

A comparison of visual and semi-quantitative analysis methods for planar cardiac I-123-MIBG scintigraphy in dementia with Lewy bodies

Running title: Visual vs semi-quantitative analysis for cardiac MIBG in DLB

Gemma Roberts^{1,2}, Joseph P.M. Kane¹, Jim J. Lloyd^{1,2}, George S. Petrides², Kim Howe², John T. O'Brien³, Alan J. Thomas¹

1: Institute of Neuroscience, Newcastle University, Biomedical Research Building, Campus for Ageing and Vitality, Westgate Road, Newcastle upon Tyne, UK. NE4 6BE

2: Nuclear Medicine Department, Leazes Wing, Royal Victoria Infirmary, Richardson Road, Newcastle upon Tyne, UK. NE1 4LP

3: Department of Psychiatry, University of Cambridge, Box 189, Level E4 Cambridge Biomedical Campus, Cambridge, CB2 0SP

Corresponding author:

Gemma Roberts

Institute of Neuroscience, Newcastle University, Biomedical Research Building, Campus for Ageing and Vitality, Westgate Road, Newcastle upon Tyne, UK. NE4 6BE

gemma.roberts@newcastle.ac.uk

ORCID ID: 0000-0002-6445-4023

Conflicts of Interest and Source of Funding

Gemma Roberts is supported by an Alzheimer's Society healthcare professional Fellowship. Infrastructure and support was provided to authors based at Newcastle by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre, a partnership between Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University. John O'Brien was supported by the NIHR Cambridge Biomedical Research Centre. GE Healthcare sponsored the visit of Edwin Poel for phantom calibration measurements.

George Petrides and Gemma Roberts have received honoraria from GE Healthcare for educational presentations. Alan Thomas has received support from GE Healthcare for investigator led research. John O'Brien has acted as a consultant for GE Healthcare and received research support for investigator led research. All other authors declare that they have no conflict of interest.

Abstract

Objectives

Cardiac I-123-MIBG imaging is an established technique for the diagnosis of dementia with Lewy bodies (DLB) but various analysis methods are reported in the literature. We assessed different methods in the same cohort of subjects to inform best practice.

Methods

Seventeen subjects with DLB, 15 with Alzheimer's disease (AD) and 16 controls were included. Planar images were acquired 20 minutes and four hours after injection. Nine operators produced heart-to-mediastinum ratios (HMRs) using freehand and 6, 7 and 8cm diameter circular cardiac regions. Inter-operator variation was measured using the coefficient of variation. HMR differences between methods were assessed using ANOVA. Seven raters assessed the images visually. Accuracy was compared using ROC analysis.

Results

There were significant differences in HMR between region methods ($p=0.006$). However, with optimised cut-offs there was no significant difference in accuracy ($p=0.2-1.0$). The sensitivity was 65% to 71% and specificity 100% for all HMR methods. Variation was lower with fixed regions than freehand ($p<0.001$). Visual rating sensitivity and specificity were 65% and 77% on early images and 76% and 71% on delayed images. There was no significant difference in HMR between early and delayed images ($p=0.4-0.7$) although a greater separation between means was seen on delayed images (0.73 vs 0.95).

Conclusions

HMR analysis using a suitable cut-off is more accurate than visual rating. Accuracy is similar for all methods, but freehand regions are more variable and 6 cm circles easiest to place. We recommend calculating HMR using a 6cm circular cardiac ROI on delayed images.

Keywords

Cardiac MIBG, HMR, Heart to mediastinum ratio, Dementia with Lewy bodies

Introduction

Dementia with Lewy bodies (DLB) is the second most common form of neurodegenerative dementia after Alzheimer's disease, accounting for 5-10% of cases [1, 2]. Accurate diagnosis is important for clinical management, prognosis, and carer wellbeing [3-5], but initial misdiagnosis outside the specialist setting is common [2, 6, 7].

Cardiac MIBG sympathetic innervation imaging is an established technique for the differential diagnosis of DLB from other dementias, which do not cause sympathetic denervation and thus have normal scan appearances. The technique has been used in Japan for the diagnosis of Lewy body disease for over a decade was recently included as an indicative biomarker in the fourth DLB consensus criteria, alongside ^{123}I -FP-CIT SPECT [6]. FP-CIT is more widely used than MIBG in clinical settings in Europe and is supported by a more substantial evidence base that has been validated against autopsy data [8-10]. However a reported sensitivity of 77% suggests many patients with DLB have normal FP-CIT findings [11].

Single centre studies of DLB diagnosis using cardiac MIBG scintigraphy reported sensitivity and specificity values of over 90% [12-16]. However, the largest multicentre trial of cardiac MIBG scintigraphy in DLB to date (133 DLB and non-DLB dementia patients scanned in Japanese centres) reported a sensitivity and specificity of 69% and 89% respectively [17]. This used ROC analysis to give optimal separation between the groups, rather than using an independent normal cardiac uptake threshold. With diagnostic revisions after three year follow up of 65 of the subjects, these results have been updated to 77% and 97% [18], although the optimal cardiac uptake cut-off between DLB and non-DLB was updated based on re-analysis of this subgroup.

Several authors have commented on the need to standardise cardiac MIBG image acquisition and processing parameters in order for cardiac uptake thresholds to be applicable between centres, e.g. [19-24]. A method to correct for differences in image acquisition parameters between centres (predominantly caused by different gamma camera and collimator models) has been developed by Nakajima *et al.* [25] was used in the multicentre study mentioned above [17, 18]. However centres outside Japan still tend to publish results generated using local acquisition parameters and local normal uptake thresholds [26, 27], omitting information known to affect apparent cardiac uptake such as gamma camera model and collimator type, which limits the generalisability of the results. In addition, there is currently no generally accepted method for processing and interpreting the images. Okuda *et al.* have developed a semi-automated image analysis method [28], used in the multicentre study [17], but outside Japan there remains a variety of methods in use for interpreting planar cardiac MIBG images for Lewy body disease, including visual analysis and semi-quantification. These issues are discussed in a review article by Chen *et al.*, where it is noted that further studies comparing the accuracy and repeatability of different analysis methods are required [29]. To our knowledge only two studies utilising visual assessment of cardiac MIBG images in Lewy body disease have been published [27, 30], neither directly comparing the accuracy of visual and semi-quantitative methods using statistical analysis, and thus there is a need for further work to clarify how the performance of visual rating compares with semi-quantitative analysis.

The majority of studies using cardiac MIBG for the assessment of Lewy body disease acquire both early images (acquired 15-20 minutes after MIBG administration) and delayed images (3-4 hours after administration). The HMR is usually calculated on the delayed images since these are thought to reflect the function of the cardiac sympathetic nerves and early images the distribution of the nerves [31]. However, three recent studies carried out in Japan by separate groups [30, 32, 33] demonstrated no significant difference between diagnostic accuracy of early

or delayed imaging for the diagnosis of Lewy body disease, suggesting that early images alone may be sufficient. Acquiring early images alone would benefit both patients and clinical departments and would also make the test much more convenient to perform than FP-CIT, which requires a 3-6 hour uptake period before scanning.

In this project we compare planar cardiac MIBG analysis methods for discriminating between DLB and Alzheimer's disease (AD) using dementia and healthy control subjects recruited and scanned locally as part of a research study. We compare visual assessment for both early and delayed imaging with various semi-quantitative analysis methods, including one as close as possible to that used by Nakajima *et al.* [25], made possible by participating in a European multicentre phantom calibration exercise [22]. The overall aim is to determine the optimal method for processing and interpreting planar cardiac MIBG images for DLB diagnosis, and the best time point for imaging. This is achieved by considering the variation between operators and raters for each method as well as the accuracies given by ROC analysis.

Methods

Subjects

Dementia subjects were recruited as part of the Newcastle University study into the use of cardiac MIBG in a representative UK population [34]. Thirty-two of the 33 patients included in our previous publication were included in this analysis: 15 with probable AD and 17 with probable DLB (one AD subject was excluded here because the early image was not available). Diagnoses were made by an expert panel of three old-age psychiatrists (AT, JK, JOB) blinded to scan results, based on clinical assessment and applying standard research criteria [6]. This consensus panel method has previously been validated against autopsy and is accepted by regulatory authorities as the clinical gold standard for living patients [11, 35]. Sixteen healthy controls were taken from a separate study recruiting healthy older people. They had undergone the same detailed neurological examination by a research physician as the dementia subjects, including rating for parkinsonism with UPDRS, a thorough neurocognitive examination which confirmed they had normal cognition and normal MRI brain scan.

Image acquisition

Subjects were administered 111 MBq I-123-MIBG via slow intravenous injection. Potassium iodate tablets (170mg) were given before and after injection for thyroid blockade. Ten minute anterior planar images were acquired at 20 minutes ("early") and 4 hours (\pm 30 minutes, "delayed"). Images were acquired on a Siemens Symbia T series or Siemens Symbia Intevo gamma camera with medium energy low penetration (MELP) collimators and processed on a Hermes workstation. Medium energy collimators are recommended for cardiac MIBG imaging by the EANM task group on standardisation [36] since they minimise septal penetration of the high energy photons that reduce image contrast. The energy window was $159 \text{ keV} \pm 10\%$, matrix size was 128×128 and no zoom was applied. The same settings were applied when participating in the European phantom calibration study [22]. We used the phantom calibration results for our centre to apply a correction factor to the HMR cut-off for DLB diagnosis determined by Yoshita *et al.* [17]. This small correction increases the cut-off for both gamma cameras from 2.10 to 2.25.

Visual rating

Images were anonymised and rated blind to diagnosis on a 4 point scale: definitely abnormal, probably abnormal, probably normal or definitely normal. Seven individuals experienced in nuclear medicine reporting (two consultant radiologists and five nuclear medicine physicists)

were recruited as raters. The raters were provided with a set of reference images for each category, including example images for both early and delayed imaging. The modal (most frequent) rating was taken as the consensus result. The results were dichotomised into normal or abnormal ratings for the accuracy study and comparison with HMR.

Heart-to-mediastinum ratio processing

Nine operators experienced in processing nuclear medicine studies took part in this part of the study (six nuclear medicine physicists and three nuclear medicine technologists). Five of the physicists also rated the images visually (see above). This was carried out in a randomised order and before the quantitative results were generated to avoid bias. Operators were provided with comprehensive instructions including example images showing region placement.

Both the early and delayed anterior images were processed. We used four different cardiac region placement methods: freehand regions delineated manually around the myocardium and fixed circular regions of 6, 7 and 8 cm in diameter placed over the centre of the visible myocardium. Smaller regions were not tested because the ventricular cavity apparent on some images would lead to an under sampling of counts. A fixed sized rectangular mediastinum region of 4 cm x 3 cm was placed between the apices of the lungs, in an area containing only background counts. The mediastinum region size was chosen to be consistent in shape and position within the mediastinum with Okuda *et al.* [37] but simplified by using a fixed sized region placed manually. The size is small enough not to be subject to interference from the lungs and large enough to give a reasonable sample of background counts. Since there is little variability uptake throughout the mediastinum we did not investigate mediastinum region size or placement, or automate placement. The same mediastinum region was used with all four cardiac ROIs for HMR calculation. The heart and mediastinum values were exported and all HMR values were calculated by the author using a macro to minimise typographical errors. An example image with manual and circular cardiac ROIs and resulting HMR values is shown in Figure 1

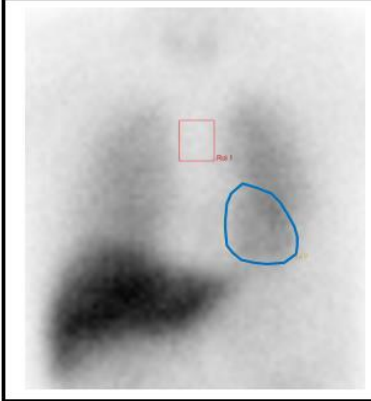
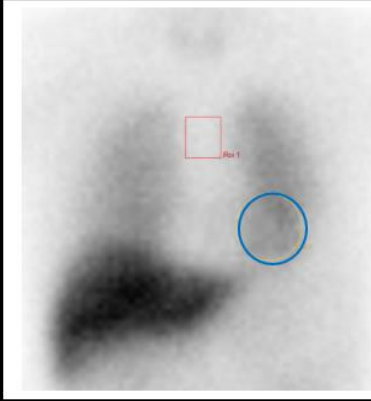
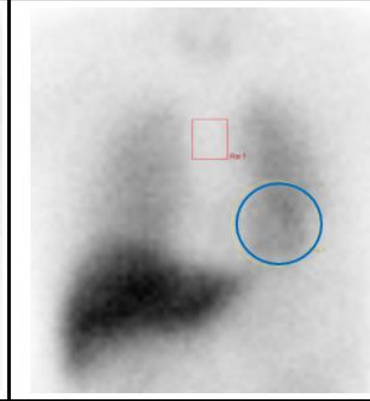
| Freehand region HMR = 2.5 | 6cm diameter circular region HMR = 2.6 | 8cm diameter circular region HMR = 2.4 |
|---|---|---|
|  |  |  |

Fig 1: Example normal uptake cardiac MIBG image with HMR values for freehand, 6 cm and 8 cm circular cardiac regions.

Analysis

Difference in HMR between cardiac region methods

The effect of region drawing method on HMR was tested using a generalised linear model ANOVA analysis applying the Greenhouse-Geisser method for conditions where sphericity

cannot be assumed (SPSS Statistics v23.0.0.3 (IBM Corporation)). A p-value of <0.05 was taken to indicate a significant dependence of HMR on region method.

Difference in inter-operator variability between cardiac region methods

The average coefficient of variation (CoV) was taken as a measure of variability with which to compare the region drawing methods. The CoV was calculated for each subject using the result for each of the nine operators and these CoV values were averaged to give the mean. This was repeated for each region method and time point. Paired t-tests were used to test for differences in coefficient of variation.

Difference in accuracy between cardiac region methods

The HMR value for each subject was taken as the mean of nine operators and these mean values used for comparing cardiac region methods. SigmaPlot v.13 (SystStat Software, San Jose, CA) was used to generate ROC curves including all subjects for each method by varying the normal/abnormal cut-off, using the consensus diagnoses (with the DLB, AD and control subject diagnoses dichotomised as either DLB or non-DLB) as the gold standard. SigmaPlot output was used to check for significant differences between the area under the ROC curves. This allows a comparison of overall accuracy of each HMR method to be made. The best method (time point and region combination) was selected based on the area under the curve (accuracy), difference between mean DLB and non-DLB HMRs and the average CoV between operators determined above.

The HMR values for the DLB and non-DLB subjects obtained with the method identified as optimal were used to determine the sensitivity and specificity for this method. This was done within SigmaPlot by applying the optimal cut-off given by the ROC analysis.

Difference between early and delayed HMRs

Paired t-tests were used to test for difference in HMR between early and delayed imaging again using a significance level of 0.05 (two-tailed test). The mean HMR with 95% confidence interval was calculated for the DLB and non-DLB groups for each method using SPSS and the mean differences calculated.

Comparison between visual rating and HMR analysis

The sensitivity and specificity for early and delayed consensus visual ratings were calculated from the true and false positive and negative results, using the consensus clinical diagnoses as the gold standard. As above, sensitivity and specificity results were calculated with the DLB, AD and control subject diagnoses dichotomised as either DLB or non-DLB.

To assess agreement between raters, the intraclass correlation coefficient was calculated using SPSS for all subjects. The concordance between early and delayed ratings was calculated as well as the number of control images rated as abnormal.

Results

Subject characteristics

The age, sex, MMSE and ACE scores and MDS-UPDRS scores [38] for all subjects are given in Table 1. As expected, there were significantly higher levels of Parkinsonism in the DLB group but otherwise the dementia groups were well matched for age and cognition. The control and AD subjects taken as a single non-DLB group were well matched for age and gender mix with the DLB group.

Table 1: Clinical characteristics of the AD, DLB and healthy control groups. In this study we grouped the AD and control subjects together to form a single group of “non-DLB” subjects that would be expected to have normal scan appearances.

| | Number of subjects (no. female) | Mean age (min – max) | Mean MDS-UPDRS score (min – max) | Mean MMSE (min – max) | Mean ACE total (min – max) |
|---------------------------|---------------------------------|----------------------|----------------------------------|-----------------------|----------------------------|
| Alzheimer’s Disease | 15 (4) | 76.2 (62 – 85) | 5.0 (0 – 27) | 22.0 (12 – 26) | 64.8 (23 – 85) |
| Dementia with Lewy bodies | 17 (2) | 77.5 (60 – 89) | 33.4 (2 – 89) | 21.8 (14 – 27) | 66.0 (34 – 87) |
| Healthy controls | 16 (6) | 72.3 (61 – 86) | 5.4 (0 – 16) | 28.6 (26 – 30) | 93.6 (85 – 99) |
| p-value AD vs DLB | 0.31 | 0.63 | <0.001 | 0.95 | 0.83 |
| p-value non-DLB vs DLB | 0.09 | 0.17 | <0.001 | n/a | n/a |

Effect of cardiac region method on HMR

The HMR values averaged over the nine operators for each processing method are shown in Table 2. The ANOVA repeated measures test demonstrated a significant difference in HMR values between the four region drawing methods ($F=7.196$, $p=0.006$). Pairwise comparisons showed the largest differences were between the 6cm and freehand regions and the 6 cm and 8 cm regions. For fixed circular regions, HMRs tend to decrease with increasing region size, as expected (Table 2).

Table 2: Mean and standard deviation of HMR values over all subjects calculated with freehand and circular cardiac regions on early and delayed images. The p-values test for significant differences between early and delayed HMR with each method, using paired t-tests.

| Region method | Mean early HMR (SD) | Mean delayed HMR (SD) | p |
|---------------|---------------------|-----------------------|------|
| Freehand | 2.22 (0.55) | 2.19 (0.66) | 0.37 |
| 6cm circle | 2.32 (0.60) | 2.31 (0.74) | 0.70 |
| 7cm circle | 2.26 (0.57) | 2.25 (0.69) | 0.74 |
| 8cm circle | 2.19 (0.53) | 2.18 (0.64) | 0.67 |

Difference in inter-operator variability between cardiac region methods

The fixed sized circular methods have a lower average coefficient of variation (CoV), than the freehand method (Table 3, $p < 0.001$). There was no significant difference in variation between any two fixed circle HMR methods (p values between 0.14 and 0.84). All methods have a lower CoV on delayed images than early images but this was not statistically significant (Table 3).

Table 3: Average coefficients of variation for each region drawing method used for HMR calculation. All subjects are included.

| | Region method | | | |
|----------------------------|---------------|------|------|------|
| | Freehand | 6 cm | 7 cm | 8 cm |
| Early CoV | 5.8% | 4.9% | 4.9% | 4.9% |
| Delayed CoV | 5.7% | 4.3% | 4.2% | 4.2% |
| p value (early vs delayed) | 0.80 | 0.37 | 0.17 | 0.15 |

Display of ROIs superimposed for all operators showed closer agreement between operators with the circular regions than with the freehand region, although this was not analysed using statistical testing as region overlap data was not available. An example image with normal uptake and example image with reduced uptake is shown in **Error! Reference source not found.** with freehand regions and 6 cm circular regions for all operators displayed. Operators reported that the 7 and 8 cm circular regions were difficult to place on some images without including non-cardiac uptake, for example in the lung or liver.

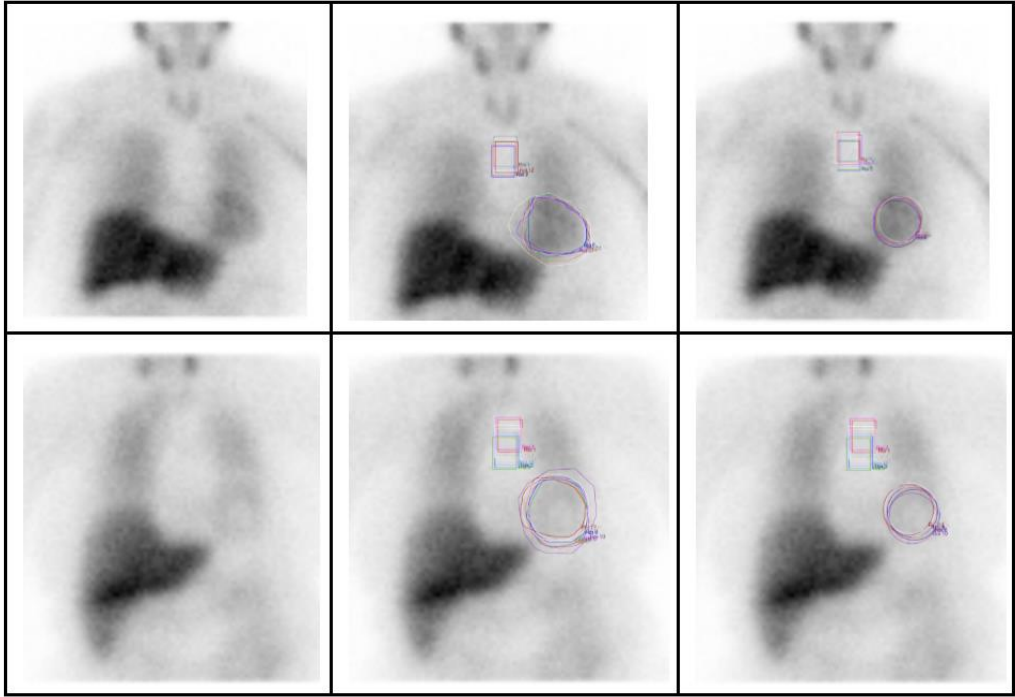


Fig 2: Top: example normal cardiac MIBG image (top row) and abnormal image (bottom row) processed by multiple operators with freehand and 6cm circular regions.

Difference in sensitivity, specificity and accuracy between cardiac region methods

ROC curves and AUC results for the early and delayed HMR methods are shown in Figure 3, with 95% confidence intervals given in Table 4. AUCs were very similar for all methods and there were no statistically significant differences in accuracy between any of the region drawing methods (p-values ranging between 0.23 and 1.00).

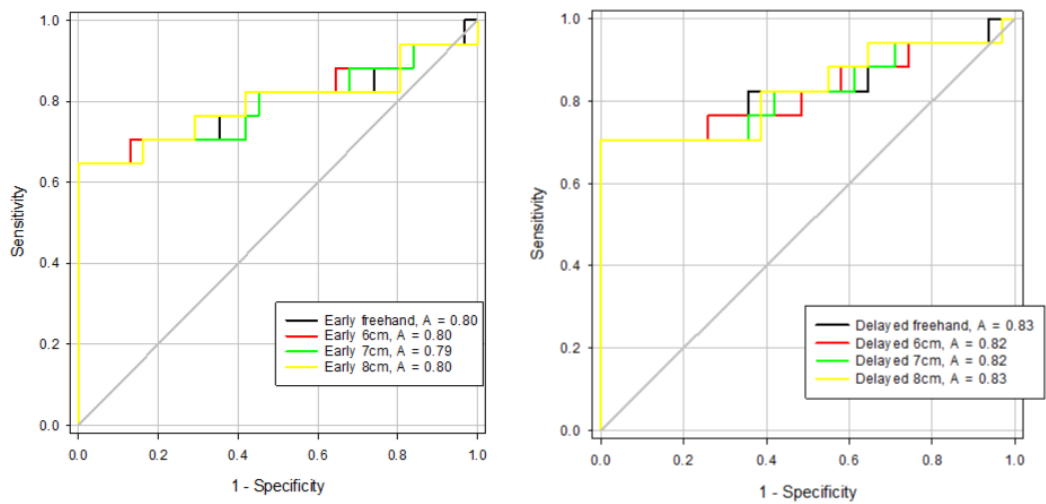


Fig 3: ROC curves for all cardiac region methods on early images (left) and delayed images (right).

The optimal cut-offs for each method were applied to the HMR values for the DLB and non-DLB groups. This gave sensitivities for distinguishing between DLB and non-DLB subjects ranging between 65% and 71% for the different region methods. The specificity with the optimal cut-offs applied was 100% for all methods.

The results for the 6cm circle region HMR on delayed images **Error! Reference source not found.** were 71% sensitivity (95% binomial confidence interval: 44% to 90%) and 100% specificity (89% to 100%) using the optimal cut-off of 1.66. The overall accuracy of this method was 90% (77% to 97%).

Effect of time point on HMR

Table 5 shows the difference in mean HMR for the DLB and non-DLB groups for each method, with 95% confidence intervals around the means. The difference between group mean HMRs is higher for delayed imaging than early imaging for all region sizes, with the delayed 6cm method giving the largest difference.

Table 4: ROC curve areas and 95% confidence intervals for all regions tested.

| | Early freehand | Early 6cm | Early 7cm | Early 8cm | Delayed freehand | Delayed 6cm | Delayed 7cm | Delayed 8cm |
|-------------------------|----------------|--------------|--------------|--------------|------------------|--------------|--------------|--------------|
| ROC curve area | 0.80 | 0.80 | 0.79 | 0.80 | 0.83 | 0.82 | 0.82 | 0.83 |
| 95% confidence interval | 0.63 To 0.96 | 0.63 To 0.96 | 0.63 To 0.96 | 0.63 To 0.96 | 0.68 To 0.98 | 0.67 To 0.98 | 0.67 To 0.98 | 0.68 To 0.98 |

Table 5: Difference in mean HMR between DLB and non-DLB groups for the HMR methods tested, with 95% confidence intervals around the mean HMR values given in brackets

| Time point | Region method | Mean HMR DLB (95% CI) | Mean HMR non-DLB (95% CI) | Difference |
|------------|---------------|-----------------------|---------------------------|------------|
| Early | Freehand | 1.80 (1.49 to 2.12) | 2.45 (2.33 to 2.58) | 0.65 |
| | 6cm circle | 1.84 (1.51 to 2.19) | 2.57 (2.43 to 2.70) | 0.73 |
| | 7cm circle | 1.88 (1.51 to 2.25) | 2.49 (2.36 to 2.61) | 0.61 |
| | 8cm circle | 1.81 (1.50 to 2.12) | 2.40 (2.28 to 2.52) | 0.59 |
| Delayed | Freehand | 1.63 (1.28 to 1.98) | 2.50 (2.35 to 2.65) | 0.87 |
| | 6cm circle | 1.69 (1.28 to 2.10) | 2.64 (2.48 to 2.81) | 0.95 |
| | 7cm circle | 1.68 (1.30 to 2.06) | 2.56 (2.41 to 2.72) | 0.88 |
| | 8cm circle | 1.65 (1.31 to 2.00) | 2.47 (2.32 to 2.61) | 0.82 |

However, there are no statistically significant differences in HMR between the early and delayed images for any region method ($p=0.37$ to 0.74 , Table 2). The wider range of HMRs on delayed imaging (over all subjects) is shown for the 6cm circle region method, which gave the largest separation between group means, in Figure 4.

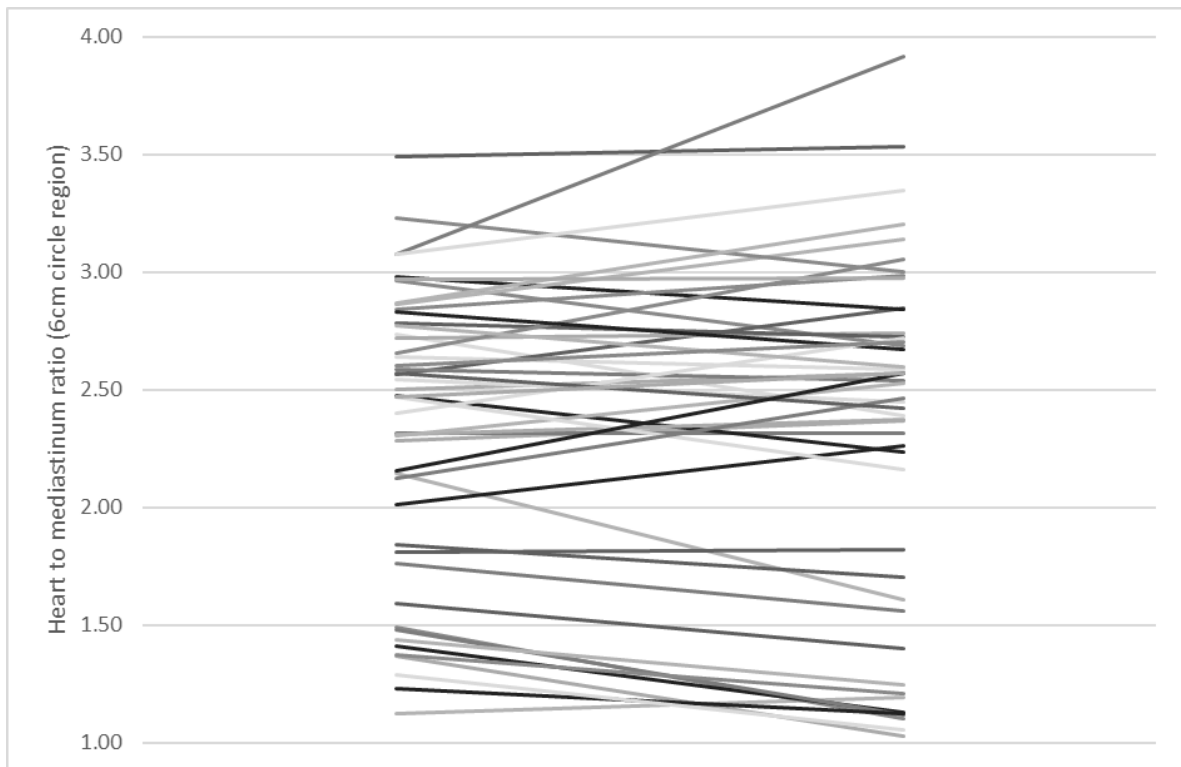


Fig 4: Comparison of early (left) and delayed (right) HMR values for all subjects, showing a wider variation in HMR values on the delayed images

Comparison between visual rating and HMR analysis

Four of the 32 dementia cases were rated by the panel as normal on early images and abnormal on delayed images or vice versa giving a concordance of 88% between early and delayed ratings. Concordance between early and delayed image rating was 100% for controls, although two of these control cases were rated as abnormal. Sensitivity was higher for visual rating of delayed images than early images but specificity higher for early image rating; the overall accuracy, defined as the percentage of correct designations, was the same for both time points (Table 6). The agreement between visual raters was very good for both early and delayed images (intraclass correlation coefficient of 0.96 (0.94 to 0.98) for both). Overall, accuracy was lower for visual rating (73%) than for HMR analysis (90%), due to the lower specificity.

Table 6: Sensitivity, specificity and accuracy for distinguishing between DLB and non-DLB subjects using consensus visual rating, with binomial confidence intervals in brackets.

| | SENSITIVITY | SPECIFICITY | ACCURACY |
|-----------------------|------------------|------------------|------------------|
| EARLY VISUAL | 65% (38% to 86%) | 77% (59% to 90%) | 73% (58% to 85%) |
| DELAYED VISUAL | 76% (50% to 93%) | 71% (52% to 86%) | 73% (58% to 85%) |

The overall concordance between consensus visual rating and mean HMR using the 6cm circle method for the controls was 12/16 (75%) for early images and 13/16 (81%) for delayed. Overall the concordance between the 6cm circle HMR method and visual rating for delayed images for all subjects was 38/48 (79%).

Discussion

Effect of cardiac region method on HMR

We have demonstrated that the region drawing method had a significant impact on HMR values and that ROI methods should not be used interchangeably – the appropriate cut-off depends on the processing method. In this dataset, the 6cm circle HMR results were significantly higher than the freehand results, likely due to the freehand regions of interest being larger on average and so including more non-specific background counts.

A fixed sized region minimises variation between operators processing the same patient, but could in theory lead to inaccuracies for individual patients. The size of the visible myocardium varies between patients, as does the size of the ventricular cavity, often visible on images as a “cold” area with low counts. However our results show very similar accuracy for all methods suggesting that this effect is minimal. Given the higher variation between operators for freehand region HMRs and no benefit in accuracy, we recommend that fixed sized circular regions be used over freehand.

The 8cm circle method gave the lowest inter-operator variability. However, the difference in both accuracy and variability between the 8cm and 6cm circle method is very small and not statistically significant. Furthermore, slightly smaller ROIs are less likely to be influenced by non-cardiac uptake in the liver, lungs or bowel and at the same time less likely to include background counts surrounding the LV. Our study supports the use of the 6cm fixed circle method used in several previous publications, and we therefore suggest that this be adopted as standard.

Accuracy of visual rating compared with HMR analysis

In our study, consensus visual rating with binary outcome (i.e. probable/definite normal vs probable/definite abnormal) performed less well than using an HMR cut-off, with a lower specificity for detecting DLB and two of the 16 control subjects rated as probably abnormal. Of the two controls rated as probably abnormal on both early and delayed imaging, one had borderline HMR values on delayed imaging using the 6cm circle method (2.16), but not on early (2.47). The other control rated as abnormal had normal HMR values on both early and delayed imaging (2.47; 2.57).

Tiraboschi *et al.* [27] used binary (normal / abnormal) visual image assessment (consensus of three raters) as their primary analysis method in their study comparing cardiac MIBG and FP-CIT for DLB diagnosis, with HMR analysis as a secondary method. In contrast to our study, their results show 100% specificity for both visual and HMR techniques (95% confidence interval 88-100%). The sensitivity is slightly higher for the HMR method than visual consensus (97% vs 93%) but the confidence intervals overlap considerably and the authors state the results were similar [27]. We found that extending our normal categorisation to include the “probably abnormal” category improved specificity without having a large impact on sensitivity, suggesting that any level of cardiac uptake above background is likely normal. However such a system may not be useful in detecting early disease, where we might expect to see subtly reduced uptake.

Although the heart-to-mediastinum ratio is intended to be objective, it may be affected by the amount of lung and liver uptake and proximity to the heart. In publications using HMR results, visual inspection of images for confounding factors potentially affecting HMR is not typically mentioned. The presence of lung uptake on MIBG scans is normal and is due to uptake by pulmonary endothelial cells [39, 40]. Reduced lung uptake has been reported in Lewy body disease [41] and has been linked to treatment with L-dopa leading to an increase in plasma homocysteine levels [42]. Reduced lung uptake in DLB due to medication could potentially

exaggerate the difference in HMR between DLB and non-DLB groups, due to reduced scatter from the lungs into the cardiac ROI in those taking L-dopa. However, although ours is a small sample, we found no evidence of any correlation between L-dopa dose and lung uptake for DLB patients.

We recommend HMR analysis be used as the primary method of assessment for planar cardiac MIBG images for DLB diagnosis, rather than visual rating. However the visual appearance of the study should still be taken into account, particularly in borderline cases where extra-cardiac uptake could give confounding results.

Difference between early and delayed imaging

Our results show no significant difference between HMR values taken on early and delayed images for dementia subjects or healthy controls. Examining the literature on early versus delayed MIBG in dementia and Lewy body disease we found two articles that stated early MIBG imaging was more accurate, 13 that stated delayed imaging was more accurate and 18 that reported no difference. However, only five of these compared early and delayed statistically, four of which showed no significant differences [30, 32, 33, 43]. Yoshita *et al.* [44] demonstrated that HMR values in Parkinson's disease were significantly lower on delayed images than early images, whereas control HMR values were the opposite and were higher on the delayed images. The differences between early and late HMR were statistically significant.

It is possible that there is a true difference in the accuracy of early and delayed imaging but that our dataset is too small to reach statistical significance. Several of our subjects had quite different uptake levels on delayed imaging compared to early; this did not affect normal/abnormal categorisation with the optimal ROC cut-offs applied, but would have done with the higher prior cut-off of 2.25 applied. Furthermore, we tended to see greater differences in mean HMR between our DLB and non-DLB groups on delayed imaging compared with early, which is in keeping with previous studies, including those that demonstrated no statistically significant difference in diagnostic accuracy between early and delayed imaging [33, 43]. Given the uncertainty in the literature, the theoretical advantage of delayed images, which are thought to demonstrate function rather than distribution of the cardiac sympathetic nerves, we do not recommend using early images alone in research studies. However, in clinical practice the delayed image could be omitted if the early image showed a non-equivocal result, saving time in the majority of cases.

Use of an HMR cut-off from a published multicentre study

The results of a recent study at our centre [34] suggest that our threshold of 2.25 adapted from the Japanese multicentre study [17] is too high, as several patients with clear Alzheimer's disease and without clinical signs suggestive of Lewy body disease have HMR values below this cut-off. Similarly, the lowest control HMR value is 2.16 (6cm circle method on delayed images). The optimal cut-off for this dataset calculated using ROC analysis was only 1.66, giving 71% sensitivity and 100% specificity. With the prior cut-off of 2.25 applied the sensitivity would remain 71% but specificity drop to 87%, due to three AD subjects and one control subject becoming false positives.

Since calibration actually increased the cut-off obtained in the multicentre study from 2.10 to 2.25 for our centre, we conclude that the discrepancy in specificity is likely to be due to differences in the patient population studied, rather than due to differences between acquisition parameters. Indeed, different optimal HMR cut-offs were found in the three-year follow up to the Yoshita study [18], which used 65 of the original 113 subjects.

Conclusion

HMR values depend on the method used to draw the cardiac region, with the smallest region (6cm diameter) giving the highest values. Accuracy is similar for all HMR methods but inter-operator variation lower for fixed sized ROI methods than freehand. Visual rating specificity is lower than HMR methods so HMR should be used as the primary reporting method, provided an appropriate cut-off for the gamma camera(s) used has been established. We found no significant difference in HMR between early and delayed images. However, the separation between the mean HMR of DLB and non-DLB groups did tend to be greater for delayed methods; further studies using both early and delayed imaging would be needed to confirm this.

Given that all fixed sized region HMR methods give similar results, we recommend using a 6cm diameter circular region for research studies; this is in keeping with published methods and may reduce the effect of non-cardiac uptake compared to larger ROIs. As there are theoretical advantages to delayed imaging, we recommend continuing to acquire delayed images for research studies even though the accuracy of early images appears to be similar. The sensitivity and specificity of the 6cm ROI method on delayed images with optimal cut-off of 1.66 were 71% and 100%.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Newcastle University and Newcastle upon Tyne NHS Foundation Trust research committee 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.

Acknowledgements

We thank the clinical technologists, scientists and radiologists who processed and rated scans for this study: Tamir Ali, Elizabeth Howell, Elizabeth Jefferson, Janice Johnston, David McCulloch, Richard Peace and Timothy Powell.

We thank Edwin Poel and Kenichi Nakajima for their kind assistance with gamma camera calibration.

References

1. Vann Jones, S.A. and J.T. O'Brien, *The prevalence and incidence of dementia with Lewy bodies: a systematic review of population and clinical studies*. Psychol Med, 2014. **44**(4): p. 673-83.
2. Kane, J.P.M., et al., *Clinical prevalence of Lewy body dementia*. Alzheimers Res Ther, 2018. **10**(1): p. 19.
3. Hanyu, H., et al., *Differences in clinical course between dementia with Lewy bodies and Alzheimer's disease*. Eur J Neurol, 2009. **16**(2): p. 212-7.
4. Lee, D.R., et al., *Examining carer stress in dementia: the role of subtype diagnosis and neuropsychiatric symptoms*. Int J Geriatr Psychiatry, 2013. **28**(2): p. 135-41.
5. Galvin, J.E., et al., *Lewy body dementia: the caregiver experience of clinical care*. Parkinsonism Relat Disord, 2010. **16**(6): p. 388-92.
6. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium*. Neurology, 2017. **89**(1): p. 88-100.
7. Donaghy, P.C. and I.G. McKeith, *The clinical characteristics of dementia with Lewy bodies and a consideration of prodromal diagnosis*. Alzheimers Res Ther, 2014. **6**(4): p. 46.
8. Thomas, A.J., et al., *Autopsy validation of 123I-FP-CIT dopaminergic neuroimaging for the diagnosis of DLB*. Neurology, 2017. **88**(3): p. 276-283.
9. Watson, R. and S.J. Colloby, *Imaging in Dementia With Lewy Bodies: An Overview*. J Geriatr Psychiatry Neurol, 2016. **29**(5): p. 254-60.
10. Colloby, S.J., et al., *Neuropathological correlates of dopaminergic imaging in Alzheimer's disease and Lewy body dementias*. Brain, 2012. **135**(Pt 9): p. 2798-808.
11. McKeith I, et al., *Sensitivity and specificity of dopamine transporter imaging with (123)I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study*. 2007.
12. Yoshita, M., et al., *Value of 123I-MIBG radioactivity in the differential diagnosis of DLB from AD*. Neurology, 2006. **66**(12): p. 1850-4.
13. Estorch, M., et al., *Cardiac (123)I-metaiodobenzylguanidine imaging allows early identification of dementia with Lewy bodies during life*. Eur J Nucl Med Mol Imaging, 2008. **35**(9): p. 1636-41.
14. Hanyu, H., et al., *The role of 123I-metaiodobenzylguanidine myocardial scintigraphy in the diagnosis of Lewy body disease in patients with dementia in a memory clinic*. Dement Geriatr Cogn Disord, 2006. **22**(5-6): p. 379-84.
15. Slaets, S., et al., *Diagnostic value of MIBG cardiac scintigraphy for differential dementia diagnosis*. Int J Geriatr Psychiatry, 2015. **30**(8): p. 864-9.
16. Treglia, G. and E. Cason, *Diagnostic performance of myocardial innervation imaging using MIBG scintigraphy in differential diagnosis between dementia with lewy bodies and other dementias: a systematic review and a meta-analysis*. J Neuroimaging, 2012. **22**(2): p. 111-7.
17. Yoshita, M., et al., *Diagnostic accuracy of 123I-meta-iodobenzylguanidine myocardial scintigraphy in dementia with Lewy bodies: a multicenter study*. PLoS One, 2015. **10**(3): p. e0120540.
18. Komatsu, J., et al., *(123)I-MIBG myocardial scintigraphy for the diagnosis of DLB: a multicentre 3-year follow-up study*. J Neurol Neurosurg Psychiatry, 2018.
19. Nakajima, K., et al., *Standardization of 123I-meta-iodobenzylguanidine myocardial sympathetic activity imaging: phantom calibration and clinical applications*. Clin Transl Imaging, 2017. **5**(3): p. 255-263.

20. Verberne HJ, et al., *Vascular time-activity variation in patients undergoing ¹²³I-MIBG myocardial scintigraphy: implications for quantification of cardiac and mediastinal uptake*. Eur J Nucl Med Mol Imaging, 2011. **38**(6): p. 7.
21. Verschure DO, et al., *¹²³I-MIBG heart-to-mediastinum ratio is influenced by high-energy photon penetration of collimator septa from liver and lung activity*. Nucl Med Commun, 2015. **36**(3): p. 7.
22. Verschure, D.O., et al., *A European myocardial 123I-mIBG cross-calibration phantom study*. J Nucl Cardiol, 2017.
23. Slomka, P., et al., *Quantification of I-123-meta-iodobenzylguanidine Heart-to-Mediastinum Ratios: Not So Simple After All*. J Nucl Cardiol, 2014. **21**: p. 979-983.
24. Klene, C., et al., *Influence of ROI definition on the heart-to-mediastinum ratio in planar 123I-MIBG imaging*. J Nucl Cardiol, 2016.
25. Nakajima, K., et al., *Standardization of metaiodobenzylguanidine heart to mediastinum ratio using a calibration phantom: effects of correction on normal databases and a multicentre study*. Eur J Nucl Med Mol Imaging, 2012. **39**(1): p. 113-9.
26. Treglia, G., et al., *Iodine-123 metaiodobenzylguanidine scintigraphy and iodine-123 ioflupane single photon emission computed tomography in Lewy body diseases: complementary or alternative techniques?* J Neuroimaging, 2014. **24**(2): p. 149-54.
27. Tiraboschi, P., et al., *(123) I-2beta-carbomethoxy-3beta-(4-iodophenyl)-N-(3-fluoropropyl) nortropane single photon emission computed tomography and (123) I-metaiodobenzylguanidine myocardial scintigraphy in differentiating dementia with lewy bodies from other dementias: A comparative study*. Ann Neurol, 2016. **80**(3): p. 368-78.
28. Okuda, K., et al., *Semi-automated algorithm for calculating heart-to-mediastinum ratio in cardiac Iodine-123 MIBG imaging*. J Nucl Cardiol, 2011. **18**(1): p. 82-9.
29. Chen, W., Q. Cao, and V. Dilsizian, *Variation of Heart-to-Mediastinal Ratio in 123I-mIBG Cardiac Sympathetic Imaging: Its Affecting Factors and Potential Corrections* Curr Cardiol Rep, 2011. **13**: p. 6.
30. Sakamoto, F., et al., *123I-MIBG myocardial scintigraphy for the evaluation of Lewy body disease: are delayed images essential? Is visual assessment useful?* Br J Radiol, 2016: p. 20160144.
31. Orimo, S., et al., *(123)I-meta-iodobenzylguanidine (MIBG) cardiac scintigraphy in alpha-synucleinopathies*. Ageing Res Rev, 2016. **30**: p. 122-33.
32. Maruyama, Y., et al., *Comparison of the diagnostic performance of H/M ratio between early and delayed phases for Lewy body disease*. Nucl Med Commun, 2015. **36**(5): p. 477-80.
33. Manabe, Y., et al., *123I-metaiodobenzylguanidine myocardial scintigraphy with early images alone is useful for the differential diagnosis of dementia with Lewy bodies*. Psychiatry Res, 2017. **261**: p. 75-79.
34. Kane, J.P.M., et al., *(123)I-MIBG scintigraphy utility and cut-off value in a clinically representative dementia cohort*. Parkinsonism Relat Disord, 2019.
35. McKeith, I., et al., *Predictive Accuracy of Clinical Diagnostic Criteria for Dementia with Lewy Bodies: A Prospective Neuropathological Validation study*. 54, 1050-1058. Neurology, 2000(54): p. 1050-1058.
36. Flotats, A., et al., *Proposal for standardization of 123I-metaiodobenzylguanidine (MIBG) cardiac sympathetic imaging by the EANM Cardiovascular Committee and the European Council of Nuclear Cardiology*. Eur J Nucl Med Mol Imaging, 2010. **37**(9): p. 1802-12.

37. Fujishiro, H., et al., *REM sleep without atonia in middle-aged and older psychiatric patients and Lewy body disease: a case series*. Int J Geriatr Psychiatry, 2016.
38. Goetz, C.G., M. Emre, and B. Dubois, *Parkinson's disease dementia: definitions, guidelines, and research perspectives in diagnosis*. Ann Neurol, 2008. **64 Suppl 2**: p. S81-92.
39. Lee, K., et al., *Characteristics and regulation of 123I-MIBG transport in cultured pulmonary endothelial cells*. J Nucl Med, 2006. **47**: p. 437-42.
40. Arao, T., N. Takabatake, and M. Sata, *In vivo evidence of endothelial injury in chronic obstructive pulmonary disease by lung scintigraphic assessment of 123I-metaiodobenzylguanidine*. J Nucl Med, 2003. **44**: p. 1747-1752.
41. Nakae, I., et al., *Clinical significance of lung iodine-123 metaiodobenzylguanidine uptake assessment in Parkinson's and heart diseases*. Ann Nucl Med, 2013. **27**(8): p. 737-47.
42. Miller, J.W., et al., *Effect of l-dopa on plasma homocysteine in PD patients: Relationship to B-vitamin status*. Neurology, 2003. **60**(7).
43. Shin, D.H., et al., *Clinical Implications of Cardiac-MIBG SPECT in the Differentiation of Parkinsonian Syndromes*. J Clin Neurol, 2006. **2**(1): p. 51-7.
44. Yoshita, M., M. Hayashi, and S. Hirai, *Decreased myocardial accumulation of 123I-meta-iodobenzyl guanidine in Parkinson's disease*. Nucl Med Commun, 1998. **19**(2): p. 137-42.