., Burn, D.J., Costa, D., Del Ser, T., Dubois, B., Galasko, D., Gauthier, S., Goetz, C.G., Gomez-Tortosa, E., Halliday, G., Hansen, L.A., Hardy, J., Iwatsubo, T., Kalaria, R.N., Kaufer, D., Kenny, R.A., Korczyn, A., Kosaka, K., Lee, V.M., Lees, A., Litvan, I., Londos, E., Lopez, O.L., Minoshima, S., Mizuno, Y., Molina, J.A., Mukaetova-Ladinska, E.B., Pasquier, F., Perry, R.H., Schulz, J.B., Trojanowski, J.Q., Yamada, M., Consortium on, D.L.B., 2005. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology 65, 1863–1872.

McKeith, I., O'Brien, J., Walker, Z., Tatsch, K., Booij, J., Darcourt, J., Padovani, A., Giubbini, R., Bonuccelli, U., Volterrani, D., Holmes, C., Kemp, P., Tabet, N., Meyer, I., Reininger, C., 2007. Sensitivity and specificity of dopamine transporter imaging with 123I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study. Lancet Neurol. 6, 305–313.

McKeith, I., Taylor, J.P., Thomas, A., Donaghy, P., Kane, J., 2016. Revisiting DLB diagnosis: a consideration of prodromal DLB and of the diagnostic overlap with alzheimer disease. J. Geriatr. Psychiatry Neurol. 29, 249–253.

McKeith, I.G., Boeve, B.F., Dickson, D.W., Halliday, G., Taylor, J.P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C.G., Bayston, A., Beach, T.G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., Duda, J.E., El-Agnaf, O., Feldman, H., Ferman, T.J., Ffytche, D., Fujishiro, H., Galasko, D., Goldman, J.G., Gomperts, S.N., Graff-Radford, N.R., Honig, L.S., Iranzo, A., Kantarci, K., Kaufer, D., Kukull, W., Lee, V.M.Y., Leverenz, J.B., Lewis, S., Lippa, C., Lunde, A., Masellis, M., Masliah, E., McLean, P., Mollenhauer, B., Montine, T.J., Moreno, E., Mori, E., Murray, M., O'Brien, J.T., Orimo, S., Postuma, R.B., Ramaswamy, S., Ross, O.A., Salmon, D.P., Singleton, A., Taylor, A., Thomas, A., Tiraboschi, P., Toledo, J.B., Trojanowski, J.Q., Tsuang, D., Walker, Z., Yamada, M., Kosaka, K., 2017. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. Neurology 89, 88–100.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944.

Morton, R., Guy, M., Clauss, R., Hinton, P., Marshall, C., Clarke, E., 2005a. Comparison of different methods of DaTSCAN quantification. Nucl. Med. Commun. 26, 7.

Morton, R.J., Guy, M.J., Marshall, C.A., Clarke, E.A., Hinton, P.J., 2005b. Variation of DaTSCAN quantification between different gamma camera types. Nucl. Med. Commun. 26, 7.

O'Brien, J.T., Colloby, S., Fenwick, J., Williams, E.D., Firbank, M., Burn, D., Aarsland, D., McKeith, I.G., 2004. Dopamine transporter loss visualized with FP-CIT SPECT in the differential diagnosis of dementia with Lewy bodies. Arch. Neurol. 61, 919–925.

Poli, G., Bianchi, C., Guerra, U., 2013. Use of the BasGan algorithm for [123I]FP-CIT SPECT quantification: a phantom study. J. Nucl. Med. Mol. Imaging 57, 10.

Roth, M., Tym, E., Mountjoy, C.Q., Huppert, F.A., Hendrie, H., Verma, S., Goddard, R., 1986. CAMDEX A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br. J. Psychiatry 149, 698–709.

Slomka, P., Radau, P., Hurwitz, G., Dey, D., 2001. Automated three-dimensional quantification of myocardial perfusion and brain SPECT. Comput. Med. Imaging Graph. 25, 12.

Thomas, A.J., Attems, J., Colloby, S.J., O'Brien, J.T., McKeith, I., Walker, R., Lee, L., Burn, D., Lett, D.J., Walker, Z., 2017. Autopsy validation of 123I-FP-CIT dopaminergic neuroimaging for the diagnosis of DLB. Neurology 88, 276–283.

Tossici-Bolt, L., Hoffmann, S.M., Kemp, P.M., Mehta, R.L., Fleming, J.S., 2006. Quantification of [123I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio. Eur. J. Nucl. Med. Mol. Imaging 33, 1491–1499.

Tossici-Bolt, L., Dickson, J.C., Sera, T., de Nijs, R., Bagnara, M.C., Jonsson, C., Scheepers, E., Zito, F., Seese, A., Koulibaly, P.M., Kapucu, O.L., Koole, M., Raith, M., George, J., Lonsdale, M.N., Munzing, W., Tatsch, K., Varrone, A., 2011. Calibration of gamma camera systems for a multicentre European I-123-FP-CIT SPECT normal database. Eur. J. Nucl. Med. Mol. Imaging 38, 1529–1540.

Tossici-Bolt, L., Dickson, J., Sera, T., Booij, J., Asenbaun-Nan, S., Bagnara, M.C., Borght, T.V., Jonsson, C., de Nijs, R., Hesse, S., Koulibaly, P.M., Akdemir, U.O., Koole, M., Tatsch, K., Varrone, A., 2017. [(123)I]FP-CIT ENC-DAT normal database: the impact of the reconstruction and quantification methods. EJNMMI Phys. 4, 8.

Vann Jones, S.A., O'Brien, J.T., 2014. The prevalence and incidence of dementia with Lewy bodies: a systematic review of population and clinical studies. Psychol. Med. 44, 673–683.

Varrone, A., Dickson, J.C., Tossici-Bolt, L., Sera, T., Asenbaum, S., Booij, J., Kapucu, O.L., Kluge, A., Knudsen, G.M., Koulibaly, P.M., Nobili, F., Pagani, M., Sabri, O., Vander Borght, T., Van Laere, K., Tatsch, K., 2013. European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur. J. Nucl. Med. Mol. Imaging 40, 15.

Walker, R.W., Walker, Z., 2009. Dopamine transporter single photon emission computerized tomography in the diagnosis of dementia with Lewy bodies. Mov. Disord. 24 (Suppl. 2), S754–S759.

Walker, Z., Costa, D.C., Walker, R.W., Lee, L., Livingston, G., Jaros, E., Perry, R., McKeith, I., Katona, C.L., 2004. Striatal dopamine transporter in dementia with Lewy bodies and Parkinson disease: a comparison. Neurology 62, 1568–1572.

Walker, Z., Moreno, E., Thomas, A., Inglis, F., Tabet, N., Rainer, M., Pizzolato, G., Padovani, A., Da, T.D.L.B.P.S.G., 2015. Clinical usefulness of dopamine transporter SPECT imaging with 123I-FP-CIT in patients with possible dementia with Lewy bodies: randomised study. Br. J. Psychiatry 206, 145–152.