



Can we improve the early diagnosis of Lewy
body disease with more accurate quantification
of nuclear medicine scans?

Gemma Roberts

Institute of Neuroscience, Newcastle University

&

Nuclear Medicine department, Newcastle upon Tyne
Hospitals NHS Foundation Trust

Thesis submitted for the degree of Doctor of Philosophy

September 2019

In memory of Doreen Rix

1947 – 2017

Abstract

This thesis investigates the quantification of two scintigraphic biomarkers used for the diagnosis of dementia with Lewy bodies (DLB): ^{123}I -FP-CIT (^{123}I -N- ω -fluoropropyl-2 β -carbomethoxy-3 β -(4-iodophenyl) nortropane), commonly known as DaTSCAN™, and cardiac ^{123}I -MIBG (^{123}I -metaiodobenzylguanidine). Accurate quantification is critical as we increasingly move towards diagnosis at the earlier mild cognitive impairment (MCI) stage, where more subtle changes from normality are expected. A range of novel approaches have been examined to overcome technical limitations that have previously been barriers to accurate quantification. Uniquely, this has been studied in cohorts of highly characterised dementia and MCI subjects as well as older adults with normal cognition recruited as age matched controls. The subject studies have been complemented by work using advanced anthropomorphic phantoms. Throughout, the innovative methods have been compared with the established ones. Results are presented in detail and clinical and research relevance is discussed together with proposals for optimal usage. Briefly, the key findings are:

FP-CIT key findings

- Specific binding ratio values (SBR) for FP-CIT images calculated by different software packages are systematically different, although give similar diagnostic accuracy.
- Striatal uptake does not decrease with age in healthy older adults, as previously reported, indicating potential misdiagnosis if age correction is applied.
- Absolute quantification separates normal and abnormal subjects less well than relative-quantification with SBR.
- Advanced FP-CIT reconstruction methods using SPECT-CT and collimator modelling improve the accuracy of activity concentration measurements in a phantom.
- Advanced FP-CIT reconstruction methods affect relative quantification with SBR, but not clinical interpretation.

Cardiac MIBG key findings

- Different methods of planar MIBG analysis are operator dependent and give systematically different results – recommendations are provided for an optimal method.
- Establishing a normal threshold is critical. This thesis shows that previously published values may not be valid in a UK population and proposes a suitable alternative.
- Images obtained soon after injection give similar accuracy as those obtained at 3.5 hours (the standard delayed method), and the latter scans could be omitted in the majority of cases.
- Planar cardiac MIBG semi-quantification is significantly dependent on subject size. Using SPECT-CT gives greater separation between normal and abnormal scans than planar.

In summary, an in-depth and comprehensive study of technical aspects of Nuclear Medicine biomarker quantification using ^{123}I labelled radiopharmaceuticals for the diagnosis of Lewy body disease is presented in this thesis. This provides a solid foundation for clinical and research application of these techniques in both early and established disease.

Acknowledgments

I would like to thank my supervisors Professor Alan Thomas and Dr. Jim Lloyd for their guidance and support both throughout this doctoral fellowship and leading up to it. Additional thanks are due to Dr. Lloyd for reviewing this manuscript in detail and suggesting ways to make it clearer. I am grateful for computing and statistical advice from my co-supervisor Dr. Andrew Sims and from Dr. Michael Firbank, and for further helpful advice and suggestions from Professors John O'Brien and John-Paul Taylor and from Dr. George Petrides. I had helpful discussions by email and at conferences with Prof Kenichi Nakajima (Kanazawa University Hospital, Japan), Dr. Colin Brown (NHS Greater Glasgow and Clyde), Dr. John Dickson (University College London Hospitals), Dr. Livia Tossici-Bolt (University Hospitals Southampton NHS Trust) and Dr. David Raffel (University of Michigan, USA).

It has been a pleasure to work with Dr. Jim Lloyd, Elizabeth Jefferson, Dr. George Petrides, Kim Howe, Dr. Ifti Haq, Andrew Curry and the entire team of clinical scientists, technologists and support staff in the NuTH nuclear medicine department. I am particularly grateful to consultant radiologists Dr. Petrides and Dr. Tamir Ali for their additional efforts to support the SUPeRB study, without which some the CT data collection for this project would not have been possible. I would also like to thank and acknowledge the MIDAS and SUPeRB study teams at Newcastle University, led by Professor Thomas - in particular Dr. Joseph Kane, Dr. Rory Durcan, Helen Kain, Dr. Sean Colloby, Dr. Paul Donaghy, Dr. Sarah Lawley, Sally Barker, Calum Hamilton and Joanna Ciafone.

I thank the Alzheimer's Society for awarding me the fellowship, for featuring the project on their website and magazine and showcasing my work at their annual conferences. Melanie Howard, Keith Newbegin and Isobel Common, the volunteer monitors appointed by the Alzheimer's Society to support my work have been fantastic to work with and I thank them for their helpful and insightful feedback throughout. Finally I would like to acknowledge the substantial time and effort of everyone who volunteered as a patient or healthy control for our studies and that of their family members who accompanied them. It's fantastic that people want to contribute to clinical research by volunteering for studies and we are incredibly grateful for their support.

Declaration

This PhD thesis was submitted in September 2019. The viva examination was held in December 2019 with the outcome that the degree should be awarded pending minor corrections. This is the corrected version of the manuscript, accepted by the internal examiner in January 2020.

The work submitted for this PhD was carried out between September 2016 and September 2019 as part of a full-time Clinician and Healthcare Professional Training Fellowship, funded by the Alzheimer's Society. In accordance with the terms of the clinical fellowship, I spent up to one day a week continuing to work as a Clinical Scientist within the Nuclear Medicine department of the Newcastle upon Tyne Hospitals NHS Foundation Trust (NuTH), in order to maintain state registration. Much of the PhD work was also carried out in the NuTH Nuclear Medicine department, under an honorary NuTH contract. Prior to the fellowship I was awarded part-time Research Capability Funding by the National Institute for Health research, which released me from clinical duties for two days a week for a year and enabled me to carry out pilot work related to this thesis and submit doctoral fellowship applications.

The patient scan data analysed in this thesis was acquired in the NuTH Nuclear Medicine department for Newcastle University and NuTH research studies (LewyPro, MIDAS and SUPeRb), all of which were led by my primary supervisor Professor Alan Thomas. LewyPro and MIDAS were funded by the Newcastle NIHR biomedical research centre and SUPeRb by Alzheimer's Research UK. GE Healthcare funded the FP-CIT scans in LewyPro and SUPeRb, provided the DaTQUANT software package and funded two of the phantom studies. All study participants gave informed consent prior to participating, including consent for their data to be used in other Newcastle studies, so I was able to use the data for this thesis. I was involved in setting up the nuclear medicine SPECT-CT protocols for two of the studies (MIDAS and SUPeRb), prior to data collection. I was not involved in any patient recruitment, patient assessment or scanning, other than as an observer. I was directly involved in image processing, image interpretation and data analysis for the MIDAS and SUPeRb studies. I used data from the LewyPro study for this thesis but otherwise had no direct involvement in this study.

I carried out all the analysis in this thesis, with support from my supervisors and mentors. In 2018 Professor Thomas and I supervised a Medical Sciences MSc student, Jamie Reyes, who carried out a project related to the work in section 8.4. I devised the MSc project and later re-did the analysis using different methods for this thesis and publication.

In 2017 and 2018 I gave various talks on the FP-CIT quantification work in sections 4.1 and 4.2 at workshops organised by GE Healthcare, for which I received honoraria. Details of published work are given in the next section.

Published work

The following first-author papers are based directly on work contained in this thesis

1. **Roberts G**, Lloyd JJ, Petrides GS, Donaghy PC, Kane JPM, Durcan R, Lawley S, Howe K, Sims AJ, Taylor JP, O'Brien JT, Thomas AJ. 123I-FP-CIT striatal binding ratios do not decrease significantly with age in older adults. *Ann Nucl Med*. 2019 Jun;33(6):434-443. **(Section 9.1)**
2. **Roberts G**, Kane JPM, Lloyd JJ, Petrides GS, Howe K, O'Brien JT, Thomas AJ. A comparison of visual and semiquantitative analysis methods for planar cardiac 123I-MIBG scintigraphy in dementia with Lewy bodies. *Nucl Med Commun*. 2019 Jul;40(7):734-743. **(Section 7.1)**
3. **Roberts G**, Lloyd JJ, Petrides GS, Donaghy PC, Kane JPM, Durcan R, Lawley S, Howe K, Sims AJ, Taylor JP, O'Brien JT, Thomas AJ. Cardiac 123I-MIBG normal uptake values are population-specific: results from a cohort of controls over 60 years of age. *J Nuc Cardiol*. 2019 Sep 16. [Epub ahead of print], to be published with editorial piece by Prof. Kenichi Nakajima **(Section 7.2)**
4. **Roberts G**, Lloyd JJ, Petrides GS, O'Brien JT, Thomas AJ. Cardiac 123I-MIBG planar heart to mediastinum ratios depend on patient size; phantom studies suggest SPECT-CT could improve quantification. *Biomed. Phys. Eng. Express* 2019, in press **(Section 8.1)**
5. **Roberts G**, Lloyd JJ, Jefferson EJ, Kane JPM, Durcan R, Lawley S, Petrides GS, Howe K, Haq I, O'Brien JT, Thomas AJ. Uniformity of cardiac 123I-MIBG uptake on SPECT images in older adults with normal cognition and patients with dementia. *J Nucl Cardiol*. 2019 Dec 9. [Epub ahead of print] **(Section 8.4)**

The following publications are not directly part of this thesis, but contain work that I contributed to during this fellowship

1. Kane JPM, **Roberts G**, Petrides GS, Lloyd JJ, O'Brien JT, Thomas AJ. 123I-MIBG scintigraphy utility and cut-off value in a clinically representative dementia cohort. *Parkinsonism Relat Disord*. 2019 May;62:79-84.
2. Lloyd JJ, Petrides G, Donaghy PC, Colloby SJ, Attems J, O'Brien JT, **Roberts G**, Thomas AJ. A new visual rating scale for Ioflupane imaging in Lewy body disease. *Neuroimage Clin*. 2018;20:823-829.
3. Thomas AJ, Donaghy P, **Roberts G**, Colloby SJ, Barnett NA, Petrides G, Lloyd J, Olsen K, Taylor JP, McKeith I, O'Brien JT. Diagnostic accuracy of dopaminergic imaging in prodromal dementia with Lewy bodies. *Psychol Med*. 2019 Feb;49(3):396-402.

Abbreviations and nomenclature

<i>Term</i>	<i>Description</i>
^{123}I	Iodine-123; radionuclide used to label both radiopharmaceuticals included in this thesis
^{123}I -FP-CIT	^{123}I -N- ω -fluoropropyl-2 β -carbomethoxy-3 β -(4-iodophenyl) nortropane; ioflupane
^{123}I -MIBG	^{123}I -metaiodobenzylguanidine
$^{99\text{m}}\text{Tc}$	Technetium-99m; most common radionuclide used in nuclear medicine
ACE	Addenbrooke's cognitive examination; used for identifying cognitive impairment
ACSC	Attenuation and scatter correction
AD	Alzheimer's disease (the dementia stage is usually implied)
AF	Atrial fibrillation
Benamer	Author of well-known FP-CIT visual ratings scale developed for PD
BMI	Body mass index
BRASS	Hermes software for brain scan quantification, including FP-CIT SBR analysis
CAD	Coronary artery disease
Cardiac MIBG	Throughout this thesis, the term "cardiac MIBG" is used to refer to the cardiac sympathetic imaging technique using ^{123}I -MIBG
CD_{norm}	Normalised count density; quantity analogous to SUV, used in this thesis for cardiac MIBG absolute quantification
CoV	Coefficient of variation; measurement of spread of data or noise in dataset
CSPECT	Coronal SPECT images, summed to give image similar to planar scan
CSPECT-CT	Summed coronal SPECT images, with CT ACSC applied
DaTQUANT	Software for performing SBR analysis, manufactured by GE Healthcare
DaTSCAN™	GE Healthcare's tradename for the ^{123}I -FP-CIT radiopharmaceutical; commonly used in clinical practice to refer to the scan itself
DLB	Dementia with Lewy bodies
EANM	European Association of Nuclear Medicine
ENC-DAT	EANM multicentre project acquiring FP-CIT scans of healthy controls
FBP	Filtered back projection; commonly used SPECT reconstruction method
FCM	Full (Monte Carlo) collimator model
FP-CIT	Throughout this thesis, the term "FP-CIT" is used to refer to the dopaminergic imaging technique using ^{123}I -FP-CIT
Hermes	Hermes Medical Solutions Ltd; manufacturer of nuclear medicine image reconstruction and processing software used at NuTH
HMR	Heart to mediastinum ratio; H/M also used in the literature

HMR cut-off	Threshold value, below which HMRs are considered abnormal
HybridRecon	Hermes software for reconstructing SPECT images
Intevo	Siemens hybrid gamma camera with SPECT-CT functionality
kBq/ml	Unit of activity concentration in phantom or tissue
LBM	Lean body mass; alternative normalisation method to patient weight
LEHR	Low energy high resolution; refers to collimator properties
LewyPro	Newcastle study into the use of FP-CIT in MCI patients
MBq	10^6 Becquerels - unit of radioactivity; used to specify radiopharmaceutical activities administered to patients
mCi	10^{-3} Curies - unit of radioactivity; used in USA. 1 mCi = 37 MBq
MCI-AD	Mild cognitive impairment due to Alzheimer's disease
MCI-LB	Mild cognitive impairment with Lewy bodies
MCRR_CT	Reconstruction protocol used for FP-CIT images incorporating full Monte Carlo model of collimator interactions, resolution recovery and CT ACSC;
MCRR_uniform	Reconstruction protocol used for FP-CIT images incorporating full Monte Carlo model of collimator interactions, resolution recovery and uniform ACSC;
MDS-UPDRS	Movement disorders society unified PD ratings scale; used to quantify parkinsonian features
MEGP	Medium energy general purpose; refers to collimator properties
MELP	Medium energy low penetration; refers to collimator properties
MI	Myocardial infarction
MIDAS	Newcastle pilot study into the use of cardiac MIBG in a typical NHS setting
MLEM	Maximum likelihood expectation maximisation; iterative reconstruction method
MMSE	Mini mental state examination; used for identifying cognitive impairment
MRI	Magnetic resonance imaging, a functional and anatomical imaging technique
mSv	10^{-3} Sieverts – unit of effective radiation dose
NHS	National Health Service – public funded healthcare provider in the UK
NIA-AA	National Institute on Aging and the Alzheimer's Association (US public bodies)
NICE	National Institute for Clinical Excellence, UK public body that provides guidance on evidence based healthcare
NuTH	Newcastle upon Tyne Hospitals NHS Foundation Trust
OSEM	Ordered subsets expectation maximisation; fast implementation of MLEM
OSEM_RRACSC	Reconstruction protocol used for FP-CIT images analysed with BRASS incorporating resolution recovery and uniform ACSC
PD	Parkinson's disease
PDD	Parkinson's disease dementia
PET	Positron emission tomography, a radionuclide imaging technique

PPMI	Parkinson's progression markers initiative; multicentre PD study, from which the DaTQUANT healthy control database is taken
PSF	Point spread function; describes blurring of point source due to finite resolution
PubMed	Database of scientific literature related to medicine
QPS	Quantitative perfusion SPECT
QRISK2	Online tool used in NHS for estimating 10 year risk of cardiovascular event
REM	Rapid eye movement, as in REM sleep behaviour disorder
ROC	Receiver operating curve; analysis used to determine accuracy of a test
RR	Resolution recovery; geometric correction for collimator resolution using PSF
RR_CT	Reconstruction protocol used for FP-CIT images incorporating resolution recovery and CT ACSC
RR_uniform	Reconstruction protocol used for FP-CIT images incorporating resolution recovery and uniform ACSC; same as OSEM_RRACSC
RVI	Royal Victoria Infirmary; NuTH site and location of nuclear medicine department where study patients were scanned
SBR	Specific binding ratio; striatal to background ratio also used in literature
SNM	Society of Nuclear Medicine (US body)
SNR	Signal to noise ratio
Southampton method	Resolution-independent method for calculating SBR using large box-like striatal VOIs to capture all counts attributed to striata
SPECT	Single photon emission computed tomography
SPECT-CT	Hybrid imaging technique combining SPECT and x-ray computed tomography
Subset	Refers to dividing raw projections in OSEM method into multiple groups of fewer projections per iteration to speed up convergence
SUPERB	Newcastle study into the use of FP-CIT and cardiac MIBG in MCI patients, with older adults with normal cognition recruited as controls
SUV	Standardised uptake values, used for absolute quantification of tracer uptake
SUV _{max}	SUV of maximum voxel in VOI
SUV _{mean}	Mean SUV in VOI
SUV _{peak}	Mean SUV of maximum cm ³ within VOI
SUV-SPECT	Technique for determining standardised uptake values from SPECT or SPECT-CT images
Symbia	Siemens gamma camera model (no CT)
TEW	Triple energy window; empirical scatter correction method
VoD	Tracer volume of distribution within the body
VOI	Volume of interest; volume on SPECT image used for quantification

Contents

Chapter 1: Introduction.....	1
Chapter 2: Clinical background.....	4
2.1 Dementia and mild cognitive impairment.....	5
2.2 Nuclear medicine imaging techniques for Lewy body disease	9
2.3 Radionuclide imaging studies in DLB and MCI-LB at Newcastle University	13
Chapter 3: Technical background.....	18
3.1 Overview.....	19
3.2 Radionuclide imaging (nuclear medicine).....	20
3.3 Factors affecting quantification of ¹²³ I uptake in radionuclide imaging.....	25
3.4 Relative and absolute quantification	33
3.5 ¹²³ I-FP-CIT image analysis overview.....	35
3.6 Cardiac ¹²³ I-MIBG image analysis overview.....	38
Chapter 4: Development of imaging methods for research studies: FP-CIT.....	40
4.1 Comparison and validation of two commercial ¹²³ I-FP-CIT software packages	41
4.2 Comparison between BRASS and DaTQUANT in a large MCI research cohort	52
4.3 Comparison between DaTQUANT and BRASS in healthy ageing and dementia.....	55
4.4 Effects of reorientation and pixel size on quantification with BRASS.....	62
Chapter 5: Development of imaging methods for research studies: cardiac MIBG	63
5.1 Cardiac ¹²³ I-MIBG planar imaging protocol review	64
5.2 Phantom calibration for HMR normalisation between centres.....	68
5.3 Is 111 MBq an appropriate activity for cardiac MIBG imaging?.....	73
Chapter 6: ¹²³ I-FP-CIT image reconstruction and analysis.....	80
6.1 Comparison between visual rating and BRASS SBR results in SUPeRB MCI cohort.....	81
6.2 Absolute quantification of FP-CIT striatal uptake in MCI patients.....	91
6.3 Advanced FP-CIT reconstruction methods: a phantom study.....	104
6.4 Can SBR using advanced FP-CIT reconstruction methods improve early diagnosis?.....	116
6.5 Test-retest variation of FP-CIT SBR and SUV in healthy older adults.....	125
6.6 Summary of chapter 6.....	135
Chapter 7: Planar cardiac ¹²³ I-MIBG imaging for the diagnosis of Lewy body disease	137

7.1	Analysis methods for planar cardiac ¹²³ I-MIBG scintigraphy in DLB.....	138
7.2	Are cardiac ¹²³ I-MIBG normal uptake values population-specific?	148
7.3	Can we use early phase cardiac MIBG images to diagnose Lewy body disease?	158
7.4	Summary of Chapter 7.....	168
Chapter 8: Tomographic cardiac ¹²³ I-MIBG analysis.....		169
8.1	Planar cardiac MIBG HMRs depend on patient size – can SPECT-CT reduce dependence? 170	
8.2	Does cardiac ¹²³ I-MIBG SPECT-CT improve the diagnosis of Lewy body disease?	182
8.3	Absolute quantification of ¹²³ I-MIBG cardiac uptake using SPECT-CT imaging.....	195
8.4	Regional cardiac ¹²³ I-MIBG uptake in dementia and healthy ageing	207
8.5	Summary of chapter 8.....	219
Chapter 9: FP-CIT and MIBG uptake in healthy ageing		220
9.1	Do ¹²³ I-FP-CIT specific binding ratios decline with age in older adults?.....	221
9.2	Does cardiac ¹²³ I-MIBG uptake decrease with age in healthy older adults?.....	232
9.3	Summary of chapter 9.....	240
Chapter 10: Discussion and conclusions		241
10.1	Overview of main findings and novel contributions	242
10.2	Comparison between FP-CIT and cardiac MIBG.....	248
10.3	Strengths and limitations	250
10.4	Future directions	253
10.5	Concluding remarks.....	254
Appendices		254
Appendix A: Key technical and practical differences between DaTQUANT and BRASS.....		255
Appendix B: Factors affecting FP-CIT Z scores.....		256
Appendix C: Effect of scan reorientation and zoom on FP-CIT quantification with BRASS		259
Appendix D: The effect of pixel size on the quantification of FP-CIT images.....		268
Appendix E: Cardiac MIBG imaging protocol details from the literature.....		277
Appendix F: Consistency of FP-CIT processing for semi-quantification		280
Appendix G: Simulation work using Monte Carlo modelling		281
References.....		283

Chapter 1: Introduction

Lewy body disease is one of the most common causes of neurodegeneration and cognitive decline, second only to Alzheimer's disease. There are two nuclear medicine scans that have been shown to be useful in the detection of neurodegeneration at the relatively late dementia stages of the disease: the ^{123}I -FP-CIT scan (^{123}I -N- ω -fluoropropyl-2 β -carbomethoxy-3 β -(4-iodophenyl) nortropine) (DaTSCAN™) and the cardiac ^{123}I -MIBG scan (^{123}I -metaiodobenzylguanidine). Both scans are indicated as biomarkers in the 2017 consensus criteria for the diagnosis and management of dementia with Lewy bodies (DLB) [1], reflecting the extensive evidence base in the literature for their use at the dementia stage. However, they have both been much less frequently studied in the early stages of Lewy body disease.

Evidence from pathology studies suggests that Lewy bodies are present in the cardiac nerves at an early stage in the disease, and this can occur prior to brain involvement [2, 3]. Cardiac MIBG therefore has the potential to be the more sensitive test for distinguishing Lewy body disease from Alzheimer's disease at an early stage. However, the current standard methods for cardiac MIBG analysis are more basic than FP-CIT and could be subject to more technical inaccuracies. If these are not addressed, a potentially valuable biomarker test for early DLB diagnosis might not be used to its full potential.

Both scans have technical limitations that suggest that, whilst very accurate in distinguishing between DLB cases and other neurodegenerative conditions at the dementia stage, they may not be ideal for detecting early neurodegenerative change using current methods. Neither scan is routinely corrected for attenuation or scatter, and the radionuclide used, iodine-123 (^{123}I), poses further challenges due to its high energy emissions that reduce image contrast and complicate accurate quantification. In addition, the interpretation of quantification data is limited by the availability of reliable data from age-matched controls with similar co-morbidities to the clinical population.

In this doctoral fellowship project we sought to address these issues in a practical manner, using techniques that could easily be introduced into routine NHS practice, as well as more advanced methodology that could be implemented in a state-of-the-art centre of excellence setting, where more complex techniques for research and clinical use are appropriate. Throughout the thesis these standard techniques are compared with the more complex ones to evaluate their potential benefit.

We used both advanced anthropomorphic phantom data and patient data from joint Newcastle University and Newcastle Hospitals clinical research studies in Lewy body disease. These were the MIDAS study on the use of cardiac MIBG for DLB diagnosis in a typical UK population, which recruited 48 patients with dementia, the LewyPro study on the use of FP-CIT in 76 people with mild cognitive impairment, and the SUPeR study, which expands and extends LewyPro to include cardiac MIBG, recruiting 100 MCI patients and 31 age-matched controls.

An important feature of these studies is the recruitment of cohorts of highly characterised dementia and MCI subjects, with diagnoses confirmed by a panel of expert NHS Consultants. This approach has significant advantages over the more common approach in nuclear medicine, often used in peer reviewed publications, of retrospectively reviewing the scan data of patients referred to the service as part of clinical care. The clinical diagnoses of patients referred for a scan are intrinsically linked to the results of the scans and not independent of them as they are in our research studies. Our approach makes it possible to assess the accuracy of the scans without this bias.

Another important feature of the SUPeR study is the inclusion of a relatively large sample of 31 older adults with normal cognition recruited as age matched controls, all of whom are over the age of 60. The scans of these individuals were used to set normal thresholds with our scanning techniques that are appropriate for the study population and to investigate whether FP-CIT and cardiac MIBG uptake measurements decline in healthy ageing. It is common in nuclear medicine imaging studies to use patients with normal scan results to generate normal databases retrospectively; our approach is independent of the scan result and avoids possible bias, allowing the full range of image appearances in healthy older adults to be assessed. The control data was also used to investigate and optimise image reconstruction protocols, particularly for cardiac MIBG SPECT-CT, which is a novel technique.

The overall objective of the project is the optimisation of acquisition, processing and interpretation methods for two nuclear medicine biomarkers in order to minimise variability and maximise accuracy for clinical and research use in both established and early Lewy body disease.

Structure of thesis

Chapter 2 provides background information on diagnosing dementia and mild cognitive impairment, outlines the current status of FP-CIT and cardiac MIBG imaging for DLB diagnosis and introduces the research studies at Newcastle University used in this thesis. Chapter 3 consists of an overview of methods for ^{123}I quantification in nuclear medicine imaging and provides more detail on current quantitative methods for FP-CIT and cardiac MIBG image analysis used in the literature.

Chapters 4 and 5 consist of work done towards the beginning of the fellowship (and in some cases as pilot work for funding applications). This work explores standard FP-CIT and cardiac MIBG methods and develops the nuclear medicine imaging protocols to be used within the SUPeRB study.

Chapter 6 compares visual rating of FP-CIT scans with standard and more complex methods of uptake quantification, including absolute quantification with SUVs (standardised uptake values). The test-retest repeatability of each method is assessed using follow up scans in healthy controls.

In chapter 7, planar cardiac MIBG analysis methods are covered in more detail, using data from the MIDAS and SUPeRB studies to assess repeatability and accuracy of visual and quantitative methods, the effect of imaging at 20 minutes post MIBG administration instead of at 3.5 hours and whether normal cardiac uptake thresholds from the literature are suitable for use in a UK population with comorbidities.

Chapter 8 focuses on more advanced tomographic methods of cardiac MIBG quantification, using SPECT and SPECT-CT with corrections for attenuation and scatter. The effect of patient size on apparent uptake is assessed with phantom data and with the SUPeRB controls, for both planar and advanced image reconstruction methods. The diagnostic accuracy and the separation of normal and abnormal scans into clusters is compared for planar, SPECT-CT relative quantification and SPECT-CT absolute quantification. The uniformity pattern of MIBG uptake within the myocardium is assessed for controls and dementia patients, as a feasibility study to see if this might be useful for diagnosing Lewy body disease in patients with global uptake within the normal range.

Chapter 9 uses the scan data for the controls with normal cognition to explore ageing effects for both FP-CIT and cardiac MIBG. Multivariate analysis using linear mixed effects models is performed to determine whether there is an association between uptake and ageing in adults with normal cognition between 60 and 90 years of age.

The final chapter summarises the main findings of the thesis and evaluates whether or not it provides evidence that applying advanced image reconstruction and quantification methods to ^{123}I -FP-CIT and cardiac ^{123}I -MIBG scans gives rise to an improvement in the accuracy of the diagnosis of Lewy body disease at the mild cognitive impairment stage. The strengths and limitations of the project are discussed here as well as future directions.

Chapter 2: Clinical background

This thesis focusses on the application of nuclear medicine imaging techniques to the diagnosis of Lewy body disease at the mild cognitive impairment stage. In this chapter an overview of the neurodegenerative causes of cognitive impairment and how they are diagnosed is given, focussing on Lewy body disease and Alzheimer's disease since they are the clinical groups included in this thesis. An introduction is given to two nuclear medicine imaging biomarker tests included in the international DLB consortium criteria [1] for the diagnosis of dementia with Lewy bodies: FP-CIT and cardiac MIBG and the evidence base for each reviewed, including autopsy validation studies. An overview is then given of the three clinical studies in Lewy body disease at Newcastle University from which data has been used in this thesis: MIDAS, LewyPro and SUPeRb.

2.1 Dementia and mild cognitive impairment

2.1.1 Causes of dementia

Dementia is a clinical syndrome of cognitive impairment caused by brain diseases [4]. It presents one of the biggest health and social care challenges of the 21st century [5]. In 2014 there were estimated to be 850,000 people living with dementia in the UK [5] and over 47 million worldwide [6]. It is expected that by 2021 there will be more than a million people living with dementia in the UK [5]. Dementia has now become the biggest cause of death in the UK [7] with cases projected to double by 2050 [8].

Dementia is not a disease in itself, but rather an umbrella term for a syndrome. The symptoms associated with dementia, including problems with memory, cognition and language, can be caused by various neurodegenerative diseases of the brain [9], which include Alzheimer's disease, dementia with Lewy bodies (the clinical focus of this work), vascular dementia, fronto-temporal dementia and many others¹. The neurodegenerative diseases are all progressive, so get worse over time and eventually lead to death. Although medication can be used to alleviate some of the symptoms, there are currently no drugs available that can stop the progression of any of the neurodegenerative diseases causing dementia.

Alzheimer's disease (AD) is the most common illness causing dementia, affecting approximately two thirds of people with dementia [9]. In AD, abnormal structures made from amyloid and tau proteins build up in the brain, damaging the nerves and leading to loss of brain tissue. As the hippocampus is generally affected early on in AD, memory problems are often one of the earliest symptoms. However, the specific symptoms an individual experiences vary widely and are thought to depend on the part of the brain affected (see Diagnosing Dementia, below). After Alzheimer's disease, vascular dementia and Lewy body disease (comprising dementia with Lewy Bodies and Parkinson's disease dementia) are the most common causes of dementia. Vascular dementia (VaD) is caused by problems with the blood supply to the brain, often due to strokes or chronic small vessel disease [10]. Since any part of the brain can be damaged by a vascular event, the symptoms vary and memory loss is not necessarily an early feature in VaD [11].

Dementia with Lewy Bodies (DLB) is caused by abnormal alpha-synuclein protein structures (Lewy bodies) that damage neurons in the brain, and elsewhere in the nervous system, including the sympathetic nerves to the heart. DLB is characterised by fluctuations in cognition, motor features of Parkinsonism, REM sleep behaviour disorder and complex visual hallucinations (see section on Diagnosing Dementia below). Dementia with Lewy bodies (DLB) is the second most common form of

¹ Less commonly, dementia can be caused by reversible conditions such as space occupying lesions or can be drug-induced.

neurodegenerative dementia after Alzheimer's disease, accounting for 5-10% of clinically diagnosed cases [12, 13]. Accurate diagnosis is important for clinical management, prognosis, and carer wellbeing [14-16], but initial misdiagnosis outside the specialist setting is common [1, 13, 17]. Schneider *et al.* found that about 15% of dementia cases at autopsy have Lewy bodies in the brain [18], further evidence that many people with DLB or mixed dementia are being misdiagnosed during life.

Parkinson's disease (PD) is also caused by Lewy body pathology and PD patients often go on to develop dementia, termed Parkinson's disease dementia (PDD) [19]. These disorders are often grouped together as Lewy body dementia or Lewy body disease, and are thought of as a spectrum of essentially the same disease process. DLB and PDD are common and account for 15-20% of all neurodegenerative dementia cases [12]. DLB can be difficult to diagnose, particularly in the early stages, because the symptoms often overlap with AD. Early and accurate diagnosis is important for all people with dementia so that they can access relevant care, support and treatment services and plan for the future. Diagnosing DLB at an early stage is especially important as DLB patients can have severe, sometimes fatal, reactions to neuroleptic drugs that might be given if the disease is not recognised [20, 21], and they experience more debilitating symptoms and give rise to more carer stress than other forms of dementia [15, 22, 23].

Accurate early diagnosis of the neurodegenerative diseases leading to dementia is essential for research studies, particularly in studies looking to understand the disease pathogenesis, develop or test treatments.

2.1.2 Diagnosing dementia

Dementia is usually diagnosed following a referral to a specialist memory clinic, where interviews and tests of memory and cognitive function take place. The National Institute on Aging and the Alzheimer's Association, recommend that dementia should be diagnosed if the person's cognitive problems represent a decline from previous levels of function and cause significant interference with normal living [24]. Delirium or a major psychiatric illness must be excluded and the impairment must include two or more of the following [24]:

- Short-term memory problems
- Impaired reasoning and ability to perform complex tasks
- Impaired visuospatial ability – e.g. not recognising common objects
- Language problems – e.g. word finding difficulties
- Changes in personality or behaviour.

If the decline in ability is not judged to be a significant barrier to normal day to day function, Mild Cognitive Impairment (MCI) would be diagnosed rather than dementia (see section 2.1.3 below).

The particular disease causing the dementia can be challenging to diagnose. The symptoms of AD, VaD and DLB can overlap significantly and autopsy studies show that it is not uncommon to have more than one disease present [25]. The only definitive way to diagnose the specific disease causing dementia is by examining the brain at autopsy. However, there are different clinical criteria for the diagnosis of the major causes of dementia based on a combination of symptoms and tests that have been shown to have good accuracy. These are devised by international working parties of experts and are revised from time to time to take new knowledge into account. The most recent criteria for DLB diagnosis published in 2017 by the international DLB consortium [19] are summarised in Table 1 alongside the current National Institute on Aging and the Alzheimer’s Association criteria for AD diagnosis in Table for comparison (AD being the most likely differential diagnosis for DLB) [24]. At least one of the core features of fluctuations, visual hallucinations, parkinsonism and REM sleep behaviour disorder must be present for probable DLB to be diagnosed; at least two if no positive biomarker evidence is available. Possible DLB may be diagnosed if only one core feature is present and there is no indicative biomarker evidence or if at least one biomarker is positive in the absence of core features (see Table 1). The indicative biomarkers may therefore contribute towards both possible and probable DLB diagnoses, but cannot be used to diagnose probable DLB alone. Supportive clinical features and biomarkers include autonomic dysfunction, postural instability, apathy and normal medial temporal lobes on CT or MRI. These features are often present in DLB but are not recommended to be relied on for diagnosis due to low specificity.

Table 1: Fourth DLB consortium consensus clinical criteria for diagnosing dementia with Lewy bodies [1]

Probable DLB	Possible DLB
<p>CENTRAL FEATURE: Progressive cognitive decline that interferes with normal functioning. <i>Memory impairment may not occur at an early stage but usually occurs as the disease progresses.</i></p>	
<p>Either: Two or more core clinical features (and any number of suggestive features) Or: One core feature and one or more indicative biomarkers Any number of supportive features or biomarkers</p>	<p>Either: One core feature with no indicative biomarkers Or: One or more indicative biomarkers but no core features Any number of supportive features or biomarkers</p>
<p>CORE CLINICAL FEATURES</p> <ul style="list-style-type: none"> • Fluctuating cognition, attention and alertness • Recurrent visual hallucinations • Spontaneous features of parkinsonism • REM sleep behaviour disorder 	
<p>INDICATIVE BIOMARKERS</p> <ul style="list-style-type: none"> • Reduced dopamine transporter uptake in basal ganglia demonstrated by SPECT or PET • Abnormal (low uptake) ¹²³Iodine-MIBG myocardial scintigraphy • Polysomnographic confirmation of REM sleep behaviour disorder 	

In the AD diagnostic criteria, AD is diagnosed as the probable cause of dementia if the main symptom is related to memory, language, spatial awareness or impaired reasoning and no other causes are identified. A diagnosis of possible, rather than probable AD is made if symptoms were of sudden onset or it is not certain that they are progressive.

Table 2: NIA-AA clinical criteria for diagnosing Alzheimer's disease [24].

Probable AD	Possible AD
Meets general criteria for dementia	Atypical course: Meets criteria for probable AD except there is either sudden onset or uncertainty regarding progression
Gradual onset of symptoms	Mixed dementia: Meets all criteria for probable AD except that there is evidence of other disease such as DLB and / or cerebrovascular disease
Worsening cognition	
Most prominent symptom is one of the following: <ul style="list-style-type: none"> • Impaired learning and recall • Language difficulties • Impaired spatial cognition • Impaired reasoning and judgement 	
Does not have evidence of <ul style="list-style-type: none"> • Cerebrovascular disease • Features of other dementias including DLB • Another neurological disease • Use of medication affecting cognition 	

2.1.3 Diagnosing mild cognitive impairment

As discussed above, a diagnosis of mild cognitive impairment is made if a patient is experiencing problems with cognition in two or more domains, but these are not severe enough to impact their day to day function. MCI can be caused by the same disease processes that lead to dementia and in most cases patients will go on to develop dementia at a later stage. These cases are also known as prodromal dementias, though in the case of dementia with Lewy bodies there are other types of prodromal presentation, e.g. psychiatric onset [26]. In this thesis MCI is the only type of prodromal dementia syndrome considered. Although it often precedes a dementia diagnosis, MCI is non-specific and can be caused by other disorders such as depression, so it is not certain that someone labelled as having MCI will go on to develop dementia. Clinical trials of disease modifying drugs for dementia due to Alzheimer’s disease have suggested that to be effective any intervention would need to be delivered

at the very early stages of the disease, making prodromal dementia extremely important to diagnose and characterise [17, 27, 28]. Clinical studies on patients with mild cognitive impairment with Lewy bodies (MCI-LB) will be needed to test potential new treatments for DLB and so early accurate diagnosis will be crucial for patient selection. Imaging biomarkers will play an important role in both early diagnosis and in monitoring response to treatment [28].

Currently there are published consensus guidelines for identifying people with MCI due to Alzheimer's disease (MCI-AD) but not yet for MCI-LB - although such guidelines are expected to be published by the DLB consortium imminently. Several authors have put forward the view that further work is needed on the characterisation of the various presentations of prodromal DLB [17, 26-34], including the investigation of the utility of various biomarker tests available for DLB in the early stages. Biomarker utility in MCI-LB is the topic of the LewyPro and SUPERB studies at Newcastle University (see below), which this PhD on image analysis integrates with.

2.2 Nuclear medicine imaging techniques for Lewy body disease

^{123}I -FP-CIT (dopaminergic imaging of the striatum) and cardiac ^{123}I -MIBG imaging (sympathetic imaging of the myocardium) are both established biomarker techniques for distinguishing between DLB and AD, but neither has been thoroughly evaluated for prodromal diagnosis [35]. The evidence base for the use of these radiopharmaceuticals in dementia diagnosis was reviewed by searching on PubMed for original articles containing the following terms: ("MIBG" OR "metaiodobenzylguanidine" OR "FP-CIT" OR " ^{123}I -N-omega-fluoropropyl-2beta-carbomethoxy-3beta-(4-iodophenyl) nortropane)" OR "DaTSCAN") AND ("lewy" OR "dementia"). This returned 253 entries in March 2017 increasing to 343 in June 2019, showing the current interest in this topic. None of the articles retrieved in March 2017 focussed the use of FP-CIT or MIBG for diagnosis at the MCI stage (excluding case reports), but by June 2019 five such articles had been published [36-40], including the Newcastle report on the LewyPro study [37].

A brief review of the clinical evidence base for FP-CIT and cardiac MIBG and the autopsy validation studies carried out for each biomarker is given below. Standard image processing and analysis methods will be introduced in Chapter 3.

2.2.1 FP-CIT (DaTSCAN) imaging

^{123}I -FP-CIT imaging demonstrates the density of dopamine transporter in neurons in the basal ganglia, with reduced striatal uptake in both PD and DLB and normal uptake in AD [41]. Typical patterns of uptake can be seen in Figure 1. Images are usually displayed on a relative colour scale, with the maximum voxel shown as white and all other voxels assigned based on their percentages relative to maximum (Figure 1). This means that the background colour scaling is used to assess for global loss of

uptake in the striata, with the background appearing to increase when striatal uptake is globally reduced. Other displays may be used, and the potential advantages and disadvantages of the use of an absolute uptake scale is discussed in Chapter 6.

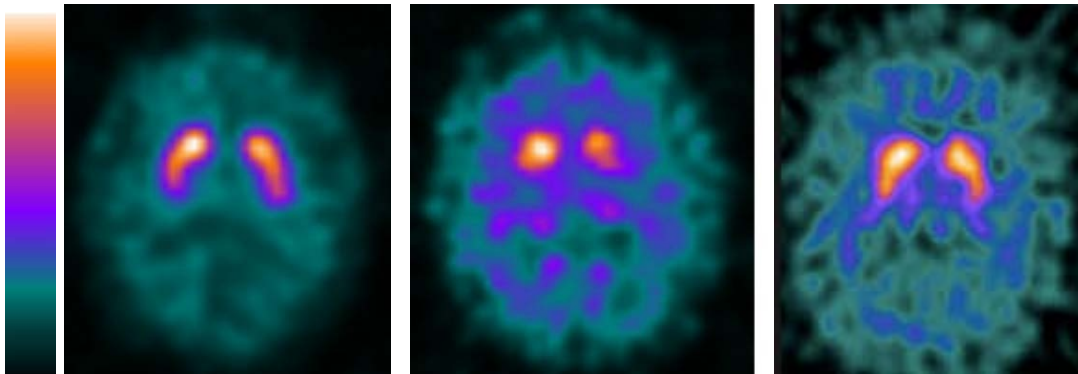


Figure 1: Left: normal FP-CIT scan showing the classic “comma” appearance. Middle: abnormal scan showing severe loss of dopaminergic function in the putamen bilaterally and more moderate loss in the caudate nuclei (the classic “dot” appearance). Right: mild reduction in global uptake affecting the caudate and putamen, identified by the raised background appearance.

FP-CIT is well established in the UK for distinguishing between probable DLB and AD [42-48] and is a biomarker in the international consensus criteria for DLB diagnosis [19]. A recent pooled analysis of four clinical trials found a sensitivity of 92% and specificity of 84%, based on visual analysis [49]. The sensitivity is likely to be much lower in MCI-LB as neurodegeneration is likely to be at an earlier stage and in some cases pathology may not (yet) be present in the dopaminergic system. Indeed, in the LewyPro study into prodromal DLB (see below) we found that 54% of MCI-LB subjects had positive FP-CIT scans [37].

2.2.2 Cardiac MIBG imaging

Cardiac ^{123}I -metaiodobenzylguanidine (MIBG) sympathetic innervation imaging is now included as an indicative biomarker in the fourth DLB consensus criteria, alongside ^{123}I -FP-CIT SPECT [1] and was subsequently recommended in the 2018 UK National Institute for Clinical Excellence (NICE) guideline NG97 for dementia diagnosis and management [50], to be considered when ^{123}I -FP-CIT SPECT is not available. FP-CIT is more widely used than MIBG in clinical settings in Europe although a reported sensitivity of 89% shows that some patients with DLB have normal FP-CIT findings [49], which has been confirmed by autopsy studies on patients scanned during life [48]. The NICE guideline NG97 advises clinicians not to rule out a diagnosis of DLB based solely on normal FP-CIT or cardiac MIBG imaging [50].

Cardiac MIBG imaging demonstrates the integrity of neurons in the sympathetic nervous system, including the myocardium. Sympathetic denervation occurs in DLB (and PD) and results in low cardiac MIBG uptake [51] as demonstrated in Figure 2.

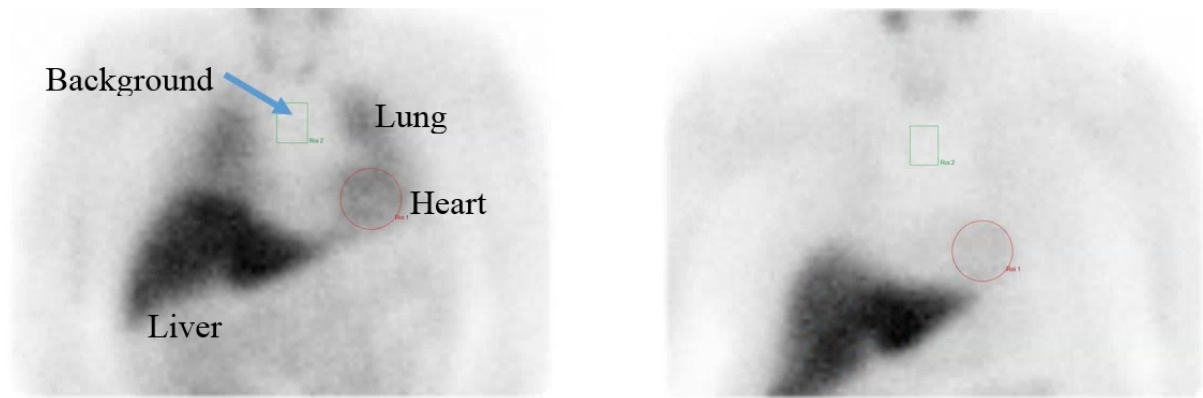


Figure 2: Left: Normal planar cardiac MIBG image. The heart is outlined with a red circle and the mediastinum background region in green. Note the overlap between heart and lung. Right: Abnormal cardiac MIBG scan, with reduced uptake seen in the heart.

Cardiac MIBG is widely used in Japan in the diagnosis of dementia with Lewy bodies, and to some extent in other European countries [51-57], but is not routinely established in the UK. The majority of published studies on cardiac MIBG for distinguishing between AD and DLB were single centre studies carried out in Japan, and showed very high sensitivities and specificities of over 90% [54, 55, 58-60]. However most excluded cardiac conditions, which can also cause reduced uptake and therefore affect the specificity. A meta-analysis of these studies determined a pooled sensitivity in detecting DLB of 98% and specificity of 94% [61]. 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 [53]. This used ROC (receiver operating curve) 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% [62], although the optimal cardiac uptake cut-off between DLB and non-DLB was updated based on re-analysis of this subgroup. Such findings led to the inclusion of MIBG imaging in the 2017 consensus criteria for DLB as a biomarker alongside FP-CIT imaging [1].

Cardiac MIBG scintigraphy is also used for the estimation of risk of sudden cardiac death in heart failure patients and predicting those who could benefit from an implantable cardiac defibrillator (ICD) [63, 64] and this is currently the principal application in Europe and the US [65]. Cardiac MIBG may also be used to distinguish between Parkinson's disease (PD) and other parkinsonian syndromes such as multi-system atrophy and progressive supranuclear palsy, in which the sympathetic nervous system is not affected [66].

2.2.3 Comparison between FP-CIT and cardiac MIBG

There have been few published studies comparing FP-CIT and cardiac MIBG directly, even in established DLB. Tiraboschi *et al.* [67] used visual image assessment in their study comparing cardiac

MIBG and FP-CIT for DLB diagnosis, with HMR analysis as a secondary method. This was the first prospective study comparing the two biomarkers in DLB. Their cardiac MIBG results in 30 DLB patients and 29 non-DLB show 100% specificity for both visual and HMR techniques with 97% sensitivity. ¹²³I-FP-CIT had a sensitivity of 90 and specificity of just 76%, lower than seen in the literature [67]. However, prior to this, a smaller Italian study in 31 patients found the same high sensitivity (90%) and specificity (91%) for both techniques [68]. Our study at Newcastle University (see section 2.3.1 below), published during this doctoral fellowship, reported a sensitivity and specificity of 71% and 81% for cardiac MIBG and 88% and 75% for FP-CIT [69] – our results were therefore similar to Tiraboschi *et al.* for FP-CIT but much lower for cardiac MIBG. The first Japanese study to compare both scans in 76 DLB and 57 AD patients found the sensitivity and specificity of differentiating DLB from AD were 72% and 94% for cardiac MIBG and 88% and 89% for FP-CIT, with slightly better diagnostic accuracy for FP-CIT [70].

Although the above points towards FP-CIT being slightly more sensitive in established DLB, there is some limited evidence from pathology studies that in MCI-LB cardiac MIBG may be a more sensitive biomarker than FP-CIT. Pathology studies in Lewy body disease have shown degeneration of cardiac sympathetic neurons prior to dopaminergic neurons in the basal ganglia [71, 72]. Furthermore, previous longitudinal studies have shown that some subjects with normal appearances on FP-CIT imaging were diagnosed as having probable DLB after follow up [73-75], and that follow up FP-CIT imaging in these patients with normal scans may become abnormal over time [76]. However, a recent review article on imaging in the diagnosis of prodromal dementia with Lewy bodies [77] found that dopaminergic deficits have a high specificity for Lewy body disease at the pre-dementia stage, although longitudinal follow up is lacking, whereas for cardiac sympathetic denervation it remains to be determined whether imaging abnormalities precede DLB [77].

2.2.4 Autopsy validation of FP-CIT and cardiac MIBG

A recent systematic review by Sonni *et al.* [35] looked at the evidence base for both FP-CIT and MIBG for the differential diagnosis between AD and DLB. They found that for both biomarkers, whilst there is good evidence of ability to distinguish between AD and DLB, autopsy validation is limited and no studies have yet performed autopsy validation of both biomarkers in the same group of patients [35]. Confirmation of disease subtype at autopsy is important to support the validity of these scans and is limited in both cases, but stronger in FP-CIT than in cardiac MIBG.

Walker *et al.* [78] showed that in eight patients with DLB and 12 with other dementias the sensitivity of FP-CIT with respect to neuropathological confirmation was 88% and specificity 100%, with abnormal scans defined as those with quantification results more than two standard deviations below the mean of controls. Colloby *et al.* [79] also correlated quantitative FP-CIT uptake with neuropathology and

found that image uptake was associated with dopaminergic neuronal density in the substantia nigra, but not the presence of alpha-synuclein.

Two studies have validated visual rating of FP-CIT scans, common in clinical practice, with autopsy findings. Jung et al. [80] performed autopsy validation of FP-CIT in 18 patients with dementia, including 12 who had DLB. The sensitivity and specificity of abnormal FP-CIT SPECT by visual rating for predicting Lewy body disease were 92% and 83% respectively. The largest autopsy validation study to date, by Thomas et al. [48], included 33 patients with DLB and 22 with Alzheimer disease who underwent FP-CIT scanning as part of a research study and donated brain tissue on death. FP-CIT was found to have a sensitivity of 80% and specificity of 92% with respect to autopsy findings. There were three DLB cases (10%) who met pathologic criteria for Lewy body disease but had normal FP-CIT imaging.

There have been several autopsy studies from Orimo and colleagues reporting on the degeneration of the cardiac sympathetic nervous system in people who had DLB, as well as in asymptomatic individuals [2, 3, 71]. Beach et al. showed that alpha-synuclein pathology is found throughout the sympathetic nervous system at autopsy in patients with Lewy body disease and is predominant in the sympathetic ganglia [81], which suggests that cardiac sympathetic innervation imaging should be an effective diagnostic tool.

There is only one autopsy study on patients who had cardiac MIBG scanning in life, including 15 DLB patients, published by the same group as the Japanese studies above [82]. This is the only direct evidence to link reduced cardiac MIBG uptake and loss of sympathetic cardiac neurons [82]. Of the 15 patients confirmed to meet the pathological criteria for DLB, cardiac MIBG uptake was found to be decreased in 13 cases (sensitivity 87%), with no decrease of MIBG uptake observed in two of the DLB cases with confirmed pathology in the brain. Of these two cases, one had moderate cardiac sympathetic denervation at autopsy, two years following the scan, and one had only mild denervation.

Both the cardiac MIBG autopsy study of Takahashi et al. and the FP-CIT studies of Thomas et al. and Jung et al. show that a normal scan result does not rule out a diagnosis of DLB, suggesting it is possible to have significant Lewy body pathology without dopaminergic or sympathetic denervation.

2.3 Radionuclide imaging studies in DLB and MCI-LB at Newcastle University

There are three studies completed or currently recruiting subjects for cardiac MIBG and / or FP-CIT studies in Newcastle in order to assess the utility of imaging as a diagnostic tool: MIDAS, LewyPro and SUPeRB. All subjects in these studies agreed to their data being used for subsequent studies within Newcastle University when their consent to take part was first obtained. The images and the clinical diagnoses from these studies were therefore available for use in this doctoral fellowship. The imaging

protocols used in the studies are described in more detail in the main body of thesis. The information below is included to give a broader clinical overview of the studies.

2.3.1 MIDAS study

The MIDAS study (^{123}I -MIBG in dementia with Lewy bodies as a marker for sympathetic denervation) is a single centre pilot study at Newcastle University, with the aim of determining whether cardiac MIBG is a useful tool in typical subjects presenting for dementia assessment in the NHS [83]. MIDAS also aimed to assess more difficult cases of possible DLB and to compare MIBG with FP-CIT. In most published studies patients with even mild cardiac disease were excluded because heart failure and cardiovascular disease can cause abnormal cardiac MIBG uptake and give rise to confounding results. This means that the positive results seen in the literature may not translate to a typical elderly UK population, where the prevalence of cardiovascular disease and heart failure is high. In this study inclusion criteria have been deliberately kept as broad as possible and patients with mild cardiovascular disease are not excluded.

The MIDAS study aimed to recruit and scan a total of 90 subjects with dementia – 30 with probable DLB according to the consensus criteria [19], 30 with probable or possible AD according to the revised NIA-AA criteria [24] and a further 15 subjects with possible DLB and 15 with probable DLB who have negative FP-CIT results. Seventeen DLB patients completed the protocol, 16 AD patients and 6 additional FP-CIT negative DLB. All subjects underwent MIBG and FP-CIT imaging at baseline (if not already done), as well as clinical and neuropsychometric assessment. The cardiac MIBG scans were carried out with SPECT, but the SPECT-CT scanner was not yet installed when the study started, so CT data for attenuation correction is not available.

2.3.2 LewyPro study

LewyPro was an NIHR Biomedical Research Unit study at Newcastle and was the first prospective assessment of subjects with MCI-LB [37, 84]. This study was a predecessor to SUPeRB (see below) and carried out by the same principal investigator. The methodology was similar to SUPeRB, but with some features, such as cardiac MIBG scanning, not included. LewyPro closed in 2016, but subjects continue to be followed up and many have agreed to enter the SUPeRB study.

The team recruited patients presenting for assessment at NHS memory clinics with mild cognitive impairment, rather than with dementia. The clinical assessment and cognitive tests were used to judge whether the subjects had probable or possible MCI with Lewy bodies (MCI-LB), or probable MCI due to AD (MCI-AD). Of the 90 patients recruited, 75 completed the assessments and scans. Of these 33 were judged to have probable MCI due to Lewy body disease, 15 possible MCI due to Lewy body disease and 27 had probable MCI due to AD. We found that 54% of patients with MCI-LB had an

abnormal FP-CIT scan, based on visual assessment of the scans by an experienced panel reporters, a significantly lower sensitivity than seen in dementia cohorts. The specificity was 89%, which is comparable to results in dementia (see section 2.2 above).

2.3.3 SUPeR study

SUPeR (Scintigraphy Utility in Prodromal dEmentia with Lewy Bodies) is an extension of the LewyPro study with a larger cohort of 100 subjects with MCI and 30 age matched controls, which is ongoing at the time of writing. It is a single centre prospective cohort observational study of people over 60 years of age recruited from memory clinics with symptoms not sufficient for a dementia diagnosis, or recruited as volunteers with normal cognition. MCI subjects have baseline MIBG imaging, FP-CIT and high resolution hippocampal MRI, an in-depth clinical and neuropsychometric assessment and then annual review for up to four years. At annual review the clinical assessments are repeated and the diagnoses reviewed. The control volunteers undertake the same detailed assessments as patients presenting with MCI, so they are thoroughly evaluated as having normal cognition.

A key feature of the SUPeR study for this thesis is that the imaging protocols include SPECT-CT for both FP-CIT and cardiac MIBG, which was not the case for MIDAS and LewyPro. This enables the full benefits of quantitative imaging with CT based scatter and attenuation correction to be studied in this patient and control cohort and the results compared the standard methods that do not use SPECT-CT. After consenting to participate, subjects have a baseline assessment, consisting of five sessions in which they complete the tests outlined below:

- 1) Consent and neurocognitive assessment consisting of the following items
 - a. Addenbrooke's Cognitive Examination
 - b. Standard MMSE
 - c. Graded Naming Test
 - d. Rey AVLT
 - e. Trail Making Test
 - f. FAS
 - g. Neuropsychiatric Inventory (NPI)
 - h. Epworth sleepiness scale
 - i. Visual Hallucinations Scale (VHI)
 - j. Geriatric Depression Scale
 - k. Questionnaire for Symptoms Suggestive of Lewy Body Disease
 - l. National Adult Reading Test
 - m. Digit Symbol Substitution tasks

- 2) A diagnostic interview, medical examination and neuropsychiatric assessment including medication review, an assessment of olfactory function, height, weight, pulse, lying and standing blood pressure, blood sample. Measurement of supine and standing heart rate variability (HRV) and ECG for MCI patients only.
- 3) Cardiac MIBG scan
- 4) FP-CIT scan
- 5) 3T MRI scan and EEG; optional lumbar puncture.

MCI diagnoses in SUPeRB

A diagnosis of either possible or probable MCI-LB or probable MCI-AD will be given following review of the assessments by a panel of experienced old-age psychiatrists, following the international DLB consortium criteria given in Table 1 above, except that polysomnography is not included as a biomarker in the study. To summarise, patients with two or more core features will be classed as probable MCI-LB, even if their FP-CIT and cardiac MIBG biomarker tests are normal. Patients with one core feature will be classed as possible MCI-LB if their scans are normal and probable MCI-LB if one or both scans are abnormal. Patients with no core clinical features are classified as probable MCI-LB if one or both scans are abnormal and MCI-AD if both scans are normal. In the SUPeRB study an abnormal FP-CIT scan is determined by a consensus panel of five raters and an abnormal cardiac MIBG scan is defined as planar HMR result lower than two standard deviations below the mean of the controls. A cut-off at two standard deviations below the mean of normal subjects is a standard method throughout radionuclide imaging, used by the manufacturers of FP-CIT software for example and in the autopsy studies above [78, 79].

In this thesis we are assessing the imaging biomarkers, which makes using the scan results for diagnosis problematic. One approach would be to use the core clinical features only, resulting in the same clinical groupings throughout the thesis. However, we chose to include either FP-CIT or cardiac MIBG results for diagnosis, depending on which scan was under evaluation, with the aim of increasing diagnostic certainty. So for the chapters assessing FP-CIT methods using the SUPeRB MCI patients we use the cardiac MIBG result as part of the diagnostic criteria and for the chapters assessing MIBG methods we use the FP-CIT result. Possible MCI-LB subjects are excluded in both scenarios to increase diagnostic certainty. The clinical symptoms at one year follow up assessment are used for diagnosis of the SUPeRB subjects in this thesis, resulting in 59 MCI patients and 31 controls with follow up completed in January 2019 being considered for inclusion. Flow charts depicting the decision making processes leading to the clinical groupings for SUPeRB MCI patients are shown in Figure 3 for FP-CIT assessment (i.e. using cardiac MIBG results) and in Figure 4 for cardiac MIBG assessment. The resultant groups consist of 27 probable MCI-LB and 18 MCI-AD patients to use in the FP-CIT chapters and 28 probable MCI-LB and 18 MCI-AD patients to use in the MIBG chapters.

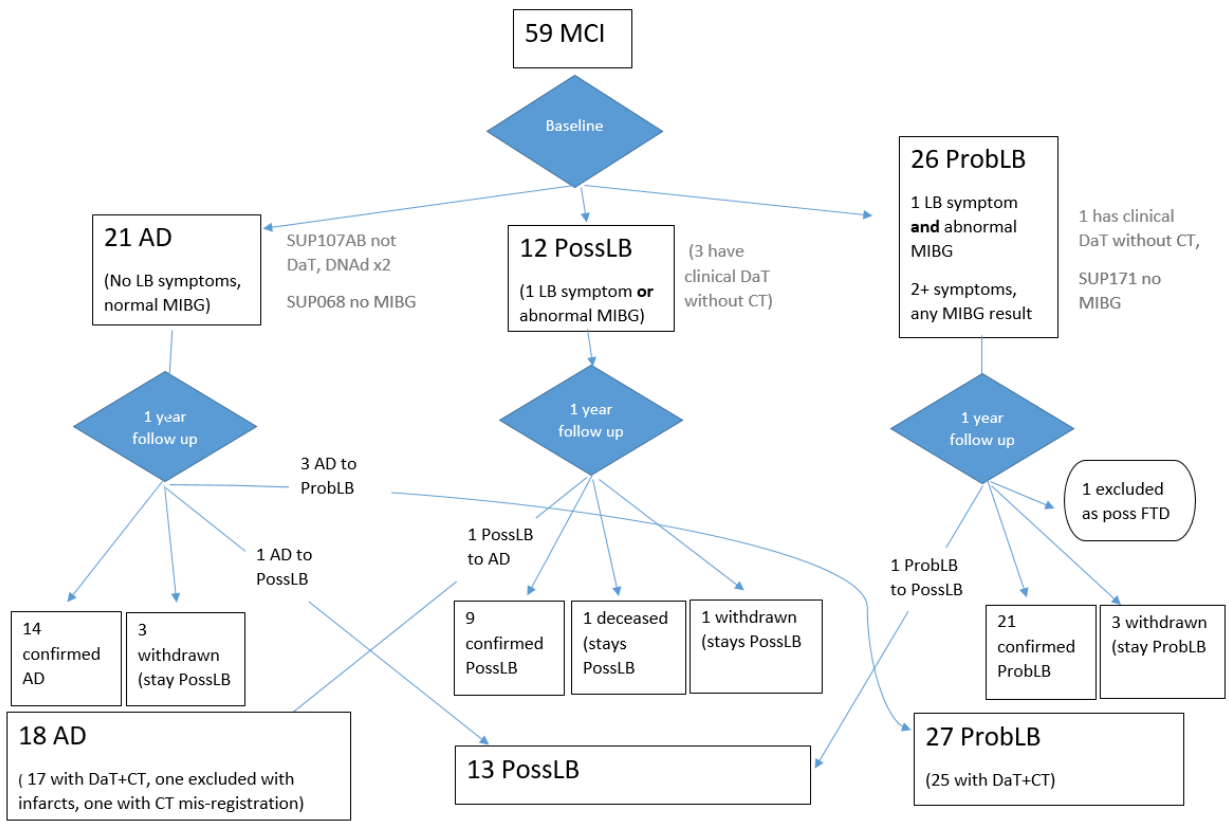


Figure 3: Decision flow chart for clinical groupings of SUPERB MCI subjects for the FP-CIT chapters in this thesis, based on core clinical features at 1 year follow up and MIBG result at baseline.

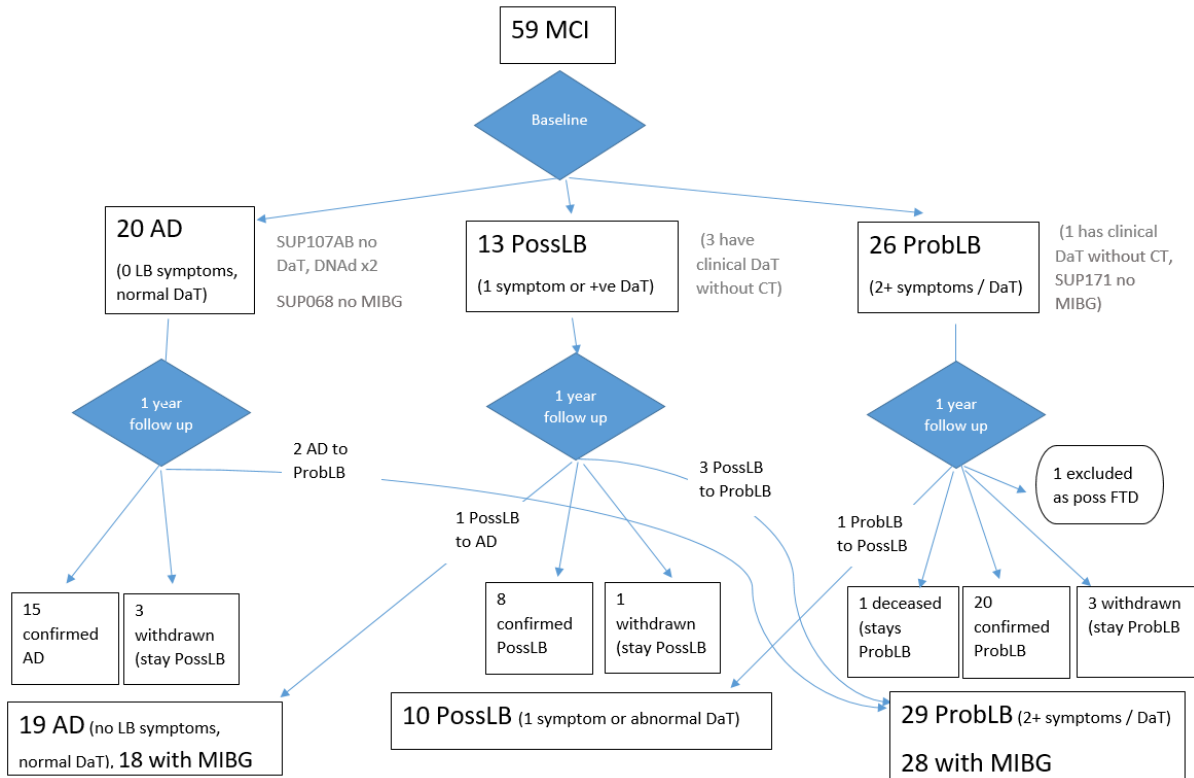


Figure 4: Decision flow chart for clinical groupings of SUPERB MCI subjects for the cardiac MIBG chapters in this thesis, based on core clinical features at 1 year follow up and FP-CIT result at baseline. Two of the cases did not have MIBG imaging so are of course excluded.

Chapter 3: Technical background

Recent developments in nuclear medicine quantification have the potential to improve the diagnostic accuracy of FP-CIT and MIBG for dementia diagnosis, maximising the value of the acquired scan data. More accurate quantification has the potential to improve diagnosis for DLB, both in routine practice and in clinical research - particularly in early disease. This thesis implements and assesses such improved quantification methods, which include corrections for signal lost due to scatter and attenuation and for the effect of high energy gamma photons that penetrate through the collimator. In this chapter the fundamental underlying issues related to ^{123}I image quantification are discussed, and the specific issues in FP-CIT and cardiac MIBG quantification addressed in this thesis are introduced.

3.1 Overview

Nuclear medicine images are considerably degraded by photon scatter and attenuation and so do not inherently provide an accurate measure of tracer uptake within a subject. For example, signal from the centre of the brain is subject to 70% more attenuation than the cortex and corrections for attenuation and scatter have a significant effect on the appearance of ^{99m}Tc perfusion brain scans [85]. Conventional nuclear medicine imaging relies on visual detection of characteristic patterns of uptake and so does not require quantification. However, for nuclear medicine scans to truly realise their potential as imaging biomarkers, quantitative techniques are required and corrections for image degradation are essential for these to be as accurate as possible.

Lack of the ability to measure the absolute uptake of a radiotracer has meant that traditionally ratios between the counts in the organ of interest and counts in a background region are used for semi-quantification [86]. This is the case for standard analysis of both FP-CIT and cardiac MIBG scans, where the striatum and myocardium respectively are compared to a background region. Absolute measurement of activity concentration has in recent years become possible due to advances in attenuation and scatter correction and SPECT reconstruction methods, which all contribute to providing a more accurate representation of the activity distribution within a subject, if careful calibration methods are used [87, 88]. A particular enabler for these advances has been the ability to measure attenuation and scattering properties with a simultaneously acquired CT scan (SPECT-CT scanning) rather than relying on unrealistic assumptions about these properties [89].

^{123}I radionuclides emit a principle gamma photon at 159 keV, of a similar order of magnitude as the ^{99m}Tc photopeak at 141 keV, but with additional high energy emissions between 440 and 784 keV (3% of total emissions) that make accurate quantification even more challenging. The advances discussed in this chapter aim to improve the accuracy of semi-quantification of ^{123}I images and also enable true absolute uptake quantification after normalising for patient size and injected activity. The improvements in semi-quantification and the potential benefits of absolute quantification for cardiac MIBG and FP-CIT will be evaluated in this thesis. In 2018 a quantitative SPECT meeting was held by the Institute of Physics and Engineering in Medicine, at which the author presented some of the work discussed in this thesis. The overall feeling of the delegates was quantitative SPECT imaging is ready to be implemented for widespread clinical use, as a subsequent review article by the organisers demonstrates [88].

3.2 Radionuclide imaging (nuclear medicine)

3.2.1 Introduction

Nuclear medicine is a functional imaging technique using radioactive tracers to obtain information about physiological function. The aim is to visualise or quantify the distribution of uptake of a radiopharmaceutical within an organ of interest to aid in the diagnosis and staging of disease or response to treatment. Typically, radiopharmaceuticals are labelled with technetium-99m (^{99m}Tc), which emits a gamma photon of 141 keV energy, with a half-life of 6 hours. However, both the techniques used in this thesis, FP-CIT and cardiac MIBG are labelled with iodine-123 (^{123}I), which has a principal gamma emission at 159 keV and half-life of 13.2 hours.

The radiopharmaceutical is administered via intravenous injection and allowed to distribute within the body and be taken up by organs of interest, before being imaged using a gamma camera (Figure 5).

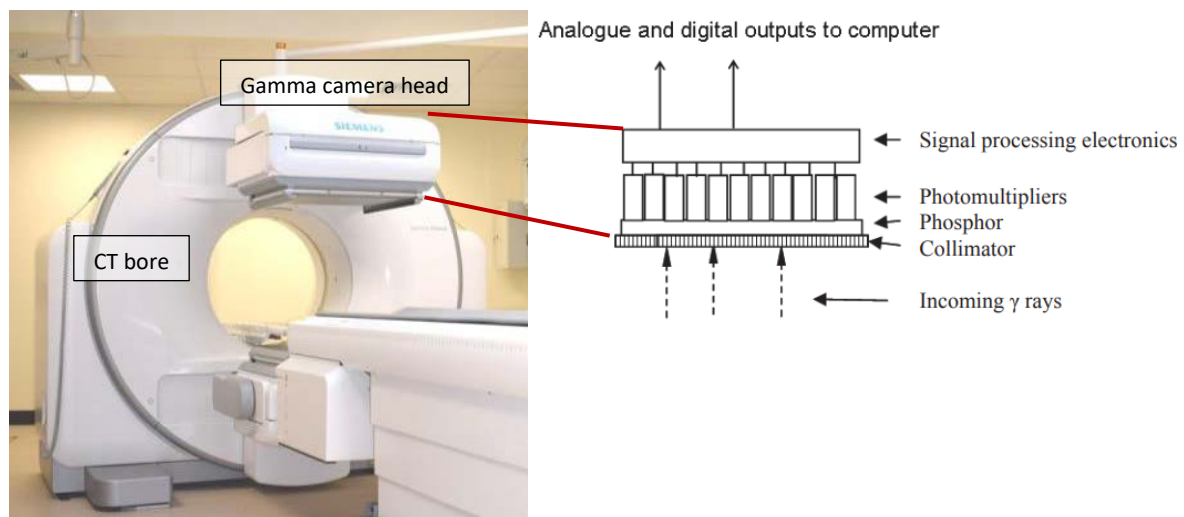


Figure 5: Left: Siemens Intevo SPECT-CT gamma camera at the Royal Victoria Infirmary. Right: Schematic diagram of a gamma camera detector head [90]

The main components of the gamma camera design used in this thesis are the parallel hole collimator, sodium iodide (NaI) scintillation crystal, photomultiplier tubes and processing circuits (Figure 6). The collimator is a lead grid of parallel holes designed to allow only photons incident within a small angle of acceptance to pass through to the scintillation crystal². Gamma photons incident from oblique angles (the vast majority of the total emissions) are absorbed by the collimator septa. Without the collimator, a meaningful image could not be formed as the path of the photons would be unknown.

² Other collimator designs, such as high sensitivity pinhole collimators are available, as are novel designs of gamma camera that do not use scintillation detectors. These are not used in our centre and consideration of them is beyond the scope of this work.

Gamma photons that reach the scintillation crystal interact via the photoelectric effect and emit thousands of visible photons, which are then detected as photoelectrons and amplified by an array of photomultiplier tubes. The relative signal detected by each photomultiplier tube is used to determine the likely position of the interaction within the crystal, and the overall magnitude of the signal determines the energy of the gamma photon that caused the interaction. An image is formed from many thousands of detected gamma photons, or “counts”, each assigned to a pixel within an image matrix. This process is subject to stochastic uncertainty, which gives rise to uncertainties in both spatial localisation and energy resolution.

Due to absorption and scatter within the patient, many photons originally emitted in a perpendicular direction to the gamma camera fail to reach the detector (Figure 6), resulting in a loss of signal. Attenuation causes structures located deeper within the body to appear less active than those nearer the surface, with the signal reduction following an exponential relationship within a uniform attenuating medium.

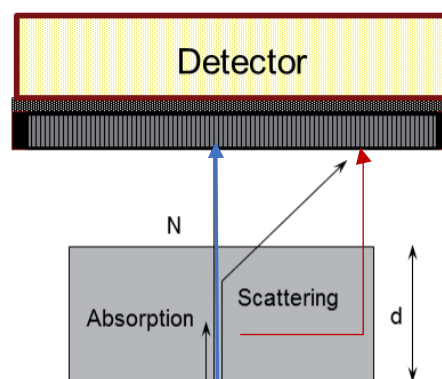


Figure 6: Illustration of possible photon trajectories, adapted from [91]. The blue arrow depicts the ideal scenario. The black arrows represent attenuation due to absorption and scattering. The red arrow depicts a scattered photon originating from the centre of the source that will be detected as if it came from the edge.

It is also possible for photons not originally emitted perpendicularly to the collimator to undergo Compton scattering within the patient, resulting in the scattered photon reaching the crystal and being detected in the wrong location (Figure 6). Photons lose energy via Compton scattering, so the amount of scatter detected is minimised by only accepting photons within a certain energy window around the photopeak, typically $\pm 10\%$. The photopeak energy is broadened due to statistical uncertainty, with most 159 keV ^{123}I photons distributed between 140 and 170 keV (see ^{123}I spectra in Figure 13, section 5.3.4). The width of the energy window is set to obtain a balance between accepting as many primary photopeak photons as possible (to maximise sensitivity) and rejecting as many scattered photons as possible. However, even photons that have scattered by 90 degrees, as shown in Figure 6, may still carry enough energy to be detected within a $\pm 10\%$ photopeak window [92], which means that the

detection of a significant proportion of scattered photons is unavoidable. Scatter causes an increase in incorrectly positioned background counts, reducing image contrast.

The probability of these processes occurring is patient-specific, as it depends on the composition and amount of tissue between the source and the detector. Accurate compensation for attenuation and scatter improves image quality and is essential for quantification of tracer distribution [92, 93]. A recent review by Ljungberg concludes that the most important factors that affect SPECT quantification and require compensation for accurate results are attenuation, scatter and poor collimator resolution [94]. Common methods for performing these compensations within tomographic SPECT imaging are described below.

3.2.2 SPECT imaging and reconstruction

Images obtained with the gamma camera detector in a static position over the patient are called planar images. Planar images are a simple method for visualising the three dimensional activity distribution in the patient as two dimensional projections. However, they are inherently subject to attenuation and scatter, with deeper structures more affected, and this cannot be accurately corrected for. Nevertheless they are often used in nuclear medicine, including for cardiac MIBG imaging.

The summation of activity from overlying structures into a two dimensional projection results in a significant loss of contrast, making planar imaging particularly unsuitable for brain imaging. This is resolved by looking at the activity distribution within a slice through the brain using single photon emission computed tomography (SPECT). SPECT allows more accurate visualisation and quantification of tracer uptake and is used for both brain and torso imaging. SPECT images are reconstructed from a set of multiple planar images taken at different angles as the detector rotates around the patient, typically 120 or 128 views in total for brain imaging. These raw planar images are called projections and are used to estimate the activity at each point within the field of view. The simplest method to do this is via back projection of the signal from each projection, which is illustrated in Figure 7 for a simplified scenario of just four orthogonal projections.

In simple back projection, the raw signal is projected back over the whole field of view, which leads to blurring of structures and star artefacts from the overlapping projections (Figure 7). To correct for this blurring, a ramp filter is applied in frequency space, to suppress low spatial frequencies (which correspond to large structures). High spatial frequencies (corresponding to small structures) are also suppressed to reduce image noise, typically with an additional Butterworth filter. The resultant technique is called filtered back projection and has been widely used in nuclear medicine for decades.

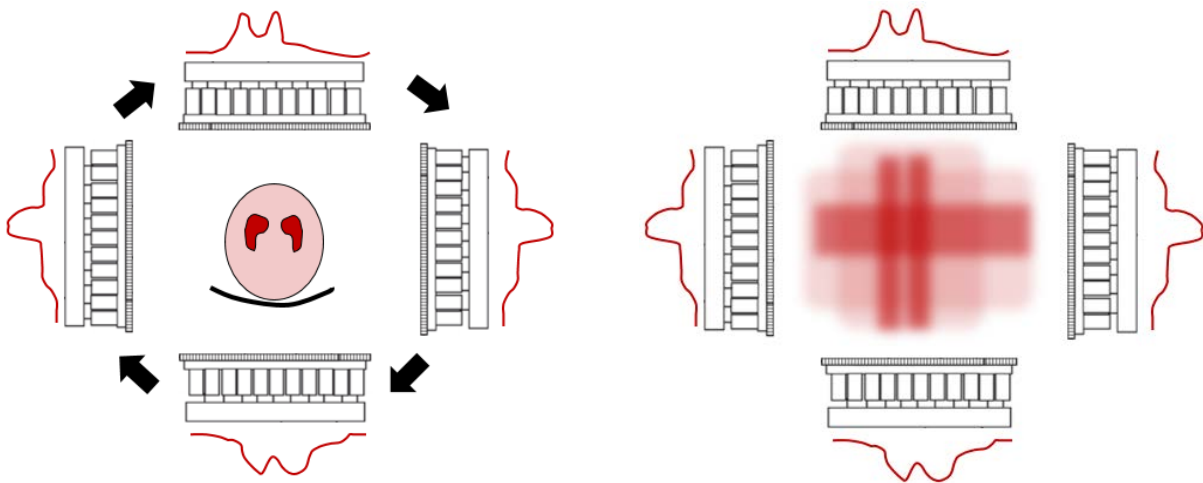


Figure 7: Left: Illustration of gamma camera detector rotating around patient, acquiring projections. Right: Simple back projection of acquired projections to form image estimate. The radial blurring is corrected by a filter in FBP, which reduces the “star artefact”. This simplified example shows four projections – typically 64 to 128 are used to build up a much more accurate image estimate.

An alternative method is iterative reconstruction, where an initial image estimate is refined and updated via a series of comparisons between the raw projections and the estimated projections that would give rise to the estimated image as shown in Figure 8. The forward projection step (shown outlined in red in Figure 8) is a model of the acquisition process, and can incorporate effects such as attenuation and scatter and poor collimator resolution [91] (see section 5.3). The better the forward projection step model, the closer the resultant image estimation is to the true activity distribution. However, artefacts can be introduced if the assumptions used are incorrect. The ability to compensate for scatter and attenuation makes the iterative method much more suitable for quantification than filtered back projection.

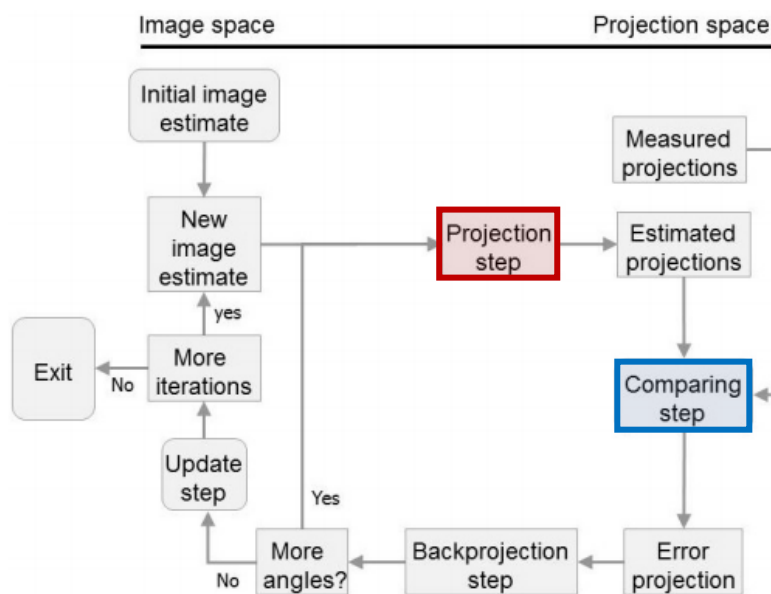


Figure 8: Overview of iterative reconstruction process, adapted from [91]

There are different algorithmic approaches to implementing the iterative method illustrated in Figure 8, all of which have different ways of implementing the key step (shown in blue) of comparing estimated and measured projections in order to converge on a solution that represents the true activity distribution. In the 1990s Gilland et al. used phantom studies to show that iterative reconstruction using the maximum likelihood expectation maximisation algorithm (MLEM) gave quantitative accuracy comparable to FBP for ^{123}I imaging with the benefit of less noisy images [95]. However, MLEM is computationally demanding and long reconstruction times prevented adoption for clinical use. A variation of the MLEM algorithm, ordered subsets expectation maximisation (OSEM) was developed by Hudson et al. to speed up the iterative process by including only a subset of the projections in each iteration [96], resulting in more rapid convergence to the final estimate. The combination of OSEM and modern processing speeds means iterative reconstruction is now commonly used in routine nuclear medicine imaging, including for FP-CIT SPECT [97].

However, a disadvantage of OSEM is that it is not clear when adequate convergence is achieved, and the algorithm may continue updating the estimate with only small improvements in activity distribution at the cost of high frequency noise being introduced [90]. In clinical practice it is not feasible to optimise individual images so a fixed number of iterations is generally used to avoid amplification of noise, with the number of iterations and subsets tailored to each type of scan. However, convergence rates are likely to depend on the individual patient activity distribution, with Dickson et al. showing that for ^{123}I striatal phantom scans different numbers of iterations were required for different uptake scenarios, and high striatal uptake images did not converge [98]. Nevertheless, the authors concluded that OSEM was more suitable for FP-CIT quantification than FBP because corrections, such as attenuation and scatter correction, can be incorporated.

Filtered back projection remains widely used in clinical practice and the adoption of iterative methods is still far from universal. The key advantage of iterative methods is that they support the incorporation of attenuation and scatter correction and collimator resolution modelling. The benefit of OSEM without image corrections is unproven and implementation in centres without access to advanced image correction methods could be unhelpful. Changing the reconstruction method from FBP to OSEM is not straightforward as it requires medical physics support for optimisation and a period of retraining for reporters, since the visual appearance of the images is affected. In this thesis the OSEM method with corrections for resolution, scatter and attenuation for visual rating of FP-CIT images is compared with the Newcastle upon Tyne Hospitals (NuTH) standard method of filtered back projection. OSEM is then used for FP-CIT throughout the thesis, rather than the local departmental FBP method. For cardiac MIBG SPECT, OSEM with corrections for resolution, scatter and attenuation is used throughout thesis, except in section 8.4 on regional cardiac innervation, where FBP is used instead due to concerns about possible artefacts introduced by the resolution recovery model.

3.3 Factors affecting quantification of ^{123}I uptake in radionuclide imaging

3.3.1 Attenuation

The current “gold standard” method for attenuation correction makes use of transmission scans to provide an attenuation map, using a CT scan acquired immediately after SPECT on a hybrid SPECT-CT scanner [89]. This method has been available since SPECT-CT scanners were introduced in 1999 [99]. The SPECT scan is performed first and the patient couch then moves through to the CT gantry, located behind the gamma camera head (see photograph of SPECT-CT scanner in Figure 5). The CT scan provides patient specific attenuation maps for attenuation correction via the iterative reconstruction algorithm and also an anatomical image that can be used for localisation. As the gamma photopeak varies depending on the radionuclide used, and generally the CT energies are lower, the measured CT transmission is not a true reflection of the attenuation properties at the gamma photon energy of interest. A look up table based on empirical data is used to modify the attenuation coefficients [99, 100].

An alternative method that does not require a SPECT-CT scanner is the uniform, or “Chang” method, where a patient contour is automatically defined on an initial estimate of the reconstructed image, and all the tissue inside the contour is assumed to be the density of water for the purposes of the attenuation map [101]. Uniform attenuation correction is often used in brain imaging and has been shown to be of similar accuracy to CT methods for both $^{99\text{m}}\text{Tc}$ perfusion scans [85] and for ^{123}I -FP-CIT dopaminergic imaging [102]. The assumption of uniform density is adequate for brain tissue and the issue of potential mis-registration between SPECT and CT due to patient motion is avoided. The uniform method also requires no additional radiation dose to the patient.

However, SPECT-CT is required for the attenuation correction of torso SPECT images because the uniform density assumption is inappropriate [100]. The CT image provides a detailed map of attenuation coefficients throughout the patient, which is used for attenuation correction and can also be used for Monte Carlo simulation of scatter within the imaged volume.

3.3.2 Scatter

In this context, scatter refers to photons that are detected within the photopeak but attributed to the wrong position, as depicted by the red arrow in Figure 6. Photons that scatter out of the field of view or are detected outside the energy window contribute to attenuation rather than detected scatter. Photons can undergo a single Compton scattering event of up to 90 degrees and still be detected in the photopeak energy window [92], demonstrating the importance of scatter correction. A simple method for correcting for scatter within the photopeak, the triple energy window method (TEW), involves setting narrow energy windows just either side of the photopeak and using the signal in these windows to estimate the amount of signal due to scatter within the photopeak [91]. The scatter

fraction is estimated by interpolating between the signal in the upper and lower scatter windows as illustrated in Figure 9. The counts attributed to scatter within the photopeak are then subtracted from the total.

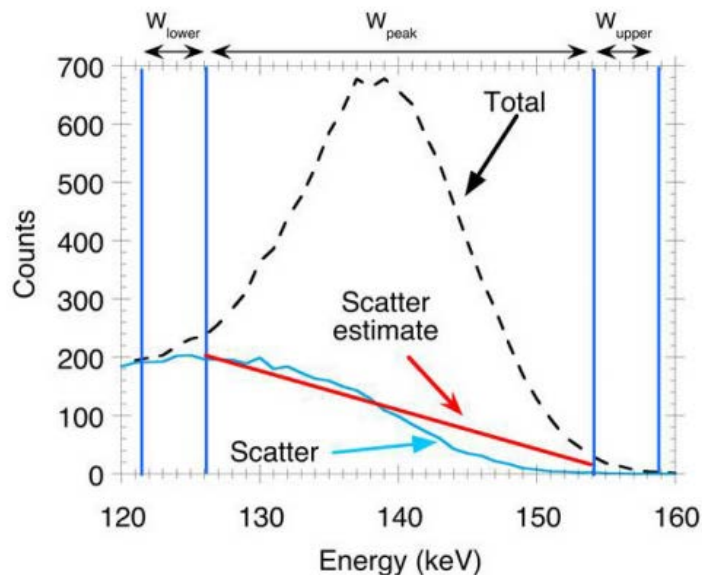


Figure 9: Illustration of the triple energy window scatter estimation technique [91]

A fundamental problem with the TEW method is that the spatial distribution of scatter is energy dependent, with the photopeak window more likely to contain photons deflected by smaller angles than the lower energy window [92]. The windows are kept narrow to minimise this effect. A further disadvantage of scatter window subtraction methods is that the projection data are noisy due to the necessity for narrow scatter windows, and subtraction of a noisy estimate may result in negative pixels within the projections [92]. Papanastasiou et al. state that the use of TEW scatter correction for cardiac ^{123}I imaging provides a semi-quantitative assessment, based on results in a phantom [103]. Lagerburg et al. investigated scatter window subtraction methods using an image quality phantom with different ^{123}I source to background ratios [104] and found that scatter correction was effective for low energy collimators but made quantification less accurate with medium energy collimators. However, Inoue et al. found that the triple energy window could not correct for ^{123}I scatter in patients using low energy high resolution collimators [105]. Collimator choice for ^{123}I imaging in relation to scatter and septal penetration is discussed in section 5.3.4.

An alternative approach to scatter window estimations is to simulate the scatter within the patient using Monte Carlo methods. Full estimation of scatter for an individual patient was traditionally considered too computationally demanding to be practical [92]. However, recent work in optimising

Monte Carlo methods has demonstrated that it is feasible to compute scatter estimates in sufficient time to be practical for inclusion in image reconstruction [93, 106, 107]. Shcherbinin evaluated one such algorithm showing accuracy to within 5% in a phantom for ^{123}I using SPECT-CT [108]. Robinson et al. have shown that the triple energy window is less accurate than Monte Carlo methods of scatter correction [109]. Scatter window techniques, such as TEW, will not be investigated further in this thesis as the literature shows both theoretical and empirical evidence that they are sub-optimal. Monte Carlo methods are not widely adopted in clinical practice, but show promise and are supported by evidence from the literature. Monte Carlo methods are supported within the reconstruction software used in NuTH (see below) and are used throughout this thesis.

3.3.3 Resolution and partial volume effect

The resolution of nuclear medicine images is inherently limited by the uncertainty in localisation of the gamma photon interaction in the scintillation crystal. A further major factor affecting resolution is the collimator geometry (Figure 10) with longer, narrower holes providing better resolution but reduced sensitivity [110]. Geometric resolution, R_{coll} is defined based on the physical dimensions of the collimator of hole length, l , hole width, d , and the distance of the source, b :

$$R_{coll} = d(l + b)/l$$

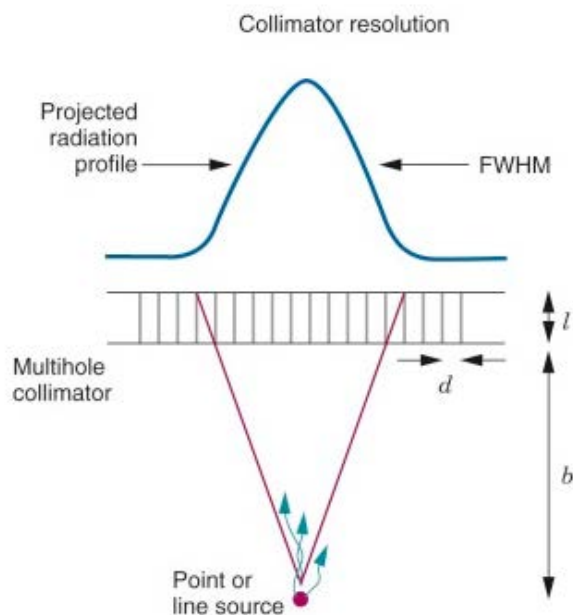


Figure 10: Illustration of the dependence of collimator resolution on geometric properties of hole length and width and source distance [110]. The resolution can be improved by moving the source closer or selecting a collimator with narrower and longer collimator holes, the latter of which is at the expense of reduced sensitivity.

Table 3 gives the properties of the two collimators used in this thesis, the Siemens low energy high resolution (LEHR) collimator and medium energy low penetration (MELP) (Siemens Healthcare, Munich, Germany). The holes in the MELP collimators have lead septal walls that are 7 times as thick

as the LEHR, which minimise septal penetration (see next section). Medium energy general purpose collimators are used in some centres, and have similar properties to the MELP collimators below.

Table 3: Geometrical properties of the Siemens LEHR and MELP collimators used in this thesis.

Collimator	LEHR	MELP
Hole Shape	Hexagonal	Hexagonal
Number of Holes	148,000	14,000
Hole Length	24.05 mm	40.64 mm
Septal Thickness	0.16 mm	1.14 mm
Hole Diameter	1.11 mm	2.94 mm
Geometric Resolution at 10 cm	6.4 mm	10.8 mm

The point spread function (PSF) is the probability function describing the spatial distribution of counts detected from a point source [91]. The measured activity distribution is the convolution of the true activity distribution and the point spread function, so in principle the true activity distribution can be recovered by de-convolution (Figure 11). The PSF is depth dependent, due to collimator geometry, with higher resolution achieved closer to the detector. As gamma camera collimators have known dimensions of hole length, hole diameter and septal thickness, the acceptance angle of photons can be determined. As the distance of the collimator from the centre of rotation is known iterative reconstruction algorithms can model the point spread function, with the width of the function increasing with distance from the detector. This technique is known as resolution recovery (RR).

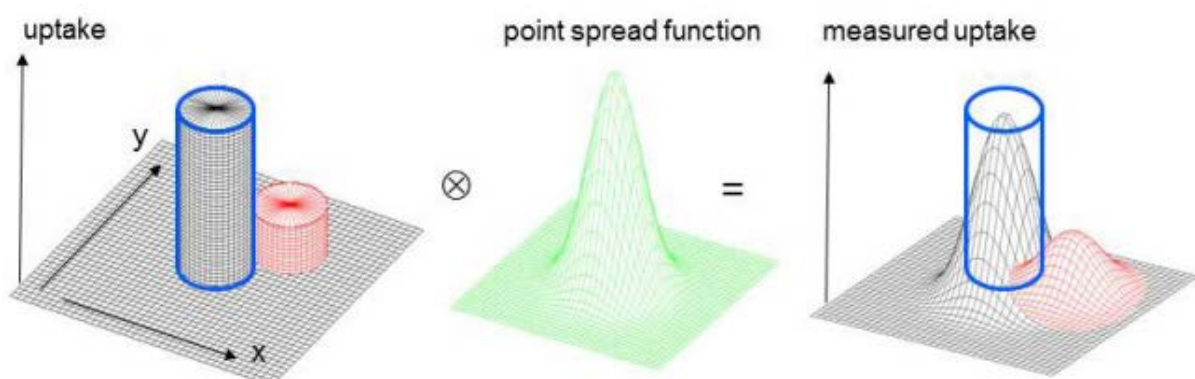


Figure 11: Illustration of blurring of structures and loss of peak signal due to the partial volume effect [91]

The partial volume effect refers to the loss of signal in a voxel of interest due to limited resolution and finite voxel size [110]. This effect is reduced by improving resolution using resolution recovery

modelling. In objects much larger than the full-width half maximum of the PSF (approximately equivalent to the collimator geometric resolution) the centre will reflect true uptake, provided the voxel size is adequately small, and the edges will be blurred. The impact of the partial volume effect is greater for small structures, where even the maximum voxel is underestimated. In this thesis, this has particular relevance to FP-CIT imaging of the striata, which are only around 11 cm³ in volume [111], with maximum dimension around 10 mm. The resolution of MELP collimators at a distance of 10 cm is 10.8 mm, making MELP collimators unsuitable for resolving striatal anatomy, especially at a typical imaging distance of 15 cm. In FP-CIT imaging, low energy high resolution collimators are used, despite the effects of septal penetration (see next section), because high resolution is more important for diagnosis than accurate quantification. It is important to be able to resolve the striata and distinguish between the caudate and putamen. LEHR collimators are recommended by GE Healthcare and the European and SNM guidelines [112, 113].

3.3.4 Septal penetration and collimator choice

The decay scheme of ¹²³I is complex (Figure 12) with a predominant emission at 159 keV (97% probability) and several higher energy emissions of lower probability.

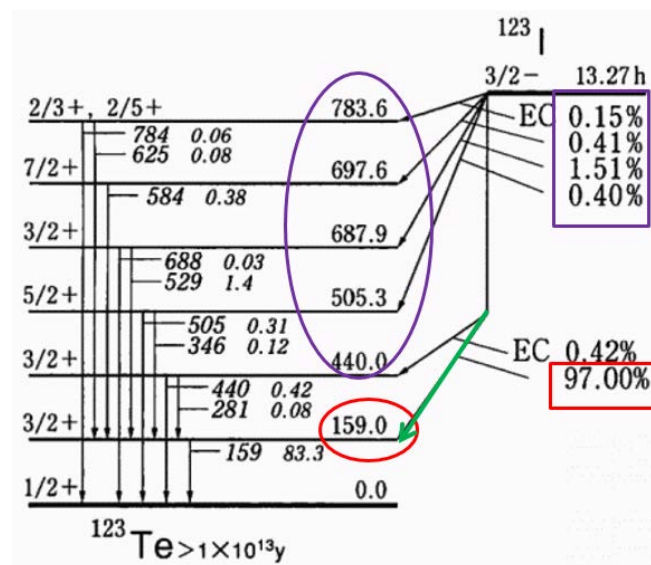


Figure 12: Adapted from http://www.nucleide.org/DDEP_WG/Nuclides/I-123_tables.pdf

Although the higher energies represent a lower proportion of emissions, due to septal penetration and subsequent loss of energy in the detector they can cause a considerable increase in down scatter into the photopeak window. Measurements of ¹²³I energy spectra obtained at our centre with low energy collimators and medium energy collimators (with properties in Table 3) are shown in Figure 13. The reduction in scatter due to septal penetration is clear by examining the area just above the photopeak.

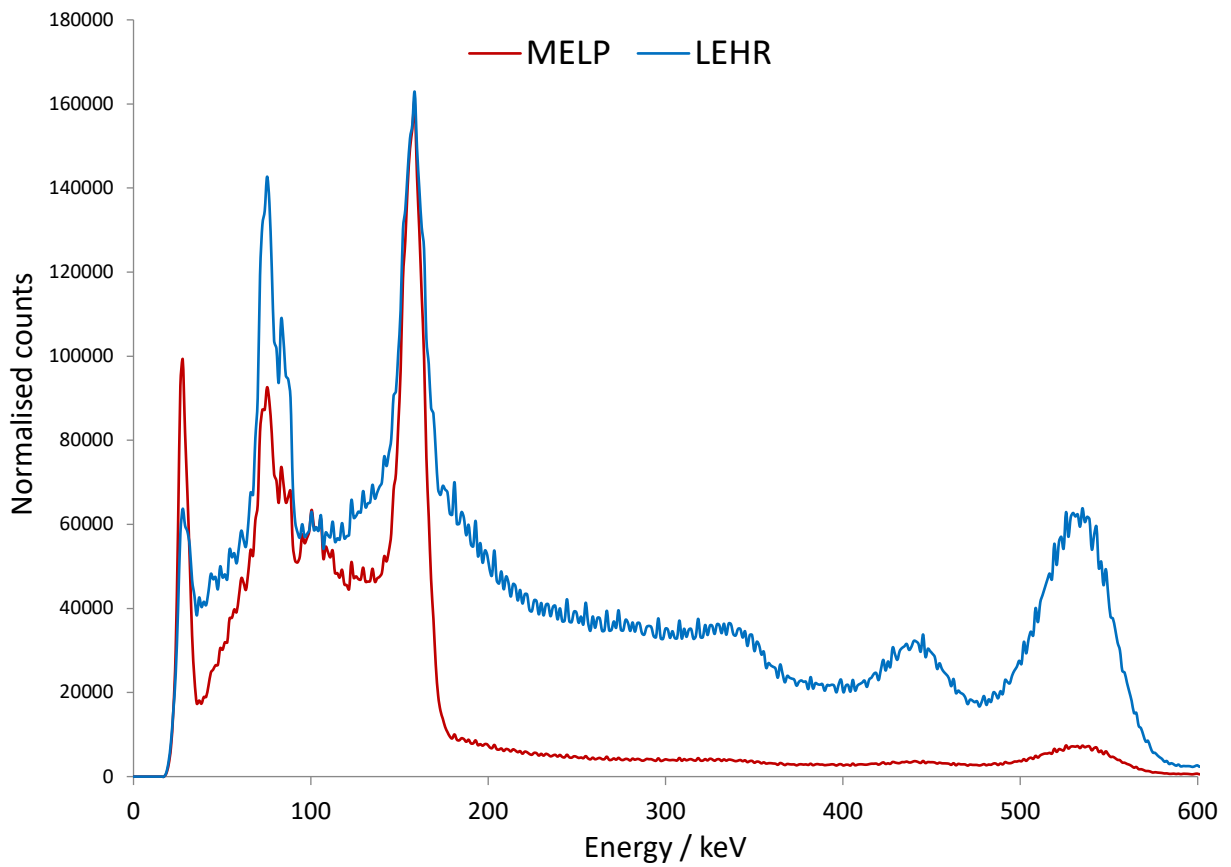
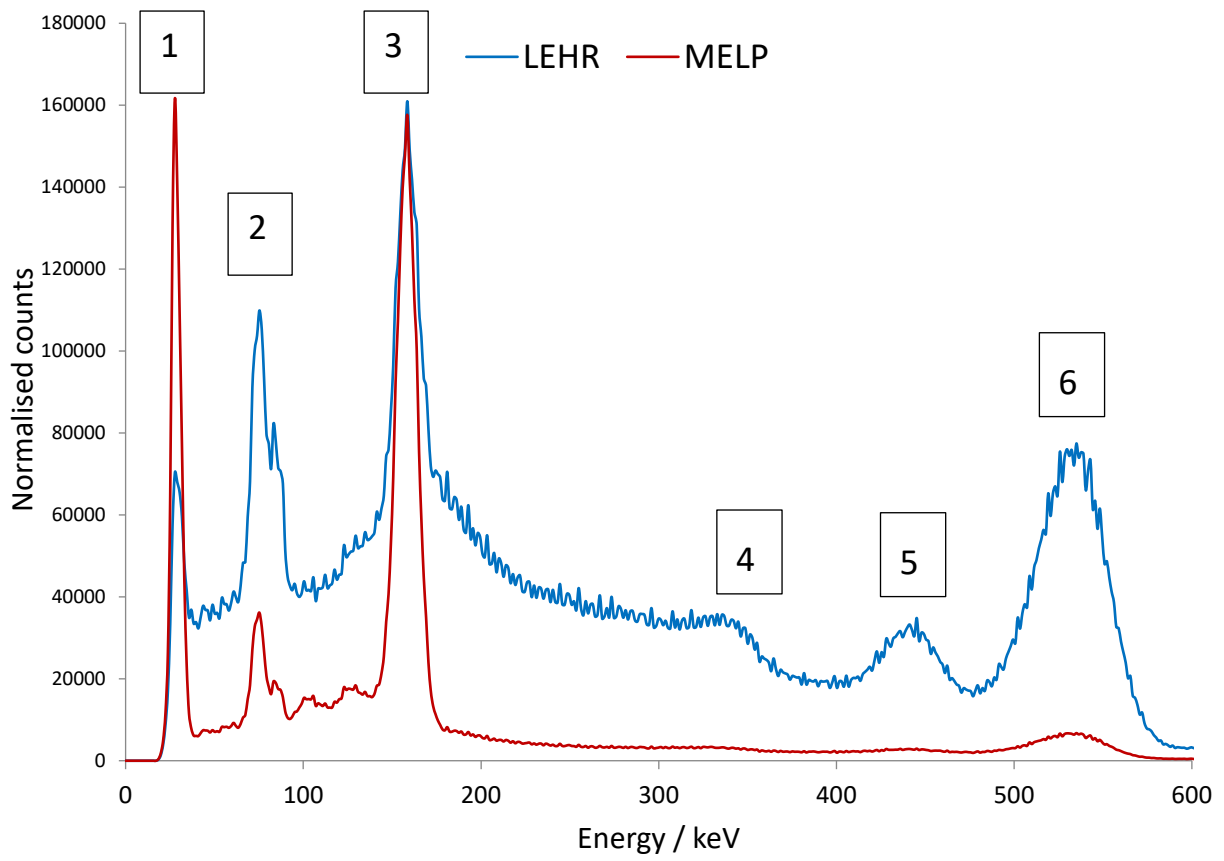


Figure 13: ^{123}I energy spectra acquired by the author with LEHR and MELP collimators for a point source (top) and cylindrical phantom (bottom). Peaks: 1= iodine k-edge, 2= compton scatter peak, 3 = photopeak at 159 keV, 4-6 = high energy emissions

Macey et al. advised the use of medium energy collimators for ^{123}I imaging to reduce septal penetration in the 1980s (for all imaging applications) and this is reiterated by several authors [114] including Dobbeleir et al. [115] and Fletcher et al. [116]. Inoue et al. demonstrated that uptake ratios in an image quality phantom are lower with LEHR [117] and that septal penetration is low with medium energy collimators and of no clinical significance [118].

In cardiac MIBG imaging resolution is less critical as the heart is a larger structure than the striata, so quantification is less susceptible to the partial volume effect. Cardiac MIBG studies are usually quantified by the heart-to-mediastinum ratio (H/M ratio, or HMR) calculated from regions of interest placed on anterior planar images [51]. The HMR is affected by acquisition parameters, most notably the collimator used [119]. Low energy collimators give lower HMRs than medium energy collimators for the reasons discussed above [120] [121]. The difference in image appearance between medium energy and low energy collimators is shown in Figure 14 for torso phantom images acquired in our centre. Figure 15 shows a clinical example, where the increase in background scatter with low energy collimators is clear.

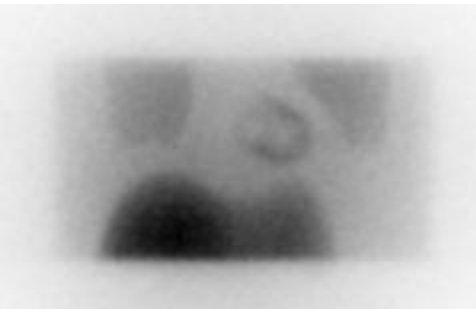

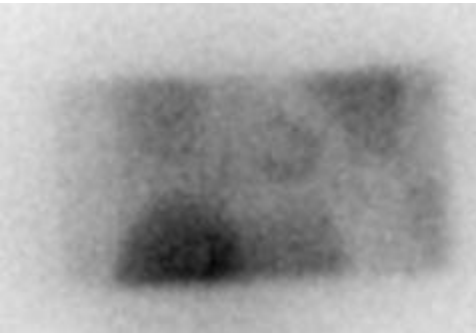
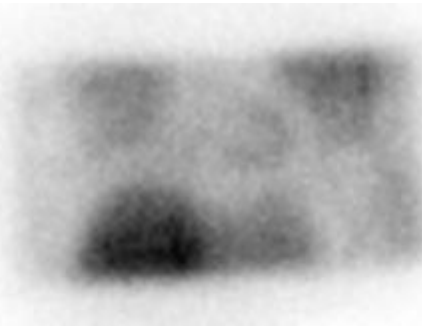
LEHR collimator	MELP collimator
	
	

Figure 14: Torso phantom ^{123}I images acquired by the author using LEHR and MELP collimators. The bottom row shows the phantom covered with attenuating material. The reduction in downscatter from high energy photons with MELP is clear from the reduced background signal outside the phantom.

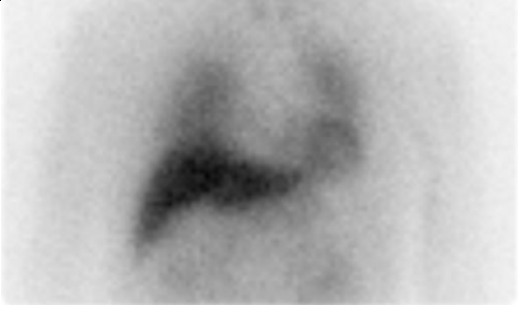

	
Image with LEHR collimator HMR = 1.77 (below MELP normal range)	Image with MELP collimator HMR = 2.79 (within normal range)

Figure 15: Planar cardiac MIBG image from the MIDAS study with incorrect low energy collimator (left) and correct medium energy collimator (right).

3.3.5 Quantitative SPECT reconstruction algorithms used in this thesis

All SPECT reconstructions in this thesis are performed with HybridRecon software (Hermes Medical Solutions, Stockholm, Sweden). The HybridRecon reconstruction algorithm uses the OSEM method discussed above, allowing image corrections to be applied during the iterative process. This includes the implementation of attenuation correction and Monte Carlo scatter correction [122]. The Monte Carlo simulator tracks the path of simulated scattered photons to the detector as part of the forward projection step discussed above. The algorithm supports either a uniform or CT attenuation map, used for both attenuation and scatter correction [123]. Depth-dependent resolution recovery is implemented in the algorithm, using the collimator geometry specified by the user to define the point spread functions. This has been validated for myocardial perfusion imaging with ^{99m}Tc but not for ^{123}I [123]. However, resolution recovery is recommended by Hermes for ^{123}I -FP-CIT reconstruction, as well as uniform scatter and attenuation correction, and is used throughout this thesis.

Septal penetration is not modelled in the standard algorithms used in this thesis – only the 159 keV ^{123}I photopeak is taken into account. However, a more advanced Monte Carlo algorithm incorporating septal penetration and subsequent scatter of high energy emissions for LEHR collimators has been developed [124, 125] and was accessed under a research agreement with Hermes. This algorithm is called full collimator modelling and is applied to FP-CIT imaging in chapter 6. Hermes have not implemented the full collimator model algorithm for ^{123}I imaging with medium energy collimators, as septal penetration is already minimised, as shown in the energy spectra in Figure 13 above.

3.4 Relative and absolute quantification

Until relatively recently, SPECT was not thought to be a fully quantitative imaging modality, unlike PET [86]. Quantification of activity in organs of interest has traditionally been estimated by comparison with a reference region assumed to contain non-specific uptake. Specific binding ratios in FP-CIT and heart to mediastinum ratios are both examples of this method, known as semi-quantification (see FP-CIT and cardiac MIBG sections below), which continues to be used for most radionuclide imaging applications. However, the reliance on a background region and the assumption that this represents non-specific uptake introduces uncertainty – the regions used are often noisy due to low count density and there may be issues with image registration and uniformity of signal. These background regions may also be affected by disease, for example atrophy may cause reduced signal, or may vary between patients due to fluctuations in physiological uptake. It may also be difficult to ensure there is no element of specific uptake contained within the background region.

Absolute quantification of the activity concentration within a volume of interest (VOI) is therefore attractive because it removes the reliance on the background region. A calibration factor is used to convert between reconstructed image counts and activity concentration in kBq/ml; this is obtained using a phantom containing measured activity concentration. Accurate quantification relies on a reconstruction method that is able to model the factors that cause image degradation and loss of signal – gamma ray attenuation, scatter, and the finite resolution of the system, which leads to substantial partial volume effects. For ^{99m}Tc , absolute quantification has been shown to be accurate to within 10% in a multicentre phantom setting [126], but this has not been explored for ^{123}I , which is more complex to quantify. In the case of ^{123}I , septal penetration of high energy emissions also affects quantification, for low energy collimators but is less important for medium energy [104]. The absolute activity concentration also depends on the tracer volume of distribution, which is related to patient size, so some form of normalisation is necessary. The simplest way to normalise the images is to use patient weight, but this may be an oversimplification.

Crucially, software must be available for performing the absolute uptake calculations. Lack of suitable methods for attenuation and scatter correction has traditionally been a barrier to absolute quantification methods being developed for nuclear medicine, but in recent years several manufacturers have released software supporting SUV (standardised uptake value) calculations, including within HybridRecon (Hermes Medical Solutions), the image reconstruction software used in this thesis. SUV is a dimensionless quantity defined as the observed activity concentration within a volume of interest, normalised to the administered activity and patient weight (as a surrogate for tracer volume of distribution).

$$SUV = \frac{\text{Activity concentration (kBq/ml)}}{\text{Administered activity (MBq)/Patient weight (kg)}}$$

SUV values are calculated for each voxel, allowing images to be displayed using an absolute SUV scale rather than the standard relative scale. SUV has been used in PET imaging for many years and been shown to be useful in a wide range of clinical settings, despite issues regarding the normalisation method [88, 127, 128].

SUV information for a volume of interest can be reported in a number of ways. The SUV of the maximum voxel within a volume of interest, SUV_{\max} is the simplest technique and commonly used in PET imaging, but can be an unreliable marker due to statistical noise [129-131] and will not be used in this thesis. SUV_{mean} is the mean SUV within a volume of interest. SUV_{peak} is defined as the SUV_{mean} of the most active 1cm^3 cubic volume within a given volume of interest.

3.5 ¹²³I-FP-CIT image analysis overview

FP-CIT studies are often reported using visual analysis alone, although a 2015 Cochrane review concluded that semi-quantitative measures alone may be more accurate than both visual ratings and clinical diagnoses [132]. In DLB there is often a more uniform reduction in striatal uptake, rather than the predominantly putaminal loss seen in PD [133, 134]. This “balanced loss” is more difficult to recognise visually, because images are scaled to the maximum pixel, which may account for some of the scans reported as normal in DLB patients.

FP-CIT semi-quantification usually involves calculating the specific binding ratio, SBR, from the mean counts per voxel (CPV) in striatal and background volumes of interest:

$$SBR = \frac{CPVmean_{striatum} - CPVmean_{background}}{CPVmean_{background}}$$

Here the count density in the background volume is assumed to be representative of the non-specific binding in the striatal volume. The SBR is often referred to as the striatal to background ratio but this is not strictly accurate as it is only the specific binding in the striatal region that is used in the calculation.

As an adjunct to clinical reporting, semi-quantification with SBR has been shown to reduce equivocal reporting of ¹²³I-FP-CIT studies in clinical practice [135] and increase repeatability [136] and reporter confidence [137] with a small increase in diagnostic accuracy reported for DLB patients [138]. However, a recent large single centre study comparing visual interpretation with semi-quantification in 382 patients in a movement disorders setting concluded that visual read alone was at least as accurate as semi-quantification [139]. However, changes in PD may be more obvious visually than in DLB, and in early Lewy body disease changes could be more subtle and benefit from quantification.

The introduction of sophisticated fully automated commercial ¹²³I-FP-CIT software packages [111, 135, 140-142], combined with greater awareness in the nuclear medicine community of “balanced loss” of dopaminergic function, particularly in DLB, has led to increased interest in quantification for FP-CIT imaging [137]. Balanced loss can be challenging for the non-expert to identify visually as images are scaled to maximum uptake rather than to a predefined absolute uptake level; low specific binding ratio (SBR) values for striata with otherwise normal appearance are therefore useful for identifying global loss of uptake. Z scores, defined as the number of standard deviations the subject SBR is away from the mean SBR of a database of healthy control subjects, are often used to help show how normal or abnormal an SBR value is, with a Z score below -2 typically raising concern for an abnormal scan.

There is a wide range of SBR calculation methods available for identifying reduced striatal uptake, ranging from automated 3D image registration methods to simple yet repeatable manual programs.

The Southampton method uses large volumes of interest encompassing counts from the entire striata to minimise partial volume effects [143, 144], whereas most other methods calculate count density in a volume of interest registered to the striata [144]. The accuracy of these two approaches are similar, although a study from Japan, where the Southampton method has been adopted for clinical use, suggested GE Healthcare's automated method (DaTQUANT) was slightly more accurate [145]. A recent large multicentre comparison of standard SBR methods (striatal-sized VOIs) and the Southampton method concluded that the standard method is slightly more accurate for single centre use and the Southampton method for multicentre [146]. However, the SBR method used was an in-house method so it is not certain that these results would apply with commercially available software packages.

Crespo et al. showed that correction for the partial volume effect using a freely available method, BasGan, improves quantification [147, 148]. However, there are no CE marked commercially available software packages with PVE correction implemented, so it is not available in clinical practice. Recent interesting results from Japan suggest that SBR accuracy improves when corrections for low signal ventricle contribution are corrected for [149-151], however this is outside of the scope of this thesis.

Regardless of the quantification method used to calculate the SBR, the results are subject to reported variation due to age and sex [152, 153] as well as technical factors [154-156] including variations in sensitivity and resolution between gamma camera models, SPECT reconstruction method, use of attenuation and scatter correction and the partial volume effect. Substantial variation between centres was demonstrated by a recent multicentre phantom study [157]. This means a normal range for striatal uptake is only usually valid if acquired in the same centre as the scan being analysed. It is challenging for an individual centre to acquire scans of sufficient normal subjects for confident application of semi-quantitative analysis, so visual reporting has remained common in clinical centres. The European ENC-DAT project has produced a multicentre normal database of control subjects by accounting for the differences between gamma cameras via phantom calibration data [158]. Hermes have used the data to provide normal ranges within their BRASS software [159, 160], which is assessed in Chapter 4. We assess the benefits of a normal range developed with local controls in Chapter 4 and 6.

Lange et al. showed that both CT-based AC and uniform AC had minimal impact on the interpretation of FP-CIT scans, both on visual analysis, where no difference was seen and on quantification where diagnostic accuracy was not improved significantly [102]. A small retrospective study by Lapa et al. also showed no difference in diagnostic accuracy in thirty patients with suspected PD [161]. Warwick et al. showed SPECT-CT to be effective in a striatal phantom study but not in patients undergoing FP-CIT scanning [162]. However, some authors have found evidence of an improvement in relative quantification of FP-CIT scans when SPECT-CT is used for image registration, rather than attenuation correction. Maebatake et al. found that in phantom studies a volume of interest defined on registered

CT images [163]; however the study failed to investigate CT attenuation correction and instead used the uniform method. Yokoyama found similar results for clinical studies reconstructed with uniform corrections and registered to a template using the CT data [164]. It is unclear why these publications did not consider CT attenuation correction. The benefit of CT attenuation correction for MCI-LB diagnosis will be evaluated in Chapter 6.

3.6 Cardiac ^{123}I -MIBG image analysis overview

Despite the technical advantages of medium energy collimators, low energy collimators continue to be used in publications and in clinical practice. The European Association of Nuclear Medicine guidelines for standardisation of cardiac MIBG imaging state that the “extensive availability of low energy, high-resolution (LEHR) parallel-hole collimators determine their common use for ^{123}I studies” but recommend that medium energy collimators be used for cardiac MIBG if practical [165]. Medium energy collimators have limited use for other studies as $^{99\text{m}}\text{Tc}$ radionuclides are imaged using low energy collimators; it may therefore be difficult for a small nuclear medicine department to justify the additional cost. Collimator choice for cardiac MIBG imaging is considered further in Chapter 5, as part of a literature review of clinical protocols.

Calibration of HMRS between gamma cameras to overcome collimator effects using a phantom has been shown to be feasible [166] and was implemented in a Japanese multicentre study [53]. More recently, a European calibration study was published that included results for the gamma cameras at the Newcastle Royal Victoria Infirmary used for the cardiac MIBG studies discussed in this thesis [167]. This is discussed in more detail in Chapter 5.

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. [167-172]. 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.* [166] was used in the multicentre study mentioned above [53, 62]. However centres outside Japan still tend to publish results generated using local acquisition parameters and local normal uptake thresholds [67, 68], 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 [173], used in the multicentre study [53], 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 [120]. To our knowledge only two studies utilising visual assessment of cardiac MIBG images in Lewy body disease have been published [67, 174], 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. This will be covered in Chapter 7.

The phantom calibration procedure is useful in reducing camera-specific factors, but the HMR calculated from planar images is also affected by patient specific factors such as the amount of non-cardiac uptake overlying the myocardium, scattered radiation from the liver and lungs and the degree of scatter and attenuation due to differences in body habitus [170]. This is discussed further in Chapter 8. The variation in HMR means planar analysis may not be a robust method for quantifying small reductions in innervation that may be seen in prodromal cases – the magnitude of the variation may mask any true difference between MCI-LB and other subjects.

Although planar imaging is used as standard throughout the literature, use of SPECT has the potential to improve DLB diagnosis by increasing contrast between the heart and background. SPECT-CT could lead to further improvements by reducing the effects of scatter and attenuation, potentially allowing more subtle reductions in uptake to be identified. Tomographic imaging may allow identification of regional areas of low uptake due to heart disease which may improve specificity in a population with a high prevalence of coronary arterial disease. The additional benefit of SPECT MIBG imaging over planar imaging for DLB has not been fully evaluated in the literature, but regional assessment of denervation is used routinely in heart failure imaging [175-177] and is beginning to be used for Lewy Body diseases [56]. Alvi et al. have shown that HMR results from a dedicated SPECT cardiac camera correlate with planar HMR [178]. SPECT has been shown to improve the sensitivity of PD diagnosis primarily by reducing the number of false negative scans caused by non-cardiac uptake in the heart region [179]. However, corrections for scatter and attenuation were not evaluated in this study and could bring further improvements. The use of the mediastinum as a reference region has been questioned, with studies coming to different conclusions about its validity [169, 180-182]. This will be explored in Chapter 8 and 9.

Chapter 4: Development of imaging methods for research studies: FP-CIT

The work presented in this chapter deals with establishing an optimised standard method to use in the SUPeR study for quantitative and visual analysis of FP-CIT scans. ^{123}I -FP-CIT scans of normal clinical patients, control subjects, MCI and dementia subjects recruited and scanned locally are used to explore a range of technical aspects related to DaTQUANT and BRASS and to determine which is the most appropriate to use for research studies using FP-CIT, in particular the SUPeR study.

We show that BRASS and DaTQUANT give similar results and outline our reasons for selecting BRASS for the SUPeR study. Concerns over age-correction are uncovered, to be investigated further later in the thesis. MCI subject scans and phantom studies are used to show that the effects of acquiring images with a smaller pixel size and applying zoom and reorientation during image reconstruction have little impact on BRASS quantification results but may affect DaTQUANT results.

4.1 Comparison and validation of two commercial ^{123}I -FP-CIT software packages

4.1.1 Introduction

The ^{123}I -FP-CIT radiopharmaceutical is patented by GE Healthcare (Chalfont, UK) and sold under the brand name DaTSCAN. GE Healthcare have released a software package for semi-quantification, DaTQUANT, which calculates SBRs and compares the results to a normal database. Similar functionality is available in the FP-CIT module within the BRASS brain analysis software produced by Hermes Medical Solutions (Stockholm, Sweden). Both programs perform an image registration between the subject FP-CIT image and a normal template generated from over 100 control scans. They then apply fixed volumes of interest to the striata and the occipital background region in order to calculate SBRs. These values are expressed relative to the control database via Z scores, which indicate how the subject result compares to age-matched controls in terms of number of standard deviations from the control mean. Example output from DaTQUANT (v1.01) and BRASS (v2.5) for the same normal scan is given in Figure 16.

Despite the apparent similarity between DaTQUANT and BRASS there are several technical and practical differences between them, which are listed in Appendix A. These include different approaches to image registration, the use of different normal databases, whether corrections for the camera used are applied and whether the images are reconstructed within the program or externally. BRASS has several technical advantages over DaTQUANT but DaTQUANT is quicker and easier to use, and produces results in a useful graphical form. Both apply age correction factors to the Z score output, although this can be removed or modified in BRASS, boosting the results in old age to compensate for a reduction in normal SBR values with age. In this thesis the default age corrected results are used, unless otherwise specified.

Objectives

- To acquire striatal phantom scan data for known striatal and background activity concentrations and produce a camera correction factor for BRASS.
- To validate the two software packages using FP-CIT scans from clinical patients unlikely to have Lewy body disease – average age corrected Z scores should be close to zero
- To compare SBR and age corrected Z score results given by each package

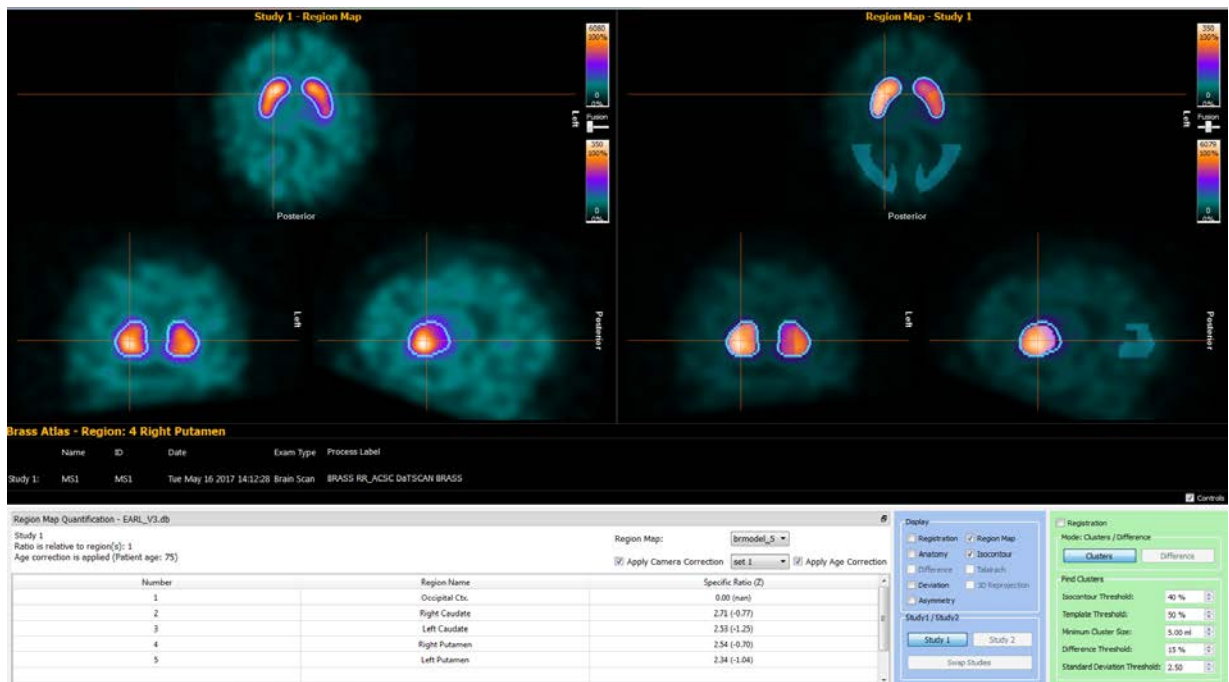
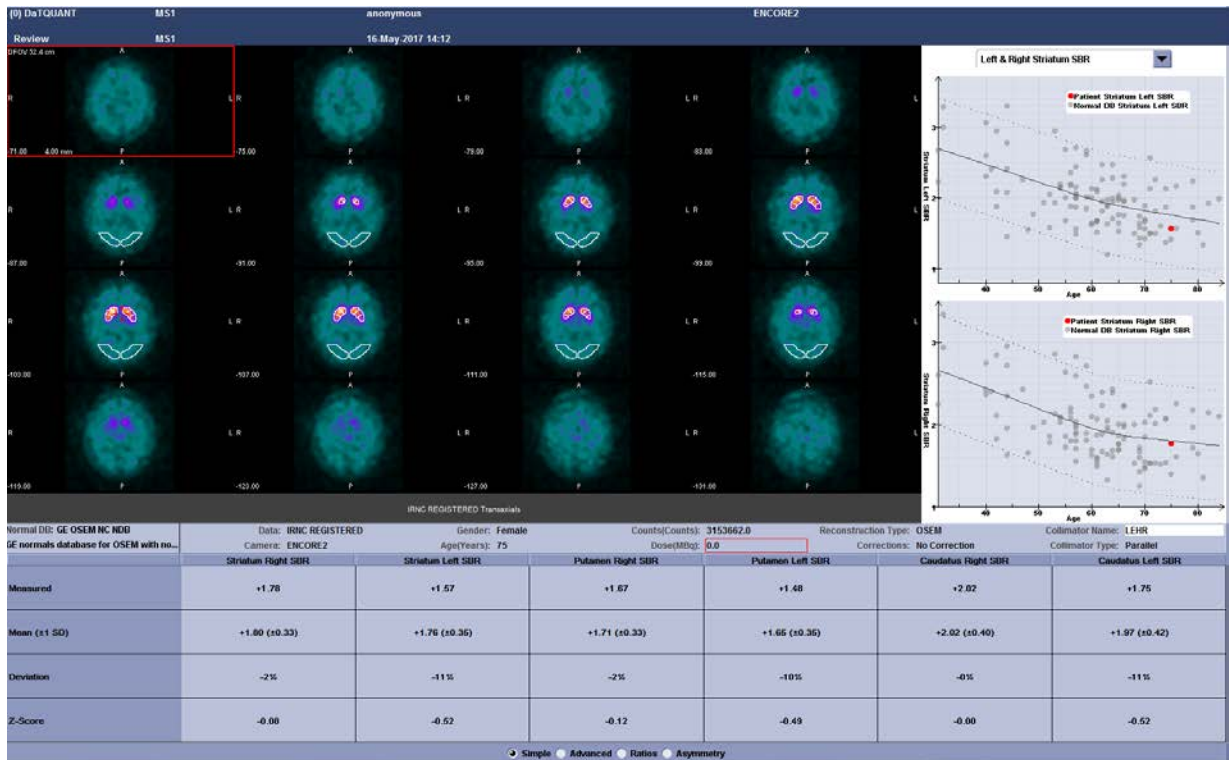


Figure 16: Results screens for DaTQUANT (top) and BRASS (bottom) for the same subject with normal uptake.

4.1.2 Methods

The software packages are compared here using data from clinical studies rather than true healthy control subjects because such control data was not yet available at the time the initial comparison was made (autumn 2016). Fifteen anonymised patients with normal scans were selected from an RVI clinical audit of FP-CIT scans. These patients had been referred for suspected PD or DLB but had had a normal FP-CIT scan report and at least three years follow up post-scan, confirming that there was no indication that any of them had Lewy body disease. Patient records were accessed by NHS staff forming part of the clinical care team, and not by the author, to conform to clinical governance guidelines and protect patient confidentiality. Whole striatum, caudate and putamen SBRs and Z scores were calculated with BRASS (after camera calibration – see below) and with DaTQUANT.

Acquisition parameters

All individuals included in this thesis were scanned three to six hours following a bolus intravenous injection of 185 MBq of ¹²³I-FP-CIT (DaTSCAN, GE Healthcare, UK) using a dual headed gamma camera fitted with low energy high resolution (LEHR) parallel hole collimators. 120 x 25 second views over a 360° orbit were acquired on a 128x128 matrix with a zoom of 1.23x giving a pixel size 3.9mm x 3.9mm.

BRASS: camera calibration

BRASS applies a camera correction factor to the SBR results, which is calculated by Hermes using scans of a striatal phantom of six different known SBRs. The known ratios are compared to the uncorrected SBRs given by BRASS for these scans³. We obtained the required image data using an anthropomorphic striatal phantom (RSD Radiology Support Devices, Long Beach, CA, USA), which consists of a perspex brain shell with fillable caudate, putamen and background compartments inserted into a skull model for scanning (Figure 17).

The phantom was filled with ¹²³I activity in the striatal and background sections – with the right striatum more concentrated than left, giving activity ratios of approximately 7:1 and 5:1. The exact ratios were verified later using striatal and background samples measured in a gamma counter. The phantom was imaged three times with increased background activity concentration each time to obtain scans with six different known striatal to background activity ratios (approx. 7:1, 5:1, 3.5:1 and 2.5:1, 1.7:1 and 1.2:1).

³ The calibration factor is intended to correct for differences between camera models that can introduce small systematic differences in SBR. The ENC-DAT normal database scans come from a range of different camera models with all results corrected using this phantom technique. A similar calibration must therefore be applied to our image data to be consistent with the normal database.

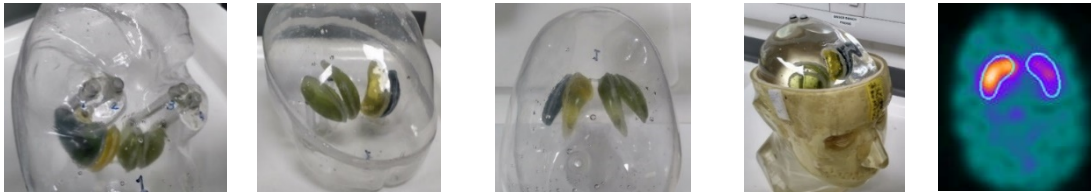


Figure 17: Striatal phantom used for FP-CIT quantification work. The images show the brain shell and assembled phantom being scanned and an example transaxial phantom image analysed in BRASS

The uncorrected SBRs given by BRASS were recorded for each of these known ratios by running the software on the reconstructed phantom images. The performance of the system can be evaluated by calculating the gradient of a linear fit of the BRASS SBRs against the known activity ratios. This gradient is always less than 1 due to count losses from partial volume effect and any residual losses due to attenuation and scatter. For our camera the gradient was 0.46 (see Figure 18). The gradient for the reference camera model used in the normal database is 0.42. These results were used by Hermes to produce a camera specific calibration factor applied to adjust the SBR values within the BRASS software so that they would be the same as those given by the reference gamma camera. The factor obtained was 0.9265 and was supplied via an updated system file that BRASS refers to when displaying SBRs. Both the uncorrected and corrected SBRs can be viewed in the software. The factor is multiplicative so a correction factor of less than one corrects for SBRs given by our camera model (Siemens Intevo) being higher than the reference camera.

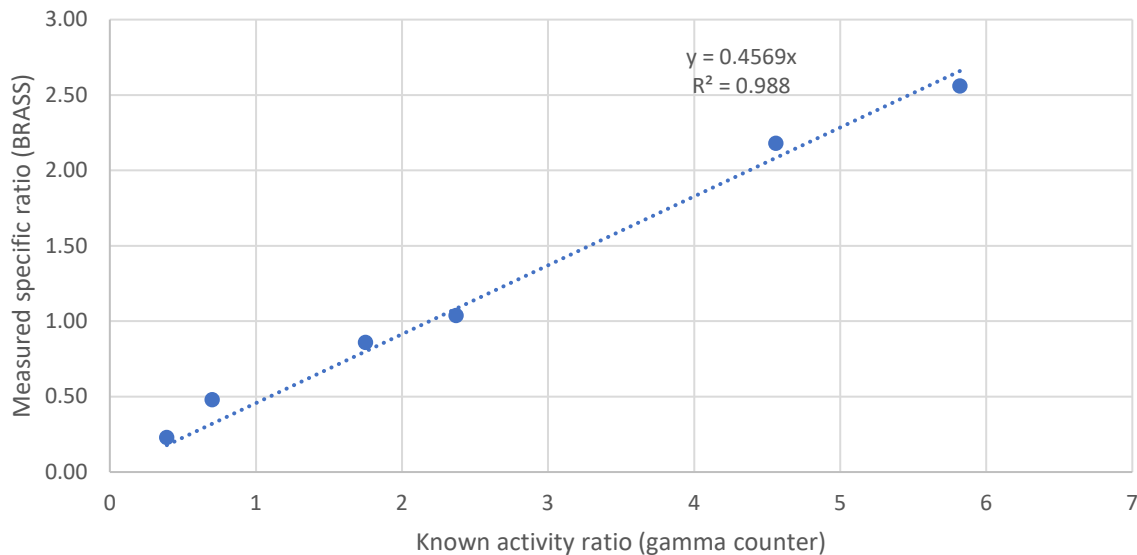


Figure 18: BRASS specific ratios (SBR -1) against activity ratios measured with a gamma counter, showing a linear fit with gradient 0.46.

BRASS: normal patient processing

The raw patient data was reconstructed within Hermes HybridRecon Neurology, using iterative reconstruction with resolution recovery, uniform attenuation correction and Monte Carlo scatter correction (OSEM_RRACSC). Reconstruction parameters specified by Hermes (Figure 19) were used so that the BRASS normal database results, derived from images reconstructed with the same parameters, would be valid. These parameters were also used for the phantom acquisitions discussed above. The resultant images were processed with BRASS and the SBR results and age corrected Z scores for the left and right whole striatum, caudate and putamen were recorded. The mean SBR and mean Z score for each region was calculated. This work led to the detection of a software error, affecting multiple centres and causing BRASS Z scores to be underestimated by around 1 standard deviation, which was subsequently rectified by the manufacturer.

DaTQUANT: camera calibration

DaTQUANT does not have a camera correction feature as the study that the normal database scans were taken (PPMI – Parkinson’s Progression Markers Initiative) used multiple cameras but did not take differences between them into account.

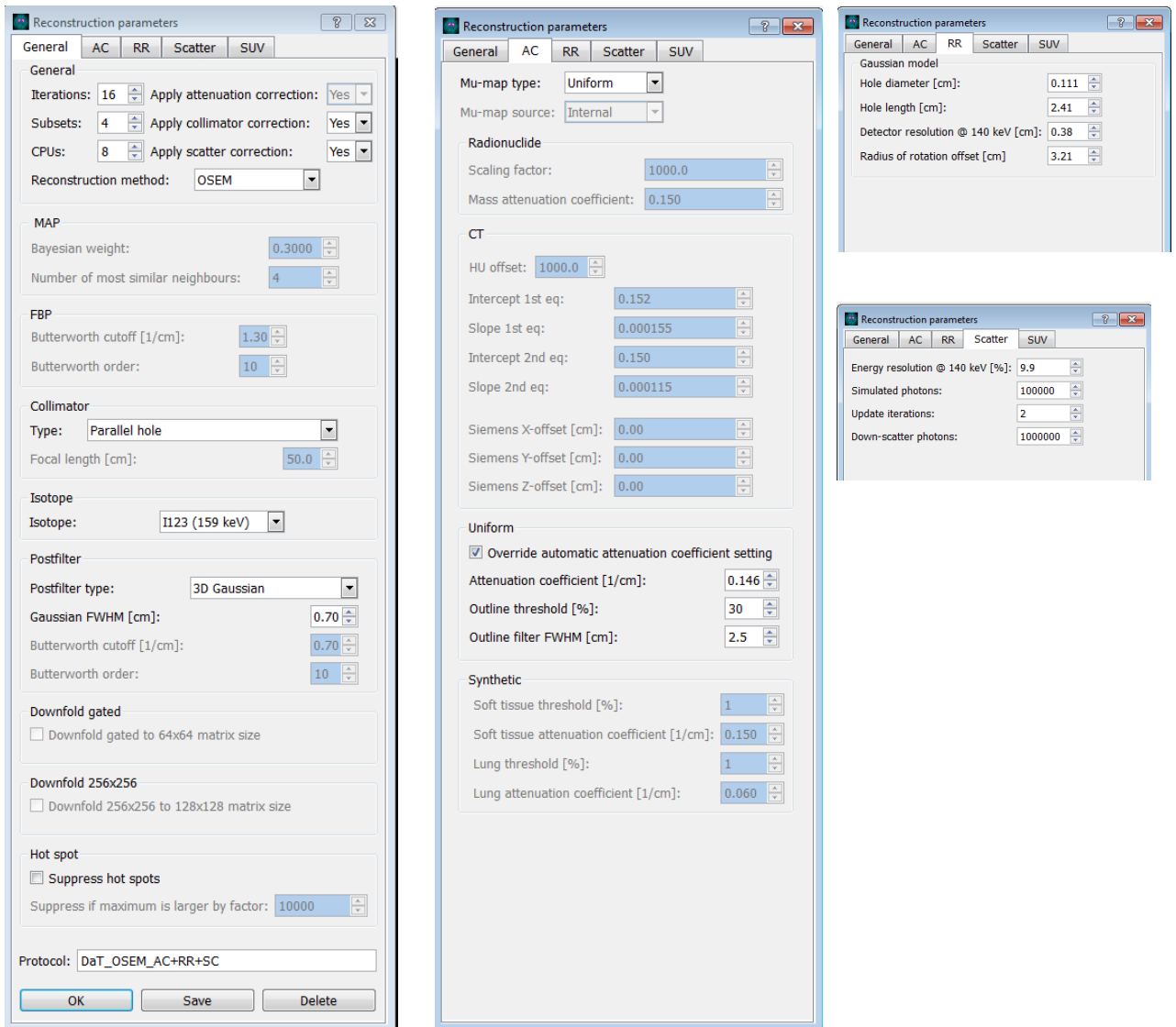


Figure 19: BRASS reconstruction parameters for FP-CIT scans

DaTQUANT: normal patient processing

The raw FP-CIT data was imported into DaTQUANT and reconstructed using iterative reconstruction (OSEM) with no attenuation correction⁴ with the parameters shown in Figure 20. This is one of the four reconstruction options recommended by GE (OSEM with and without AC and FBP with and without AC). The PPMI healthy control database has been reconstructed with each of these sets of parameters and the database matching the reconstruction settings chosen for the input image is automatically selected. SBR results and age corrected Z scores for the left and right whole striatum, caudate and putamen were recorded for each patient. The mean SBR and mean Z score for each region was calculated.

⁴ Uniform (Chang) attenuation was found to be unreliable in DaTQUANT with the default threshold applied, leading to discontinuities between transverse slices. AC was therefore not used with DaTQUANT.

DaTQUANT Customization	
Input Customization Reconstruction Hybrid QC Non-Rigid Reg. Review	
Skip Spect Options screen	Yes
** GROUP NAME	I123 (or not Tc99m or In111)
Reconstruction Type	OSEM
Prefilter type	BUTTERWORTH
Prefilter Param1	0.6
Prefilter Param2	10.0
3D Postfilter type	BUTTERWORTH
3D Postfilter Param1	0.6
3D Postfilter Param2	10.0
Reconstruction Ramp Filter	Quantitative
Number of OSEM iterations	2
Maximum number of OSEM sub...	10
** CORRECTIONS	
Scatter Correction	Off
Attenuation Correction	Off
Reconstructed Sets	Non-Corrected + Corrected
Center Slice Strategy	Brain
Scatter weight for Tc99m	1.1
Scatter weight for Tl201	1.4
Scatter weight for Ga67	1.0
Default scatter weight	1.0
Confirm Cancel Request	Yes
Find Originals CT	Enable
Chang Attenuation Threshold	15
Chang Attenuation Coefficient	0.15

Figure 20: DaTQUANT reconstruction parameters

Direct comparison between DaTQUANT and BRASS SBRs

Because of the different standard reconstruction parameters used by DaTQUANT and BRASS it is not possible to use the above data to determine whether there are any technical differences between the SBR calculation methods of the two software packages. To compare BRASS and DaTQUANT SBRs directly, the OSEM_RRACSC reconstructed images used for BRASS were exported from the Hermes system and imported into DaTQUANT. DaTQUANT accepts externally reconstructed images for SBR calculation but cannot give a Z score as the reconstruction parameters are different from the DaTQUANT normal database. If the two programmes are comparable there should be no significant difference between the BRASS SBR results without camera correction applied and the DaTQUANT SBR results for the same input images.

4.1.3 Results and discussion

One patient was excluded from analysis because there was significant atrophy in the background region, giving a much higher SBR than the other normal subjects on DaTQUANT, with Z score showing this was more than 4 standard deviations above the database mean (Figure 21). The SBR was not high for BRASS, even though the region placement was similar. The discrepancy was investigated and followed up with Hermes, who confirmed that BRASS compensates for lower than average voxel values in the background VOI, to avoid this artificial increase in SBR. This seems sensible, but as it is not mentioned in the extensive manual it is somewhat misleading. Expert users will recognise that atrophic background regions could lead to overestimated uptake, and if they are unaware of this compensation they may discard results that are fine or mentally adjust them without needing to. We found it is possible to indirectly tell when this compensation has occurred as the results can be presented with the number of voxels in the VOI alongside the SBR result. A typical value for the background is 8000 voxels and in this case less than 7000 were included.

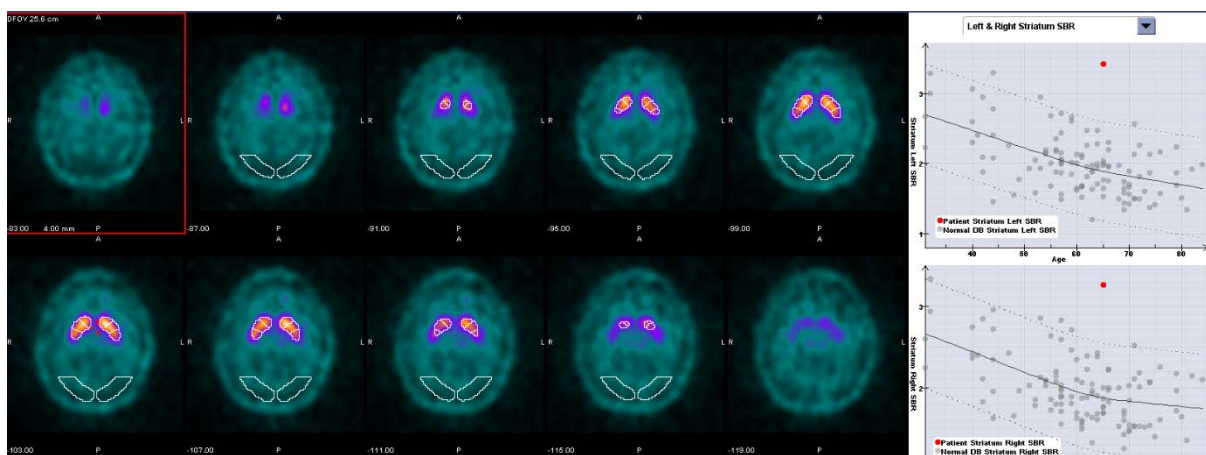


Figure 21: DaTQUANT output for the excluded normal patient with low signal in the occipital background region.

The average SBRs and Z scores for the right and left whole striatum, caudate and putamen regions for the remaining 14 patients are given in Table 4 for BRASS and DaTQUANT. The SBR results are much higher for BRASS than DaTQUANT, due to the different image reconstruction methods used for each software package. Attenuation and scatter corrections are applied to the BRASS input images but not for the images reconstructed within DaTQUANT, which is a likely reason for the SBRs to be higher with BRASS. However, the Z scores are also quite different and this cannot be explained by differences in reconstruction parameters as Z scores are based on comparing the images with a normal database reconstructed with the exact same parameters. The average Z scores for the normal subjects analysed with BRASS are close to zero (0.08), as expected. The Z scores for DaTQUANT were higher than expected, 0.65 standard deviations higher than the mean of the normal database.

The Z scores for DaTQUANT and BRASS are shown against patient age in Figure 22. GE Healthcare was contacted to discuss the higher than average Z score with DaTQUANT but no clear reason was identified. The technical specialist commented that it is possible to see very high SBRs in normal subjects and that this can skew Z score results since a normal distribution is assumed.

Table 4: Mean normal patient SBR and Z score results for BRASS and DaTQUANT

	BRASS		DaTQUANT	
	Mean SBR	Mean Z score	Mean SBR	Mean Z score
Right striatum	3.2	0.1	2.2	0.72
Left striatum	3.19	0.07	2.21	0.57
Average striatum	3.2	0.08	2.2	0.65
Right caudate	3.33	0.14	2.51	0.79
Left caudate	3.4	0.16	2.49	0.62
Right putamen	3.06	0.05	2.09	0.66
Left putamen	2.97	-0.04	2.04	0.51

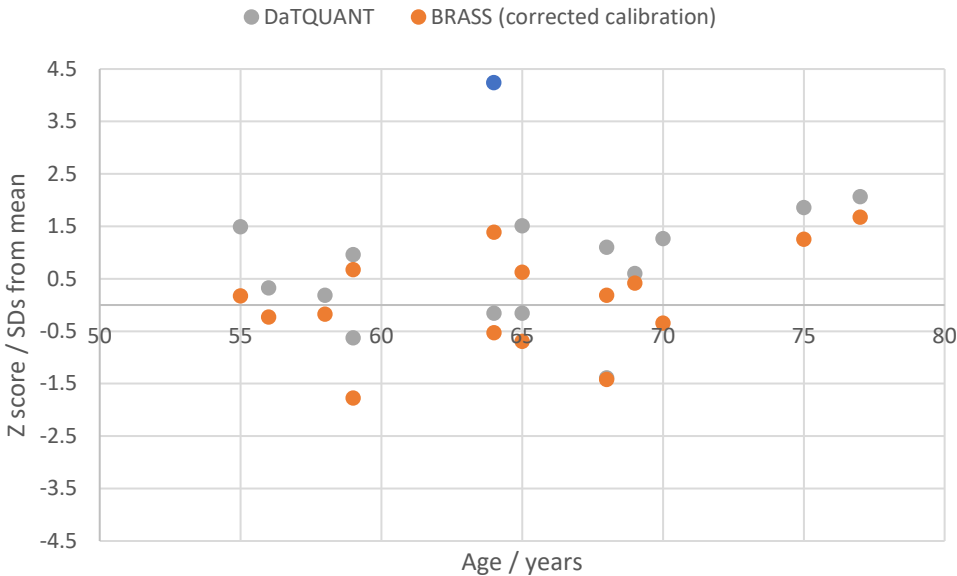


Figure 22: Z scores for the 15 patients for BRASS and DaTQUANT. The patient with the very high result with DaTQUANT result (Z score > 4 SD, shown in blue) was excluded from analysis. The results for BRASS are equally distributed around zero (with the correct calibration applied). Most of the DaTQUANT results are above zero.

Direct comparison between DaTQUANT and BRASS SBRs

The average SBR results for the 14 patients reconstructed with Hermes OSEM_RRACSC and processed on both BRASS and DaTQUANT are given in Table 5. The difference between the mean SBR and the p-value for a paired t-test between the results are also given. The plot in Figure 23 shows that DaTQUANT SBRs are higher than BRASS for all 14 patients. Ideally a comparison in a larger dataset containing normal and abnormal studies would be done. This is covered later in section 6.2.

Table 5: Difference in SBR results between BRASS and DaTQUANT

Mean SBR	DaTQUANT	Mean BRASS SBR (no camera correction)	Mean diff BRASS - DaTQUANT	p-value (paired t-test for difference)
3.29		2.84	-0.49 (15%)	<0.001

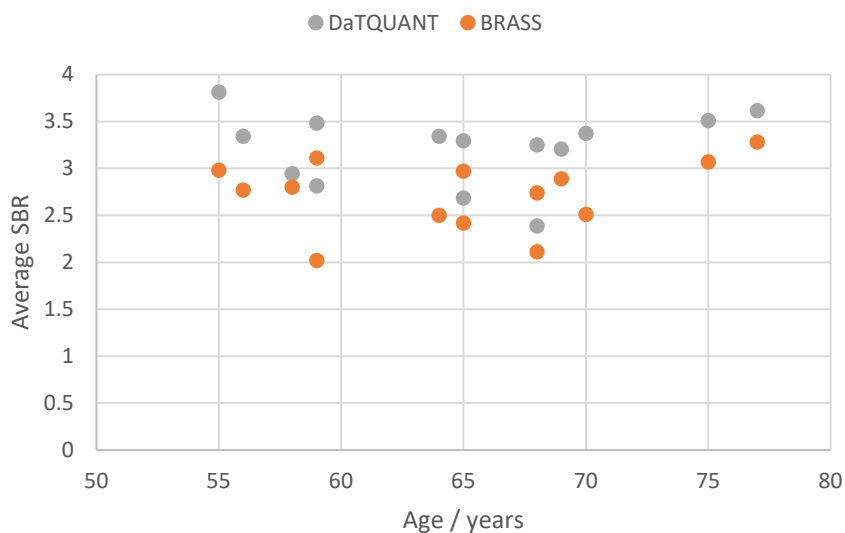


Figure 23: Direct comparison between DaTQUANT and BRASS SBR values for the same input images, indicating a systematic difference between the packages in the way SBR values are calculated.

DaTQUANT gives significantly higher SBR values than BRASS for the exact same input images. The mean difference is 0.52 (15% higher with DaTQUANT). This shows that there is a difference in the method the two packages use to produce SBR results, which is likely due to differences in image registration, differences in the volume of interest template applied to the scans or a combination of factors. However, this difference in SBR does not explain the difference in Z scores seen above because each normal database has been reconstructed and processed in the same way as the input images. The difference in Z score is more likely linked to differences in the normal databases, which could be

caused by population differences, camera differences, lack of camera correction, or differences in age correction. These potential factors are discussed in Appendix B.

4.1.4 Conclusions

In this section two commercial software packages were explored using phantom and normal patient data. Initial work discovered an error in camera correction for BRASS which was corrected for all subsequent analysis. We found a difference in Z scores between the two packages – the impact of this difference will be considered in the next sections on research subjects with well-characterised MCI and dementia diagnoses.

Based on this initial validation and comparison, BRASS is the package of choice. BRASS gives Z scores close to zero for normal patients as expected and has several technical advantages over DaTQUANT. The practical effect of the differences will be explored further in the following sections.

4.2 Comparison between BRASS and DaTQUANT in a large MCI research cohort

4.2.1 Introduction

The previous section showed some theoretical and practical differences between BRASS and DaTQUANT for FP-CIT quantification. In this section we aim to compare DaTQUANT and BRASS age corrected Z scores in a large MCI dataset from the LewyPro study consisting of approximately 50% normal and 50% abnormal scans. We will compare the sensitivity and specificity of each quantification method with consensus binary visual assessment (normal/abnormal), using follow up consensus clinical diagnoses as the gold standard.

4.2.2 Methods

Seventy-seven FP-CIT scans of subjects with mild cognitive impairment were taken from the Newcastle LewyPro study (see Chapter 2 for LewyPro study details). The raw scan data was processed and analysed with DaTQUANT and BRASS using the methods introduced in the previous section. The average Z score results were calculated for the whole striatum and compared between DaTQUANT and BRASS. The caudate and putamen results were not compared as the boundary between these structures is rather arbitrary on FP-CIT images and there are likely to be differences between the two VOI templates. The overall SBR was thought to be more meaningful, particularly as it is known that in DLB loss of signal is more uniform than in PD.

Clear abnormal or normal cases could have quite different Z scores but still be interpreted in the same way, but borderline cases with similar absolute numerical differences could be reported differently. For example, if DaTQUANT gave a Z score of +2 and BRASS +1, both would be reported as normal despite the large difference, but if the scores were -1 and -2 respectively this would make a difference. The Z scores were therefore compared for individual patients to assess the clinical impact of any difference between DaTQUANT and BRASS Z scores. Scans with both whole striatum Z scores above -0.5 on both programs are regarded as clearly normal, at least one striatum below -2 on both programs as clearly abnormal and all scans in between as low normal/borderline.

Consensus binary visual assessment had been carried out for the LewyPro study in 2014 by a panel of five raters, using images reconstructed with filtered back projection without attenuation or scatter correction. The consensus visual results were compared to BRASS and DaTQUANT results based on whole striatum Z scores. For each method the sensitivity (percentage of LB cases with abnormal scans), specificity (percentage of non-LB cases with normal scans) and accuracy (percentage of results in keeping with clinical diagnosis) were calculated. This was done with the three year follow-up diagnoses of the subjects taken as the gold standard. Note that at the time this analysis was done (2017) most

subjects still had an MCI rather than dementia diagnosis, with eighteen having gone on to develop dementia.

4.2.3 Results and discussion

One subject with very low uptake was excluded because the study did not spatially register to the template correctly on DaTQUANT. Two subjects were later excluded from the study as they did not fulfil the requirement for MCI. The Z scores for the remaining 74 scans are on average higher for DaTQUANT than BRASS, by 0.43 SD (paired t-test $p < 0.001$). However, the results are very variable, with the DaTQUANT result ranging from 0.77 SD lower to 1.93 SD higher than BRASS (Figure 24). Although slightly smaller, the difference between BRASS and DaTQUANT is probably in keeping with the results seen for the normal patient scans above, given that lower Z scores are likely to show a smaller absolute difference and the fact that LewyPro subjects are older on average than the RVI normal patients so may have lower SBRs even if normal.

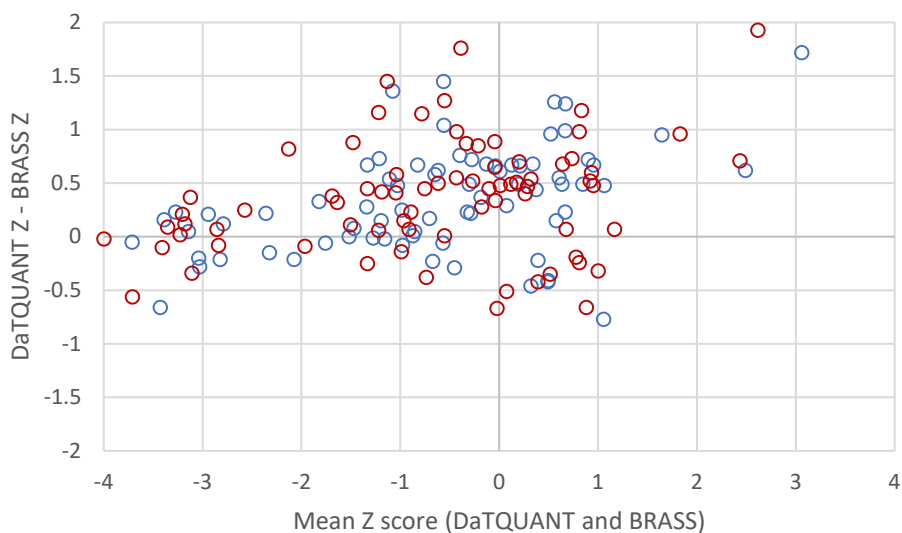


Figure 24: Bland-Altman plot of differences in Z score between BRASS and DaTQUANT for subjects with mild cognitive impairment from the LewyPro study. Right striatum shown in blue; left striatum red.

Around half of the subject scans (54%) were clear cut normal or abnormal (26/74 clear normal with Z scores over -0.5 and 14/74 clear abnormal with one or both Z scores lower than -2 on both BRASS and DaTQUANT). The other 34 subjects had low end of normal to possibly abnormal Z scores between -2 and -0.5. Of these “borderline” cases, 13 have similar scores on both BRASS and DaTQUANT (both striata within 0.5 SD), 21 are lower with BRASS by at least 0.5 SD on at least one striatum. None are more than 0.5 SD lower with DaTQUANT. However, there were only four cases where the Z scores were greater than -2 for both striata with one package and less than -2 for at least one striatum with

the other package –i.e. different clinical interpretation with a -2SD cut-off. In one of these four cases the quantification results were very similar but the DaTQUANT result was just below -2 and BRASS just above (-2.18 vs -1.97). In the other three cases the BRASS results were significantly lower than DaTQUANT. If a Z score of -2 was being used as the cut-off, as is often the case, the outcome would have been the same for the majority (95%) of the scans.

The sensitivity of both BRASS and DaTQUANT is lower than for consensus visual reporting, but specificity is higher (Table 6). Overall accuracy is lower, but it is important to note that quantification is used as an aid to visual reporting rather than as a stand-alone diagnostic tool.

Table 6: Sensitivity and specificity for visual, DaTQUANT and BRASS analysis for subjects with mild cognitive impairment.

	Visual reporting	DaTQUANT	BRASS
Sensitivity	52%	31%	31%
Specificity	88%	100%	92%
Accuracy	64%	55%	52%

4.2.4 Conclusion

In this dataset of 74 scans in MCI subjects the average Z score was 0.43 SD higher with DaTQUANT than BRASS. However, DaTQUANT and BRASS have similar low sensitivities and high specificities for diagnosis of early LB disease, with respect to the consensus clinical MCI diagnosis. Visual reporting with filtered back projection was more sensitive but less specific. However, it should be noted that most of the LewyPro subjects do not yet have validated dementia diagnoses, so it is not possible to use these results to compare the diagnostic performance of BRASS and DaTQUANT in early disease, only to provide further evidence of a difference between the packages. A comparison using subjects with known dementia diagnosis is required for this and this will be carried out in the next section.

4.3 Comparison between DaTQUANT and BRASS in healthy ageing and dementia

4.3.1 Introduction

In the previous section we confirmed that DaTQUANT gave higher Z scores than BRASS in a large cohort of MCI subjects. In this section we compare the performance of the two packages in dementia using research subjects with firm consensus dementia diagnoses and introduce controls with normal cognition from the SUPeR study to compare with the patient data analysed in section 4.1.

Objectives

- To compare DaTQUANT and BRASS age corrected Z scores in subjects from the MIDAS study with firm consensus dementia diagnoses
- To compare quantification with consensus binary visual assessment (normal/abnormal) via the sensitivity and specificity with respect to clinical diagnosis
- To explore whether a fixed SBR threshold could be more accurate than age-corrected Z score for DLB diagnosis
- To process a subset of healthy control subjects from the SUPeR study and compare with the normal RVI patient results in section 4.1

4.3.2 Methods

Patient selection

Patients with confirmed dementia diagnoses of probable DLB or probable AD were taken from the MIDAS and LewyPro studies. Although LewyPro is an MCI study, patients who had converted to dementia (based on August 2017 records) were included if they did not continue into the SUPeR study. This is to avoid overlap with future analysis of SUPeR subjects. The six MIDAS DLB cases recruited specifically because they had normal FP-CIT scans were of course excluded and a further three MIDAS patients were excluded – one because of uncertain diagnosis (possible DLB), one because of lack of raw image data due to being scanned in another hospital and one because a non-Siemens camera was used. One LewyPro subject was excluded because of image registration problems due to low uptake and two because of possible rather than probable DLB diagnosis. The first sixteen controls recruited to the SUPeR study were analysed. These older adults with normal cognition had undergone the same rigorous clinical and psychiatric tests as the SUPeR MCI subjects, with normal results – see section 2.3.3 for further details. Of these 16 controls one could not be used because of low signal in the brain background region in the occipital lobe, possibly due to atrophy. The final dataset of DLB, AD and controls is summarised in Table 7.

Table 7: MIDAS, LewyPro (subset converted to dementia) and SUPeRB subjects included in comparison

	DLB	AD	Controls
MIDAS	16	17	0
LewyPro	13	5	0
SUPeRB	0	0	15
Totals	29	22	15

The SBRs and Z scores were calculated for each dementia subject with DaTQUANT and BRASS, using the methods detailed above. Assigning a Z score of -2 or below as abnormal, the sensitivity, specificity and accuracy of BRASS and DaTQUANT for diagnosing DLB in patients with dementia was calculated, using the consensus dementia diagnosis as the gold standard. The control data was also analysed with DaTQUANT and BRASS and used to set local normal SBR cut-offs of 2 standard deviations below the control mean. ROC analysis was carried out using SPSS to give the areas under the curve for BRASS and DaTQUANT and the sensitivity and specificity for DLB diagnosis for the optimal SBR cut-offs. Controls were not included in the ROC analysis.

Comparison of local normal patients, SUPeRB study controls and PPMI database controls

The 15 SUPeRB control subject SBR results were compared to the local normal patients and to the PPMI healthy control database (subjects over 60 only). Unpaired t-tests were used to test for difference. Results were also compared with the BRASS normal database indirectly via Z scores as the BRASS normal database SBRs are not provided to the user. The SBRs were not age corrected so a comparison between the ages of the different control groups was made.

4.3.3 Results

The sensitivity and specificity for DLB diagnosis based on Z scores (designating scores less than -2 as abnormal) for BRASS and DaTQUANT is given in Table 8. The sensitivity and specificity based on control SBR cut-offs of two standard deviations below the mean of the SUPeRB controls is given in Table 9. The results show similar low sensitivity and high specificity for BRASS and DaTQUANT for both the Z score (normal database) and local SUPeRB control cut-offs.

Table 8: FP-CIT sensitivity and specificity for DLB diagnosis in dementia with BRASS and DaTQUANT based on Z score results

	BRASS		DaTQUANT	
	Right	Left	Right	Left
Cut-off (Z score)	-2	-2	-2	-2
True positives	17	17	16	16
False positives	12	12	13	13
True negatives	22	21	22	22
False negatives	0	1	0	0
Sensitivity	59%	59%	55%	55%
Specificity	100%	95%	100%	100%
Accuracy	76%	75%	75%	75%

Table 9: FP-CIT sensitivity and specificity for DLB diagnosis with BRASS and DaTQUANT based on SUPeRb control data

	BRASS		DaTQUANT	
	Right	Left	Right	Left
Cut-off (mean-2SD)	2.2	2.1	1.4	1.3
True positives	18	17	18	16
False positives	11	12	11	13
True negatives	21	21	21	20
False negatives	1	1	1	2
Sensitivity	62%	59%	62%	55%
Specificity	95%	95%	95%	91%
Accuracy	76%	75%	76%	71%

ROC analysis of the dementia cases (Figure 25; Table 10) shows the highest area under the curve is given for the right striatum analysed with DaTQUANT – although the confidence intervals are very wide so this is not statistically significant. The optimal SBR cut-off for the right striatum analysed with DaTQUANT is 1.46 giving a sensitivity of 73% and specificity of 91%. This is an improvement compared to both the Z score results and the results using a cut-off based on the control data. The right striatum results are more accurate than the left striatum with both programs, which is probably a chance result for this dataset rather than because of a systematic error in the way the programs calculate SBR.

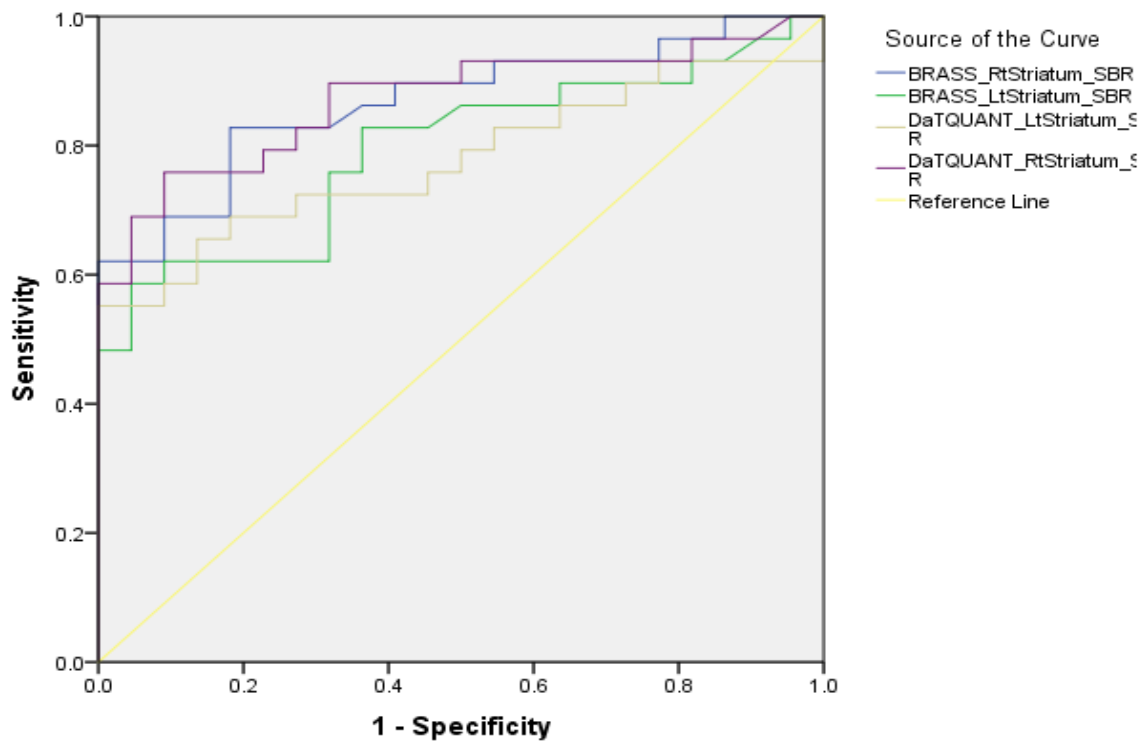


Figure 25: ROC curves for BRASS and DaTQUANT showing higher area under the curve for the right striatum

Table 10: Area under the ROC curves - accuracy comparison

Test Result Variable(s)	Area	Std. Error ^a	Asymptotic 95% Confidence Interval		Cut-off	Sens.	Spec.
			Lower Bound	Upper Bound			
BRASS_RtStriatum_SBR	0.868	0.050	0.769	0.966	2.51	83%	82%
BRASS_LtStriatum_SBR	0.793	0.063	0.670	0.916	2.70	83%	64%
DaTQUANT_RtStriatum_SBR	0.872	0.050	0.774	0.971	1.46	73%	91%
DaTQUANT_LtStriatum_SBR	0.777	0.066	0.648	0.907	1.15	55%	100%

For this dataset, slightly higher accuracy (although not statistically significant) was seen using a fixed SBR cut-off derived from ROC analysis, which was lower than the cut-off found from the control data (mean-2SD). However, that cut-off would need to be applied and tested in an independent cohort of dementia subjects before we could conclude that a cut-off based on ROC analysis of dementia subjects has higher diagnostic accuracy than a cut-off based on control data. Furthermore, the results based on control data (either local or database) all have very high specificity, higher than that with the ROC cut-off. This is more important than high sensitivity particularly given that it is known that DLB patients can have normal looking FP-CIT scans [48].

Comparison of local normal patients and SUPeRB controls

There was a large difference in the mean age of the local normal patients (65 years) and the SUPeRB controls (73 years, Table 11) with the latter being 8 years older on average ($p < 0.001$). This is because around half the local normal patients are between 55 and 60 whereas the SUPeRB controls are all over 60. In hindsight, it could be argued that patients under 60 should not have been included in this comparison, but this would have reduced local normal patient numbers significantly. SBRs for the SUPeRB controls are lower than for the local normal patients with DaTQUANT ($p < 0.001$, Table 11). Z scores are also significantly lower for the SUPeRB controls ($p < 0.001$).

Table 11: Mean whole striatum SBRs with DaTQUANT, Z score with respect to PPMI database (if applicable) and mean age for the PPMI controls, SUPeRB controls and local normal patients.

	PPMI controls over 60	SUPeRB controls	Local normal patients
n	40	15	14
Mean SBR	1.8	1.7	2.1
z score	n/a	-0.3	0.6
Mean age	67.4	72.7	64.8

With BRASS, SBRs and age corrected Z scores are again significantly lower for the SUPeRB controls than the local normal patients (Table 12). Given that there is no downward trend in SBR with age in either group it is likely that age correction is not necessary in older adults and is actually having a negative effect on the results. However, it would not make sense to use either the BRASS or the DaTQUANT normal database without age correction as it covers such a wide age range, and indeed this is not even an option with DaTQUANT. Age correction is discussed in more detail in section 9.1.

Table 12: Mean whole striatum SBRs with BRASS, Z score with respect to ENC-DAT database and mean age for SUPeRB controls and local normal patients. Note that ENC-DAT database data is not available.

	SUPeRB controls	Local normal patients	p-value (t-test)
n	15	14	n/a
Mean SBR	2.67	3.15	<0.001
z score	-0.77	-0.01	<0.001
Mean age	72.7	64.8	<0.001

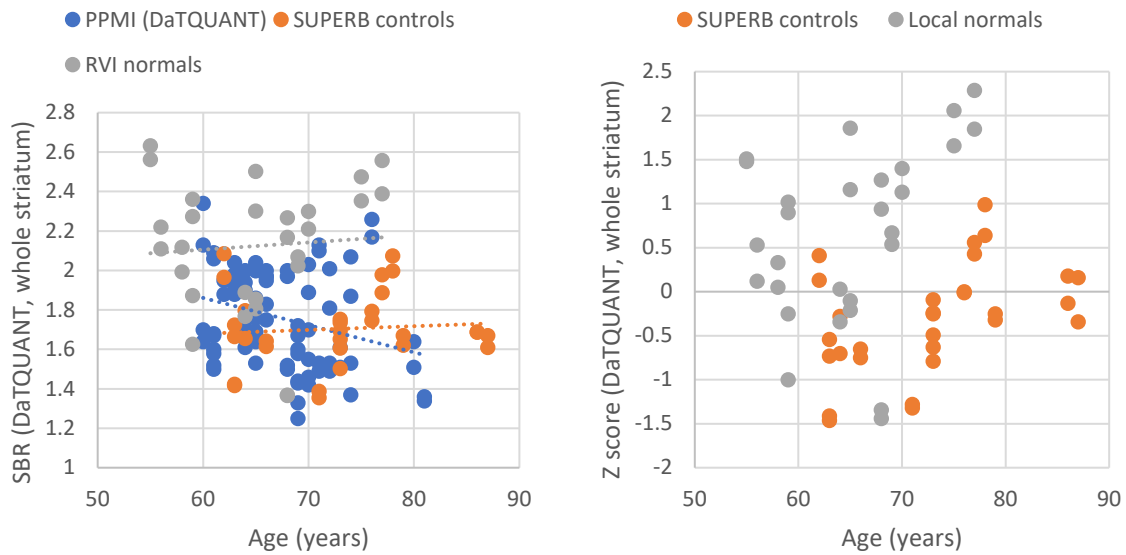


Figure 26: PPMI controls, SUPERB controls and local normal patient SBRs (left) and Z scores for SUPERB and local normal (right) calculated with DaTQUANT and plotted against age.

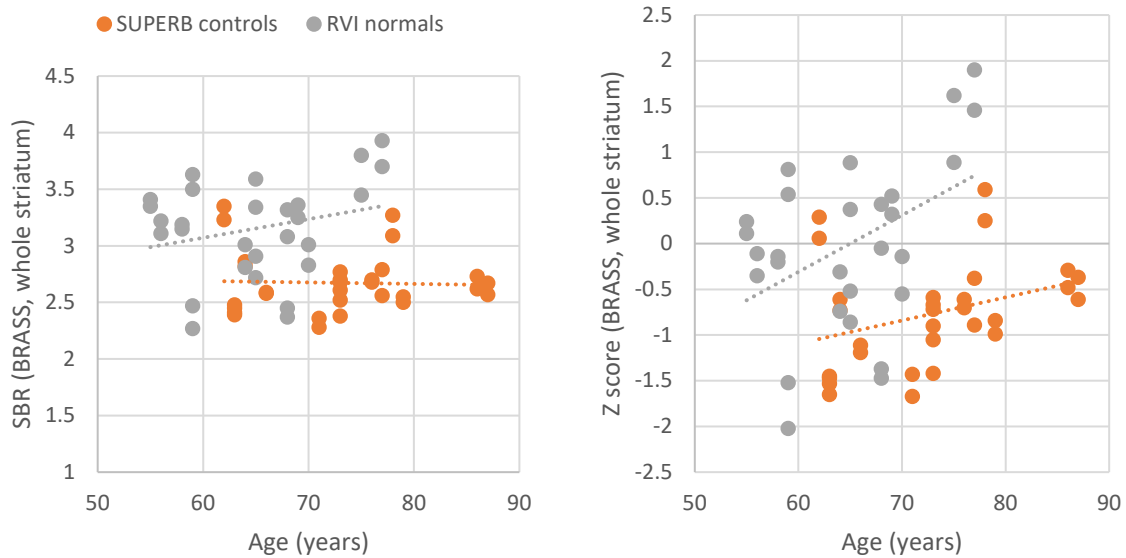


Figure 27: SUPERB and local normal patient SBRs with camera correction (left) and age corrected Z scores (right) calculated with BRASS and plotted against age.

The SBRs of local normal patients and SUPERB controls differed even though both groups were scanned in the same department. The results are higher (i.e. “more normal”) for the patients than the SUPERB controls, so inclusion of patients with Lewy body disease is unlikely to be the cause. It is likely that a bias was introduced in the local normal group during the audit by selecting patients based on normal scan report rather than clinical diagnosis, potentially excluding subtle changes that are within normal limits.

4.3.4 Conclusion

The accuracy of BRASS and DaTQUANT is very similar. In this dataset it did not make a significant difference whether interpretation of the results is based on Z scores relative to large multicentre healthy control databases (age-corrected) or relative to local SUPeRB control data (without age correction). There is no downward trend in SBR with age in either the normal patient dataset (aged over 55 years) or the SUPeRB control dataset (over 60 years). In this dataset a lower cut-off based on ROC analysis gave higher overall accuracy results, but lower specificity.

The local normal patients have higher SBRs and age-corrected Z scores than the SUPeRB controls with both BRASS and DaTQUANT. A good option for future studies, based on this evidence, would be to use BRASS SBRs with a local cut-off based on age-matched controls. The full set of SUPeRB control scans could be used to set this cut-off (not all had been recruited at the time of writing so this was not a practical option for the study itself).

4.4 Effects of reorientation and pixel size on quantification with BRASS

The purpose of this piece of work was to assess whether reorientation, image cropping, zoom and pixel size have a significant effect on BRASS quantification results. These factors were investigated because the use of these parameters at NuTH differs from the recommendations for use with BRASS.

FP-CIT scans from LewyPro subjects were used to show that reorientation and post-processing zoom applied during reconstruction both have a small statistically significant effect on BRASS quantification results, but that these are both clinically insignificant and furthermore act in opposite directions. This work is detailed in Appendix C.

We used a striatal phantom to investigate the effect of changing pixel size, via the application of hardware zoom. We also acquired a series of repeat scans with no changes in zoom to assess random variation. Full details are given in Appendix D. The data showed that changing the zoom has no more effect on BRASS quantification results than simply repeating the scan with the phantom repositioned, which suggests that the variation is random and could be due to slight changes in the input image. This work showed that we could continue to use the same pixel sizes for clinical scans and for research scans to be processed with BRASS.

Chapter 5: Development of imaging methods for research studies: cardiac MIBG

This chapter deals with establishing methods for planar cardiac ^{123}I -MIBG imaging to use in the MIDAS and SUPeR research studies.

In section 5.1 a systematic review of imaging protocols used in studies applying cardiac MIBG imaging to dementia diagnosis is presented, including the administered activities, collimator and imaging time point. A cardiac MIBG imaging protocol was set up for research studies at Newcastle, based on guidelines and evidence from this literature review.

In section 5.2 a method for calibrating cardiac uptake results with a phantom to correct for imaging hardware factors is presented. This method was applied to the scanners to be used in our research studies so that normal threshold heart to mediastinum ratio cut-off values from Japan could be applied to our data.

In section 5.3 the administered activity used in our research studies is explored in more detail. Phantom studies with different imaging times and retrospective evaluation of counting statistics from research studies are used to show that the combination of 111 MBq activity and 10 minutes imaging time used in our research protocol gives adequate image quality.

5.1 Cardiac ¹²³I-MIBG planar imaging protocol review

5.1.1 Introduction

Very few cardiac MIBG scans were carried out at NuTH for the diagnosis of DLB prior to the start of our research studies since the FP-CIT scan, recommended in NICE dementia guidelines, was usually requested. We were aware that the majority of clinical and research use of cardiac MIBG for this purpose was in Japan, with some research studies from Europe published, and that the activities administered, collimators and time points for imaging varied. We were not aware of any centres in the UK carrying out cardiac MIBG to distinguish between AD and DLB. We therefore reviewed the imaging protocols used in the literature in order to implement an evidence-based protocol for research studies. As well as determining the best way to acquire the images, an additional aim of the review was to determine an HMR cut-off value to use for interpreting results.

5.1.2 Methods

A PubMed search was carried out applying the same search terms used by Treglia et al. in their 2012 review, which covered articles published before May 2010. (Search ("mibg" OR "metaiodobenzylguanidine")) AND ("lewy" OR "dementia")). This brought up 132 results on PubMed on 20/4/15. The abstracts of each article were reviewed and the 16 original articles evaluating the use of cardiac MIBG for differentiating between DLB and AD or other dementia were selected and retrieved. These are listed in Appendix E. Each study is listed alongside the country or origin, the number of patients included, the activity administered and acquisition duration of the planar image, the delay between injection and imaging, the collimators used (MEGP – medium energy general purpose; LEHR - low energy high resolution; LEGP - low energy general purpose) and the heart to mediastinum ratio (HMR) applied as a cut off value for distinguishing between DLB and AD. The cut-off values were determined by ROC analysis in most cases. The sensitivity and specificity were recorded if part of the study design.

5.1.3 Results and discussion

Although most studies published on the use of cardiac MIBG in dementia diagnosis were carried out in Japan (10/16), there have been some recent studies in Europe, including a moderate sized retrospective patient study conducted in France in 2015 with 45 patients with Lewy body disease and 19 without that reported a sensitivity of 82% and specificity of 89% [56]. Most studies are small in sample size (<30 DLB patients), retrospective in nature with clinical patient scans used, and report very high sensitivity and specificity. The Japanese studies all administered a dose of 111 MBq (except Oide et al. who administered 148 MBq [183]). In the European studies the administered activity range was broad: 74 MBq in Belgium [59], 111 MBq in Italy [68, 184], a 148-296 MBq weight based protocol in France and 370 MBq in Spain.

The EANM Cardiovascular Committee guidelines state typical activities range between 111 MBq and 370 MBq and do not recommend a specific activity as “solid evidence documenting superior results with certain activities is not available” [165]. In the USA the FDA approved dose is 370 MBq, significantly higher than in Japan [65, 185]. The UK ARSAC guidance also suggests 370 MBq for sympathetic innervation imaging of the heart (procedure code 123I-117-136). Although these levels sound precise, they originate from activity measured in millicuries, where 1 mCi = 37 MBq. The average effective dose associated with the administration of 111 MBq (3 mCi) of MIBG is 1.4 mSv, with 185 MBq (5 mCi) and 370 MBq (10 mCi) of course giving higher effective doses of 2.4 mSv and 4.8 mSv. These higher doses would seem unnecessary if the same information can be obtained with 111 MBq.

It is of interest to note that even though the Japanese centres use a lower activity than the standard NuTH clinical protocol (370 MBq), many also use a shorter imaging duration of five or even two minutes. Even taking the greater sensitivity of the low energy collimators used in many of these studies into account, the count density will be reduced by several fold compared to the NuTH protocol. It therefore seems that the higher activity of 370 MBq is unlikely to be necessary and the decision was taken to move to an activity of 111 MBq for our research protocols, keeping a 10 minute image duration. The validity of this approach is considered further in section 5.3, using phantom and MIDAS clinical data to assess counting statistics.

The vast majority of published cardiac MIBG studies in dementia have used low energy high resolution collimators, including the European studies conducted after the EANM guidance on standardisation to medium energy were published [165]. However, the most recent article by Yoshita et al., a large multicentre study carried out in Japan, presents a cut off value determined for MEGP collimators [53]. Not all the centres in the study used MEGP collimators, but all camera and collimator combinations were calibrated with a phantom to enable HMR results for non-MEGP collimators to be scaled to the equivalent of MEGP values. MEGP was chosen as the reference collimator because calibration work

carried out by Nakajima et al. showed that this collimator gave heart to mediastinum ratios in phantom studies that were both higher than those obtained with low energy collimators and closest to the true values [166, 186]. This was later confirmed for our centre using the phantom method, which is outlined in the next section. As discussed in Chapter 3, collimators designed for medium energy gamma ray imaging give better image quality (and higher heart to mediastinum ratios) than those designed for low energies because of the reduction in septal penetration of the higher energy ^{123}I photons, which reduces image contrast. The main argument against using them in the MIDAS study is that most previous studies have used LEHR, and we would want to apply HMRs derived in those studies. However all camera and collimator combinations are expected to give slightly different ratios, hence the need for the large calibration exercise in Japan, so it would not necessarily be valid to apply a value from another study. The more accurate HMR results obtained with medium energy collimators may lead to more accurate diagnosis, and the wider HMR range could give finer grained distinctions, which may allow subtle changes in early disease to be detected more easily than with low energy collimators. We therefore chose to continue using MELP collimators for cardiac MIBG research studies, despite the majority of publications using LEHR.

A wide range of HMR cut-off values were reported in the literature, likely due to the different collimators used. Since the Japanese multicentre study [53] is the largest body of data, and reports a value for medium energy collimators, we will use the value of 2.1 obtained via ROC analysis as our initial cut-off, adjusted via phantom calibration if necessary (see next section).

Since this review was conducted, publications of note include a prospective Italian multicentre study comparing cardiac MIBG and FP-CIT in dementia diagnosis [67] and a three year follow up paper reporting on around half the patients from the Japanese multicentre study [62]. The Italian study, by Tiraboschi et al. included 30 DLB patients, 16 AD and 13 FTD from five Italian centres and reported 93% sensitivity and 100% specificity [67]. At first glance this would suggest that the multicentre aspect does not reduce accuracy, but in fact the patients were scanned in only two nuclear medicine centres, both with the same model gamma camera and collimators (model not stated). The follow up to the Japanese multicentre study [62] assessed 65 of the original 133 patients and found a higher sensitivity and specificity in this subset (sensitivity 77%, specificity 97%; 69% and 89% respectively for the original study). However, there was still considerably more overlap between the HMRs in the AD and DLB groups than was seen in earlier single centre Japanese studies reporting almost perfect accuracy. It is not clear how much of this is due to the inclusion of different centres and how much to differences in inclusion criteria and sample selection. Importantly, in both the original and follow up study, the early image obtained at 20-30 minutes post injection gave similar results to the delayed image at 3-4 hours post injection (area under the ROC curve 0.90 for early and 0.92 for delayed). We therefore included

an early image at 20 minutes post-injection to our research protocol to allow comparison with the standard 3-4 hour results at the end of the study. This is explored further in section 7.3.

5.1.4 Conclusion

We determined the optimum cardiac MIBG imaging protocol for the MIDAS and SUPeR research studies based on a literature review. Key parameters of the protocol are given in Table 13.

Table 13: Settings used in the research cardiac MIBG protocol at Newcastle upon Tyne Hospitals

Factor	Specification in NuTH protocol
Thyroid blockade	170mg potassium iodate, 1 hour before injection and 8 hours later
Administered activity	111 MBq \pm 10%
Interval post injection	20 minutes (early) and 3-4 hours (delayed)
Mode	Planar; anterior chest
Collimator	MELP
Patient positioning	Feet first supine, detector as close to chest as possible
Energy Window	20% centred on 159 keV
Matrix size	128 x 128
Zoom	None
Acquisition duration	10 minutes
HMR cut-off	2.1 (taken from Japanese multicentre study)

In the next section we outline the method for calibrating our camera using the same method as the Japanese multicentre study [53], so that we can either apply the 2.10 cut-off ratio obtained in said study or correct it for our camera if necessary.

5.2 Phantom calibration for HMR normalisation between centres

5.2.1 Introduction

Calibration of HMRs between gamma cameras using a phantom has been shown to be feasible [166] and was implemented in a Japanese multicentre study [53]. Before starting cardiac ^{123}I -MIBG imaging at our centre we calibrated our gamma cameras, with help from colleagues in Japan and the Netherlands. We used the method developed by Nakajima *et al.* [166, 168, 186] which allowed us to check whether it is valid to use the HMR cut-off of 2.1 from the multicentre DLB study by Yoshita *et al.* [53, 62] for our research studies. The calibration results for our Siemens gamma cameras with MELP collimators were later included in the 2017 European multicentre calibration paper by Verschure *et al.* [167] alongside 26 centres in the Netherlands, Belgium and Austria.

The calibration method uses a phantom with heart, liver, lung and background sections filled with ^{123}I solution [166]. The relative thickness of the different cavities determines their relative activities on planar imaging (Figure 28). The phantom used in the European project was designed to give a true heart to mediastinum ratio of 2.60 when scanned in one orientation and 3.60 when turned over and scanned from the other side. These known values are compared with the values measured on planar images in order to determine a calibration coefficient. The calibration coefficient is simply the slope of the linear fit between the measured and true HMRs. An additional data point of (1,1) is included in the fit since the measured ratio would be 1 if there were no difference between the heart and mediastinum counts. However, the method is complicated slightly by the decision to standardise to a medium energy general purpose collimator, rather than to the true HMR values. This was taken so that centres using MEGP collimators, commonly used in Japan, would not need to apply a correction factor in order to use the published multi-centre threshold value of 2.1 [53]. At NuTH we use Siemens MELP collimators, which have similar properties to MEGP.

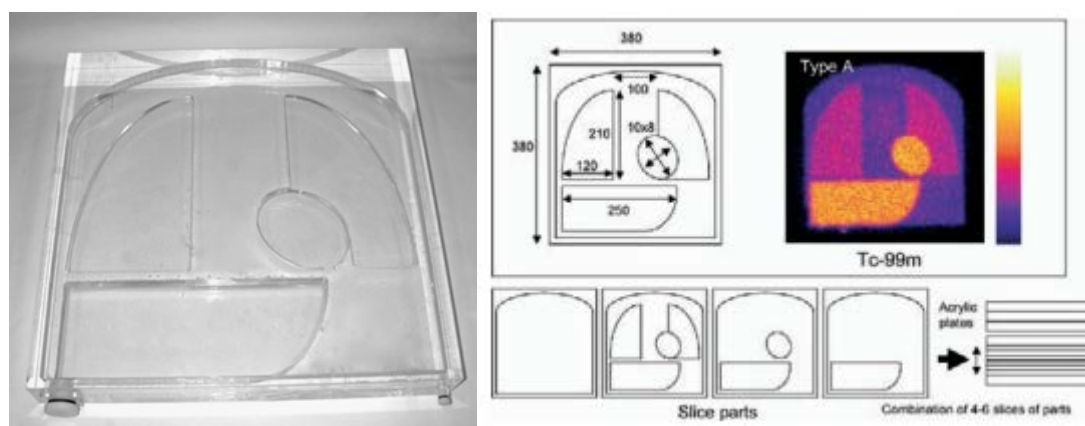


Figure 28: Photograph of planar calibration phantom used in Japan and diagram of the structure [187]. The cavities are filled with uniform ^{123}I solution through a single filling port and the depth of each compartment gives rise to the known ratios. The phantom used in the European study is a light-weight version using water-filled slices instead of solid PMMA plates.

5.2.2 Methods

Phantom calibration scans were acquired on the Siemens Intevo and Siemens Symbia gamma cameras at the Royal Victoria Infirmary, Newcastle upon Tyne Hospitals. Example phantom images and HMR region placement are shown in Figure 29.

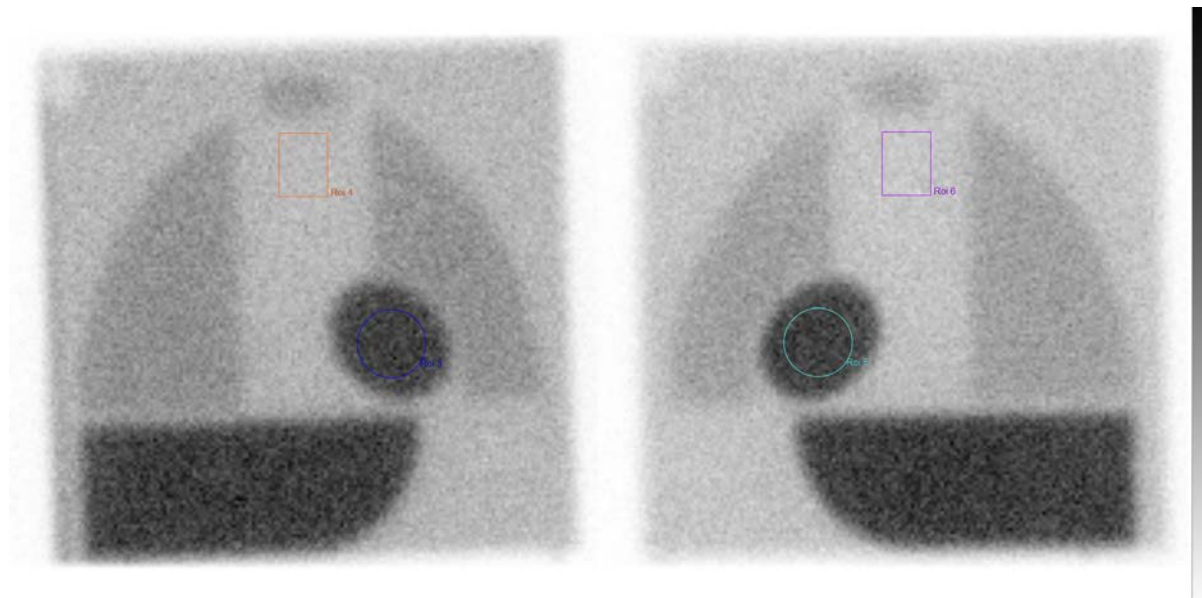


Figure 29: Phantom images for the Siemens Intevo, showing the 2.5 HMR side (left) and 3.6 HMR side (right). The ROI placement for HMR measurement is shown.

5.2.3 Results

The measured HMRs for each detector with medium energy low penetration (MELP) collimators are given in Table 14. Plots used for calibration are shown in Figure 30 and the calibration factor results in Table 15. The calibration was independently verified by Prof. Nakajima from Kanazawa University, Japan.

Table 14: True and measured HMR values for phantom scanned on Siemens Symbia and Siemens Intevo gamma cameras

	Symbia detector 1	Symbia detector 2	Intevo detector 1	Intevo detector 2
True HMR 2.50	2.69	2.68	2.70	2.72
True HMR 3.60	3.32	3.31	3.35	3.37

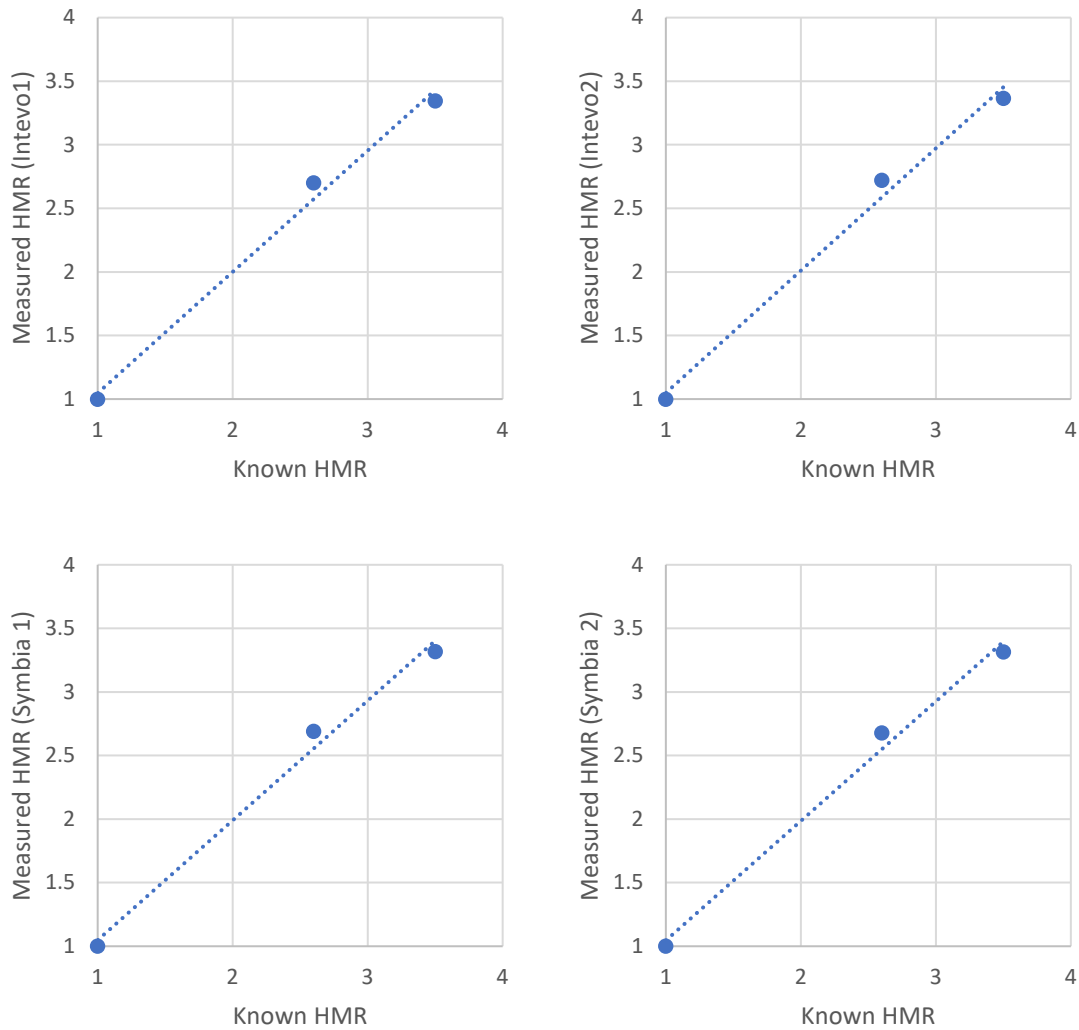


Figure 30: Plots used to determine calibration coefficients between measured and true HMRs.

Table 15: Calibration coefficients determined from phantom calibration for our cameras compared to the Japanese and European average values for MELP collimators.

	Newcastle	MELP average (Europe)	MELP average (Japan)
Intevo	0.996	0.895	0.950
Symbia	0.985		

The results indicate that our scanners with MELP collimators give values close to the true HMR, however, in Japan data was standardised to MEGP collimators, which give HMR values that are slightly lower. To compare our results to Japan an adjustment is therefore required. The Yoshita multicentre study adjusted the HMRs from each centre so that data could be pooled and used with the same normal HMR cut-off. However, we chose to adjust externally derived HMR cut-off values so that they are applicable to our data. The equation provided by Verschure *et al.* can be rearranged for this purpose to give:

$$Local\ cutoff = 1 + (standardised\ cutoff - 1) \times \frac{Local\ coefficient}{Standardised\ coefficient}$$

The mean value for the calibration coefficient for cameras with MEGP collimators in Japan was found to be 0.88; this value is taken as the standardised coefficient.

If we were to adapt the cut-off of 2.1 found by Yoshita et al. [53] using the Japanese coefficient of 0.88 and local coefficient of 1.00 we would obtain a local cut-off of 2.25 for use with our Siemens Intevo:

$$\begin{aligned} Local\ cutoff &= 1 + (2.1 - 1) \times \frac{1.00}{0.88} \\ &= 1 + (1.1 \times 1.14) \\ &= 1 + (1.25) = 2.25 \end{aligned}$$

Similarly, the HMR cut-off for our Siemens Symbia is very similar at 2.24 allowing the cameras to be used interchangeably in practice. For comparison, the average LEHR collimator coefficient is 0.55 [188] so the cut-off converted for use with LEHR collimators would be lower at 1.69.

5.2.4 Discussion

The HMR values obtained with the phantom for our setup show good agreement with the true values (a slight over estimation for the 2.5 ratio and underestimation for 3.6), resulting in high calibration coefficients of 1.00 and 0.99 (to two decimal places). The two Siemens scanners to be used in the MIDAS and SUPeRB study show very close agreement with each other. Phantom calibration with respect to the Japanese MEGP standard, using published calibration coefficients 0.88 resulted in a slight increase in the Japanese cut-offs being required for use with our scanners of around 7%.

It was much later discovered, through extensive email communication with Prof. Nakajima, of Kanazawa University, Japan, that Siemens medium energy general purpose (MEGP) collimators in Japan appear to have similar properties to Siemens medium energy low penetration (MELP) collimators in Europe, and therefore similar calibration coefficients. The observed differences between our calibration factors and those in the multi-centre study were found to be within the range of observed differences with a range of phantoms used in Japan and Europe, which are assumed to be partly due to variation in phantom manufacturing. We therefore decided to apply the multicentre cut-off without adjustment. However the above calibration method is still required if a different collimator or camera combination were to be used, for example to compare results with another centre using LEHR collimators.

5.2.5 Conclusion

The similarity between the imaging setup at the Royal Victoria Infirmary and the standard setup in Japan, both of which use medium energy collimators, means we can apply HMR cut-offs from Japan directly without the need for correction. The cut-offs from the multicentre study are 2.10 for both early and delayed images [53].

5.3 Is 111 MBq an appropriate activity for cardiac MIBG imaging?

5.3.1 Introduction

All subjects in the SUPeRB study receive an intravenous injection of 111 MBq \pm 10 % of ^{123}I -MIBG, the activity set out in the research protocol, which is approved by the Sponsor (Newcastle upon Tyne Hospitals) and by the UK Administration of Radioactive Substances Advisory committee (ARSAC). The activity of 111 MBq planned for this study (and used for the MIDAS study) is in keeping with the activity used in Japan, where most published DLB research studies have been carried out [53, 166, 188, 189]. Our review of the literature in section 5.1 suggested this lower activity would be adequate, but here we wanted to test this retrospectively using data from the MIDAS study.

In theory the HMR should not be sensitive to changes in cardiac count density due to factors such as a lower dose, extravasation (tissuing) or shortened scan time. This is because it is a ratio and any increase in image count density should affect both the heart and mediastinum equally. However, the signal to noise ratio (SNR) in a region of interest depends on the total counts and so will be affected. The SNR is proportional to the square root of the total counts, as image counting statistics follow a Poisson distribution. This means increasing the administered activity or image duration by a factor of 4, for example, will only improve the SNR by a factor of 2. Noise clearly increases the uncertainty in the measured count density, and therefore uncertainty in HMR.

In this section we use an anthropomorphic chest phantom to investigate whether increasing or decreasing the acquisition time affects the appearance of cardiac MIBG images and whether the HMR is significantly affected. A phantom is useful for this as images are acquired under controlled conditions, without the influence of other confounding variables unavoidable in patient studies. The results will let us know whether it is valid to compare results obtained with higher activities with those from our studies.

We examine the mediastinal counts from patients scanned as part of the MIDAS study to see whether these give acceptable counting statistics (<2% uncertainty), and from this infer the minimum injected activity required for a 10 minute image.

5.3.2 Methods

Phantom study

Images were acquired with the RSD torso phantom (Radiology Support Devices Ltd, Long Beach, CA, USA), with fillable heart, lung, liver and background compartments (Figure 31). The phantom was filled with ^{123}I solution, with the activity concentration in each compartment intended to give the same count densities as observed in clinical studies. The phantom was scanned using the same planar acquisition

protocol as for MIDAS and SUPeR, i.e. the same gamma camera with MELP collimators, an energy window of $159 \text{ keV} \pm 10\%$, matrix size of 128×128 and no zoom.

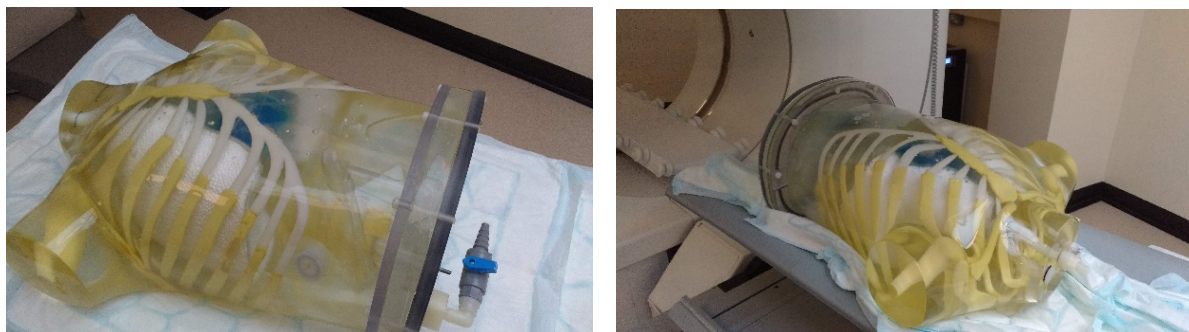


Figure 31: anthropomorphic torso phantom, with heart insert filled with ^{123}I solution coloured with blue food dye.

Further scans with twice, half and quarter the acquisition time were taken, without repositioning the phantom. These four images were then summed in various combinations to give images with a wide range of count densities, equivalent to 10 minute clinical images with injected activities between 28 MBq and 333 MBq, or alternatively 111 MBq with acquisition times between 2.5 and 30 minutes. Each image was used only once in each combination.

A freehand cardiac region of interest and rectangular mediastinum region were defined (Figure 32), and the same regions applied to all phantom images. The cardiac and mediastinal mean count densities were recorded and the heart to mediastinum ratios calculated. The percentage difference between the HMR was calculated relative to the standard phantom image, with counts equivalent to a patient administered 111 MBq and scanned for 10 minutes at 3.5 hours after injection.

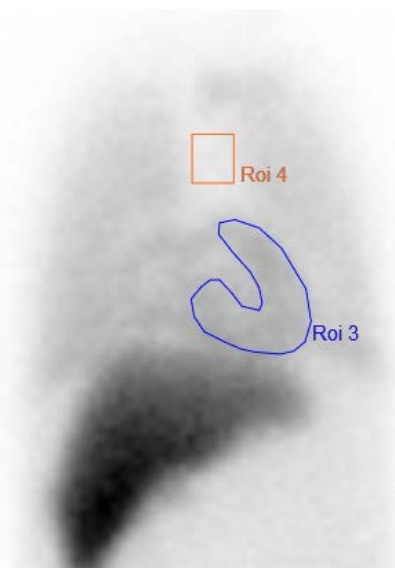


Figure 32: The standard count phantom image, with regions of interest.

Review of MIDAS patient HMR uncertainty

The mediastinal region of interest data used to calculate HMR values for MIDAS patients for our previous publications [83, 190] was used to calculate the mean number of counts in the 48 pixel region, and the 25th percentile. These patients were all administered 111 MBq \pm 10% and scanned for 10 minutes 3-4 hours after injection. The uncertainty associated with the measurement of mediastinum count density was calculated from SNR, expressed as a percentage:

$$Uncertainty (\%) = \frac{\sqrt{n}}{n} \times 100$$

where n is the total number of counts in the mediastinal region of interest (ROI). The uncertainty associated with the cardiac ROI was calculated twice: once assuming the same count density as the mediastinum (i.e. no cardiac uptake, HMR = 1) and once assuming twice the count density (HMR = 2). The larger number of pixels in the 6 cm diameter circular cardiac region compared to the 4x3 cm mediastinum region (113 vs 48) was taken into account.

The typical mediastinum and cardiac uncertainties were combined in quadrature to give an estimate of the percentage uncertainty for HMR, for cases of both no cardiac uptake and borderline-normal uptake. The minimum administered activity required to give HMR estimates of within 2% uncertainty was calculated for the different scenarios.

5.3.3 Results

Phantom study

The nine resultant images of increasing count density are shown in Figure 33. The increase in image quality due to reduced noise is apparent, particularly for the images of less than 111 MBq equivalent counts. The count densities in the heart and mediastinum regions and the HMRs for each of the images are given in Table 16. The HMR for the standard 10 minute / 111 MBq image was 2.77.

HMRs for the other images ranged between 2.79 and 2.87, which corresponds to an increase in HMR of between 0.4% and 3.4% relative to the standard image. There was no relationship between count density and HMR.

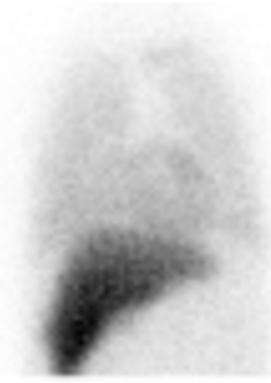








<p>2.5 min / 28 MBq</p> 	<p>5 min / 56 MBq</p> 	<p>7.5 min / 83 MBq</p> 
<p>Ref: 10 min / 111 MBq</p> 	<p>12.5 min / 139 MBq</p> 	<p>15 min / 167 MBq</p> 
<p>20 min / 222 MBq</p> 	<p>25 min / 278 MBq</p> 	<p>30 min / 333 MBq</p> 

Figure 33: Phantom images acquired with count densities equivalent to ten minutes scans with between 28 and 333 MBq administered activity or equivalent to acquisition times between 2.5 and 30 minutes with 111 MBq administered activity.

Table 16: HMRs for the nine phantom images above.

Equivalent image duration	Equivalent activity	Relative count density	Heart mean count density	Mediastinum mean count density	HMR	% difference relative to 10 mins / 111 MBq
2.5 min	28 MBq	0.25	99.4	35.0	2.84	2.2%
5 min	56 MBq	0.5	197.4	70.9	2.79	0.4%
7.5 min	83 MBq	0.75	296.8	105.9	2.80	1.0%
10 min	111 MBq	1	392.2	141.4	2.77	n/a
12.5 min	139 MBq	1.25	491.6	176.4	2.79	0.4%
15 min	167 MBq	1.5	589.6	212.3	2.78	0.1%
20 min	222 MBq	2	782.9	272.5	2.87	3.4%
25 min	278 MBq	2.5	980.3	343.4	2.85	2.8%
30 min	333 MBq	3	1175.1	413.9	2.84	2.3%

Review of MIDAS patient HMR uncertainty

The mean counts in the 48 pixel mediastinum ROI was 5926, but there was substantial variation (Figure 34). The 25th percentile value was 4867 total counts. These counts result in an uncertainty in the mediastinum count density estimate of 1.3 and 1.4% respectively. Table 17 shows that the estimated error in HMR is around 1.6% for the MIDAS study, suggesting a slightly lower activity than 111 MBq could be administered to obtain an uncertainty of within 2% for an average patient. This is achieved with a mediastinal total count of 5000, assuming an HMR value of 1, which requires an activity of 94 MBq for the average patient. However, to achieve an uncertainty of within 2% for 75% of the patients an activity of 114 MBq is required. If the administered activity was increased (3.3x) to 370 MBq the HMR uncertainty would be around 0.9% (Table 17).

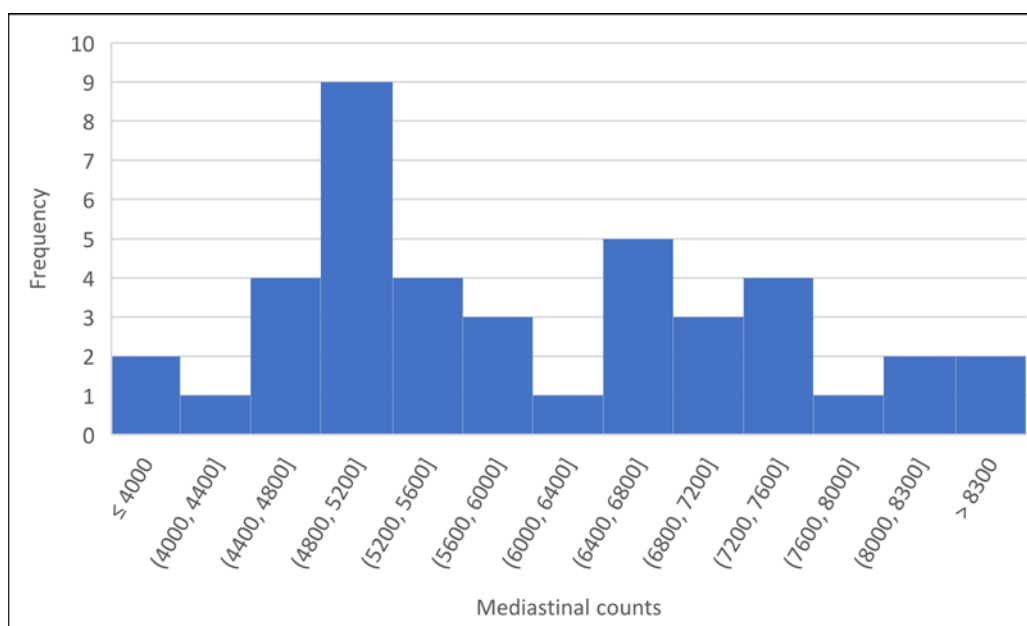


Figure 34: Histogram of mediastinum ROI counts for MIDAS patients

Table 17: Estimated uncertainty in mediastinum and heart ROIs and HMR measurement with 111 and 370 MBq

		Mediastinum counts	Mediastinum uncertainty	Estimated heart counts	Heart uncertainty	HMR uncertainty (111 MBq)	HMR uncertainty (370 MBq)
HMR = 1	Typical patient	5926	1.3%	13964	0.8%	1.6%	0.85%
	25th percentile	4867	1.4%	11469	0.9%	1.7%	0.94%
HMR = 2	Typical patient	5926	1.3%	27928	0.6%	1.4%	0.78%
	25th percentile	4867	1.4%	22937	0.7%	1.6%	0.86%

5.3.4 Discussion

Review articles and guidelines on the use of cardiac MIBG discuss the need for standardisation, including of the administered activity and image acquisition duration, suggesting that count density may affect image interpretation [64, 165, 168, 191, 192]. However, there are few studies evaluating the effect of differences in image count density on the heart to mediastinum ratio (HMR). In 2008 Verberne et al. performed a retrospective multivariate analysis on 290 heart failure patients from a European multicentre cardiac MIBG study, where large differences in acquisition protocols between the centres were present [193]. They found that acquisition duration had a small but significant effect

on HMR. However, as the administered activities ranged from 72 to 370 MBq, we would expect that this would have had more of an impact on count density - but administered activity was actually not reported as a significant variable, nor does it seem to have been combined with acquisition duration to give a single variable. Over 90% of patients were imaged with low energy collimators so the findings may not be replicated with medium energy collimators, due to reduced septal penetration.

Our results suggest that the heart to mediastinum ratio is not affected by count density within the typical clinical ranges of 111 MBq to 370 MBq seen in the literature. This is encouraging for multicentre studies and for meta-analysis. Other factors, such as collimator choice, delay between injection and imaging and patient factors are likely to be the cause of significant variation between HMR values reported in the literature. Our results with MIDAS patients confirm that 111 MBq is an appropriate activity for 10 minute delayed images, given HMR uncertainties of within 2%. The ARSAC administered activity of 370 MBq gives over three times the effective radiation dose, but reduces the uncertainty by less than half.

For the SUPeRB study, the results confirm that we do not need to correct for the exact amount of administered activity for planar imaging, or be concerned if some of the activity appears to have tissue. In the event that the scan needs to be terminated early, due to scanner fault or patient being unwell, the images can still be used.

5.3.5 Conclusion

An administered activity of 111 MBq with a 10 minute acquisition duration results in acceptable image quality and counting statistics, with 370 MBq unlikely to give clinically significant improvements. Although visual image quality and ease of region placement is affected by the administered activity and image duration, the heart to mediastinum ratio does not appear to be. Variation in image count density alone should not be a barrier to comparison between individual scans and between centres.

Chapter 6: ^{123}I -FP-CIT image reconstruction and analysis

This chapter focusses on advanced methods for reconstructing and analysing FP-CIT images, comparing them to the standard BRASS quantification methods implemented for use in the SUPeRb study in chapter 4. In section 6.1 the accuracy of visual rating of FP-CIT scans by a panel of experts with access to BRASS results is compared with the accuracy of the quantification alone. In section 6.2 we explore whether absolute quantification of striatal uptake via SUV-SPECT offers any advantages over the standard relative SBR method. In section 6.3 we use striatal phantom scans to compare the standard reconstruction method with more advanced methods using CT attenuation map and Monte Carlo modelling. In section 6.4 these advanced reconstruction methods are applied to MCI subjects from the SUPeRb study to evaluate whether they improve accuracy. In section 6.5 we assess FP-CIT test-retest variation using control subjects rescanned after one year and discuss the implications for clinical reporting of change in SBR. The results presented in this chapter are then summarised in section 6.6.

6.1 Comparison between visual rating and BRASS SBR results in SUPeRb MCI cohort

6.1.1 Introduction

In previous Newcastle University studies using ^{123}I -FP-CIT, images were reconstructed with filtered back projection, as this was the technique in clinical use at the time. In the past few years iterative reconstruction has become more widespread outside of research settings due to improved computer processing speed and the ability to incorporate corrections for attenuation, scatter and depth dependent resolution. As discussed in Chapter 4, it was decided that for the SUPeRb study the same reconstruction method required for BRASS image quantification would be used for visual ratings.

In this section we compare the diagnostic accuracy of the consensus expert panel approach using the new reconstruction method with BRASS quantification alone. Two sets of “stand-alone” BRASS quantification results are compared with the panel ratings, one with the standard ENC-DAT database and age correction and one relative to a database of 31 age-matched SUPeRb controls, without age correction.

6.1.2 Methods

Modified ratings scale

^{123}I -FP-CIT was first developed as a radioactive ligand for visualising reduced striatal uptake in Parkinson’s disease (PD) due to dopaminergic neuron loss. In PD, neurons projecting to the putamen are often lost before those projecting to the caudate, resulting in a characteristic “full-stop” appearance on transverse slices of an FP-CIT scan, compared to a “comma” appearance in healthy individuals or those with essential tremor (ET). Visual assessment can be used for distinguishing between PD and ET because of these characteristic patterns. Benamer *et al.* [194] established a FP-CIT rating scale for PD ranging from normal (Stage 0) to virtually no uptake (Stage 3), as illustrated in Figure 35:

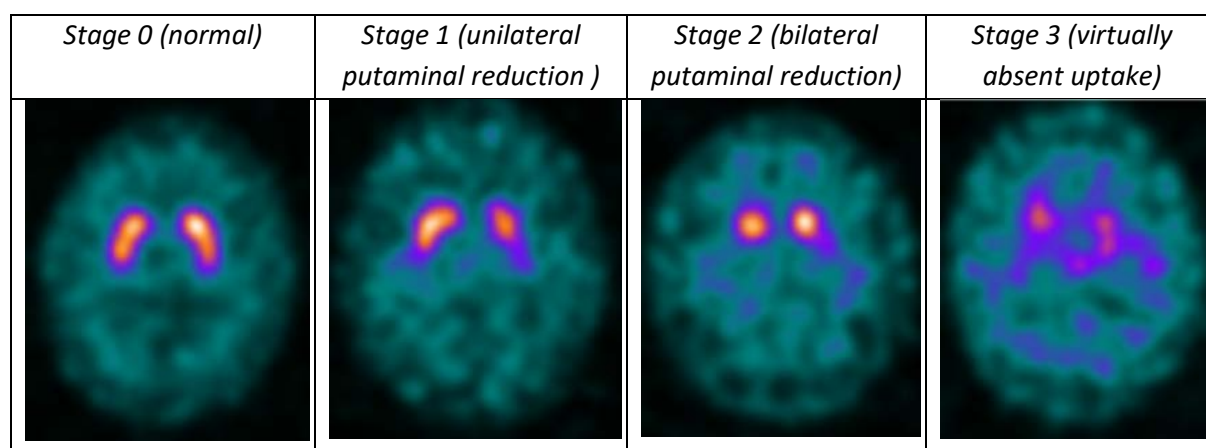


Figure 35: Visual rating scale used by Benamer *et al.* [194] Stage 0 – normal uptake, Stage 1 – asymmetric uptake with marked reduction in putamen uptake on one side, Stage 2 – activity confined to caudate nuclei bilaterally, Stage 3- virtually absent uptake bilaterally. The images are examples from Newcastle studies.

As FP-CIT came to be used for the diagnosis of dementia with Lewy bodies (DLB) it became apparent that this pattern of neuronal loss over time did not occur for all patients. O'Brien *et al.* [44] showed that the loss tended to occur at a more uniform rate in both the caudate and putamen in DLB compared to PD, leading to a “balanced loss” appearance on imaging. The rating scale introduced by Benamer does not take mild global reduction into account, so these scans are difficult to categorise. We sought to extend this simple and easy-to-use rating scale to make it suitable for use for DLB diagnosis. We added two categories to describe mild and moderate global striatal reduction and a further category for any other abnormal/atypical pattern (Figure 36). For the purposes of dichotomising scans into normal or abnormal groups, a rating of 4 (mild global reduction) counts as normal and 5 (moderate global reduction) counts as abnormal. Scans falling into the “other” category 6 were discussed on a case by case basis. The full modified scale is shown in Table 18.

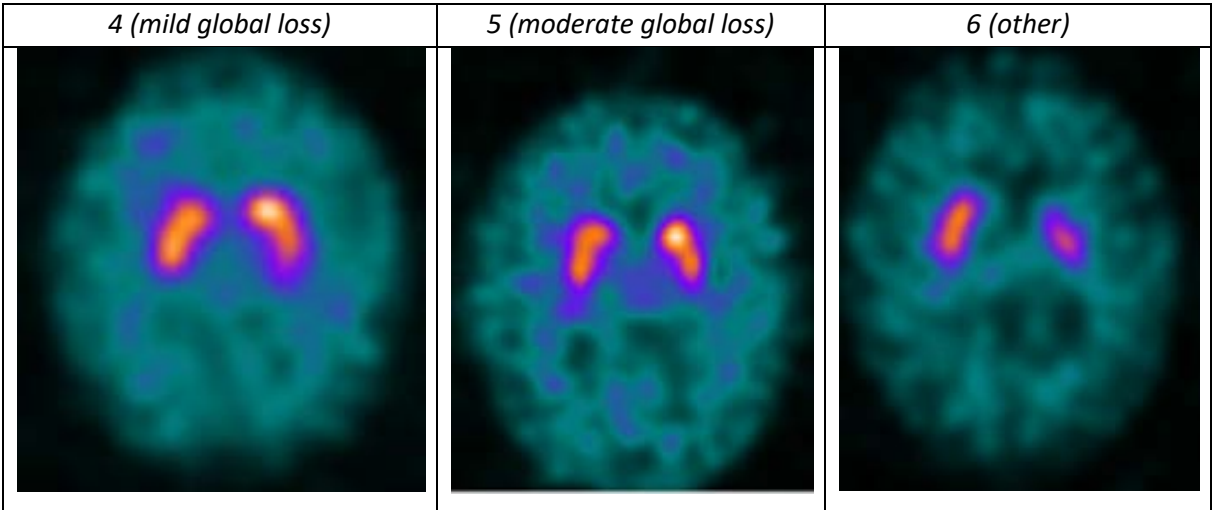


Figure 36: Additional categories added to the FP-CIT visual ratings scale to account for balanced loss and atypical appearances.

This rating scale was validated using scan data from a Newcastle University study funded by the Medical Research Council in early 2000s, using patients for which autopsy confirmed diagnoses had been obtained. In particular, it was confirmed that a rating of 4 frequently occurred in non-DLB subjects and so should be regarded as normal. This is published, alongside a further semi-quantitative ratings scale, in our 2018 publication [195].

Table 18: Modified rating scale for FP-CIT adopted for DLB. Categories 0 and 4 are considered normal, 1,2,3,5 abnormal (i.e. in keeping with Lewy body disease) and 6 is for cases that do not fit any category.

Modified rating	Description	Normal / Abnormal
0	Normal (tracer uptake bilaterally in putamen and caudate nuclei and largely symmetric).	Normal
1	Asymmetric uptake with normal or almost normal putamen activity in one hemisphere and with a more marked reduction in the contralateral putamen.	Abnormal
2	Significant bilateral reduction in putamen uptake with activity confined to the caudate nuclei.	Abnormal
3	Virtually absent uptake bilaterally affecting both putamen and caudate nuclei.	Abnormal
4	Very mild or equivocal global striatal reduction in uptake	Normal
5	Significant global striatal reduction in uptake	Abnormal
6	Other atypical pattern of uptake.	Panel to discuss

Consensus ratings panel

The consensus panel consisted of a group of five raters experienced at reviewing FP-CIT images: myself, Alan Thomas (Professor of old-age psychiatry), Jim Lloyd (Consultant Clinical Scientist), Paul Donaghy (Old-age Psychiatrist) and George Petrides (Consultant Radiologist). The panel members were sent sets of anonymized images to review in a randomised order by an independent member of the team (Sean Colloby), with the standard BRASS quantification results for each case (SBR and age-corrected Z scores). The ratings were collated by Dr. Colloby, who checked for agreement and discrepancies. Cases where 4 or 5 raters had assigned the same category were not discussed further. The remaining cases were discussed at a meeting of all the panel members, still fully blinded, and a consensus reached. For example, if two raters had rated a scan as category 3 and three as category 5 the case would be discussed and a consensus of either 3 or 5 agreed upon. A binary normal/abnormal consensus opinion was reached for each case for the purposes of diagnostic accuracy calculation. Cases with a consensus score of 6 were discussed on an individual basis and were deemed abnormal if the pattern, although atypical, was thought to be in keeping with Lewy body disease. If infarct was suspected, MRI images were reviewed retrospectively and the subject excluded if confirmed.

FBP vs iterative image reconstruction

Changing the reconstruction method has a considerable effect on the appearance of the images, with the background appearing brighter with FBP than with OSEM_RRACSC. The background scaling is often used to help decide whether there is global loss of dopaminergic function, so this change is not trivial

and background levels that would indicate reduced striatal uptake with FBP often look normal with OSEM_RRACSC to reporters unused to the wider dynamic range covered by the same colour scale. To aid reporters used to reviewing FBP images a document was produced containing instructions and a set of training/reference images reconstructed with both methods. Some example images from this document reconstructed with FBP and OSEM_RRACSC are shown in Figure 37, with tips for assessing the images using the appearance of the background in Table 19.

Given the increasing use of quantification for routine clinical FP-CIT reporting, the panel decided that the raters for the SUPeRb study should have access to the BRASS quantification results when reviewing the images individually and at consensus panel meetings. This would then reflect clinical practice for the majority of centres, in the UK at least, making the results as generalisable as possible. For similar reasons, the standard BRASS output with age corrected Z scores relative to the ENC-DAT database was used, despite the concerns with age correction models discussed in chapter 4. The role of quantification was discussed amongst the panel and it was agreed to review the images visually first and then use the quantification to identify cases of reduced striatal uptake that might otherwise be missed, as well as to assist with uncertain cases. Panel members used their professional judgement in cases where visual assessment and semi-quantification did not agree.

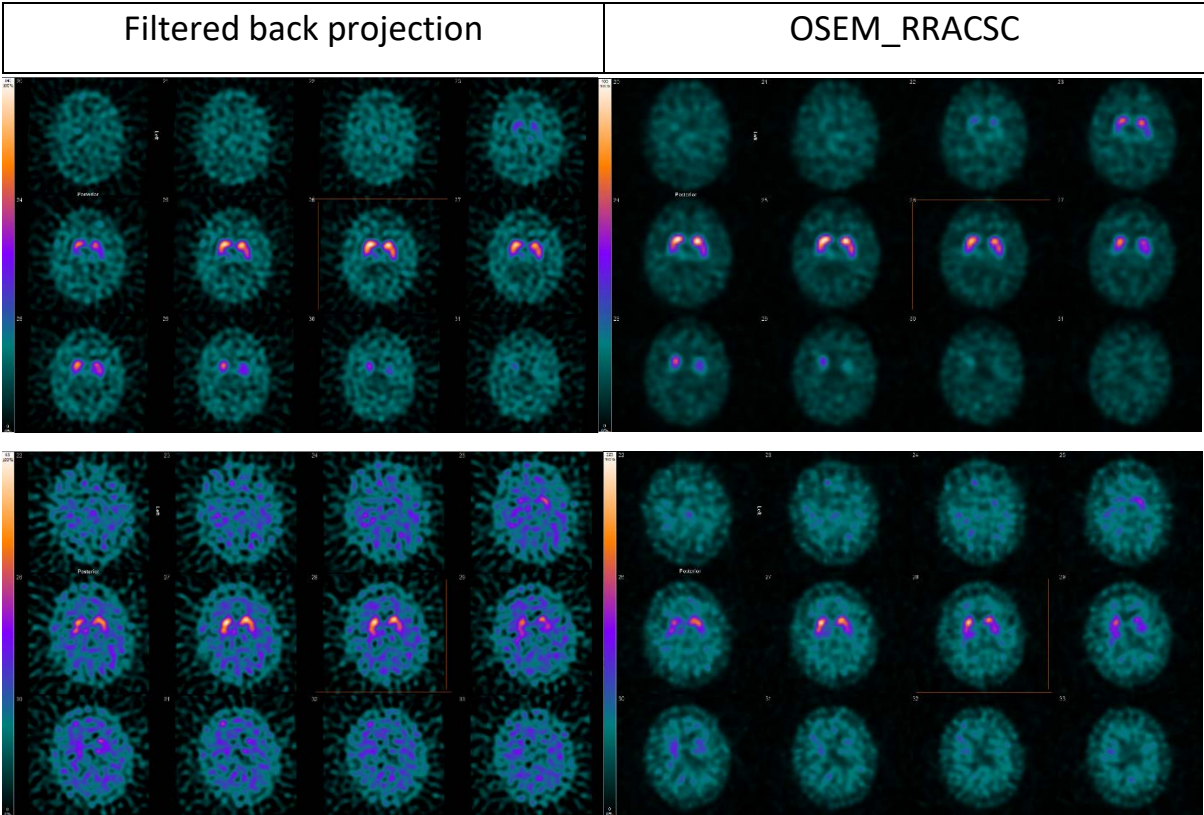


Figure 37: Normal scan (top) and abnormal scan with global striatal reduction (bottom) reconstructed with filtered back projection and the new iterative reconstruction method. The total counts in the images is around 3 times as high with the corrections, which has a corresponding effect on the colour scale, which is referenced to the maximum voxel value.

Table 19: Background appearance with the FBP reconstruction method compared to the new iterative OSEM_RRASC method.

Background	Inferred striatal uptake	Background Appearance/Colour FBP reconstruction	Background Appearance/Colour OSEM_RRASC reconstruction
Normal	Normal uptake	Mostly uniform with green or black background	Mostly uniform dark green or black background
Very mild increase	Very mild / equivocal loss	Some speckling with predominantly green/black background but with blue/purple colour coming through.	Mostly uniform dark to light green or black background, with tiny specs of blue/purple
Mild increase	Mild loss	Speckling with a mixture of blue/purple and green/black	Dark to light green or black background, with specs of blue/purple especially around the striata
Moderate increase	Moderate loss	Little Green/black remaining with predominantly purple colour	Mixed green / blue background
Marked increase	Severe loss	Extensive areas of purple with pink/yellow foci	Mixed green / blue background with specs of pink coming through

Subjects

All the SUPeRB cases, including the 31 controls with normal cognition, were rated by the panel, but only the MCI cases were used to calculate diagnostic accuracy. As discussed in section 2.3.3, the MCI cases included in this thesis are those with a one year follow up diagnosis of probable MCI-LB or probable MCI-AD only.

One MCI-AD case rated as category 6 - other was excluded because on review of the MRI brain scan obtained as part of the study there was clear evidence of significant areas of infarct in the left striatum, which corresponded to the abnormal side on quantification (Figure 38). Exclusion of this case results in a probable MCI-LB group of 27 and a probable MCI-AD group of 16 (Table 20).

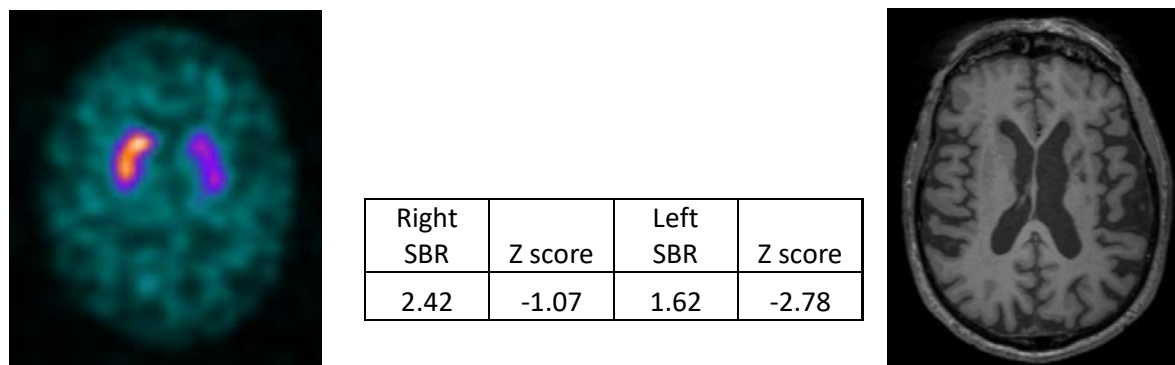


Figure 38: The probable MCI-AD case that was rated as “other” (category 6) by the panel. On review of the MRI head it was seen that there was evidence of infarcts in the left striatum, accounting for the reduced uptake. This case was excluded from analysis.

Table 20: Ages of the healthy older adults, MCI-LB and MCI-AD patients.

	Number included	Mean age (\pm SD)
Healthy older adults	31	74.1 (\pm 7.6)
Probable MCI-LB	27	73.0 (\pm 7.2)
Probable MCI-AD	16	77.5 (\pm 7.7)

Diagnostic accuracy

The 31 SUPeRB controls were used to calculate SUPeRB SBR cut-offs for BRASS so that quantification without age correction could be assessed. The SUPeRB cut-offs were taken as two standard deviations below the mean, the same method as implemented within BRASS using the factory database. The standard BRASS Z scores with respect to the ENC-DAT database with age-correction were also assessed. Sensitivity, specificity and overall accuracy was calculated for the visual ratings method and for BRASS quantification, using MCI subjects only. The results were reported with binomial confidence intervals.

6.1.3 Results

Plots of the SBR quantification results for the controls, probable MCI-AD and probable MCI-LB cases are shown in Figure 39. The control and probable MCI-AD groups are similar. The probable MCI-LB cases show a wide range of uptake values, with no clear distinction into normal and abnormal groups.

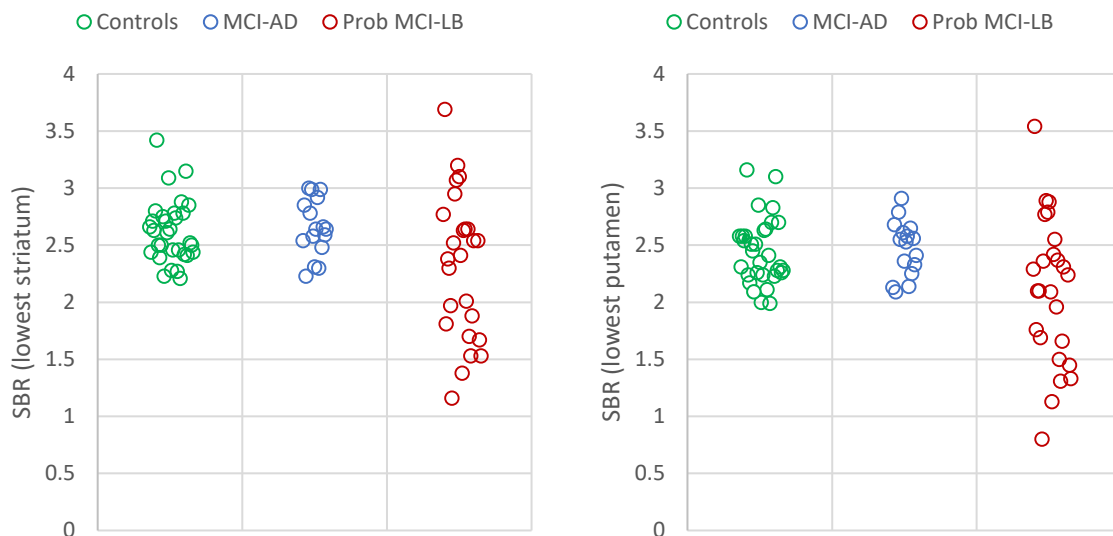


Figure 39: Plots of SBR for the control, MCI-AD and probable MCI-LB groups. The plot on the left shows the SBR for the whole striatum and the right the SBR for the putaminal sub-region. The excluded asymmetric MCI-AD case is not shown.

Diagnostic accuracy

SBR cut-offs for BRASS (without age correction) using SUPeRB controls are shown in Table 21. There were no discrepancies between the standard ENC-DAT BRASS results and the results with respect to local controls – all cases with age corrected Z scores below -2 also had SBRs below the local cut-off and vice versa. The sensitivity, specificity and overall accuracy of visual rating and quantification are given in Table 22.

Table 21: Local SBR cut-offs derived using SUPeRB age matched controls.

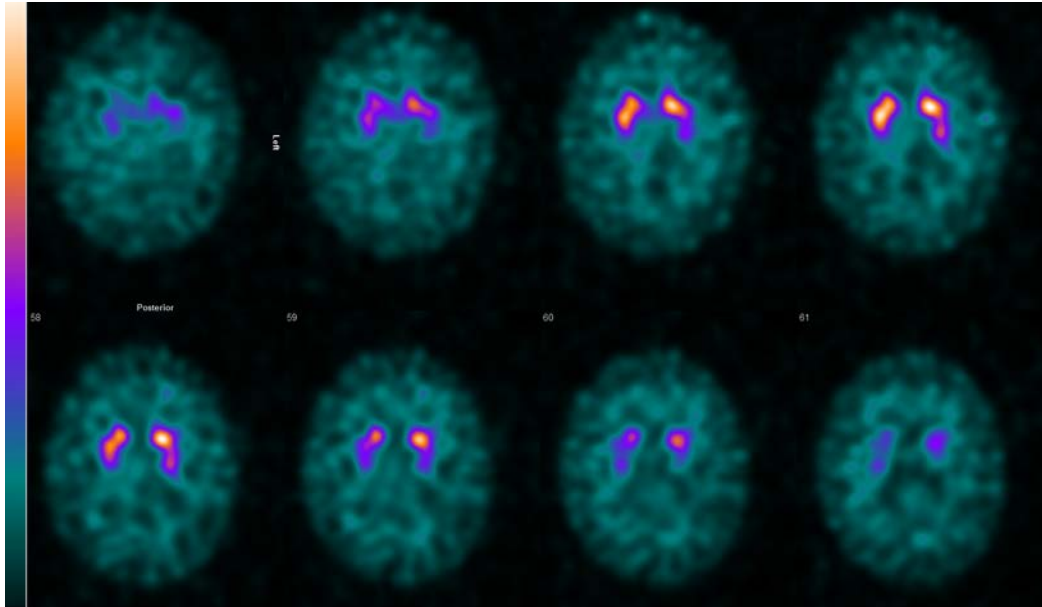
	Right striatum SBR	Left striatum SBR	Right caudate SBR	Left caudate SBR	Right putamen SBR	Left putamen SBR
Control mean	2.67	2.68	2.79	2.85	2.53	2.49
Standard deviation	0.28	0.29	0.31	0.31	0.28	0.31
Local cut-off	2.10	2.10	2.18	2.24	1.96	1.87

Table 22: Sensitivity, specificity and accuracy of visual rating and BRASS quantification for diagnosing Lewy body disease in MCI subjects

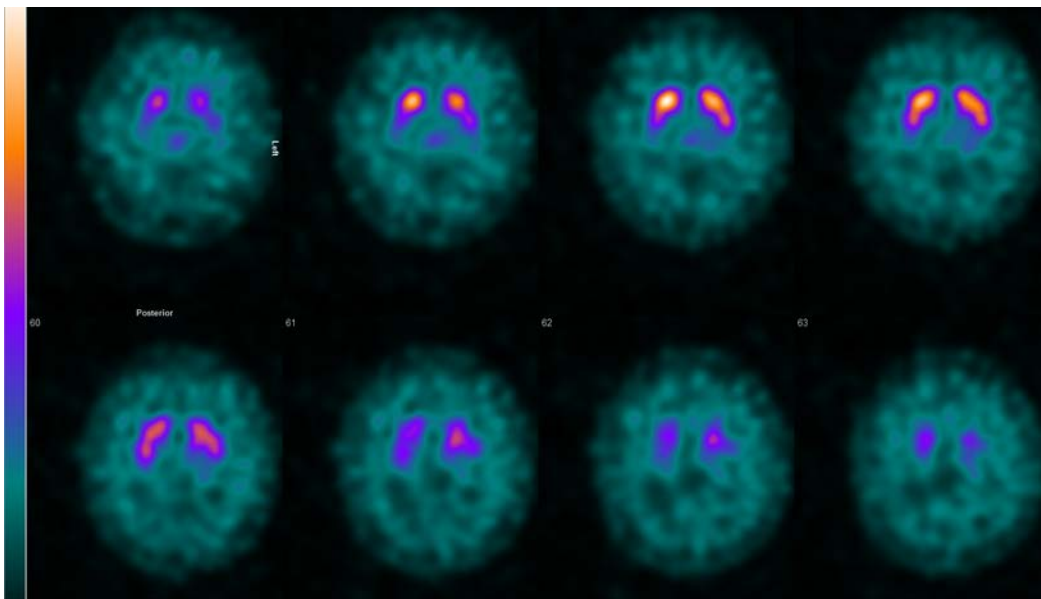
	Method	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
Visual	Consensus panel with access to BRASS (ENC-DAT results)	13/27 48% (29 to 68%)	14/16 88% (62 to 98%)	27/43 63% (47 to 77%)
BRASS (ENC-DAT)	ENC-DAT database with age correction	12/27 44% (25 to 65%)	16/16 100% (79 to 100%)	28/43 65% (49 to 79%)
BRASS (SUPeRB)	SUPeRB age matched controls	12/27 44% (25 to 65%)	16/16 100% (79 to 100%)	28/43 65% (49 to 79%)

Controls with normal cognition

Two cognitively normal controls were rated as abnormal, both with normal MR head results (part of the study inclusion criteria). These cases are shown alongside their quantification results in Figure 40. The first case, aged 73 years, has normal quantification in all regions but was thought to be abnormal because of a rather asymmetric appearance, with the left putamen never reaching normal “brightness” on any of the image slices. The second case, aged 63, had quantification results towards the bottom of the BRASS ENC-DAT database normal range, corrected for age and signs of blue speckling in the background, indicating a global reduction in striatal uptake. The image was rated as showing balanced loss (category 5).



Age at scan	BRASS Registration ok?	Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
73	y	2.63 (-0.81)	2.77 (-0.58)	2.70 (-0.85)	2.97 (-0.45)	2.57 (-0.70)	2.54 (-0.68)



Age at scan	BRASS Registration ok?	Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
63	Y	2.23 (-1.89)	2.28 (-1.88)	2.35 (-1.78)	2.35 (-1.93)	2.09 (-1.91)	2.19 (-1.68)

Figure 40: The two control subjects with FP-CIT scans rated as abnormal by the panel. The first case was rated as asymmetric (category 1) and the second as showing significant global striatal loss (category 5).

6.1.4 Discussion

Visual assessment supported by quantification shows excellent specificity for Lewy body disease diagnosis in MCI, with none of the AD cases showing low uptake on quantification and only 2/16 (12%) rated as abnormal by the panel. The specificity is in keeping with the LewyPro study result of 89% given in our previous publication, which used FBP and DaTQUANT quantification [37]. Some of the LewyPro subjects had another scan as part of SUPeRB, and are included here, so there is some overlap of cases. The visual rating sensitivity of 48%, however, is lower than the 61% for LewyPro (95% CI 42.5 to 77.4%), although the confidence intervals do overlap. If this is a real difference it may have been caused by the change in reconstruction protocol, with the OSEM_RRSCAC used in SUPeRB tending to appear more “normal”.

The specificity with quantification alone was 100% but the sensitivity slightly lower than visual rating as one MCI-LB case rated as abnormal had completely normal quantification. In general, concordance between quantification and visual rating was high. The only other cases with disagreement are two MCI-AD cases rated as normal on quantification and abnormal on visual rating. One shows the left putamen SBR reduced relative to the right, but still well within normal limits ($Z = -1.22$ left; -0.54 right). The other also shows asymmetry between the left and right striatum ($Z = -0.62$ left; -1.24 right). These cases suggest that it may be useful to quantify asymmetry using BRASS in future, a feature previously thought to give rise to too many false positive results.

There was 100% concordance between the BRASS results using the ENC-DAT database and the BRASS results interpreted using the SUPeRB control SBR cut-offs. This suggests that an age-matched local database has no clear advantage over the standard age-corrected factory database, at least in this age group. This is reassuring for other centres for whom it is likely not to be feasible to produce a local database. However, a limitation of this analysis is that there was no assessment without any use of quantification for the SUPeRB cases, and we did not attempt to quantify the subjective degree to which the panel members were influenced by the BRASS results. In the relatively young control case (63 years) rated as showing significantly reduced global striatal uptake, the age correction will have made the Z scores lower than they would have been for an older person, which likely contributed to the assessment as abnormal.

The distribution of the SBRs of the MCI-LB cases shown on the plots in Figure 39 shows that there is no clear cut-off between normal and abnormal that would give higher accuracy. There is a significant overlap with both the control and the AD groups, which suggests that many of these cases either do not have Lewy body disease affecting the substantia nigra, or not sufficient to affect dopaminergic function. Some may be misdiagnosed and not have Lewy body disease at all, but this seems less likely in this probable group who have at least two core symptoms of Lewy body disease, or one core

symptom and an abnormal cardiac MIBG scan. As part of the SUPeR study the patients are being followed up, including with repeat FP-CT scanning. This is outside the scope of this thesis but will provide valuable information in future.

6.1.5 Conclusion

The sensitivity of visual assessment of FP-CIT scans by a consensus panel of raters with access to BRASS quantification was 48% and specificity 88%. The sensitivity and specificity of BRASS quantification alone was 44% and 100%. The overall accuracy of the visual ratings method and quantification alone was very similar. Long term follow up will help to clarify the status of the cases assumed to be early Lewy body disease with normal imaging, as well as that of the three cases where there was discordance between the panel and BRASS quantification.

6.2 Absolute quantification of FP-CIT striatal uptake in MCI patients

6.2.1 Introduction

Traditionally, use of relative scaling to the maximum voxel in the image has meant that the appearance of the brain background surrounding the striata has needed to be used as a surrogate for overall striatal uptake (see the previous section). As discussed in Chapter 3, quantifying striatal uptake by determining the normalised absolute activity concentration in each striatum, rather than an uptake ratio, could in principle more closely reflect true uptake, as it removes the dependence on a low count and potentially variable background reference region in the cortex. Standardised uptake values are calculated for each voxel, allowing images to be displayed using an absolute SUV scale rather than the standard relative scale.

SUV information for a volume of interest can be reported using SUV_{max} , SUV_{peak} or SUV_{mean} . In FP-CIT imaging, SUV_{max} and SUV_{peak} have the advantage of being operator independent, but the disadvantage of being insensitive to putaminal only losses, as the highest count density is normally located in the caudate. SUV_{max} is subject to noise as it is derived from a single voxel, so is not used in this thesis. SUV_{mean} is the mean SUV within the volume of interest, so SUV_{mean} within a striatal volume is more representative of the overall striatal uptake than SUV_{peak} . However it is highly dependent on the accuracy of the striatal volume of interest.

Ideally, SUV_{mean} would be defined using accurate striatal reference regions, with the SUV images registered to a template, similar to the methods used by BRASS to determine SBR. However, BRASS does not currently support absolute quantification so does not report SUVs. We therefore developed an alternative method for determining striatal SUV_{mean} without the need for accurate striatal VOIs, based on the Southampton method [144], described in more detail in the methods. We also calculated SBRs with the Southampton method so SUV could be compared with SBR directly.

SUV-SPECT is a novel technique, with few publications on FP-CIT so far found in the literature. Two retrospective patient studies from Japan have recently been published, comparing SUV with SBR in 77 and 52 patients, and concluding that accuracy is actually lower with SUV than with SBR [196, 197]. Both studies used CT attenuation correction, however scatter correction was not performed in one of the studies [197] and the method not stated in the other [196]. There are no publications on the use of FP-CIT SUV-SPECT in mild cognitive impairment.

In this section we reconstruct SUPeR FP-CIT image data to obtain SUV_{mean} and SUV_{peak} results in controls and MCI patients. As with SBR, the controls are used to set normal thresholds and the MCI cases for estimating accuracy for diagnosing MCI-LB. We aim to compare the accuracy of SUV_{mean} , SUV_{peak} and SBR calculated with the Southampton method, assess whether SUV normalisation based on body weight is effective and explore the potential for use of an absolute scale for visual assessment.

6.2.2 Methods

Activity concentration calibration

A calibration factor to convert reconstructed count density into activity concentration was determined by scanning a uniform phantom filled with a known activity, as shown in Figure 41. The activity, volume and assay time are input into the Hermes calculation tool (Figure 42). The resultant calibration factor is then added to the reconstruction protocol for patient scans. Scan reconstructed with the calibration factor applied, but no other changes, are displayed within Hermes HybridViewer in units of kBq/ml. If the patient weight and administered activity are entered during reconstruction images may be displayed in either kBq/ml or in units of SUV.

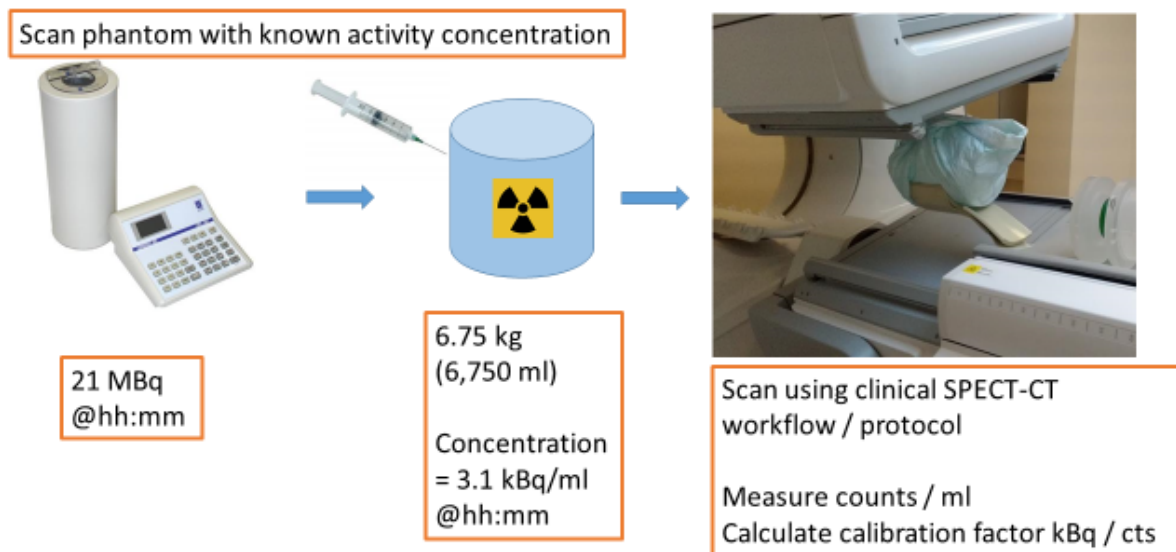


Figure 41: Illustration of steps required for deriving an activity calibration factor.

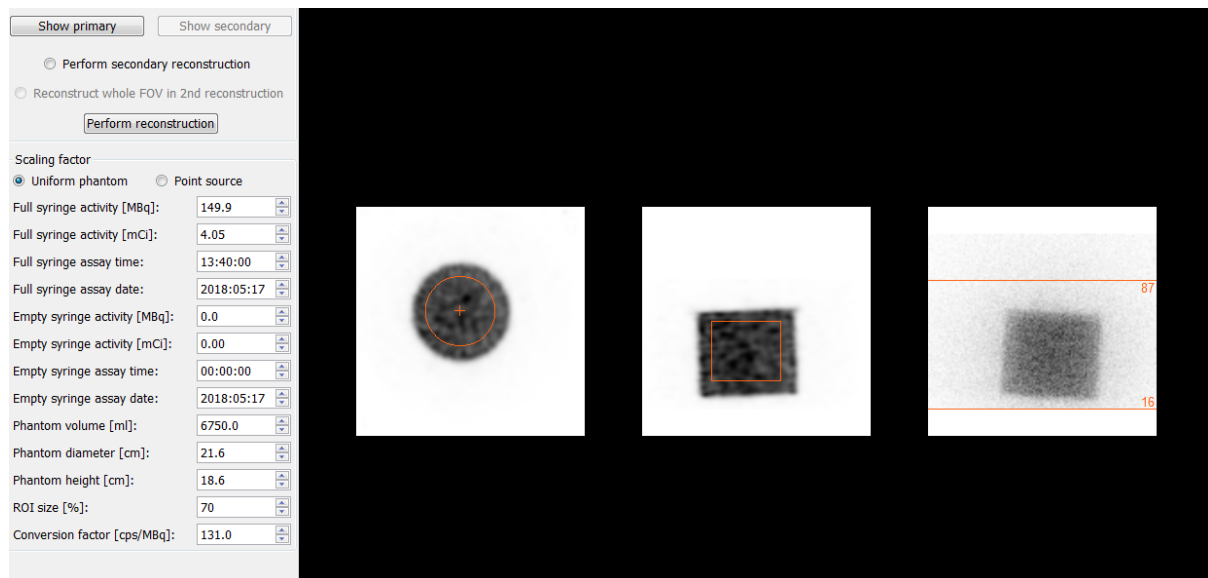


Figure 42: Example screenshot of uniform phantom activity concentration calibration tool, with the activity in the phantom, solution volume and assay times specified. There was no residue left in the syringe (to 1 d.p.).

Striatal SUVmean calculation method

The Southampton method was developed to calculate SBR, but can be adapted for SUV simply by removing the normalisation to the background. The method still requires a background VOI however, as this is used for determining the non-specific uptake in large striatal VOIs.

Briefly, the method attempts to avoid count losses due to the partial volume effect when selecting striatal regions of interest, by using large fixed regions to account for all the counts in the image that are given by radioactivity in the striata. Because a much larger volume than the actual size of the striata is used, the true volume of each striatum is assumed to be fixed at 11.2 ml, which is the volume used in the anthropomorphic striatal phantom referenced in the paper.

In the original Southampton method, large fixed sized regions of interest are placed manually over a summed transverse slice image, containing a fixed number of slices around the striatal maximum. We modified the method to use volumes of interest, which were defined on the ENC-DAT template based on the dimensions specified by Tossici-Bolt et al. [144] and then registered to the striatal phantom and patient scans. This removed much of the operator variation as striatal VOIs rarely needed adjusting. The BRASS occipital background VOIs from the template were used as the background VOI.

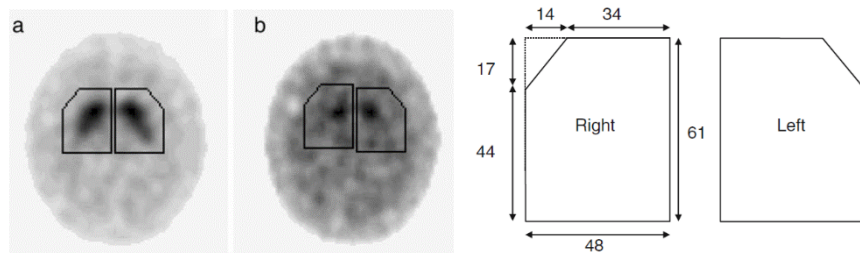


Figure 43: Diagram showing size in mm of large striatal regions of interest used in the original version of the Southampton method.

The striatal SUV_{mean} is determined on SUV-SPECT images as follows. The signal in the large striatal VOI, SUV_{total} , is attributed to the specific binding in the striatum, plus non-specific background. The non-specific background level is determined from the background VOI, and subtracted from the large striatal VOI. The volume of the large VOI, V_{VOI} , must be known so that the non-specific background can be calculated correctly. The remaining signal is the summed SUV for the whole striatum. In order to determine the SUV_{mean} the total is divided by the assumed striatal volume, $V_{striatum}$ of 11.2 ml. If SBR is required this can be simply calculated by dividing the resultant $SUV_{mean_striatum}$ by the $SUV_{mean_background}$.

$$SUV_{mean_striatum} = \frac{SUV_{total} - (SUV_{mean_background} * V_{VOI})}{V_{striatum}}$$

$$SBR = \frac{SUV_{mean_striatum}}{SUV_{mean_background}}$$

Subjects

Three of the 31 controls included in section 6.1 had to be excluded from absolute quantification due to incomplete records of injected activity. Twenty-eight healthy older adults were therefore processed using the Southampton method to give SBR, SUV_{mean} and SUV_{peak} normal ranges. These were calculated at two standard deviations below the mean, the same method used in BRASS and DaTQUANT software.

Fifteen of the 16 MCI-AD cases and 24 of 27 the probable MCI-LB cases from section 6.1 were included in the comparison between AD and LB groups and for the diagnostic accuracy calculations. Two cases were excluded due to incomplete injected activity records and two due to being scanned clinically within another NHS Trust.

SUV processing

Subject image data was reconstructed to give FP-CIT images displayed in units of SUV. The same FP-CIT OSEM protocol as used for SBR was applied (OSEM_RRACSC: 16 iterations and 4 subsets; resolution recovery, uniform scatter and attenuation correction). The images were reoriented to standard planes within HybridRecon after reconstruction to assist with VOI placement. The reconstructed images were registered to the ENC-DAT normal template, which had had the BRASS striatal VOIs removed and replaced by the Southampton striatal VOIs (Figure 44). The occipital background used in BRASS was used as the background in the Southampton method. The template VOIs were transferred to the subject scan after rigid registration and the VOI statistics exported as an xml file. SUV_{mean} values for the right and left striata were calculated within a spreadsheet following the method outlined above. SBRs obtained with the Southampton method were also calculated, by dividing striatal and background SUV_{mean} . SUV_{peak} was recorded for each striatum.

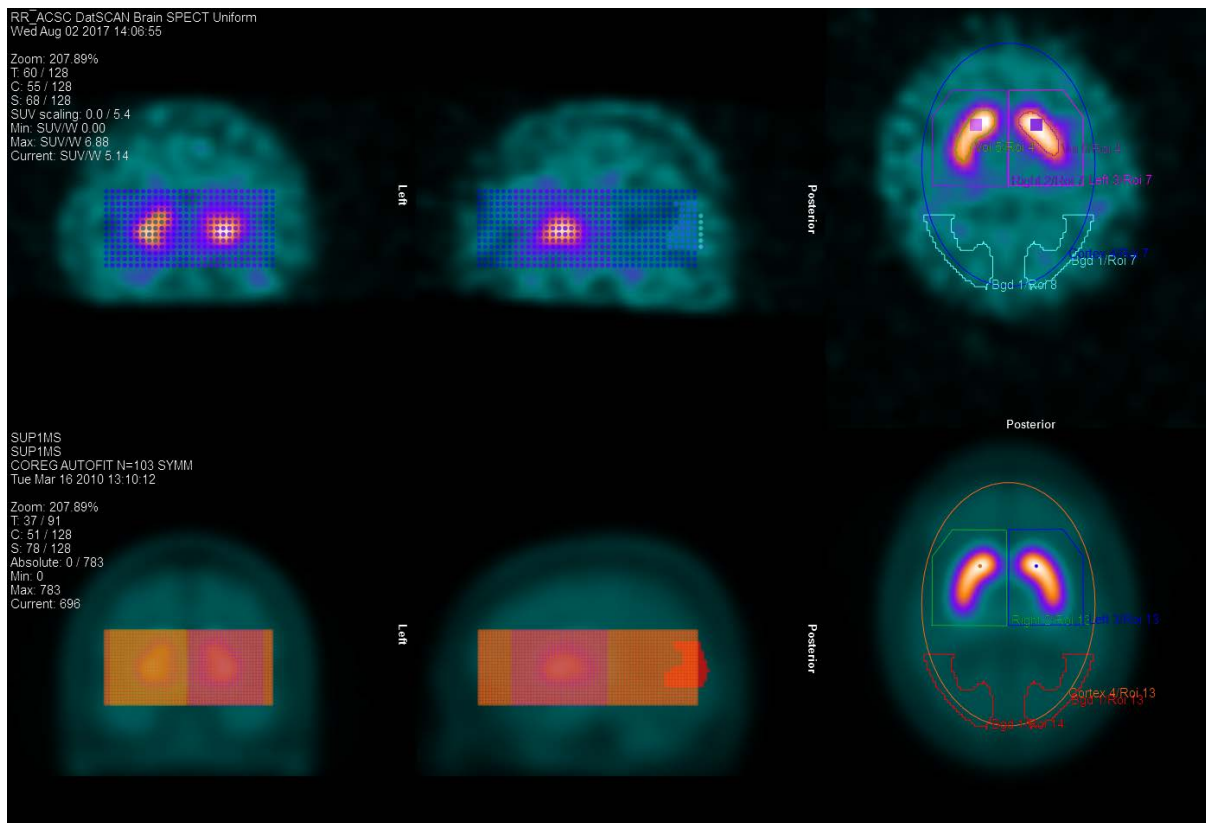


Figure 44: ENC-DAT template used in BRASS with Southampton VOIs added (bottom row), registered to a patient scan (top). The two SUV_{peak} cube VOIs are seen in the caudate nuclei on the patient scan.

Evaluation of SUV normalisation methods

The normalisation method used in SUV calculation was assessed by plotting striatal SUV_{mean} and SUV_{peak} for the controls against background SUV_{mean} . Background SUV_{mean} values should be similar for all controls, and striatal uptake should not depend on the background if normalisation is appropriate. Any residual relationship between background levels and striatal uptake is indicative of a potential problem.

Display of images with absolute SUV scale

Example control and MCI-LB subjects were displayed with the SUV scale applied, as a preliminary assessment of whether absolute scaling could be suitable for visual reporting and worthy of further evaluation in future.

6.2.3 Results

Normal thresholds

The mean, standard deviation, coefficient of variation and normal threshold cut-offs obtained from the 28 controls are shown in Table 23.

Table 23: SBR and SUV values for controls with normal threshold values (mean-2SD).

	Southampton SBR		Southampton SUV _{mean}		SUV _{peak}	
	Right	Left	Right	Left	Right	Left
Mean	5.88	5.74	8.15	7.95	6.35	6.43
SD	0.66	0.73	1.17	1.21	0.72	0.82
CoV	11%	13%	14%	15%	11%	13%
Threshold	4.56	4.27	5.81	5.54	4.90	4.79

MCI results compared with controls

Plots of SBR, SUV_{mean} and SUV_{peak} obtained with the Southampton method are shown in Figure 45, with both the right and left striata included. The SBR results with BRASS are also shown for comparison.

There are two slight outlier control subjects for SUV_{mean} (one high, one low), which are not seen with SBR. These subjects have likely contributed to the slightly higher coefficient of variation with SUV_{mean}. The high outlier is much more obvious with SUV_{peak} and was excluded from the normal range. Visually, the SUV_{mean} plot appears to show less of a distinction between MCI-LB and the other groups than the SBR plot - calculated with either the Southampton method or BRASS. SUV_{peak} shows even less distinct groups.

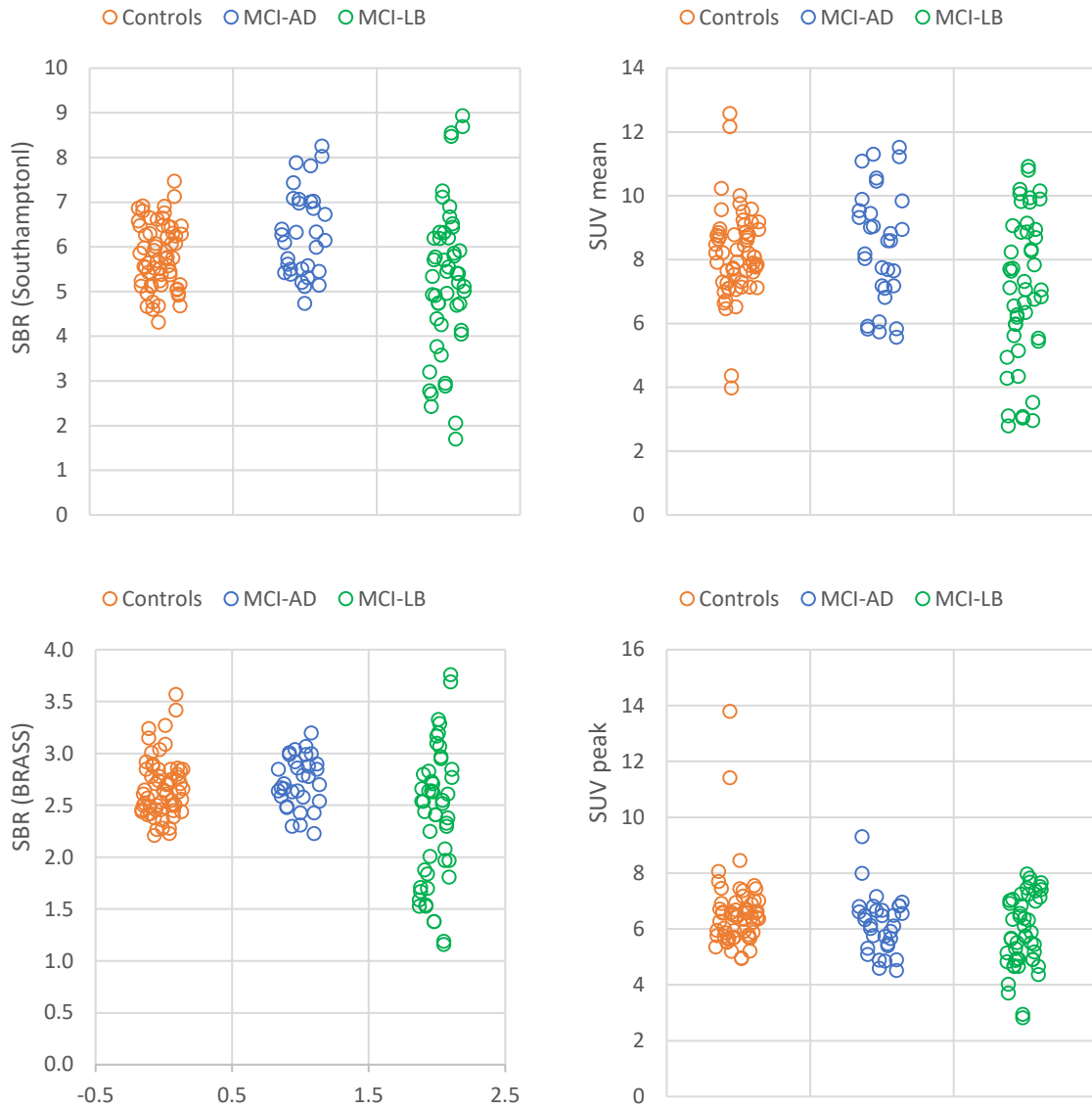


Figure 45: Whole striatum SBR, SUV_{mean} and SUV_{peak} for controls, MCI-AD and MCI-LB calculated with the Southampton method. Both left and right striata are included. BRASS SBR is also shown for comparison (bottom left). The outliers are for the right and left striata of the same control, who was excluded from the SUV_{peak} normal cut-off.

Diagnostic accuracy in MCI-LB

The diagnostic accuracy figures for the SBR and SUV methods are given in Table 24. These are calculated using the lower of the two striatal values for each MCI patient. SUV_{mean} gives the same 100% specificity as SBR with both BRASS and Southampton method. The sensitivity however is lower than BRASS SBR for both SUV_{mean} and Southampton method SBR due to only 7/24 MCI-LB cases falling below the control normal threshold. However, the 95% binomial confidence intervals are wide due to low sample sizes and none of the results for any method are significantly different from any other.

Table 24: Sensitivity, specificity and overall accuracy for MCI-LB diagnosis for SBR calculated with BRASS and Southampton methods, SUV_{mean} calculated with Southampton method and SUV_{peak}.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
BRASS SBR	10/24 42% (22 to 63%)	15/15 100% (78 to 100%)	25/39 64% (47 to 78%)
Southampton SBR	7/24 29% (12 to 51%)	15/15 100% (78 to 100%)	22/39 56% (40 to 72%)
SUV_{mean}	7/24 29% (12 to 51%)	15/15 100% (78 to 100%)	22/39 56% (40 to 72%)
SUV_{peak}	6/24 25% (10 to 47%)	13/15 87% (60 to 98%)	19/39 49% (32 to 65%)

Evaluation of SUV normalisation methods

The coefficient of variation in background mean SUV was 19%. Plots of striatal SUV_{mean}, SUV_{peak}, Southampton method SBR and BRASS SBR against background SUV_{mean} are shown in Figure 46 with slopes, p-values and R² data in Table 25. Striatal SUV shows a positive relationship with background SUV, particularly for SUV_{peak}, which is highly correlated with background SUV_{mean} (p <0.001). The SBR with the Southampton method shows a negative relationship with background SUV (p <0.001). There is no significant relationship between BRASS SBR and background SUV.

Table 25: Slopes for linear fits to SUV_{mean} background for each parameter with associated p-values for difference from zero. The estimated percentage variation explained by background SUV is given.

	SUV _{mean}	SUV _{peak}	SBR (Southampton)	SBR (BRASS)
Slope	2.82	4.49	-1.45	-0.23
p-value for slope	<0.001	<0.001	<0.001	0.09
R²	34%	87%	23%	9%

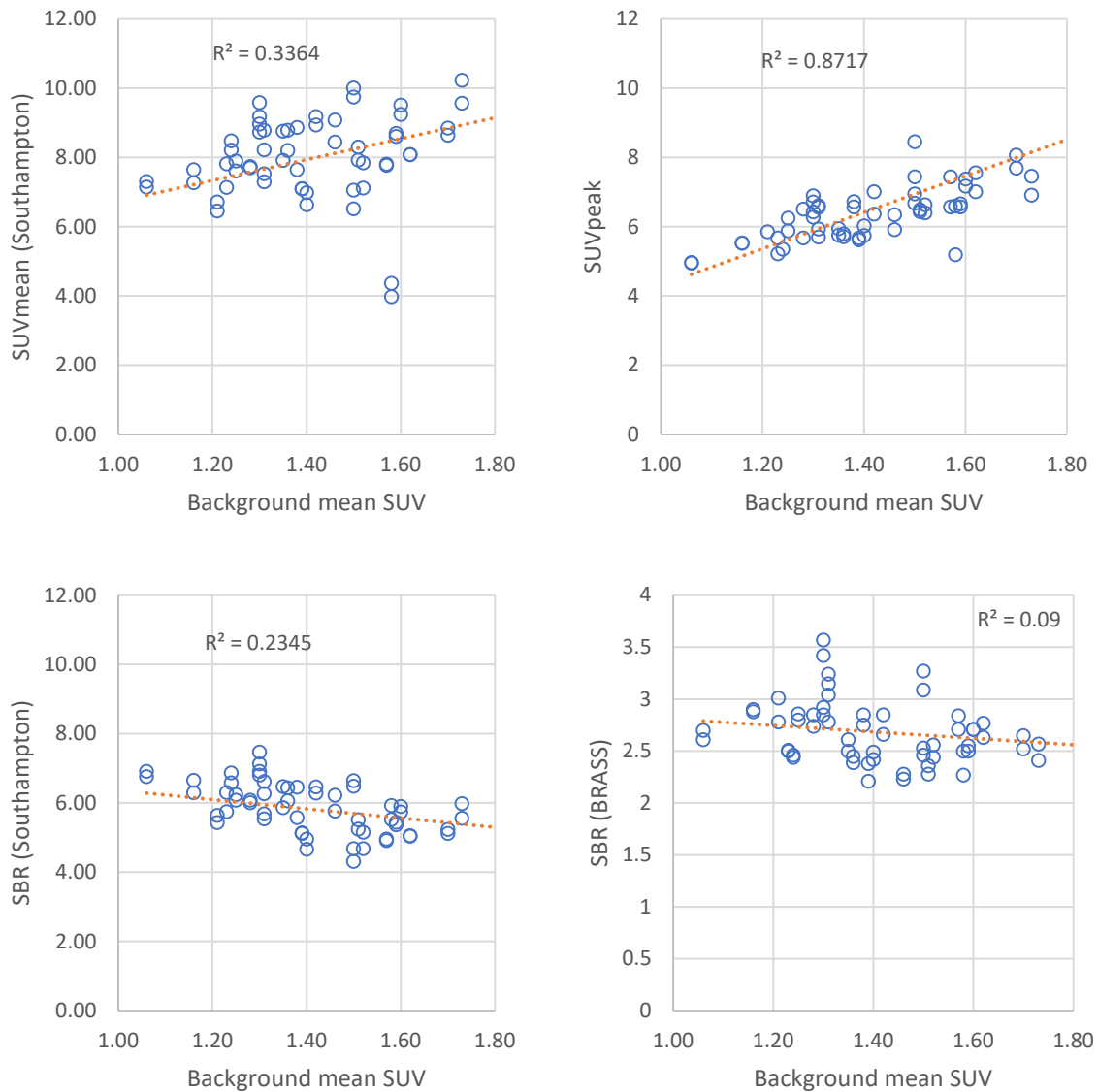


Figure 46: Striatal SUV_{mean} and SUV_{peak} plotted against background SUV_{mean} (top), showing a positive correlation. The correlation between SBR and background signal is less obvious.

Display of images with absolute SUV scale

Figure 47 shows examples of SUPeRB images displayed using the standard relative scale and the absolute SUV scale. The abnormal MCI-LB images can be identified with the relative scale by examining the background appearance, but with the SUV scale the loss of striatal uptake is immediately apparent. However, variation in the appearance of images due to the background variation seen above may limit the utility of the scale. Figure 48 shows three control examples with different background SUV levels displayed with the absolute scale, showing different striatal appearances.

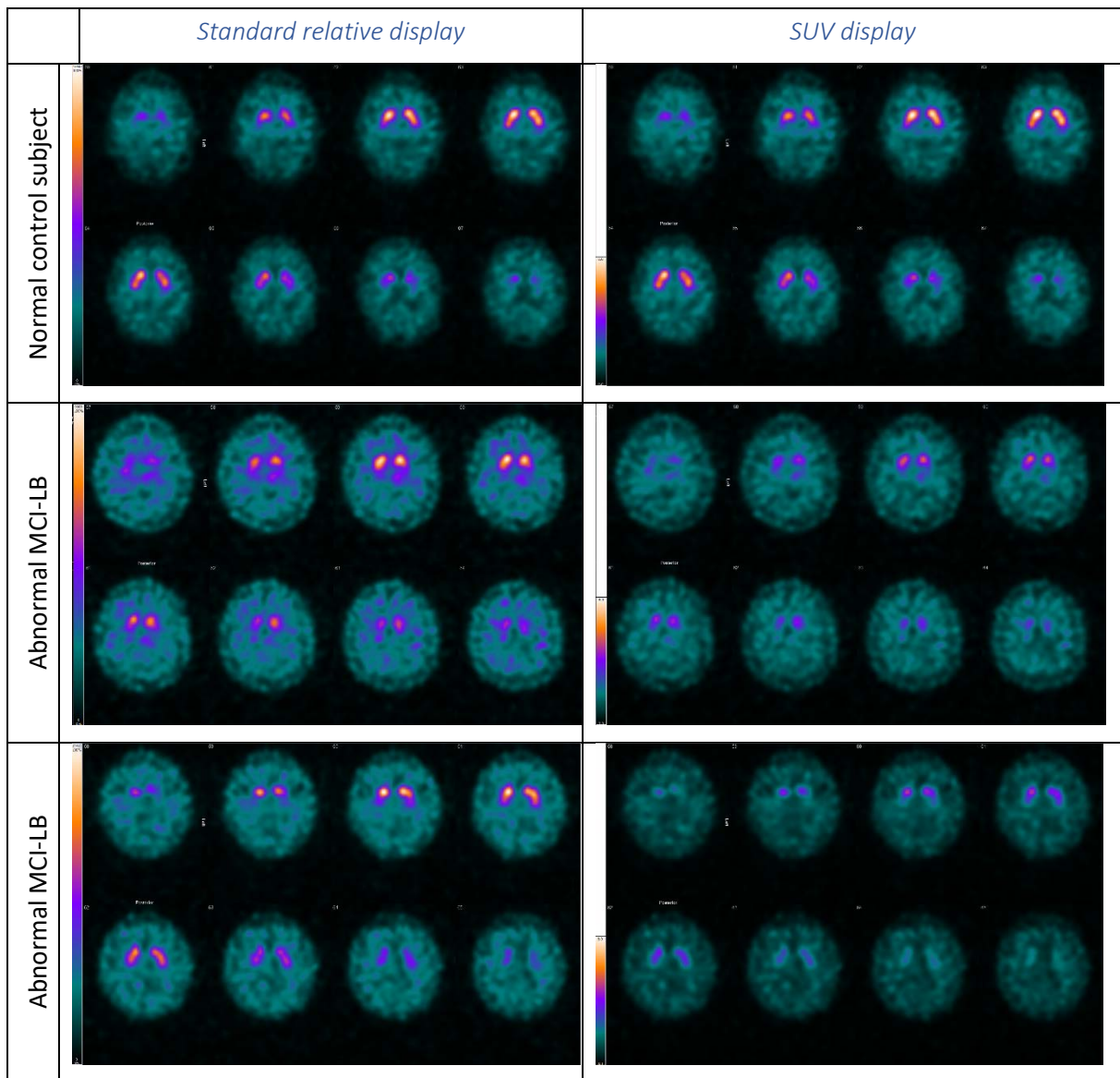


Figure 47: A normal control and two abnormal MCI-LB scans displayed with the standard relative display used in clinical practice for visual reporting (left) where the images are scaled relative to the maximum voxel and displayed using an absolute scale (right) relative to a maximum SUV of 8.

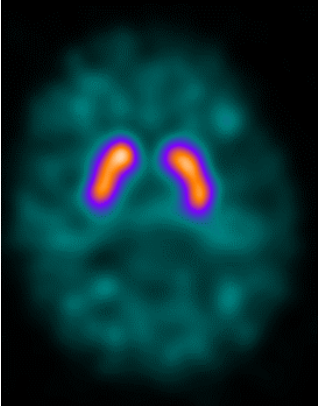
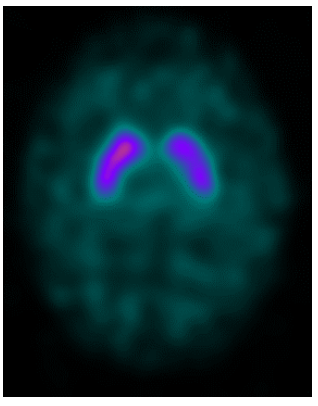
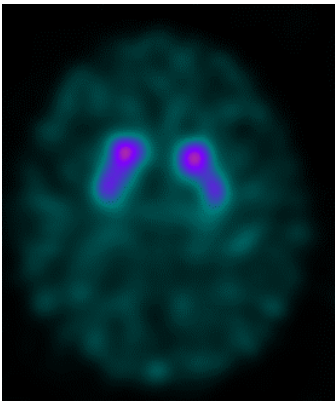
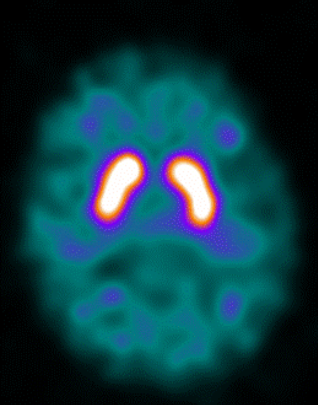
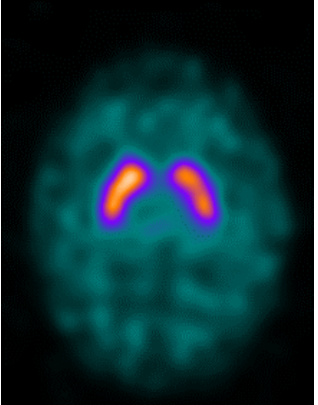
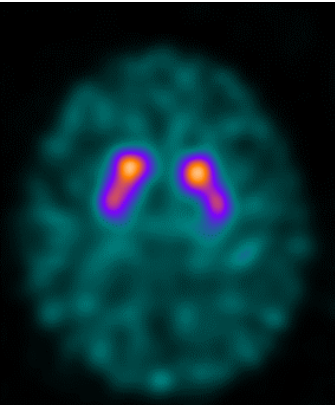
	Example 1: SUV _{peak} = 9.2 SUV _{mean} background = 1.93	Example 2: SUV _{peak} = 6.1 SUV _{mean} background = 1.55	Example 3: SUV _{peak} = 5.8 SUV _{mean} background = 1.12
Scaled to SUV 12 max			
Scaled to SUV 8 max			

Figure 48: Control images with differences in SUV showing quite dramatic differences in image appearance. The images are scaled to a maximum SUV of 12 in the top row and 8 in the bottom row. The background SUV appears linked to the striatal appearance in the first two examples.

6.2.4 Discussion

In this dataset of MCI patients, the accuracy of SUV_{mean} and SUV_{peak} were both lower than BRASS SBR, although not significantly so. At first glance this appears to be in line with the Japanese studies comparing SUV and SBR in dementia, which found slightly lower diagnostic accuracy with SUV, using ROC analysis [196, 197]. However, in our dataset, SUV and SBR calculated with the Southampton method both gave the same results, both lower than BRASS SBR. This suggests that the Southampton method, or more likely its implementation here using image registration, may not be ideal for obtaining volumes of interest for either SBR or SUV. This will be explored further with phantom studies in the next section. We would have liked to have used absolute quantification values from BRASS, to enable more meaningful comparison with BRASS SBR to be made, but this is not currently supported. It is hoped that this could be made available for research use in future. We assessed SUV_{mean} using the Southampton method as a proof-of-concept, with the possibility of requesting modifications to BRASS

in future to support SUV-SPECT without the reliance on a background region, if SUV_{mean} seems a promising parameter.

While we cannot disregard SUV_{mean} as a useful metric, the evidence here suggests that we should reject SUV_{peak} . The SUV_{peak} accuracy results were even lower than for SUV_{mean} , with some extreme outliers in the controls and possible outliers in other categories. A common pattern of dopaminergic loss is for the putamina to appear reduced compared to the caudate (Figure 49), which can lead to visually abnormal scans with normal SUV_{peak} values. This was suspected to be a limitation of SUV_{peak} from the outset, but as SUV_{peak} values are output by the software alongside SUV_{mean} this seemed a good opportunity to confirm or refute our assumptions. The SUV_{peak} results in controls were also highly dependent on background SUV levels, suggesting they cannot be an accurate reflection of dopaminergic function.

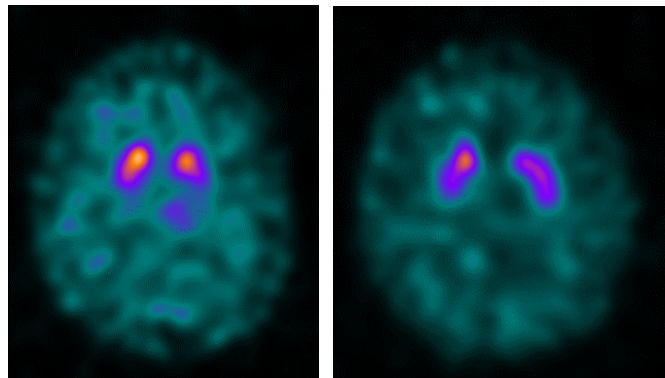


Figure 49: Left: visually abnormal MCI-LB scan with SBR, SUV_{mean} and SUV_{peak} all within normal limits. Right: visually abnormal MCI-LB scan with abnormal SBR and SUV_{mean} but right SUV_{peak} (in caudate) within normal limits.

Display of images with a relative scale has the technical disadvantage of being dependent on a single voxel (the maximum), which means the background appearance images is inherently quite uncertain, particularly in cases of low striatal uptake. The absolute scale could therefore offer advantages for visual ratings. In Glasgow, the SUV scale has been compared to the standard relative scale and raters were found to be more accurate when using it (data presented at BNMS conference 2017 by Dr. Colin Brown). However, a different reconstruction method (FBP) was used for the relative scale so the images were not compared like-for-like and the results may not hold using the more sophisticated reconstruction methods used for visual ratings in SUPeRB. This has not been explored further as part of this thesis due to the considerable time required from raters but does appear to warrant further investigation in future.

Data from our controls suggests that there is a wide variation in normal image appearances when scaled to an absolute SUV scale, more so than seen with a relative scale. This may reduce the specificity of such a scale in practice. The variation appears to be related to the background SUV demonstrated

above. Variation in background SUV may be due to factors such as incorrect recording of activity, assay times or patient weight, normal physiological variation in the amount of FP-CIT crossing the blood-brain barrier or imperfect normalisation to tracer volume of distribution using the patient weight method.

A potential limitation of accuracy of the reconstruction used in this work was that we used uniform attenuation and scatter correction, rather than CT based methods. This was so that the results were more comparable with BRASS SBRs from the previous section and also more generalisable to other centres. The impact on kBq/ml accuracy will be investigated in the next section using the striatal phantom filled with known activity concentrations.

6.2.5 Conclusion

The accuracy of striatal SUV_{mean} and SBR calculated with the Southampton method are similar. However, both appear less accurate than SBR calculated with BRASS. We hope to be able to evaluate SUV_{mean} calculated with BRASS at a later date following software modifications. Visual rating using the absolute SUV scale appears promising and worth future evaluation, although the specificity may be limited by substantial normal variation in SUV, which appears to be related to the weight-based normalisation method.

6.3 Advanced FP-CIT reconstruction methods: a phantom study

6.3.1 Introduction

As previously discussed, there are potential benefits to be gained from absolute quantification, which requires both accurate measurement of activity concentration and appropriate normalisation to patient size. Accuracy can be difficult to assess in patient studies as both these factors are combined, but the results in the previous section suggest the accuracy of striatal SUV may be limited by the normalisation step. In this section we consider the accuracy of kBq/ml quantification in a phantom study where normalisation to give SUV is not required.

So far in this thesis the method used for ^{123}I -FP-CIT reconstruction has not taken into account the effect of septal penetration of high energy photons through the collimator. Towards the end of this fellowship a more advanced reconstruction algorithm became available within Hermes HybridRecon for Siemens LEHR collimators, which Hermes term full collimator modelling (FCM). The FCM algorithm corrects for the septal penetration of high energy emissions, as described in Chapter 3. Use of the FCM reconstruction to correct for septal penetration and collimator scatter has the potential to improve the accuracy of FP-CIT quantification with LEHR collimators, whilst maintaining high resolution. Hermes use the label "MCRR_ACSC" for the reconstructed images, with MC indicating that the full Monte Carlo collimator model has been used. Note that Monte Carlo simulation is also used in the standard reconstruction used in previous sections, to estimate the scatter occurring within the patient - in the new MCRR version it is also used to model the interactions with the collimator.

Both the standard and the full Monte Carlo reconstructions can be used with either uniform attenuation and scatter correction, where the head is modelled as water, or with CT (measured) attenuation and scatter correction, where the attenuation coefficients to apply are determined from the CT scan. CT ACSC should in theory give a more accurate reconstructed image than ACSC with the uniform assumption. However, most centres, including Newcastle Hospitals, do not perform SPECT-CT imaging for FP-CIT routinely. Some authors have shown CT attenuation correction to be no more accurate than uniform correction for FP-CIT [102, 162, 198], but it should be noted that the reconstruction method were different from those considered here. CT was added to the SPECT protocol for SUPeRB so that it could be compared with uniform correction with these advanced reconstruction methods, and recommendations made for clinical practice and future research studies.

Hermes recommend that a higher number of iterations is used for ^{123}I SUV-SPECT than is recommended for BRASS: 4 iterations and 30 subsets (equivalent to 120 overall iterations). The protocol recommended for SBR calculation with BRASS uses 16 iterations and 4 subsets (equivalent to 64 overall iterations). The differences between the methods with 120 and 64 iterations was explored in the phantom with recommendations for clinical scans made.

In this section SPECT images were obtained by scanning a phantom filled with a range of different known activities, chosen to cover the range of SBRs encountered in clinical practice. The aim of this work was to determine the accuracy of SBR and activity concentration measurement on SPECT images generated from the acquired phantom data using a range of different reconstruction methods. Images reconstructed with our standard reconstruction protocol were compared with a range of advanced methods combining CT attenuation and scatter correction and Monte Carlo collimator modelling, as described below.

6.3.2 Methods

Reconstruction methods

The reconstruction methods used in this section are summarised in Table 26. Apart from the changes in ACSC method and the use of FCM, the reconstruction parameters recommended by Hermes for FP-CIT scans to be analysed with BRASS were used, i.e. 16 iterations with 4 subsets and a 0.7cm Gaussian post-filter for smoothing. As all reconstructions here use OSEM with some form of ACSC this was removed from the name for clarity. The reconstruction labelled RR_uniform in this section is therefore exactly the same as the standard reconstruction referred to as OSEM_RRACSC elsewhere in this thesis. All four reconstructions described here were also calculated with the increased number of iterations (4 iterations 30 subsets) recommended by Hermes for SUV-SPECT.

Table 26: Description of the four reconstruction methods used in this section.

	Attenuation map	Patient scatter correction	Collimator model
RR_uniform (standard method)	Uniform density of water assumed	MC simulation based on uniform density	Gaussian depth dependent resolution recovery model, no correction for septal penetration
RR_CT	Measured densities (CT)	MC simulation based on measured densities	Gaussian depth dependent resolution recovery model, no correction for septal penetration
MCRR_uniform	Uniform density of water assumed	MC simulation based on uniform density	Gaussian RR + Full Monte Carlo collimator model
MCRR_CT	Measured densities (CT)	MC simulation based on measured densities	Gaussian RR + Full Monte Carlo collimator model

Striatal phantoms

The raw planar data from the striatal phantom acquisitions described in section 4.1 were used to reconstruct sets of transaxial slices with each of the different reconstruction methods – this slice data was then input into BRASS to determine SBR values for each method.

To recap, the striatal phantom (RSD Radiology Support Devices, Long Beach, CA, USA) consists of plastic striatal compartments that are filled with ^{123}I solution, which sit within a brain background filled with more dilute solution (see Figure 17, section 6.1.2 for illustrations). The exact activity concentrations in the phantoms are determined using aliquots from each compartment measured in a sample counter. The images were reconstructed with each of the reconstructions in Table 26, with both sets of iterations. The measured SBRs and activity concentrations, are given in Table 27 and Table 28. Different activity concentrations are used for the right and left striata so that two SBRs could be acquired in the same scan. SBR is calculated as follows:

$$SBR = \frac{kBq/ml_{striatum} - kBq/ml_{background}}{kBq/ml_{background}}$$

The left and right striatal compartments are filled with solutions of different activity concentration to obtain two independent measurements per scan. The SBR values in the phantom are altered between scans by leaving the striatal compartments unchanged and increasing the background concentration by adding more ^{123}I activity. This phantom method was developed for calibrating SBR values obtained with different gamma cameras [158] and is here extended to activity concentration. The background was increased twice, resulting in three different configurations and six SBR values per phantom imaging session. For the absolute striatal compartment measurements, only the first two configurations were used as the final increase in background concentration resulted in very low SBRs. A second set of phantom images acquired on a separate occasion with different activity concentrations and SBRs was used to increase the number of striatal activity concentrations sampled. These sets of images are referred to as session A and session B.

Table 27: SBRs for the striatal phantoms, measured using a sample counter

	Right striatum	Left striatum
	SBR	SBR
Session A Background 1	5.82	4.56
Session A Background 2	2.37	1.75
Session A Background 3	0.70	0.39
Session B Background 1	13.7	8.71
Session B Background 2	5.61	3.36
Session B Background 3	2.07	0.91

Table 28: Measured phantom activity concentrations for background and striatal compartments. The third configuration with the background increased again was not used for kBq/ml measurements as the striatal to background ratios were so low.

		Session A	Session B
		Activity concentration	Activity concentration
		kBq/ml	kBq/ml
Configuration 1	Background	5.58	3.34
	Right striatum	38.0	49.2
	Left striatum	31.0	32.4
Configuration 2 (increased background)	Background	8.89	5.96
	Right striatum	30.0	39.4
	Left striatum	24.5	26.0

Calibration factor

The calibration factor used in section 6.2 to convert between image count density and activity concentration was recalculated with all the reconstruction protocols described above. This was to determine if there were differences between protocols that would be likely to affect apparent uptake.

The calibration factors are similar for different number of iterations and between RR and MCRR. Use of CT ACSC instead of uniform gives higher apparent activity concentrations (by around 5-6%). However in this section the same cps/MBq calibration factor was used throughout, to make it more straightforward to compare methods.

Calculating SBRs and activity concentration on phantom images

SBRs were calculated within BRASS and by using the Southampton method. Activity concentrations were calculated using the Southampton method only. The Southampton method, described in section 6.2, estimates striatal kBq/ml via registration of the phantom to the template (Figure 50) and

assumption of a fixed 11.2 ml striatal volume. The background activity concentration was estimated from the mean kBq/ml in the occipital VOI shown in Figure 50. As discussed in section 6.2, striatal activity concentration cannot be measured within BRASS directly because an additional scaling factor is applied. SUV_{peak} was not investigated using the striatal phantom as it is only output for SUV data, not kBq/ml.

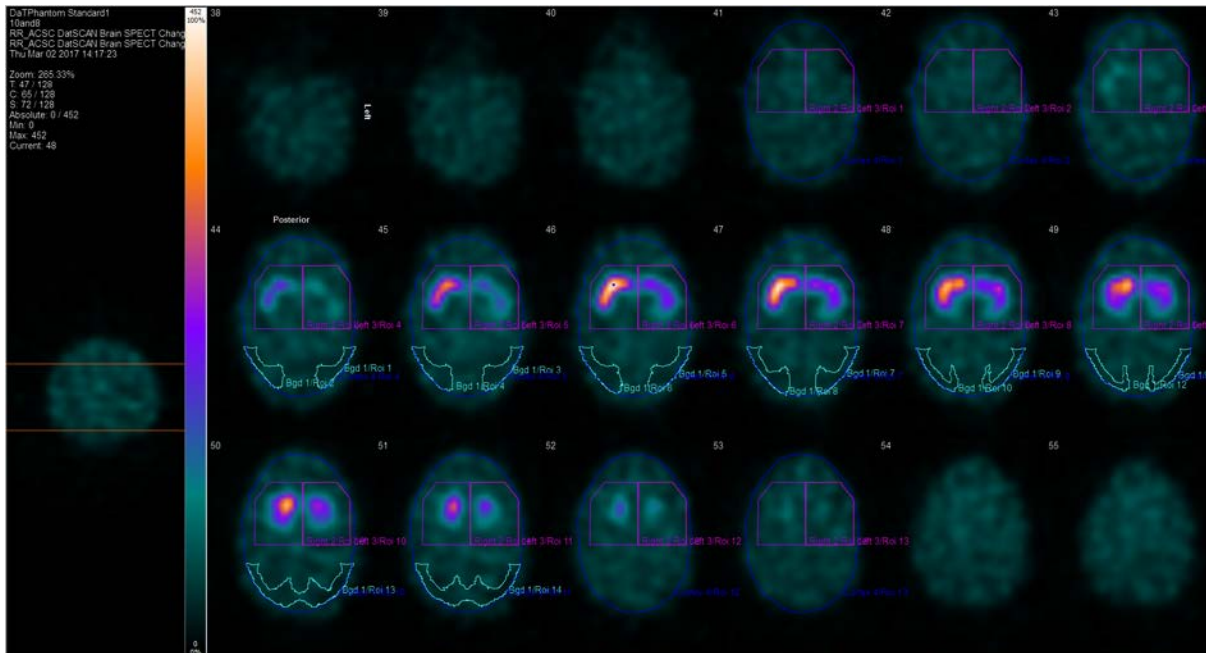


Figure 50: Southampton method VOIs registered to the striatal phantom. The occipital VOIs had to be resized because the template was slightly big for the phantom.

Effectiveness of ACSC methods

A second background measurement was taken using a 6cm diameter spherical VOI placed more centrally within the phantom (Figure 51). The occipital and central background measurements were compared as an assessment of attenuation and scatter correction.

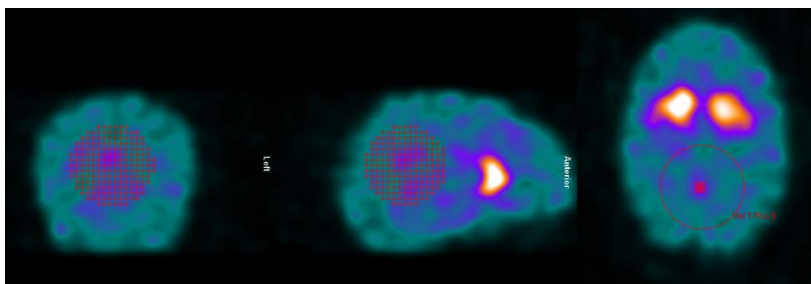


Figure 51: Additional background VOI used to compare mean kBq/ml with occipital background obtained from the Southampton method processing.

6.3.3 Results

Differences between number of iterations

There was very little difference between results obtained with 16 iterations and 4 subsets (BRASS reconstruction) and those obtained with 4 iterations and 30 subsets. All results from now on refer to the BRASS protocol of 16 iterations and 4 subsets.

BRASS SBR differences between reconstruction methods

All reconstruction methods gave BRASS SBRs that were much lower than the measured values (around 50-70%), which is expected due to the partial volume effect [111, 199]. Images reconstructed with CT ACSC give lower SBRs than with uniform ACSC, much more so than expected from the small differences in calibration factor.

A plot of the BRASS SBRs against measured SBR for each reconstruction method is shown in Figure 52, with the percentage differences between BRASS and measured SBR, and between RR_uniform and other methods in Figure 53. The highest SBRs, and therefore the closest to measured values, are given by the MCRR_uniform method, although these are still around half the measure values. There is more discrepancy between the reconstruction methods at low and borderline values than for high SBRs. There is about a 50% difference between reconstruction methods in the clinically relevant mid range (BRASS SBRs of around 1.5 to 3), but very little difference for high uptake.

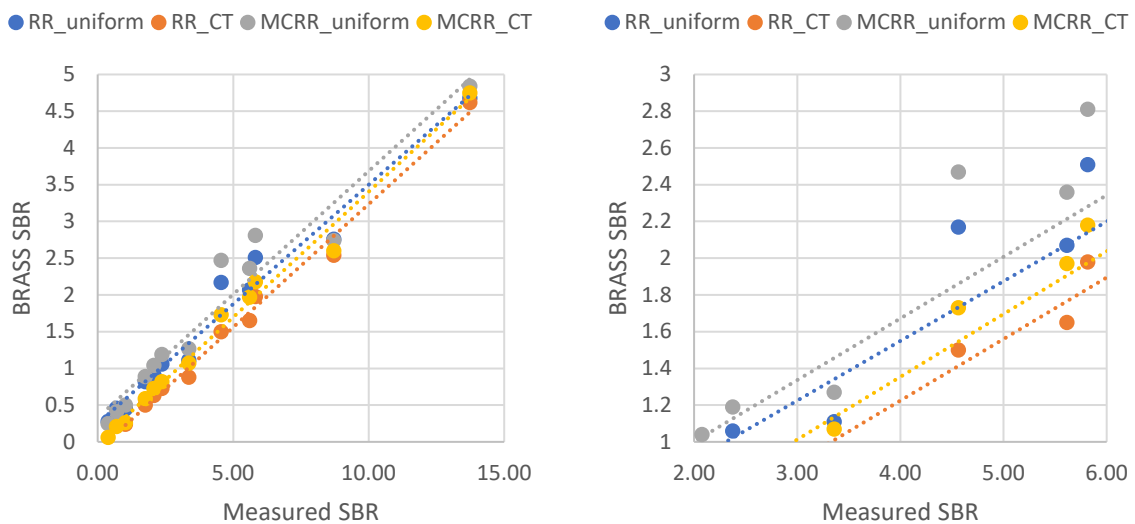


Figure 52: BRASS SBR against measured SBR for the four reconstruction methods over the full phantom SBR range (left) and restricted to a more clinically relevant range (right).

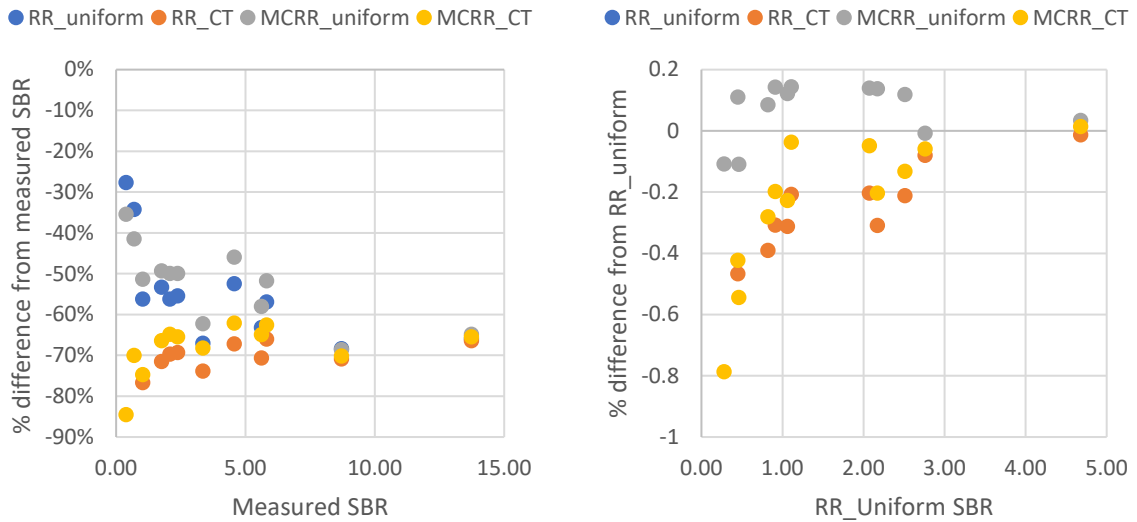


Figure 53: The percentage difference between BRASS SBR and measured SBR for all reconstruction methods (left) and between standard RR_Uniform and other methods (right).

Southampton method SBR differences between reconstruction methods

Plots of Southampton SBRs in the striatal phantoms against known values are shown in Figure 54. Both RR and MCRR methods with uniform attenuation and scatter correction appear to overestimate SBR, particularly at high SBR, but with CT corrections SBR is underestimated at all SBR values suggesting a systematic error. In the clinically relevant range (measured SBR of around 5) the RR_uniform reconstruction gives reasonably accurate SBRs.

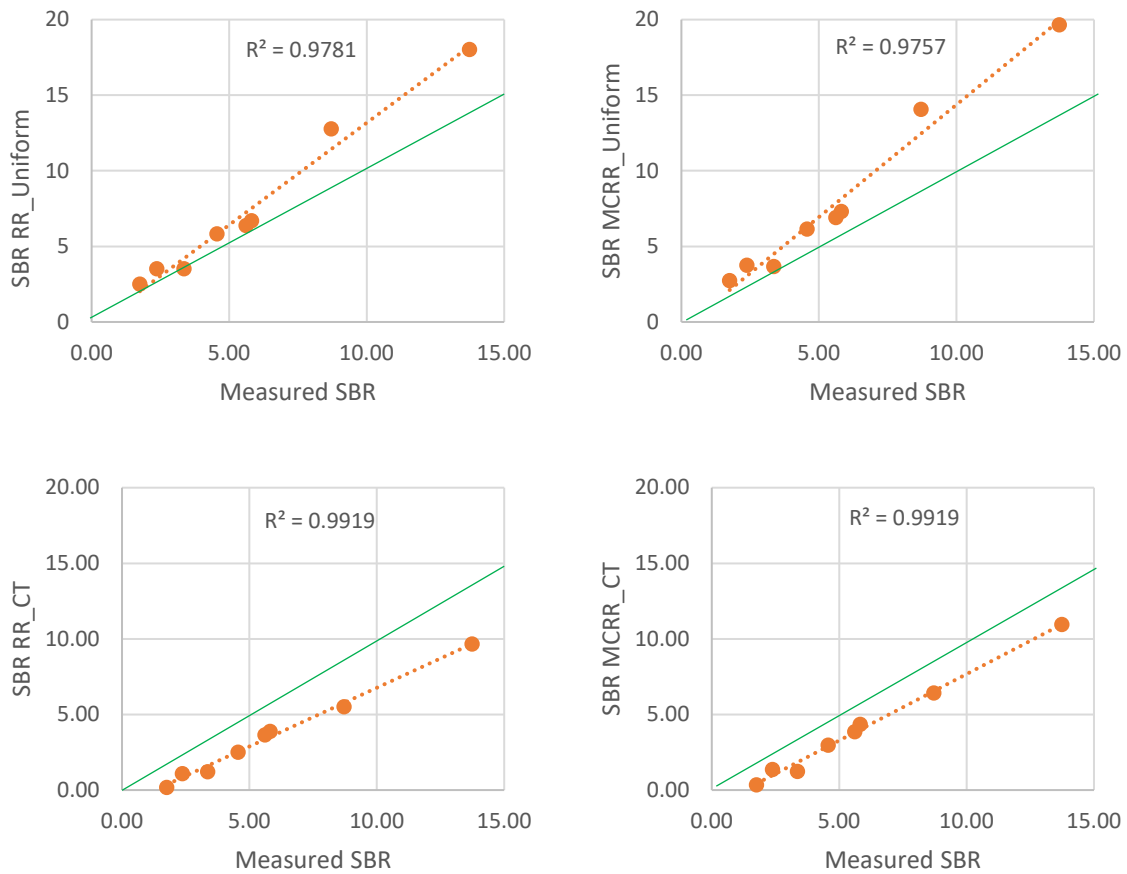


Figure 54: Plots of Southampton SBR results in striatal phantom against measured SBR for all four reconstruction methods. The line of identity is shown in green.

Activity concentration differences between reconstruction methods

Although the calibration factors obtained with the uniform phantom were similar to within 10% for the various reconstruction methods, the background activity concentrations in the striatal phantom images were much higher with CT ACSC than with uniform ACSC. The reconstructions using the full Monte Carlo collimator model (MCRR) also gave higher values for background activity concentration than the basic Gaussian collimator model (RR). Figure 55 shows the difference in background activity concentration, striatal activity concentration and SBR for each of the methods. It is demonstrated that although CT ACSC gives higher results than uniform ACSC for both background and striatal activity concentration, the difference between the methods is less for the striata. This results in CT ACSC SBRs that are lower, and less accurate, than those obtained on uniform ACSC images, as seen on the plots in Figure 54 above. Higher background with CT ACSC results in lower SBR.

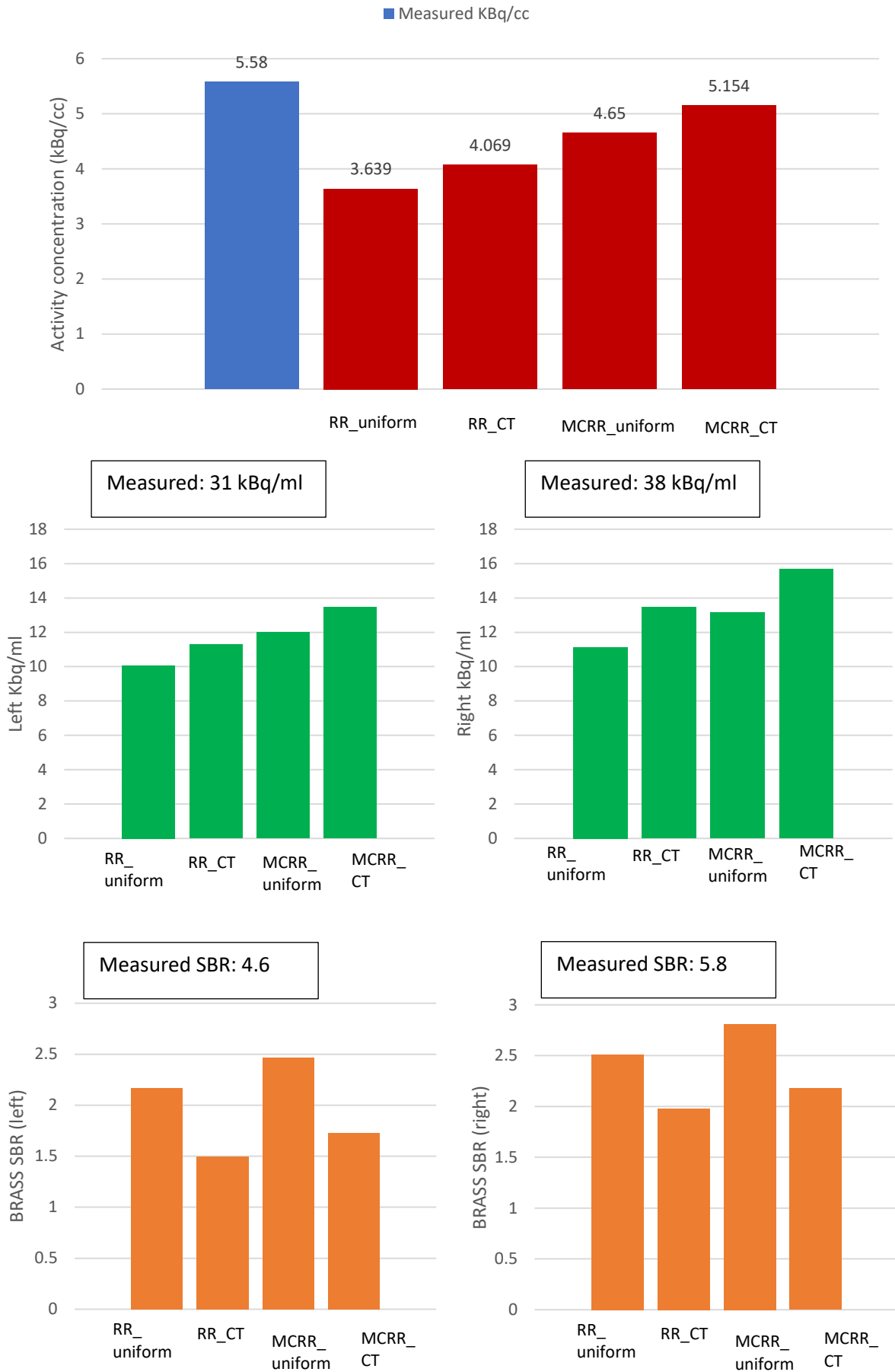


Figure 55: Background activity concentration (top), striatal activity concentration (middle) and BRASS SBR (bottom) for the different reconstruction methods. The MCRR_CT reconstruction gives the background result closest to the measured activity concentration of 5.58 kBq/ml. The true activity concentrations and SBRs are noted on the plots.

Table 29 shows the percentage difference in activity concentration for the advanced reconstruction methods compared to the RR_Uniform method. The background activity concentration differences for each reconstruction are shown for the phantom scanned on two separate occasions with different activity concentrations, alongside the striatal activity concentration differences. All the methods give higher activity concentrations than RR_Uniform, but the percentage differences are very variable.

Table 29: Percentage difference in activity concentration between standard RR_Uniform and other reconstruction methods.

		Background	Right striatum	Left striatum
Session A	RR_CT	42%	21%	12%
	MCRR_uniform	9%	18%	19%
	MCRR_CT	55%	41%	34%
Session B	RR_CT	93%	91%	81%
	MCRR_uniform	8%	11%	7%
	MCRR_CT	107%	110%	99%

Effectiveness of attenuation and scatter correction

The difference between the occipital background at the edge of the brain (i.e. the background VOI used in the Southampton method) and a 6 cm spherical background VOI placed more centrally is shown in Table 30 for uniform and CT ACSC methods. The activity concentration for the central VOI is higher than the occipital VOI, with the difference more marked for uniform ACSC, indicating possible overcompensation for attenuation and scatter. This may contribute to the SBR differences between uniform and CT methods.

Table 30: Percentage difference between occipital background activity concentration in the phantom and spherical VOI placed more centrally with different reconstruction methods.

	Session A (5.58 kBq/ml)				Session B (3.34 kBq/ml)			
	RR		MCRR		RR		MCRR	
	Uniform	CT	Uniform	CT	Uniform	CT	Uniform	CT
Occipital-central	-0.46	-0.13	-0.61	-0.22	-0.64	-0.16	-0.82	-0.20
Percentage difference	-13%	-3%	-15%	-4%	-30%	-5%	-33%	-6%

Accuracy of striatal activity concentration with the Southampton method

Plots of striatal activity concentration for the Southampton method are shown in Figure 56. Both RR_uniform and MCRR_CT methods give noisy estimates of striatal activity concentration compared to the SBR results above. The correlation is better with CT ACSC, but there is still low agreement with the measured values, with values underestimated by a factor of 2-3.

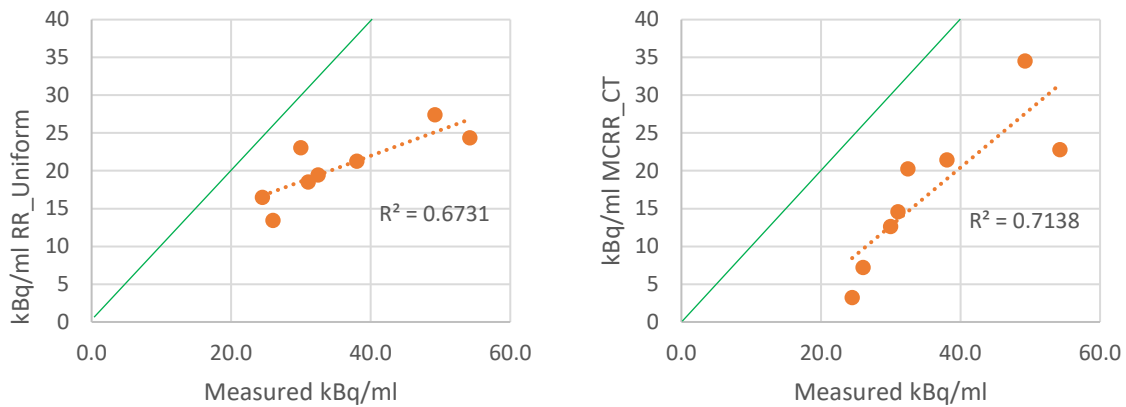


Figure 56: Plots of Southampton striatal activity concentration results in phantom against measured activity concentration for RR_Uniform (left) and MCRR_CT (right). The line of identity is shown in green.

6.3.4 Discussion

The reconstruction method had a greater impact on both SBRs and striatal activity concentrations than was expected from the calibration factor differences. The calibration factor is obtained using a uniform phantom so it seems likely that this is linked to differences in geometry.

CT ACSC gave higher activity concentration results than uniform correction in the occipital background region. Striatal values were also higher with CT ACSC than with uniform corrections, but there was less of a difference. This results in lower (and therefore less accurate) SBRs for CT ACSC than uniform. It is not clear why this is the case, but the difference between the occipital and central background VOIs with uniform correction suggests that attenuation and scatter correction differences may contribute to the differences between CT and uniform SBR. The effect of the different reconstruction methods on clinical SBR and diagnostic accuracy will be investigated in the next section.

The large differences between reconstruction methods at lower SBRs suggest that problems with image registration at lower relative striatal activity may have introduced uncertainty. The phantom “brain” is smaller than the BRASS template, so the background VOIs are closer to the edge of the “brain” than they would be in patients, which may have contributed to random error. The original paper on calibrating gamma cameras with a striatal phantom [158] uses the Southampton method with a larger background VOI and reports a random error of 14%, but does not discuss uncertainty at low SBRs. Dickson et al. [156] noted that different activity distributions in FP-CIT phantom images may require different numbers of iterations to achieve convergence. This may explain the wide variation in the phantom striatal activity concentration measurements and suggests that it would have been preferable to alter striatal activity concentration and keep background concentration constant. This is difficult to achieve in practice.

Estimating striatal uptake with the Southampton method using background reference regions appears to introduce uncertainty. If BRASS is modified in future it would be useful to see whether kBq/ml values measured directly give more accurate results.

Accurate measurement of kBq/ml values is a pre-requisite for clinical quantification with SUV-SPECT, but this alone is not sufficient to ensure accuracy of absolute quantification as there are other factors that affect SUVs. Work presented in the previous section using control SUV images showed that in practice there remain considerable obstacles to implementation due to issues with normalisation.

6.3.5 Conclusion

Advanced reconstructions methods appear to provide accurate activity concentration results, judging by the background measurements. Estimates of background activity concentration in the phantom are higher and more accurate with CT attenuation and scatter correction with Monte Carlo collimator modelling (MCRR_CT) than with uniform attenuation and scatter correction. However, SBRs are actually lower and therefore less accurate with MCRR_CT, because of the higher background. The use of advanced reconstruction methods for relative quantification in clinical images will be explored for SUPeR patients in the next section.

6.4 Can SBR using advanced FP-CIT reconstruction methods improve early diagnosis?

6.4.1 Introduction

The iterative reconstruction method used for clinical FP-CIT images so far in this thesis, OSEM with resolution recovery and uniform attenuation and scatter correction (RR_uniform), is already advanced compared to the uncorrected OSEM and FBP methods still used clinically in some departments. The Society of Nuclear Medicine FP-CIT guidelines state that “iterative reconstruction is preferred, but filtered back projection may be used” [112] and the European Association of Nuclear Medicine also lists both FBP and iterative reconstruction as options [113]. Both recommend attenuation correction, either with uniform method or CT but neither discuss additional image corrections, including scatter correction, which Hutton states is as essential for accurate quantification as attenuation correction [92]. However, these guidelines were published in 2011 and 2009 respectively, before more advanced image corrections became routinely available.

Most centres, including Newcastle Hospitals, do not perform SPECT-CT imaging for FP-CIT routinely. Some authors have shown CT attenuation correction to be no more accurate than uniform correction for FP-CIT [102, 162, 198], but it should be noted that the reconstruction methods were different from those considered here. In the previous section more advanced FP-CIT reconstruction methods using CT ACSC and Monte Carlo collimator modelling were introduced. CT was added to the SPECT protocol for SUPeRb so that it could be compared with uniform correction with these advanced reconstruction methods, and recommendations made for clinical practice and future research studies.

In this section we compare SUPeRb subjects images reconstructed with our standard protocol with those using the full collimator model, both with and without CT attenuation and scatter correction. The aim is to assess the differences between SBRs for each method and to evaluate the impact of using advanced methods on diagnostic accuracy in MCI.

6.4.2 Methods

Reconstruction methods

The four reconstruction methods used in this section are the same as those used in the phantom study in the previous section: RR_Uniform, RR_CT, MCRR_Uniform and MCRR_CT. The SUV-SPECT reconstruction parameters, with 4 iterations and 30 subsets are not included in this section as the phantom data showed little difference between the two.

For consistency, the same initial cropping of the raw data was used for each reconstruction, and no image reorientation was performed, so that the input images into BRASS would all have the same geometry. When reconstructing the CT ACSC images, the SPECT-CT alignment was checked and

adjusted if necessary. The same shifts were applied to the RR_CT as to the MCRR_CT to remove additional sources of variation.

SUPERB subjects

The 31 healthy older adults were processed to give reconstruction-specific normal ranges, since the BRASS ENC-DAT database is only valid for the RR_Uniform reconstruction. We calculated the coefficient of variation of the control SBRs for each reconstruction.

Fifteen of the 16 MCI-AD cases and 25 of 27 the probable MCI-LB cases from section 6.1 were included in the comparison between AD and LB groups and for the diagnostic accuracy calculations. The MCI-AD case was excluded because no CT was acquired due to a scanner fault. The 2 MCI-LB scans obtained at other centres were excluded as the scanner was a different model and no CT was acquired. All the SUPERB MCI patients, including 11 less certain possible MCI-LB cases were included to assess any systematic differences between reconstruction methods, but only the probable cases were included in the diagnostic accuracy calculation.

SBRs and Z scores

SBRs were calculated using BRASS, with the camera correction feature turned off as this has been set up for the standard reconstruction method. BRASS Z scores were not used for the same reason, so age correction was not applied. However, the controls were age matched to the MCI subjects (see section 6.1.2). SBRs were recorded for the left and right whole striatum and caudate and putamen sub-regions.

Systematic differences between the SBRs with each reconstruction method were assessed using Bland-Altman plots, with paired t-tests used to check for a significant difference, after checking for normality with histograms.

The mean and standard deviation of the 31 controls was calculated for each of the four reconstruction methods. This data was used to calculate Z scores for the MCI subjects, where $Z = (\text{subject mean} - \text{control mean}) / \text{control SD}$. The Z scores were used for sensitivity and specificity calculations, with a cut-off of -2. Note that this gives the same result as using 2 SD below the control mean as a cut-off, but makes the presentation of results more in keeping with standard BRASS output, and makes it easier to compare the reconstruction methods if there is a systematic difference in SBR. Any difference in accuracy between reconstruction methods was assessed via ROC analysis using the SigmaPlot statistical software package. Sensitivity, specificity and overall accuracy were calculated twice: once using just the whole striatum Z scores and one using z scores for all regions, with the scan being classed as abnormal if any Z score falls below -2. The former approach should be less dependent on mis-registration errors, but the latter more sensitive to any early changes in uptake predominantly

affecting the putamina. Typically, subregions are used in clinical practice as they are available in BRASS and DaTQUANT.

6.4.3 Results

SBR differences between reconstruction methods

The patient results show a systematic difference between the standard RR_uniform reconstruction and the same reconstruction using measured (CT) ACSC, with the RR_CT being 14% lower on average (Table 31). The full Monte Carlo model acted in the opposite direction, being 15% higher on average with uniform ACSC. These two effects cancel each other out when comparing the standard reconstruction with the most complex MCRR_CT reconstruction. A comparison between two CT reconstructions with and without the full Monte Carlo collimator model also shows an increase in SBR with the model applied. The Bland-Altman plots (Figure 57) reveal that the difference between uniform and CT ACSC does not depend on SBR. However, the difference between reconstructions with and without the full collimator model increases with SBR, suggesting that this model could increase diagnostic accuracy. However, Table 32 shows that the percentage difference in SBR between the MCI-AD and probable MCI-LB groups is the same for all the reconstruction methods. The coefficient of variation based on control data is also the same. The plots in Figure 58 show that the variation within both MCI-AD and MCI-LB data is much greater than the difference in mean SBR between the groups.

Table 31: Difference in SBR between the standard and more complex reconstruction methods, and between the two CT ACSC methods with and without full MC collimator model.

	RR_uniform vs RR_CT	RR_uniform vs MCRR_Uniform	RR_uniform vs MCRR_CT	RR_CT vs MCRR_CT
Mean difference (%)	-0.31 (-14%)	0.34 (+15%)	0.02 (+1%)	0.34 (+18%)
P value (paired t-test)	<0.001	<0.001	0.07	<0.001

Table 32: Differences in mean SBR between MCI-AD and probable MCI-LB groups for each of the reconstructions. SBRs are for the whole striatum.

SBR differences	RR_Uniform	RR_CT	MCRR_Uniform	MCRR_CT
MCI-AD mean	2.41	2.11	2.80	2.49
Prob MCI-LB mean	2.10	1.80	2.40	2.11
Difference	-0.31	-0.31	-0.40	-0.38
% difference	-13%	-15%	-14%	-15%
CoV (controls)	11%	12%	11%	12%

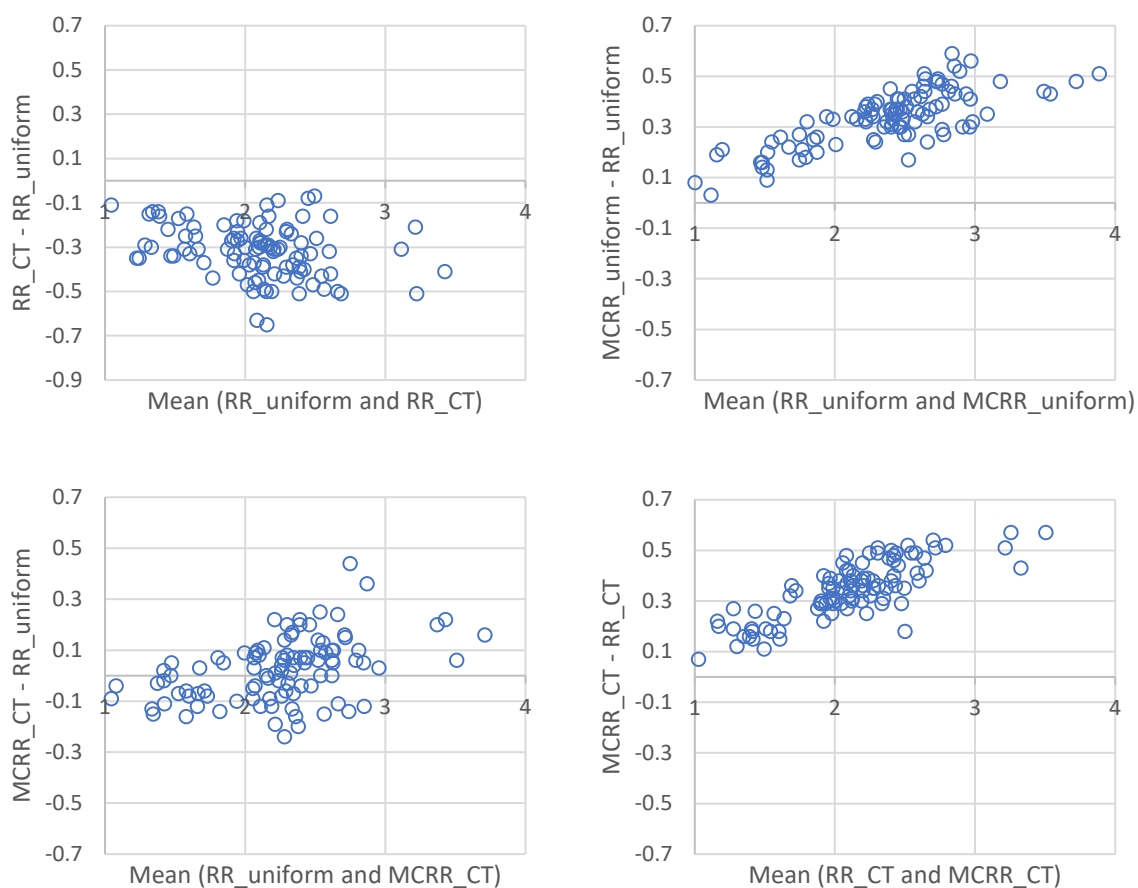


Figure 57: Bland-Altman plots showing the difference between the reconstruction methods using data for all MCI patients

Diagnostic accuracy

The normal cut-offs from the control dataset are given in Table 33 for each reconstruction. The cut-offs for the different reconstructions are in keeping with the phantom and MCI patient results.

Table 33: Mean control SBR and associated normal cut-off values for each region for the four reconstruction methods

Reconstruction method	Right striatum SBR	Left striatum SBR	Right caudate SBR	Left caudate SBR	Right putamen SBR	Left putamen SBR
RR_Uniform						
Mean	2.33	2.34	2.44	2.50	2.21	2.17
Cut-off	1.84	1.84	1.89	1.94	1.73	1.65
RR_CT						
Mean	2.05	2.02	2.16	2.16	1.93	1.87
Cut-off	1.57	1.49	1.61	1.60	1.47	1.30
MCRR_Uniform						
Mean	2.64	2.63	2.80	2.74	2.46	2.46
Cut-off	2.08	2.06	2.17	2.17	1.88	1.85
MCRR_CT						
Mean	2.40	2.37	2.57	2.49	2.22	2.24
Cut-off	1.86	1.78	1.92	1.87	1.72	1.61

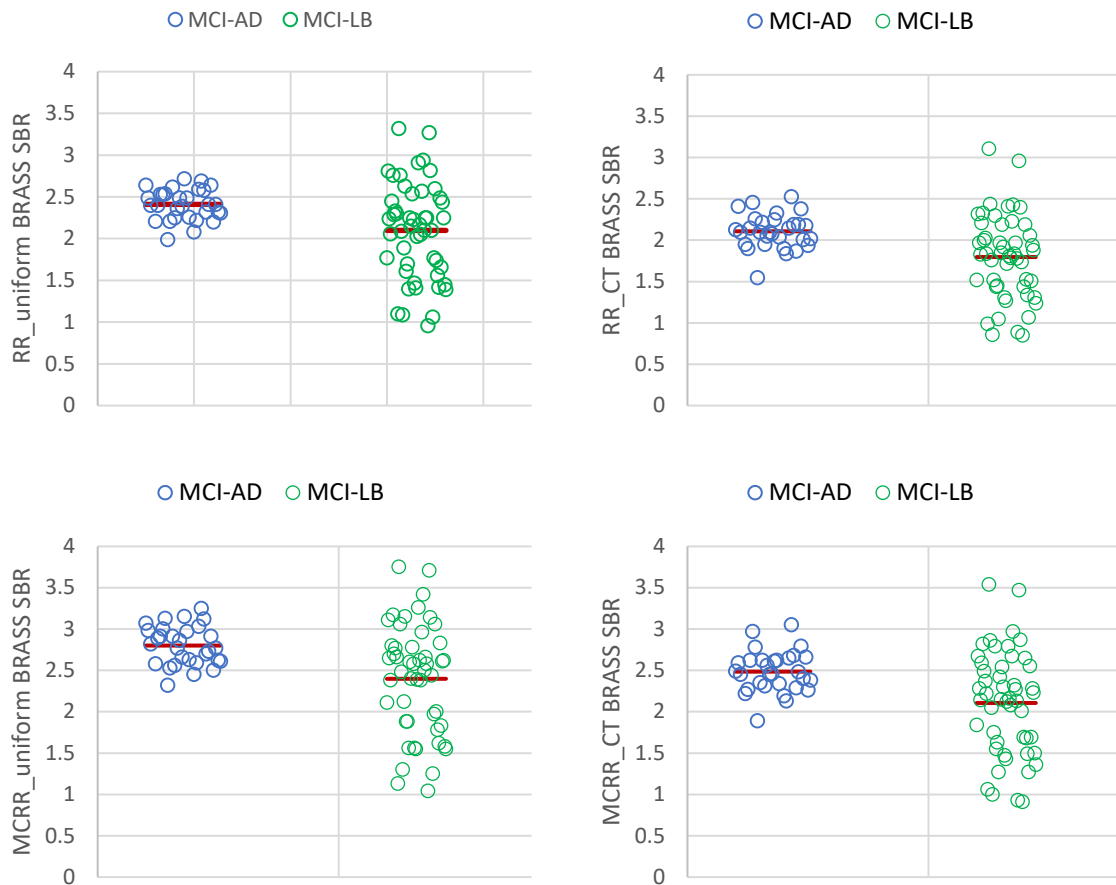


Figure 58: SBRs for the MCI-AD and MCI-LB groups for each reconstruction method, with the mean SBRs shown as red horizontal lines for each group.

The diagnostic accuracy is similar for all reconstruction methods, with only three discrepancies among the four methods (Table 34). The use of smaller caudate and putamen regions does not appear to improve accuracy. Note that the sensitivity values are slightly lower than in section 6.1 because the two abnormal MCI-LB scans acquired at other centres were excluded. The areas under the ROC curves are the same for all methods with p-values between 0.22 and 0.98 (Figure 59, Table 35).

Table 34: Left: diagnostic accuracy based on whole striatum Z scores for each of the four reconstruction methods. Right: diagnostic accuracy based on the minimum of striatum, caudate and putamen scores.

	Sensitivity	Specificity	Accuracy		Sensitivity	Specificity	Accuracy
RR_Uniform	10/25 40%	15/15 100%	25/40 63%	RR_Uniform	10/25 40%	15/15 100%	25/40 63%
RR_CT	9/25 36%	14/15 93%	23/40 58%	RR_CT	9/25 36%	14/15 93%	23/40 58%
MCRR_Uniform	9/25 36%	15/15 100%	24/40 60%	MCRR_Uniform	10/25 40%	15/15 100%	25/40 63%
MCRR_CT	10/25 40%	15/15 100%	25/40 63%	MCRR_CT	10/25 40%	14/15 93%	24/40 60%

ROC Curves

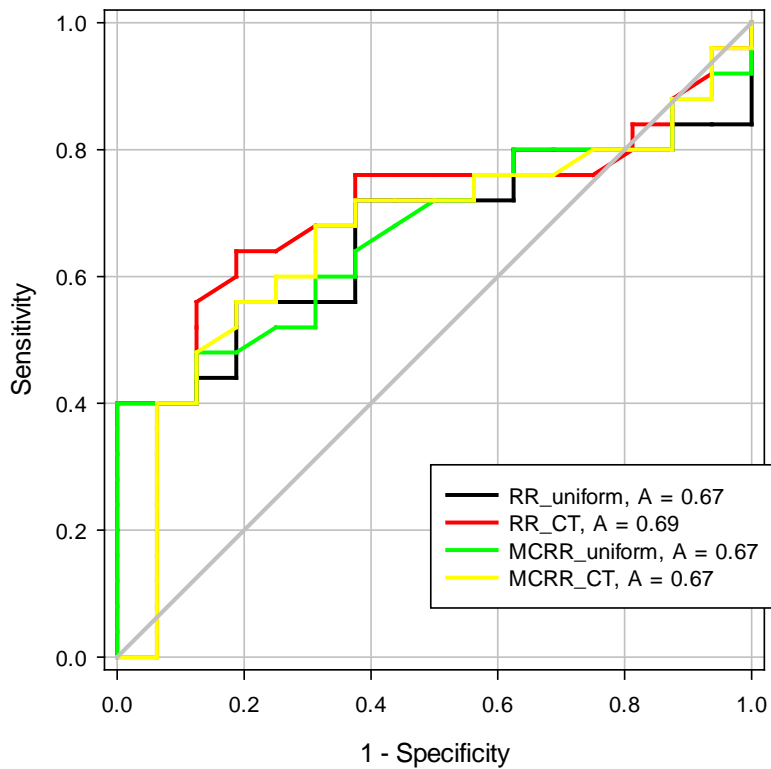


Figure 59: ROC curves for the four reconstruction methods for whole striatum SBR

Table 35: Differences between the area under the ROC curve for each reconstruction method, showing no statistically significant differences.

Pair	RR_uniform, RR_CT	RR_uniform, MCRR_uniform	RR_uniform, MCRR_CT	RR_CT, MCRR_uniform	RR_CT, MCRR_CT	MCRR_uniform, MCRR_CT
Area Difference	-0.02	0.00	0.00	0.02	0.02	0.01
Standard Error	0.04	0.02	0.04	0.04	0.02	0.04
95% Confidence Interval	-0.10 to 0.06	-0.05 to 0.04	-0.08 to 0.08	-0.07 to 0.10	-0.01 to 0.06	-0.08 to 0.09
ChiSquare, DF = 1	0.22	0.03	0.00	0.13	1.49	0.02
P Value	0.64	0.86	0.95	0.72	0.22	0.88

6.4.4 Discussion

The Bland-Altman plots with MCI data showed that CT attenuation and scatter correction gives systematically lower SBRs than standard uniform correction, which is what we saw with the phantom data in the previous section. The phantom results suggest that the SBR difference is predominantly due to higher background counts with CT ACSC than with uniform ACSC, possibly due to overcorrection

with the uniform method. Warwick et al. noted similar SBR differences in 14 patients reconstructed with uniform and CTAC, and suggested that small sub-voxel mis-registrations between SPECT and CT may be the cause of lower SBR with CTAC [162]. However it seems unlikely that this would give rise to systematically lower SBRs; random differences seem more likely. We manually registered the SPECT and CT data if it was necessary, although this can only correct for mis-registrations that are apparent on the pre-reconstruction QC tool provided by Hermes.

As the difference between SBR with CT and SBR with uniform ACSC is similar over the whole uptake range it is not expected to affect diagnostic accuracy, which did indeed seem to be the case. There was actually slightly lower sensitivity and specificity with CT than with uniform corrections so these results do not support a change to CT ACSC. However we are limited by the small sample sizes recruited into SUPeRB thus far, making it impossible to detect small differences in accuracy with certainty. Interestingly, one of the MCI-AD cases rated as abnormal due to asymmetry in section 6.1 that had normal quantification with the RR_uniform method does show abnormal right caudate uptake with both the CT methods. This reduces the specificity of the CT methods, but should not necessarily be regarded as poor performance as the scan does appear abnormal.

The difference between the standard RR_uniform reconstruction and that with the full Monte Carlo collimator model applied was of note as SBRs were greater with the full collimator model. We would expect that scatter into the background region may be being effectively suppressed with this method, leading to the increased SBRs. However with the phantom we saw that the background activity was actually higher with MCRR reconstructions than with RR. The striatal activity was even higher, leading to increased SBRs. As the difference between MCRR and RR increases with SBR, we might expect an improvement in the separation between normal and abnormal scans with the full collimator model. However the histograms in Figure 60 show that this is not the case, with the MCI subjects showing approximately normal distributions, with no division between normal and abnormal. The diagnostic accuracy results (and ROC analysis) also showed no evidence of improved accuracy with the full Monte Carlo collimator model applied. A limitation is that the standard reconstruction parameters of 16 iterations and 4 subsets (equivalent to 64 iterations), which is recommended for FP-CIT reconstruction with RR_Uniform may not be optimised for use with the full Monte Carlo collimator model as it is possible that convergence occurs after a different number of iterations. We chose to keep all parameters the same, other than the use of the full Monte Carlo model, so that comparisons could be made. In addition, there were not significant differences seen between reconstructions with 64 and 120 equivalent iterations in the phantom work.

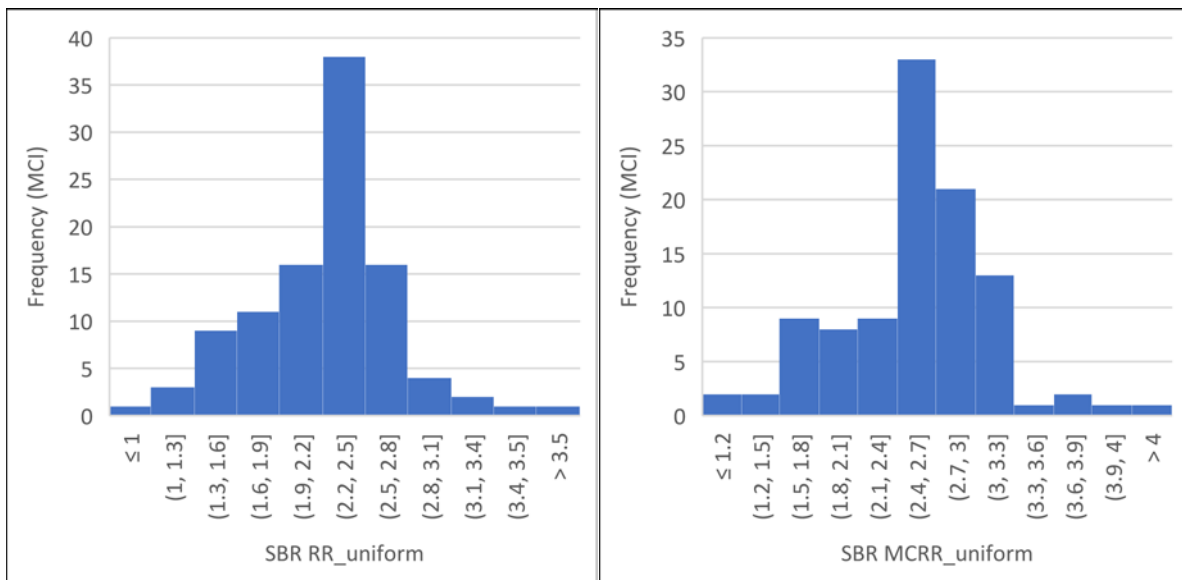


Figure 60: Histograms of standard reconstruction SBRs and full Monte Carlo collimator model, for uniform ACSC.

By coincidence, the effect of the CT ACSC and the full Monte Carlo collimator model act in opposite directions, so there is very little difference at all between the standard RR_Uniform reconstruction method and that with all the advanced corrections applied.

The much greater differences between reconstruction methods at lower SBRs seen in the phantom in the previous section are not reflected in the patient Bland-Altman plots in Figure 57. This suggests that problems with phantom image registration at lower relative striatal activity may have introduced uncertainty. The phantom “brain” is smaller than the BRASS template, so the background VOIs are closer to the edge of the “brain” than in patients, which may have contributed to random error.

Clearly, the fact that so many MCI-LB cases do not have abnormal dopaminergic function based on the FP-CIT scans makes it difficult to assess reconstruction methods. We are confident these scans are normal as they are within normal limits on both quantification and visual assessment, so the low sensitivities do not reflect the accuracy of the reconstruction methods. As shown in section 6.1 for the RR_uniform reconstruction, there is good agreement between expert visual assessment and quantification. We assume that if these subjects with normal FP-CIT scans do have Lewy body disease then the disease is not affecting the striatum and may be found in the neocortex, for example.

It is any change in interpretation of the borderline cases that is of interest, and which we do not see much of in this dataset. We can therefore assume that any increase in the quantitative accuracy of reconstructions with the use of CT ACSC or the application of the full collimator model does not have a clinically significant effect on SBR interpretation. The FCM could be applied as an additional reconstruction using the uniform ACSC assumption for further assessment as it does not inconvenience the patient and does not appear to reduce accuracy; however reconstruction times are currently still

quite long with this method (about 10 minutes) as Hermes have not yet implemented it with GPU processing. It currently is not supported by BRASS so can only be used with a local normal database.

Given the differences between SBRs with the full collimator model implemented, it is clearly important for users not to attempt to analyse FCM data with BRASS using the ENC-DAT database, as this was reconstructed with the RR_uniform method, so results are likely to be too high. Because of this clinical risk, the full collimator model has not yet been released for HybridRecon Neurology and these results were obtained under a research agreement with Hermes. In our dataset, if MCRR_uniform results are interpreted using Z scores from controls reconstructed with RR_uniform, two of the 10 probable MCI-LB cases with abnormal results (20%) would be interpreted as normal.

6.4.5 Conclusions

The use of CT attenuation and scatter correction reduces SBR with respect to uniform correction, but does not have a significant impact on diagnostic accuracy. The use of the full Monte Carlo collimator model increases SBR, with the difference between SBR values with and without FCM applied increasing with SBR. However this also does not appear to have a significant impact on the interpretation of clinical SBR results in this dataset. In clinical practice we would not recommend CT ACSC as this increases the radiation dose and scan time with no demonstrable benefit. The FCM method with uniform ACSC could continue to be used for research as in theory the greater SBR difference at higher SBRs could make the method more accurate, and if so this could perhaps be demonstrated with a larger study of borderline cases with clinical follow-up. Currently there does not seem to be strong evidence for centres using the standard BRASS reconstruction method to switch to the full collimator model.

6.5 Test-retest variation of FP-CIT SBR and SUV in healthy older adults

6.5.1 Introduction

The assessment of longitudinal change in dopaminergic function via repeat FP-CIT scanning is an important part of the SUPeRB study with the aim of assessing the rate of dopaminergic loss in early Lewy body disease. Around half of the participants with clinical diagnoses of MCI with Lewy bodies had normal FP-CIT scans at baseline. All SUPeRB patients and controls have been invited to return for a repeat scan one year after the baseline scan, and we are in the process of arranging further follow up scans. Given the considerable inter-subject variation, a subject with uptake at the top of the normal range could drop considerably and still be within normal limits at follow up. The change in uptake of an individual over time may therefore be a useful biomarker for Lewy body disease, although this will depend on the test-retest variation in the scanning technique.

In our adults with normal cognitions, we would not expect more than a 1.5% drop in the year between baseline and follow up, since the rates of decline in the literature for healthy controls are between 3% and 14% per decade [200]. This anticipated lack of true change in uptake means that the baseline and follow up scans in older adults with normal cognition can be used to give an estimate of FP-CIT test-retest variation. This is a measure of how different we might expect the uptake measurement to be due to physiological and technical sources of variation if the scan was repeated. This is something that is not often included in research studies with FP-CIT, possibly due to the considerable cost of the scan and the additional radiation dose to the participants.

An early publication by Booij et al. studied six healthy women aged 51 to 70 years, rescanning after 3-6 weeks [201]. The test-retest variability in SBR was 7% and intraclass correlation coefficient (ICC) 0.95, indicating excellent repeatability. Matsuoka et al. studied 8 healthy young men rescanned between 1 and 4 weeks, with FP-CIT scans registered to MRI images [202]. The test-retest variation was 6% for the whole striatum, with ICC 0.95 [202]. However, Tsuchida performed repeat scans after one week on 10 healthy controls aged between 25 and 70 years, reporting a higher test-retest variability of 11% and intraclass correlation coefficient (ICC) of only 0.59 for the whole striatum [203]. Differing protocols and analysis methods for striatal uptake make it difficult to compare between studies.

The most relevant study was carried out at Newcastle University in 2005, where Colloby et al. found test-retest variation of around 10% between repeat scans of 15 healthy older adults scanned a year apart, and ICC of around 0.7, indicating reasonable repeatability over a 1 year time gap [204]. The scans were acquired, reconstructed and analysed differently from the SUPeRB study, so may not reflect the repeatability of methods used in this thesis. We were unable to reanalyse this data with current methods as the raw data had not been archived so the images could not be reprocessed. We have

found no other studies of test-retest repeatability over a longer time scale that would help to inform the expected random variation in SBR that may occur in follow up scans in patients suspected of having Lewy body disease. Jacobson Mo et al. studied baseline and repeat FP-CIT at three and five years in 16 healthy controls as part of a longitudinal Swedish Parkinson's disease study [205, 206]. The mean age at baseline was 69.5 ± 7.0 years and there was no significant difference in mean uptake at follow up compared to baseline, with "a varying pattern for both a relative decline and a relative increase" in FP-CIT uptake [206]. The aim of this study was to assess age related changes, so test-retest repeatability analysis was not carried out, however the lack of change in uptake is encouraging. To our knowledge, there have been no other previous studies with repeat FP-CIT scans in healthy subjects with similar time gaps (a few years), no studies using state of the art automated quantification methods such as BRASS or DaTQUANT and no studies assessing the repeatability of absolute striatal uptake (SUV-SPECT). Although absolute quantification with SUV_{mean} at baseline appeared less accurate than SBR with BRASS, SUV may have a role for follow up studies in the same patient since accurate normalisation will be less important.

The test-retest analysis covered in this section was carried out towards the end of the doctoral fellowship, when 27 of the 31 healthy older adults had consented to and undergone a repeat FP-CIT scan. The aim of this work was to compare SBR, SUV and visual assessment in a group of cognitively normal subjects scanned at one year interval. This provides a surrogate measure of the test-retest repeatability which is important in assessing disease progression in studies in patients. Possible reasons for test-retest variation, such as operator variability were explored.

6.5.2 Methods

Subjects

The SUPeRB cohort of healthy older adults was used for this analysis, with 27 individuals with repeat scans included in this section. Seven of the repeat scans had not yet been rated by the panel, so 20 scans were included to assess the concordance between baseline and repeat visual assessment. The repeat scans were carried out between 355 and 490 days after the baseline scan (mean 390 days \pm 35).

FP-CIT imaging

The standard acquisition and reconstruction methods outlined in section 6.1 were used. Briefly, subjects were administered 185 MBq FP-CIT and scanned 3-6 hours after injection on a Siemens gamma camera fitted with LEHR collimators. Images were reconstructed with 16 iterations and 4 subsets, with uniform attenuation and scatter correction and resolution recovery applied.

Visual ratings

Visual ratings at baseline and follow up were assigned independently of one another, using the consensus panel approach outlined in section 6.1.

Consistency of processing for semi-quantification

As semi-quantification is a routine part of the SUPeRB study, the baseline and repeat studies were reconstructed and processed independently of one another, and not necessarily by the same operator. It was noted that several baseline and repeat scans had quite different BRASS quantification results, more extreme changes than expected from the literature. In section 4.4 we showed that BRASS was insensitive to reorientation of the same scan, but repeat scans were not investigated as they were not available at the time. We therefore investigated whether reprocessing of these scans would give more consistent results (Appendix F). Careful reprocessing of baseline and repeat scans by the same operator did not result in more consistent SBR values, so the original data is used in this section.

Quantification of striatal uptake

Semi-quantification was carried out with BRASS using the ENC-DAT database normal range, with camera correction and age correction applied. SUV_{mean} and SUV_{peak} results were obtained using the Southampton method, as outlined in section 6.2. As SBRs calculated using the Southampton method use the same volumes of interest as those required for absolute quantification, the repeatability of semi-quantification with the Southampton method was also tested.

Analysis

Between-subject variation was assessed with the coefficient of variation, calculated using the baseline scan results. Test-retest variation was calculated from the baseline and repeat scan for each individual. Variation was defined as the mean absolute change in uptake parameter, expressed as a percentage.

Bland-Altman plots of the differences between repeat and baseline imaging were produced for each uptake parameter. The overall mean difference between baseline and repeat imaging was calculated, and paired t-tests used to check for significance, assuming a normal distribution for striatal uptake.

6.5.4 Results

Visual ratings

Of the 20 controls rated so far, 2 were rated as abnormal at baseline and normal on repeat scanning (Table 36). None were rated as normal on baseline and abnormal on repeat. There was 75% agreement between baseline and repeat results at the scoring level, 90% if only the binary ratings were considered. The scans for the two patients with abnormal ratings at baseline are shown in Figure 61. All BRASS quantification results are within normal limits, as defined Z scores greater than -2 relative to the ENC-DAT database, with age correction. The subject with borderline-abnormal values is relatively young (64 years) so the Z scores are lower than they would be for the same SBR in an older volunteer.

Table 36: Baseline and repeat scan consensus panel ratings for 20 of the controls.

Baseline score	Baseline binary rating	Repeat score	Repeat binary rating
0	Normal	0	Normal
4	Normal	0	Normal
0	Normal	0	Normal
1	Abnormal	4	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
4	Normal	4	Normal
5	Abnormal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
4	Normal	0	Normal
4	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal
0	Normal	0	Normal

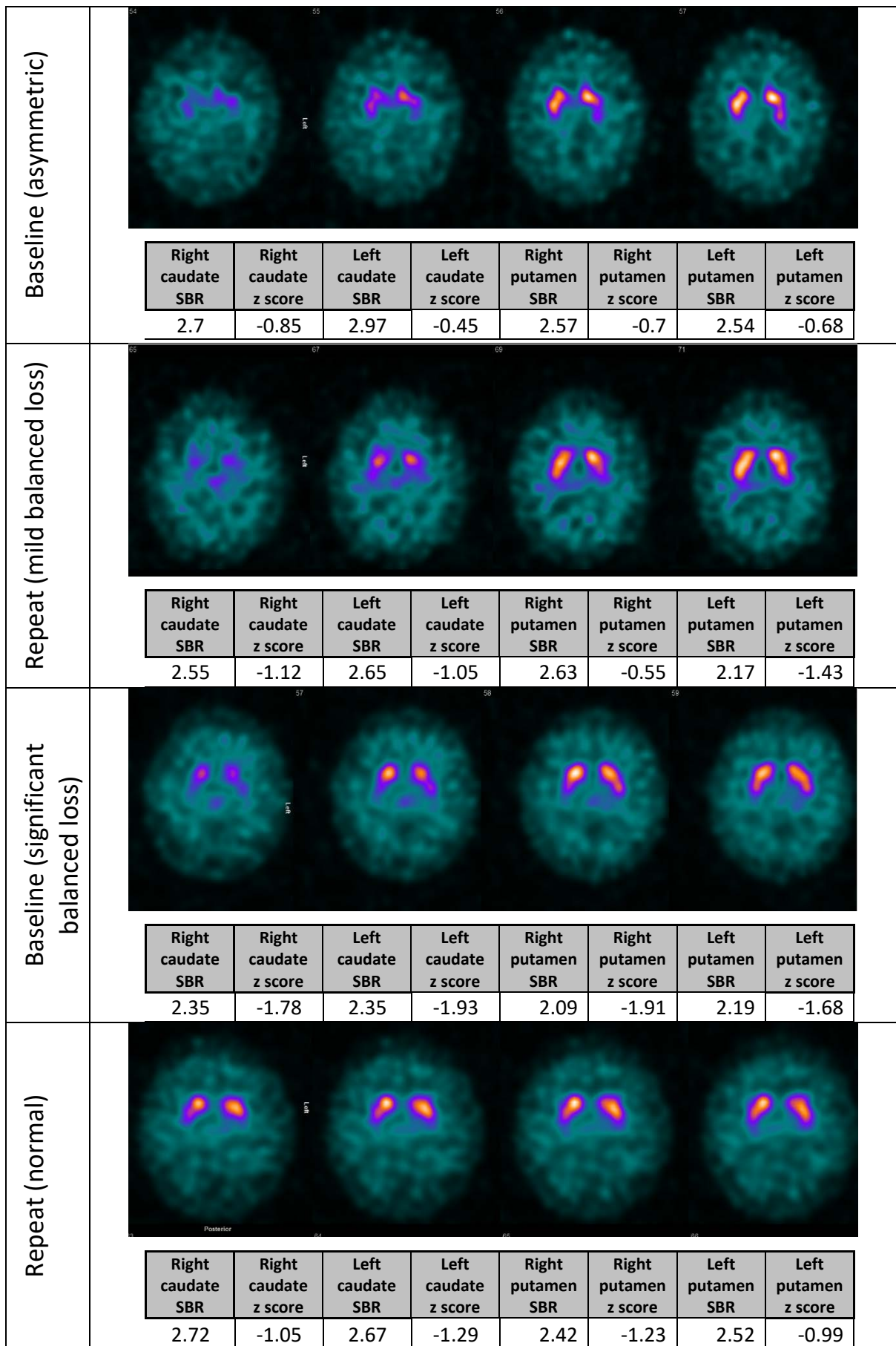


Figure 61: Baseline and repeat scans for the controls rated abnormal at baseline and within normal limits on repeat.

Absolute mean change between baseline and repeat

Our results show no drop in the mean striatal uptake of the control cohort between baseline and one year repeat (Table 37), although there is actually a suggestion of slightly increased quantification results on average for all parameters except SUV_{peak}. However these relationships are not statistically significant when correcting for multiple comparisons.

Table 37: The average change in uptake between baseline and repeat, expressed in absolute terms and as a percentage. Paired t-tests were used to assess for significance.

		Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
BRASS SBR	Mean	0.15	0.06	0.14	0.13	0.15	0.07
	SBR diff	(+5%)	(+2%)	(+5%)	(+4%)	(+6%)	(+3%)
	p-value	0.06	0.24	0.11	0.12	0.07	0.34
Southampton SBR	Mean	0.36	0.31	n/a	n/a	n/a	n/a
	SBR diff	(+6%)	(+5%)	n/a	n/a	n/a	n/a
	p-value	0.04	0.04	n/a	n/a	n/a	n/a
SUV _{mean}	Mean	0.29	0.17	n/a	n/a	n/a	n/a
	SBR diff	(+4%)	(+2%)	n/a	n/a	n/a	n/a
	p-value	0.15	0.34	n/a	n/a	n/a	n/a
SUV _{peak}	Mean	0.02 (0%)	-0.20 (-3%)	n/a	n/a	n/a	n/a
	SBR diff			n/a	n/a	n/a	n/a
	p-value	0.89	0.08	n/a	n/a	n/a	n/a

Intra-subject repeatability (baseline)

The baseline coefficient of variation results are shown in Table 38 and indicate variation of around \pm 12% for most methods, with slightly higher variation for SUV_{mean}.

Table 38: Coefficient of variation (SD / mean) striatal uptake results for each method taken between all subjects at baseline.

	Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
BRASS SBR	11.4%	12.0%	11.6%	12.2%	12.5%	13.0%
Southampton SBR	11.0%	12.2%	n/a	n/a	n/a	n/a
SUV _{mean}	14.1%	14.9%	n/a	n/a	n/a	n/a
SUV _{peak}	11.7%	12.6%	n/a	n/a	n/a	n/a

Intra-subject variation (baseline and repeat)

The mean absolute intra-subject variation expressed as a percentage of baseline and repeat, with standard deviation is shown in Table 39. The greatest drop and greatest increase in striatal uptake

between baseline and repeat for each method is also shown. For SBR methods, the variation between baseline and repeat is very similar to the inter-subject variation at baseline in Table 38. For SUV methods, the variation between baseline and repeat for individuals is lower than the variation between subjects. Bland-Altman plots showing variation in striatal uptake between baseline and repeat are shown in Figure 62.

Table 39: Intra-subject variation for each of the methods assessed. The mean percentage absolute difference between baseline and repeat scans is presented with SD. The most extreme negative and positive changes are also shown.

		Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
BRASS SBR	Mean % difference \pm SD	11.5% \pm 8.7%	11.0% \pm 8.6%	12.8% \pm 8.6%	10.6% \pm 9.1%	12.7% \pm 9.6%	11.8% \pm 7.0%
	% difference range	-26.6% to 32.8%	-23.5% to 32.7%	-32.9% to 33.0%	-25.9% to 35.2%	-34.4% to 32.8%	-18.9% to 29.7%
Southampton method SBR	Mean % difference \pm SD	12.2% \pm 8.6%	10.5% \pm 7.4%	n/a	n/a	n/a	n/a
	% difference range	-32.8 to 32.3%	-25.7 to 22.9%	n/a	n/a	n/a	n/a
Southampton method SUV_{mean}	Mean % difference \pm SD	9.3% \pm 10.7%	8.2% \pm 10.0%	n/a	n/a	n/a	n/a
	% difference range	-18.3% to 53.0%	-18.4% to 50.6%	n/a	n/a	n/a	n/a
SUV_{peak}	Mean % difference \pm SD	5.6% \pm 5.8%	6.0% \pm 5.8%	n/a	n/a	n/a	n/a
	% difference range	-19.1% to 19.3%	-20.9% to 8.8%	n/a	n/a	n/a	n/a

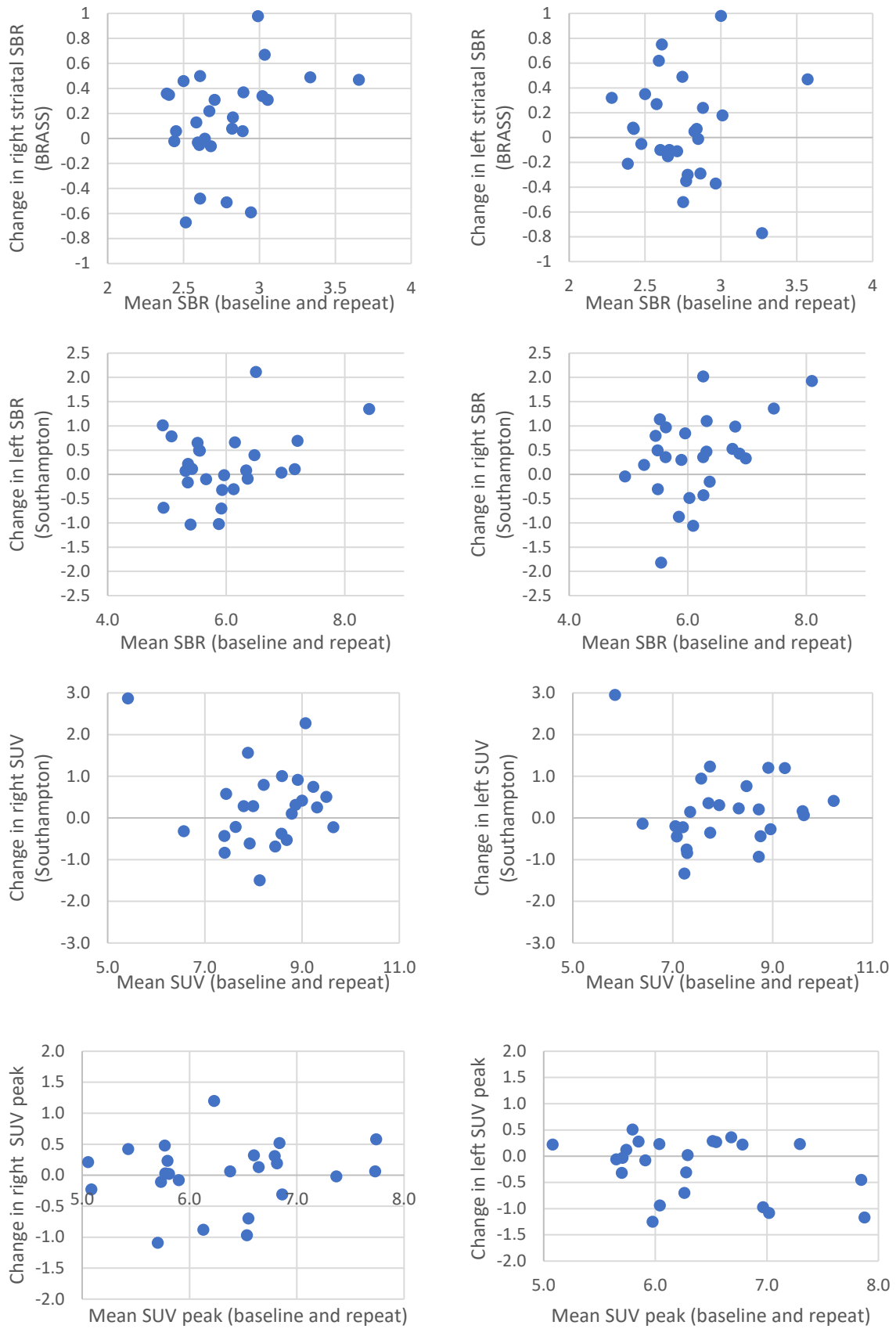


Figure 62: Bland-Altman plots for change in striatal uptake measures between baseline and repeat

6.5.5 Discussion

Our results for SBR intra-subject variation of around 12% are in keeping with those of Colloby et al. who also rescanned after a year and found around 9-10% variation. The variation between baseline and repeat is similar to the inter-subject variation at baseline, suggesting that physiological fluctuations and technical variation between baseline and repeat scanning is similar to the variation between individuals.

The fact that the FP-CIT scans of two healthy older adults were rated as abnormal at baseline but normal on repeat strongly suggests that these subjects do have normal dopaminergic function and that scans in normal patients are therefore likely to occasionally show similar borderline-abnormal appearances. Care should be taken not to over-report such scans, which was also the conclusion of our publication on FP-CIT ratings methods with autopsy validation [195], where mild balanced reduction in striatal uptake was often reported in subjects without Lewy body disease at autopsy. Both control subjects rated as abnormal had baseline quantification results within normal limits, although these did improve on repeat. The concordance between baseline and repeat ratings categories was 75%, 90% if only the binary categorisation of normal/abnormal is used. This reflects the fact that it is often difficult to decide between for example, the mild and significant balanced loss categories. A limitation is that due to the time involved in ratings, so far we have not repeated ratings on the same scans, so we do not yet have a measure of intra-rater reliability in controls or patients. However it is likely that we would see similar concordance in repeat rating of the baseline cohort just by chance as the display of the images introduces variation.

One outcome of this variation could be the presence of unusually high SBR values in a few subjects, skewing the results. The histogram in Figure 63 suggests that the distribution of BRASS SBR values is predominantly normal, with a tail towards the high end. Outliers may actually reduce the normal range SBR cut-off values used with BRASS by increasing the standard deviation. It may therefore be useful in future work to devise a probability based approach to classifying quantification results, or a machine learning approach based on the image features rather than quantification.

Our results of fairly substantial random variation in SBR with both BRASS and Southampton method, including some cases of greater than 20% apparent change, have important implications for the interpretation of repeat studies in patients. Clearly a drop in uptake of less than 10% in a patient follow up scan should not be considered significant and the results here would suggest that at least a 20% change is needed to be confident that the results are likely to represent a real decline in function.

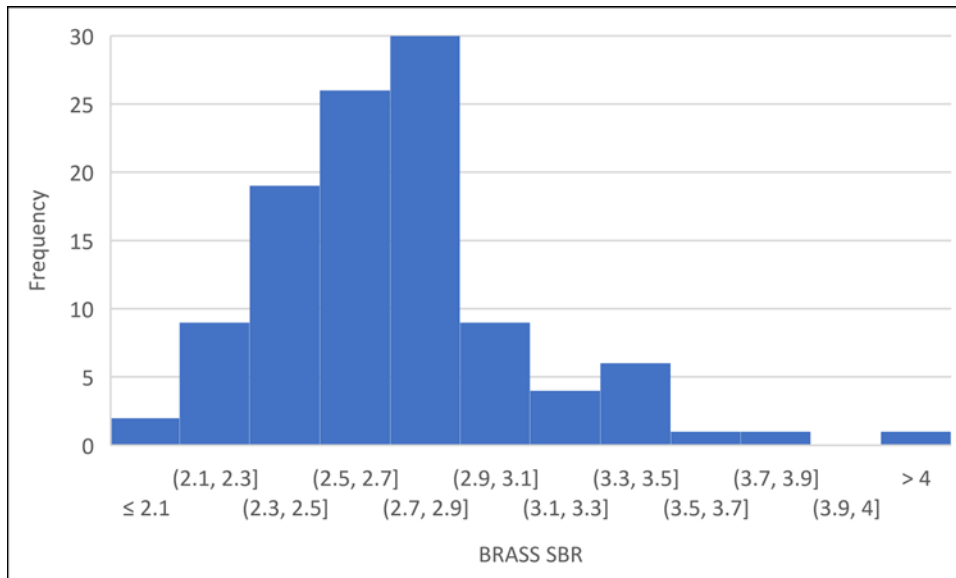


Figure 63: Histogram of BRASS SBR results, with left and right whole striatum and baseline and repeat data included

Our SUV results are interesting because, unlike with SBR, the variation between baseline and repeat scans was lower than the variation between subjects. As the variation between subjects is higher than for SBR, absolute quantification is less likely to be useful than semi-quantification for cross-sectional studies, which is what we saw in the previous section. However, the better reproducibility between baseline and repeat may make absolute quantification more suitable for longitudinal studies than semi-quantification. The issues that we assume give rise to differences in SUV between people, such as correcting for tracer volume of distribution and differences in uptake outside the brain cancel out for baseline and repeat in the same individual. It will be interesting to see whether this is reflected in the SUPeRb MCI-LB patients when all the follow up scans are completed.

6.5.6 Conclusion

The coefficient of variation in BRASS SBR among control subjects at baseline of 12% is similar to the average variation between baseline and repeat scans of $12 \pm 9\%$. Changes in SBR of less than 10% between baseline and repeat are therefore negligible and changes of up to 20% are also likely to be within normal variation. This should be kept in mind when reporting both baseline and follow-up scans. The variation in absolute striatal uptake between subjects is relatively high for SUV_{mean} at around 15%, but only around 9% between baseline and repeat suggesting SUV may be useful for longitudinal studies.

6.6 Summary of chapter 6

In section 6.1 we compared visual rating of the SUPeRb scans at baseline by a consensus panel of experts with standard BRASS quantification, using the one year clinical diagnoses (including MIBG results) as the gold standard. We compared the sensitivity and specificity of expert visual rating with BRASS quantification alone, showing that the accuracy of both methods was very similar. This has implications for efficiency in future large studies, as the consensus panel method requires considerable time and effort. Longer term follow up of the MCI cases will help to establish which method was correct for the few discrepancies between the panel and the quantification results alone. We showed that the default method of age correction in BRASS gave the same accuracy as results using a database made up of age-matched SUPeRb control subjects, suggesting that the benefits of a bespoke centre-specific normal range are limited.

The absolute quantification results in MCI presented in section 6.2 showed that the accuracy of SUV_{mean} with the Southampton method for diagnosing Lewy body disease was the same as the accuracy of relative quantification with SBR using the Southampton method, suggesting absolute quantification of striatal uptake is not beneficial. However, images presented with an absolute SUV scale may assist in visual reporting, as the striata can be assessed without reference to the background appearance. However, the variation in appearance of controls with this scale was quite large, suggesting that the method may not be very specific in practice. Full assessment of the SUV scale images by a blinded panel would be needed to confirm.

Sections 6.3 compared the standard RR_uniform reconstruction method with more advanced methods: CT attenuation and scatter correction and Monte Carlo modelling of septal penetration of high energy gamma photons. We found that the combination of CT ACSC and full collimator modelling gave background activity concentration results in a striatal phantom that were very close to the true value. However, striatal activity concentration results appeared more uncertain, and phantom SBRs were actually further from the true values with CT ACSC applied than without. In the comparison of these methods for patient SBRs in section 6.4 we confirmed that CT ACSC reduced SBR, but found that the full collimator model increased SBR compared to the standard reconstruction. However this did not result in improved diagnostic accuracy. We concluded that CT ACSC is not beneficial for relative quantification with SBR but that the full collimator model may be and should continue to be assessed in future studies.

In section 6.5 we assessed the test-retest variation in FP-CIT imaging, using baseline and one year repeat scans in the SUPeRb controls. We found that at baseline, the variation measured by the coefficient of variation of SBR measurements was 12%. The mean absolute variation between baseline and repeat scans was also 12%, with some subjects showing larger changes in SBR of 20-30%. We also

saw that with visual reporting, two control subjects were rated abnormal at baseline and normal on repeat. The clinical implications of these results are that care should be taken not to over-call scans and that SBR changes of less than 20% on follow up scans are likely to be within normal physiological or technical variation. With SUV_{mean} we saw slightly higher baseline variation, but lower test-retest variation, suggesting that absolute quantification may have a role in the assessment of longitudinal change.

In summary, we have shown that the quantification method selected for the SUPeR study, which uses BRASS SBRs with default age correction calculated on images reconstructed with the RR_uniform iterative reconstruction method, is not outperformed by any of the advanced reconstruction and analysis techniques assessed in this chapter.

Chapter 7: Planar cardiac ^{123}I -MIBG imaging for the diagnosis of Lewy body disease

This chapter focusses on the use of planar cardiac ^{123}I -MIBG imaging for the diagnosis of Lewy body disease and the changes that can be made to potentially increase accuracy or make the procedure more practical. In section 7.1 processing methods for calculating the heart-to-mediastinum ratio are assessed and compared with visual assessment, with recommendations made for clinical practice. In section 7.2 the normal-abnormal HMR cut-off value that should be used in a UK population is considered in detail, comparing results from the SUPeR study with published data for Japanese populations. In section 7.3 early imaging at 20 minutes after tracer administration is explored to see whether it can replace delayed imaging at 4 hours without affecting the sensitivity or specificity of the test.

Content from sections 7.1 and 7.3 was published in Nuclear Medicine Communications in January 2019 [190]. Content from section 7.2 has recently been accepted in August 2019 for publication in the Journal of Nuclear Cardiology, where it will be published alongside an editorial by Prof K. Nakajima, Kanazawa University, Japan.

7.1 Analysis methods for planar cardiac ^{123}I -MIBG scintigraphy in DLB

7.1.1 Introduction

As discussed in Chapter 3, there is a variety of methods used in the literature for interpreting planar cardiac MIBG images for Lewy body disease, including visual analysis and semi-quantification. In this section 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 the MIDAS and SUPeRB studies.

We compare visual assessment with various semi-quantitative analysis methods, including the one used by Nakajima et al. [166], made possible by participating in a European multicentre phantom calibration exercise [167] allowing us to apply the HMR cut-off of Komatsu et al. [53, 62] of 2.20 for delayed images. The overall aim is to determine the optimal method for processing and interpreting planar cardiac MIBG images for DLB diagnosis in routine clinical practice and research studies. This is achieved by considering the variation between operators and raters for each method as well as the accuracies given by ROC analysis. This work was published in Nuclear Medicine Communications in January 2019 [190].

7.1.2 Methods

Subjects

Dementia subjects were recruited as part of the Newcastle University MIDAS study into the use of cardiac MIBG in a representative UK population [83]. Thirty-two of the 33 patients included in our previous publication [83] 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 blinded to scan results, based on clinical assessment and applying standard research criteria [1]. This consensus panel method has previously been validated against autopsy and is accepted by regulatory authorities as the clinical gold standard for living patients [43, 207]. Sixteen older adults with normal cognition were taken from the SUPeRB study.

Image acquisition

Subjects were administered 111 MBq ^{123}I -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"). Analysis here uses delayed images only; early images and the relationship between the two time points is covered in section 7.3. 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. The energy window was 159 keV \pm 10%, matrix size was 128 x 128 and no zoom was

applied. The same settings were applied when participating in the European phantom calibration study [167]. This calibration showed that the HMR cut-off published by Komatsu *et al.* [53, 62] of 2.20 is appropriate for use with our Siemens cameras with MELP collimators (personal communication, Prof. K. Nakajima, Kanazawa University, Japan) – see sections 5.2 and 7.2 for more detail.

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. 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.

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.* [208] 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 64.

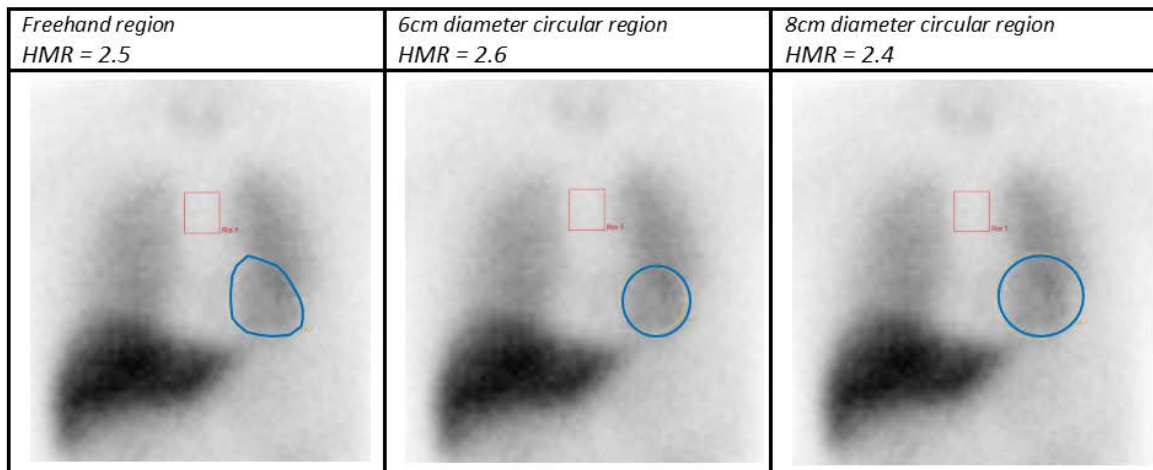


Figure 64: Example normal uptake cardiac MIBG image with HMR values for freehand, 6 cm and 8 cm circular cardiac regions.

7.1.3 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 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.

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.

7.1.4 Results

Subject characteristics

The age, sex, MMSE and ACE scores and MDS-UPDRS scores [209] for all subjects are given in Table 40. 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 40: 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.

	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)
DLB	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 41. 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 41).

Table 41: 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-value
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 42, 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).

Table 42: 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
CoV	5.7%	4.3%	4.2%	4.2%

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 Figure 65 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.

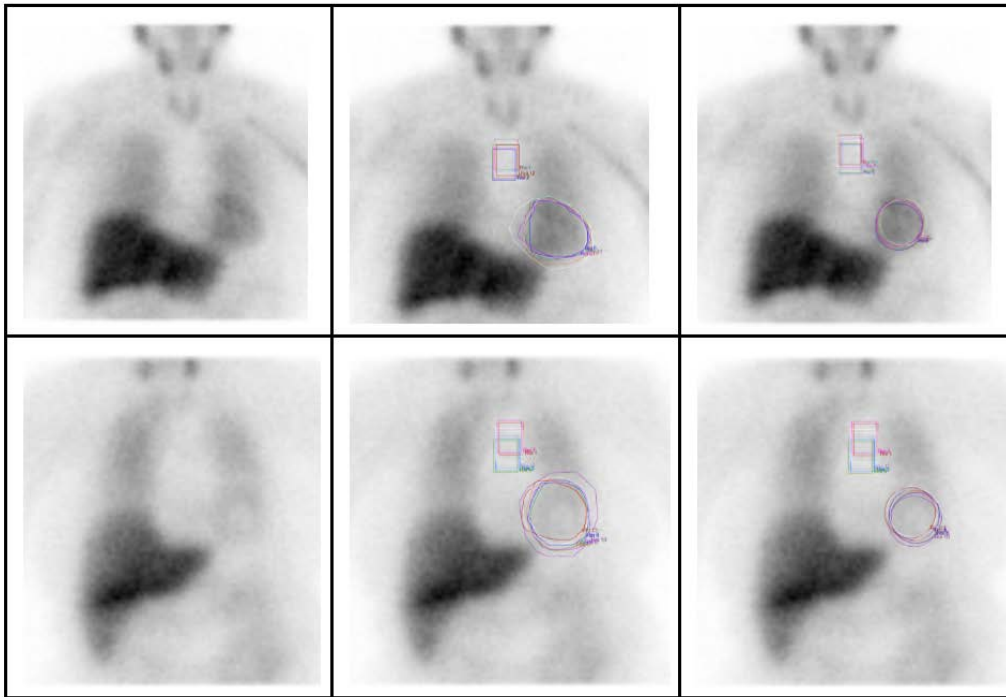


Figure 65: Top: example normal cardiac MIBG image (top row) and abnormal image (bottom row) processed by multiple operators with freehand and 6cm circular regions.

Difference in sensitivity, specificity and accuracy between cardiac region methods

ROC curves and AUC results are shown in Figure 66, with 95% confidence intervals given in Table 43. 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).

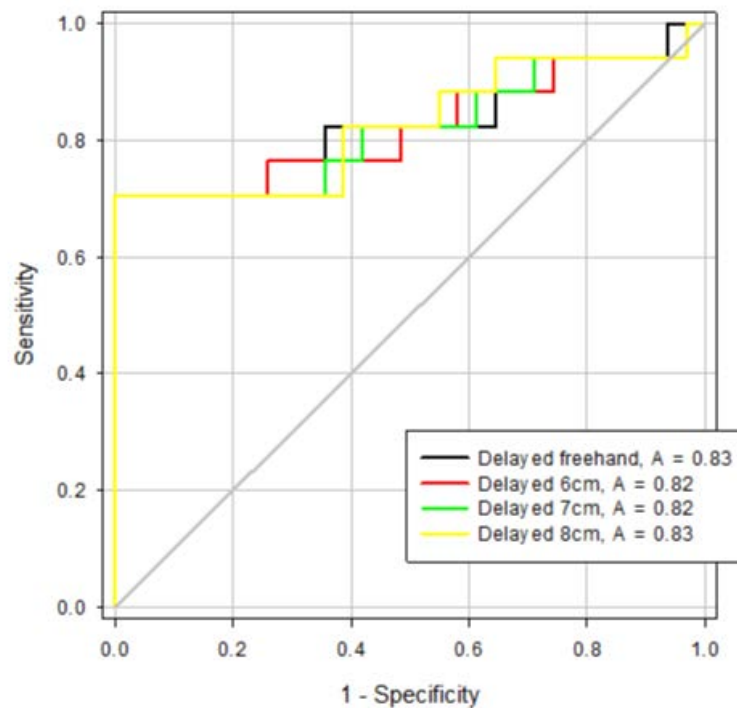


Figure 66: ROC curves for all cardiac region methods tested.

Table 43: ROC curve areas and 95% confidence intervals for all cardiac regions tested.

	Freehand	6cm circle	7cm circle	8cm circle
ROC curve area	0.83	0.82	0.82	0.83
95% confidence interval	0.68 to 0.98	0.67 to 0.98	0.67 to 0.98	0.68 to 0.98

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 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%). Table 44 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 mean HMRs is highest for the 6cm method but the 95% confidence intervals suggest this is not statistically significant.

Table 44: 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

Region method	Mean HMR DLB (95% CI)	Mean HMR non-DLB (95% CI)	Difference
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

Comparison between visual rating and HMR analysis

The agreement between visual raters was very good (intraclass correlation coefficient of 0.98). Overall, accuracy was lower for visual rating (73%) than for HMR analysis (90%), due to the lower specificity (Table 45).

Table 45: Sensitivity, specificity and accuracy for distinguishing between DLB and non-DLB subjects using consensus visual rating, with binomial 95% confidence intervals.

	Sensitivity	Specificity	Accuracy
Result	76%	71%	73%
95% confidence interval	50% to 93%	52% to 86%	58% to 85%

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

7.1.5 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. Tiraboschi *et al.* [67] 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 [67]. 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 [210, 211]. Reduced lung uptake has been reported in Lewy body disease [212] and has been linked to treatment with L-dopa leading to an increase in plasma homocysteine levels [213]. 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 a relationship between L-dopa dose and lung uptake for DLB patients (Figure 67).

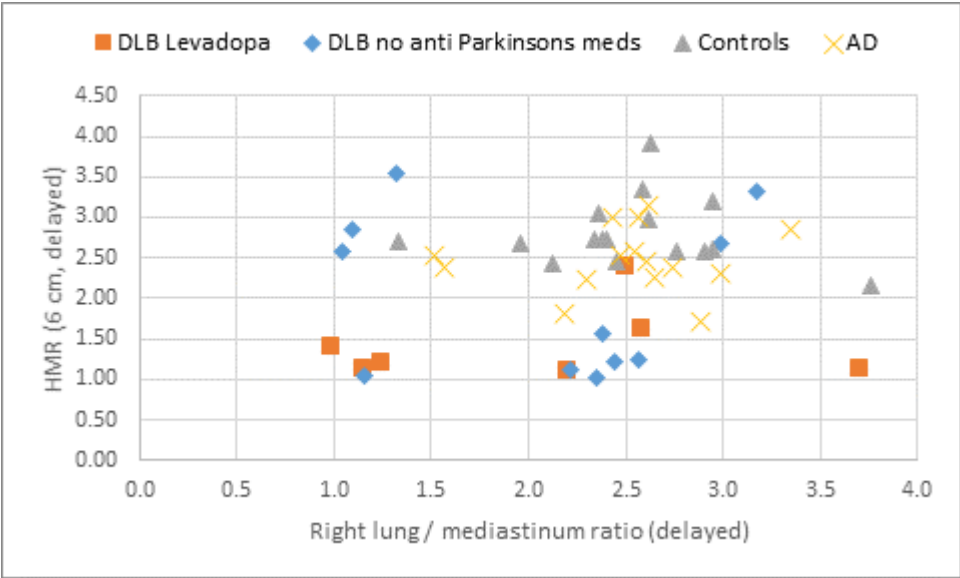


Figure 67: Cardiac uptake plotted against lung uptake for the patients and controls analysed in this section.

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.

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

The results of a recent study at our centre [83] suggest that the threshold of 2.20 taken from the Japanese multicentre follow-up study [62] 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 using the same 6cm circle method. 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 2.20 applied the sensitivity would remain 71% but specificity drop to 87%, due to three AD subjects and one control subject becoming false positives. The appropriate HMR cut-off to use in a UK population is covered in section 7.2.

7.1.6 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.

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. The sensitivity and specificity of the 6cm ROI method on delayed images with optimal cut-off of 1.66 were 71% and 100%.

7.2 Are cardiac ^{123}I -MIBG normal uptake values population-specific?

7.2.1 Background

Although planar cardiac MIBG imaging is straightforward technically to implement in any hospital with a nuclear medicine service, the interpretation of cardiac uptake depends critically on establishment of a normal range, or cut-off for HMR values. In section 7.1 we saw that, provided that an appropriate normal HMR cut-off is used, HMR analysis is more accurate than visual rating [190]. The HMR normal range will depend on:

- Acquisition factors, including the camera and collimator
- Processing factors, such as region placement
- Population specific factors, for example obesity and heart disease

We addressed processing factors in the previous section and in our corresponding publication [190], with a standardised 6 cm circular cardiac ROI recommended. We previously discussed the effects of camera calibration and population specific factors as being potential barriers to comparison of data between centres [83, 190] and now address these here using images from older adults from a typical UK older population to generate a normal range.

The phantom calibration method applied in the previous section can only adjust for technical factors, for example collimator difference, not for population differences. If typical true HMR results vary between populations phantom calibration will not be sufficient, and measurements of local controls will be required. In the MIDAS study we found that when we applied a normal HMR cut-off derived from Japanese studies to our dementia patients it seemed much too high, with several Alzheimer's disease patients falling into the abnormal category [83].

In this section we aim to:

- Establish, for the first time, a normal HMR cut-off based on older adults without cognitive impairment in a typical UK population
- Compare the diagnostic accuracy of this UK cut-off for detecting Lewy body disease in a cohort of dementia patients to that of calibrated cut-off values from the literature.

7.2.2 Methods

Phantom calibration

Before starting cardiac ^{123}I -MIBG imaging at our centre, a calibration check on our gamma cameras was performed with help from colleagues in Japan and the Netherlands, using a planar phantom, as described in section 5.2. This calibration procedure revealed that the performance of our Siemens cameras fitted with medium energy low penetration (MELP) collimators was very similar to that of

typical medium energy collimators in Japan (personal communication, Prof. K. Nakajima, Kanazawa University, Japan). The HMR cut-offs of 2.1 published by Yoshita et al. have recently been recalculated by Komatsu et al. as part of a three year follow-up study [53, 62]. The updated cut-off values of 2.51 for early images and 2.20 for delayed images are used in this section.

Subjects

Older controls without cognitive impairment:

Thirty-one individuals aged over 60 years with normal cognition were recruited as part of the SUPeRB study. Volunteers were excluded if they presented with cognitive impairment, major cerebrovascular disease on MRI brain imaging, Class II or worse heart failure according to the New York Heart Association classification, had had a myocardial infarction in the previous year, were taking labetalol or were unlikely to be able to tolerate imaging. Subjects were not excluded if they had risk factors for cardiac disease, such as a history of smoking, raised blood pressure or Type 2 diabetes because we aimed to recruit typical older adults with normal cognition. Diabetes has been shown to be associated with reduced cardiac MIBG uptake in a small proportion of patients with clinical autonomic neuropathy [214] but similar to healthy controls in another study of patients with impaired glucose tolerance [215]. In our clinical studies in Lewy body disease we aim to assess cardiac MIBG in a “real-world” setting, where diabetes and risk factors for cardiac disease are extremely common. However, a detailed medical history including assessment of cardiac risk factors was taken.

Older people with dementia:

Seventeen patients meeting the clinical criteria for probable dementia with Lewy bodies and 15 meeting criteria for dementia due to Alzheimer’s disease were taken from the MIDAS study. Briefly, patients had their dementia diagnoses confirmed by an expert panel of old-age psychiatrists. They were categorised as having probable DLB if two or more of the core consensus criteria for DLB [1] (fluctuations, visual hallucinations, REM sleep behaviour disorder and parkinsonism) were present, and AD if none of these were present. The same broad general inclusion criteria as discussed above for the controls were applied to ensure the patients were typical of the local population.

Image acquisition and processing

Ten minute anterior planar images were acquired at 20 minutes (early) and at 4 hours (\pm 30 minutes) after injection (delayed) as described in section 7.1.

All images were processed on a Hermes workstation (Hermes Medical Ltd, Stockholm). Planar images were analysed to obtain the heart to mediastinum ratio (HMR) using a 6 cm circular ROI placed over the left ventricle and 4 x 3 cm rectangular ROI between the lungs in the mediastinum, as described in our previous publication [190]. Images were processed by the author. Normal ranges based on control

images were taken to be two standard deviations below the mean value and were calculated separately for early and delayed images.

7.2.3 Results

Control and patient demographics

Of the 31 controls recruited, 2 were excluded due to their image appearance (see below). The remaining 29 people included were aged between 62 and 94 years (mean 75.2). Most were male (22/29) and white (28/29). Body mass indices (BMI) ranged from 22 to 38 kg/m², with the mean value of 27.2 in the overweight category. The demographics for the controls and patients are summarised in Table 46 alongside incidence of cardiovascular risk factors such as smoking.

Table 46: Demographics summary for the 29 controls, 17 DLB and 15 AD patients included in the study. More data is available for the controls as they are from a different study (SUPERB).

	Controls	DLB	AD
Age (Mean ± SD)	75.2 ± 8.3	77.5 ± 8.0	76.2 ± 6.8
Range	62 to 94 years	60 to 89 years	62 to 85 years
Sex	7 female (24 %) 22 male (76 %)	2 female (12 %) 15 male (88 %)	4 female (27 %) 11 male (73 %)
BMI (kg/m²)			
Mean ± SD	27.8 ± 4.2	Not measured	Not measured
Range	21.6 to 37.8		
Ethnicity	28 white (97 %) 1 Asian (3 %)	Not recorded	Not recorded
Smoking	13 never smoked (45 %) 13 ex-smokers (45 %) 2 current smokers (7 %) 1 not recorded (3 %)	Not measured	Not measured
Diabetes	0 Type I (0 %) 3 Type II (10 %)	Not recorded	Not recorded
Systolic blood pressure			
Mean ± SD	140.5 ± 17.8	Not recorded	Not recorded
Range	112 to 173		
On blood pressure medication	8 (28 %)	4 (24%)	5 (33%)
Raised cholesterol documented	11 (34 %)	Not recorded	Not recorded
Documented cardiac history	Previous MI and AF: 1 (3 %) Angina: 1 (3 %) ECG changes: 1 (3 %)	Previous MI: 2 (12%)	Previous MI: 3 (20%)

Control HMR results

Of the 31 older adults with normal cognition recruited, one had a markedly abnormal appearance on early and delayed cardiac MIBG imaging and one had an abnormal appearance on delayed imaging. The HMRs were also clearly reduced for these patients, shown as empty symbols on the plot in Figure 68. It is well recognised that occult Lewy body disease (usually termed incidental Lewy body disease) is present in many older people [81], which may explain these findings. These two outlying scans were therefore excluded from analysis to determine local normal HMR cut-offs. Figure 68 also shows the distribution of early and delayed control HMRs and the change in HMR between early and delayed.

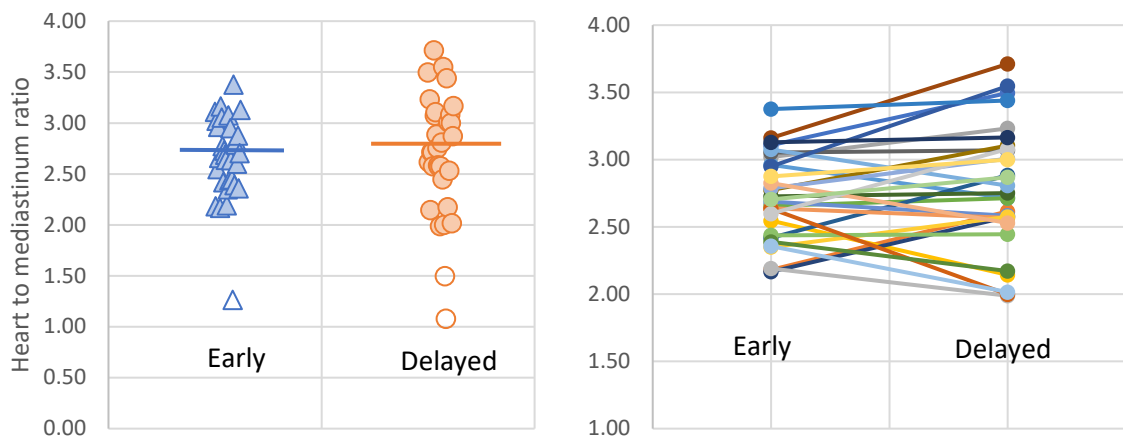


Figure 68: Left: early HMR values (blue triangles) and delayed (orange circles) with means of 2.72 (SD 0.32) and 2.78 (SD 0.46) respectively for the 29 included individuals. Excluded individuals are shown as empty symbols. Right: individual change in HMR between early (left) and delayed (right) images for each of the 29 controls.

Eight of the 29 controls had HMR values below the Japanese cut-off of 2.51 on early images and 5/29 below 2.20 on delayed images.

Means, standard deviation and normal cut-offs (mean-2SD) derived from the control HMR data are shown in Table 47. The cut-off on the early images is higher due to the lower variation between individuals on the early images – the mean HMR is actually lower on the early images than delayed. With these cut-offs applied none of the 29 control subjects included were classified as abnormal.

Table 47: HMR data from healthy control subjects with normal scan appearances

	n	Mean HMR	Standard deviation	Control normal cut-off
Early images	29	2.72	0.32	2.07
Delayed images	29	2.78	0.46	1.86

Dementia patient results

The mean HMR values for the patients with Alzheimer’s disease and those with dementia with Lewy bodies from our previous work [83] are given in Table 48 with the control results included for comparison. Figure 69 shows plots of the HMRs for controls, AD patients and DLB patients for early and delayed images, with the cut-offs determined from controls shown as red lines and the Japanese cut-offs as purple lines. Three of the AD patients have uptake below the local cut-off (i.e. unexpectedly low) on early images and two on delayed. Six of the DLB patients have uptake above the local cut-off on early images, five of which are well above and 1 borderline (HMR = 2.15). This latter patient’s HMR was 1.61, clearly abnormal, on the delayed image. The other five patients retained a normal HMR on the delayed images, and the images appeared normal visually. The corresponding sensitivity, specificity and overall accuracy for DLB vs AD diagnosis using the local cut-offs are given in Table 49, alongside the results with the higher cut-offs taken from the Komatsu *et al.* multicentre follow up study [62]. The third row shows the results with the optimal cut-off for this cohort determined via ROC analysis [83], which was 1.77 to 1.80 for early images and 1.61 to 1.70 for delayed.

Table 48: HMR results (mean and standard deviation) for DLB and AD patients.

	Early			Delayed	
	n	mean	SD	mean	SD
DLB	17	1.88	0.71	1.69	0.79
AD	15	2.47	0.41	2.48	0.41
Control	29	2.72	0.32	2.78	0.46

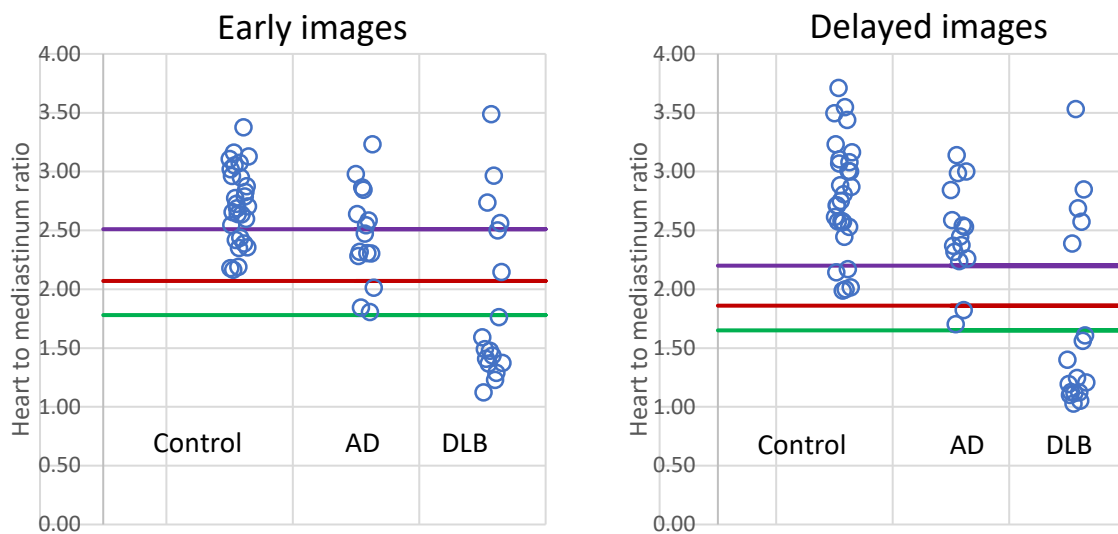


Figure 69: Heart to mediastinum ratios for controls, Alzheimer’s disease and dementia with Lewy bodies patients for early images (left) and delayed images (right). The red lines show the local normal cut-off values, the green lines the optimal DLB vs AD cut-offs for data from our previous study [83] and the purple lines the calibrated cut-off taken from the Japanese multicentre three year follow-up study [62].

Table 49: Sensitivity, specificity and overall accuracy for DLB vs AD differential diagnosis for early and delayed images with HMR cut-offs taken from local controls, Japanese multicentre study and ROC analysis of the dementia subject data.

Cut-off	Early			Delayed		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Local controls (2.07 early; 1.86 delayed)	65%	80%	72%	71%	87%	78%
Komatsu <i>et al.</i> (Japan) (2.51 early; 2.20 delayed)	76%	47%	63%	71%	87%	78%
Optimal (ROC analysis) (1.8 early; 1.7 delayed)	65%	100%	81%	71%	100%	84%

7.2.4 Discussion

Early and delayed cut-offs

Eight of the 29 controls included in our study had HMR values below the early image cut-off of 2.51 from the Japanese follow-up DLB study [62] on early images and five of 29 below 2.20 on delayed images, providing strong evidence for lower thresholds for this population. We note that the value from the Japanese normal database (patient-based) gives a normal range of 2.2 to 4.0 for early images (mean 3.1) and 2.2 to 4.4 for delayed images (mean 3.3) [188]. These HMR ranges seen in Japanese patients without Lewy body disease do not seem to fit our UK population.

Our cut-offs of 2.07 (early) and 1.86 (delayed) reflect the higher variation in normal delayed HMR results compared with early results, rather than a difference in the mean, which is in line with previous studies [58, 174, 216]. These results were obtained using Siemens cameras fitted with MELP collimators, and as such cannot be directly applied to other centres using other collimators. Using a calibration coefficient of 0.99 for our centre measured via phantom calibration we can apply the average coefficients given in the Japanese normal database publication [188] to calculate equivalent cut-off thresholds for other cameras using the method of Nakajima *et al.* [166], assuming similar patient populations and the same processing method. These average coefficients are 0.55 for LEHR, 0.65 for LEGP and 0.88 for MEGP, so lower cut-offs are required if applying our results to other collimators. These are summarised in Table 50. Note that these are average values and Verschure *et al.* showed variation of around $\pm 6\%$ (± 2 SD) for both low and medium energy collimators in the European calibration study [167]. In addition, small differences between the phantoms used in the Japanese and European study may well add to the uncertainty, as well as possible regional differences in collimator specification for collimators with the same name (personal communication, Prof. K. Nakajima, Kanazawa University, Japan).

Table 50: Cut-offs based on our controls adapted to common collimator groups. These results are only valid for the same processing protocol, i.e. 6 cm circular ROI placed over the myocardium, see [190].

	Early cut-off	Delayed cut-off
MELP (Newcastle)	2.1	1.9
LEHR	1.6	1.5
LEGP	1.7	1.6
MEGP	2.0	1.8

For both early and delayed imaging of the patients with dementia slightly higher accuracy could be achieved with the optimal ROC cut-off than with the a-priori cut-offs derived from controls, which is unsurprising as these are cut-offs tailored to this specific dataset. For early images, the use of the local control cut-off rather than the Japanese cut-off (2.07 vs 2.51) resulted in two fewer DLB patients being correctly assigned as abnormal; one of whom had a less equivocal HMR on delayed imaging. The specificity was greatly increased with the local cut-off as only seven AD patients had early HMRs above the Japanese cut-off of 2.51. However with delayed imaging, the sensitivity and specificity was the same with both the local and Japanese cut-offs (1.86 vs 2.20) as there were five DLB patients with clearly normal looking scans (lowest HMR = 2.39) and there were no AD patients falling between the two cut-offs. It is worth noting however that five control subjects have HMR values between the two thresholds, suggesting our local values are more appropriate. We note that, due to possible differences in phantom structure due to manufacturing tolerances, it is possible that the Japanese thresholds calibrated for our camera could be slightly higher or lower (personal communication, Prof. K. Nakajima, Kanazawa University, Japan).

Comparison with other studies

In the multicentre study and its three year follow up, Yoshita *et al.* used ROC analysis to produce thresholds giving optimal separation between the groups, rather than using an independent normal cardiac uptake threshold [53]. However, we note that the Japanese normal database gives similar cut-offs of 2.2 for both early and delayed images. With diagnostic revisions after three year follow up of 65 of the subjects, the sensitivity and specificity for DLB diagnosis have improved from 69% and 89% respectively to 77% and 97% [62].

At our centre we focus on the use of cardiac MIBG imaging for the diagnosis of Lewy body disease. However, as well as the differential diagnosis of Lewy body disease from other conditions [35, 51, 217], cardiac MIBG scintigraphy is also used for the estimation of risk of sudden cardiac death in heart failure patients and predicting those who could benefit from an implantable cardiac defibrillator (ICD) [63, 64] and this is currently the principal application in Europe and the US [65]. The largest study of healthy

controls recruited to a cardiac MIBG research study (n=94) was that of Jacobsen et al. [218] carried out as part of the ADMIRE heart failure study [63]. The controls were aged between 29 to 82 years (mean 58.5 ± 10.6 years, with BMI between 20.4 and 44.3 (mean 28.7 ± 5.1). These individuals were carefully screened to exclude coronary heart disease or Type 1 or 2 diabetes, making them healthier on average than the typical population. Although a substantial proportion (39/94) were aged over 60 years, the inclusion requirement of having an annual cardiac risk of less than 10% would further increase the difference in health between these individuals and a typical elderly population. We regard the inclusion of adults with diabetes and coronary risk factors as a strength of our study in relation to the diagnosis of Lewy body disease, rather than a weakness. In recruiting older adults that, aside from having no cognitive impairment, are similar in general health to that of the general population we are determining a true normal range. However, despite this, the controls from the US ADMIRE study have lower HMR values on average than those from our study, even accounting for the different collimator. Using LEHR collimators their result was 1.78 ± 0.22 , which would correspond to normal cut-off of 1.34 with LEHR (two standard deviations below the mean of 1.78). This cut-off converts to around 1.6 with our camera with MELP (determined using average coefficients of 0.55 for LEHR collimators [219] and the measured coefficient of 0.99 for our camera). One possible reason for the lower HMRs in the United States may be differences in the method used to delineate the cardiac ROI, with a freehand method used in the ADMIRE study. We found previously that this tends to give lower HMR values than a 6cm circular region [190]. Interestingly, the *a priori* cut-off used in the ADMIRE study was much higher than subsequently seen in the controls: 1.6 with LEHR collimators, equivalent to around 2.1 for our camera with MELP. The *a priori* cut-off was determined by pooling the literature on cardiac MIBG values in healthy controls (n=202) [220]. This was done without population or camera differences being taken into account, which may account for the discrepancy.

To our knowledge, the only other recent UK study in controls is by Asghar *et al.* in Manchester, again using LEHR collimators and excluding people with any history of diabetes or cardiac disease [216]. In 14 healthy adults (mean age 54.6 ± 5.4 years, mean BMI 27.0 ± 3.1) they found a mean early HMR of 1.67 ± 0.13 and delayed 1.73 ± 0.16 , which would give cut-offs (mean – two standard deviations) of 1.41 for both early and delayed images. This cut-off is equivalent to around 1.7 for our camera with MELP collimation and therefore similar to our result for delayed imaging of 1.86. Since there is thought to be little dependence of HMR on age [181, 221] we would expect any differences between our controls and this study to be due to the inclusion criteria, as it seems unlikely that there would be large differences between the population of Manchester and Newcastle, two post-industrial cities in northern England. The fact that our mean and cut-off values are slightly higher rather than lower than the Manchester study, after accounting for camera differences, provides further indirect support of

our previous findings that cardiac disease is unlikely to have a big impact on HMR [83], and further supports the small body of evidence suggesting HMR is not strongly linked to age.

Effect of cardiovascular disease on HMR

It is known that regional cardiac sympathetic denervation assessed with SPECT or PET sympathetic imaging is associated with both areas of myocardial infarction (MI) and also with ischaemia in patients without previous MI [222-224]. The impact of MI and coronary artery disease (CAD) on global uptake assessed via planar cardiac MIBG scintigraphy is not well established, however most studies in healthy controls exclude those with CAD. To our knowledge there are no large studies that evaluate the impact of regional sympathetic denervation associated with MI and CAD on planar HMR. In this study and our previous work [83], one control and three AD patients had experienced an MI, all more than 1 year prior to scanning. The control showed clear regional defects on SPECT imaging but borderline-normal delayed planar HMR (2.14). The three AD cases showed defects of varying magnitude and extent on SPECT and planar HMRs of 2.99, 2.16 and 2.59. The numbers are too low to establish a correlation between the extent of regional denervation and effect on global HMR but it seems likely that regional denervation would have to be widespread for planar HMR to be abnormally low. Given the variation between healthy individuals, it is unlikely that MI and CAD alone would give an abnormal HMR result in individual cases. However it is likely that both the mean and standard deviation of a group of subjects would be affected by heart disease, suggesting that differences in the prevalence of cardiovascular disease between populations could contribute to differences in HMR cut-offs. In the UK the prevalence of cardiovascular diseases increases with age, with 8% of people between 60 and 70 years of age and 24% of those over 80 reporting at least one symptom [225]. In England the prevalence of coronary heart disease over all ages in 2018 was reported by the National Health Service to be 3.1% [226]. Equivalent data for Japan is not published but we note that a large cohort study (Akita-Osaka study) showed 0.2% incidence of new coronary heart disease cases in urban Japanese men aged between 60 and 69 between 1996 and 2003 [227] suggesting a much lower prevalence in this age group than in the UK.

Effect of obesity on HMR

Obesity has been linked to decreased apparent cardiac MIBG uptake by two previous studies [181, 228], and is thought to be due to increased attenuation and scatter in larger individuals, rather than to true differences in cardiac uptake. The ADMIRE control study discussed above [181] had a wide BMI range of 20.4 to 44.3 kg/m² (mean 28.7) and reported a “modest” decrease in HMR with BMI. This was a study into ageing and the incidental finding of a relationship between BMI and HMR was not discussed in detail. Pellegrino *et al.* reported lower HMRs in patients with BMI>30 (n=10) compared to those with BMI<30 (n=35) [228]. This was in heart failure patients and the results have not yet been

replicated in healthy controls but do suggest patient size may affect apparent cardiac MIBG uptake. In our study the mean BMI of the controls was 27.8 kg/m² which is representative of the mean value of 27.3 for 119,230 men aged 40-69 recruited into the UK biobank study between 2006 and 2010 [229]. The Manchester study had a similar mean BMI of 27.0. In contrast, the mean BMI in a Japanese longitudinal study of 104,928 individuals was 22.6 in men and 22.9 in women [230]. A relatively higher proportion of overweight individuals may therefore contribute to the lower mean HMRs in the UK and US compared to Japan. Future work using attenuation and scatter corrected SPECT imaging may provide insight into the contribution of patient size to these population differences in apparent cardiac MIBG uptake.

Limitations

This study was retrospective in nature, utilising control and patient data recruited to our centre for wider Lewy body disease studies. The inclusion of older adults with a range of cardiac risk factors and without detailed cardiac assessment to detect undiagnosed cardiovascular disease makes it difficult to determine the impact of cardiac disease on HMR. Very large samples of older adults with and without risk factors for cardiovascular disease could help to determine this, but as seen here, older people often have cardiac risk factors, making it impractical to recruit true healthy controls. As obesity is both a risk factor for cardiovascular disease and a potential cause of artefactually reduced HMR this would also need to be accounted for. If it were feasible, a large prospective study in multiple countries recruiting cognitively normal older adults typical of the local population and scanning them using the same protocol could determine whether HMR is population specific in a more rigorous manner.

7.2.5 Conclusion

We determined cardiac ¹²³I-MIBG heart-to-mediastinum ratio cut-off values derived from a group of older adults in average health without evidence of cognitive decline for the first time. After accounting for camera differences, these values were lower than the values used in Japan, higher than in a large US study but similar to those found in another UK centre.

Our results suggest that normal cardiac ¹²³I-MIBG heart-to-mediastinum ratio values may be population specific, although processing differences cannot be excluded. Further work is required to determine the reasons for these differences, which could include contributions from heart disease, obesity and other co-morbid conditions. This work highlights the importance of deriving HMR cut-offs from controls that are representative of the population being studied.

7.3 Can we use early phase cardiac MIBG images to diagnose Lewy body disease?

7.3.1 Introduction

Many studies using cardiac ^{123}I -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 [217, 231]. However in the MIDAS study we saw little difference between heart-to-mediastinum ratios (HMR) obtained on cardiac MIBG images acquired at 20 minutes post-injection and those acquired at around 4 hours after injection [83, 190], suggesting that initial capability for uptake of MIBG by sympathetic neurons is impaired at the same time as storage mechanisms. Autopsy studies report a loss of sympathetic nerve fibres in Lewy body disease cases [3, 71], suggesting that cardiac MIBG uptake should be reduced on both early and delayed imaging. This was indeed the case in the only study to date to report autopsy findings in patients with cardiac MIBG imaging during life [82], where cardiac MIBG uptake was reduced on early images in 91% of Lewy body disease cases and on the delayed images in 96%.

Three recent cardiac MIBG studies in Lewy body disease carried out in Japan by separate groups [174, 232, 233] have demonstrated no significant difference between diagnostic accuracy of early or delayed imaging, 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 than ^{123}I -FP-CIT, which requires a 3-6 hour uptake period before scanning. This would be particularly helpful for patients with more advanced dementia symptoms, who may find a lengthy period spent in an unfamiliar environment unsettling. Even if the delayed image could only be omitted in a subset of cases, such a protocol would still be worthwhile.

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 [174, 232-234]. However, Yoshita *et al.* [235] 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.

In our publications on MIDAS we failed to find evidence for a significant difference between the diagnostic accuracy of early and delayed imaging [83, 190]. It is possible that there is a true difference in the accuracy of early and delayed imaging but that the MIDAS study dataset is too small for our results to reach statistical significance. The sample size for MIDAS was derived based on the power to

detect a difference in the mean uptake between DLB and AD groups on delayed imaging in a UK population - given the large effect size seen in previous MIBG studies in DLB, relatively small numbers were predicted to be required for MIDAS. Several of the subjects included had quite different uptake levels on delayed imaging compared to early, although this did not usually affect normal/abnormal categorisation with the optimal ROC cut-offs applied [190]. 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 [233, 234].

The objective of this section is to bring together the MIDAS dementia cases and the SUPeRB control and MCI cases used in this thesis to give larger clinical groups. The difference between early and delayed HMR are explored for the whole dataset, providing a wide range of cardiac uptake values as well as for the individual clinical groups. Diagnostic accuracy on early and delayed imaging is compared in both MCI and dementia, and for combined Lewy body (LB) disease and non-LB groups.

7.3.2 Methods

Subjects

All individuals scanned as part of MIDAS or SUPeRB up until January 2018 were used to compare early and delayed HMR values, even those subsequently excluded from the main study analyses – unless this was because of incorrect cardiac MIBG imaging protocol being followed. Two MIDAS and two SUPeRB cases had to be excluded because either the early or the delayed image was inadvertently acquired using the wrong collimator. All other cases were included: 40 MIDAS and 86 SUPeRB (including 31 controls), giving 126 data points in total.

However, to compare HMR between clinical groups and to compare diagnostic accuracy, only a well-characterised subset of 76 of the 95 cases with neurodegenerative disease was included, so as not to reduce the accuracy of gold-standard clinical diagnosis at 1 year follow-up. Patients with less certain diagnoses (i.e. possible DLB or possible MCI-LB), or diagnoses that appeared more in keeping with neurodegenerative diseases other than Alzheimer's or Lewy body disease were excluded from this part of the analysis. The 31 controls were also excluded so that specificity would reflect that of clinical cases. The patients included therefore have a 1 year follow up consensus diagnosis of either probable DLB (prob DLB) or dementia due to Alzheimer's disease (AD) for MIDAS and either probable MCI with Lewy bodies (MCI-LB) or probable MCI due to Alzheimer's disease (MCI-AD) for SUPeRB. The numbers in each group are given in Table 51.

Table 51: Number of patients included in the Lewy body disease and Alzheimer’s disease groups for each category

	Lewy body disease	Alzheimer’s disease
Mild cognitive impairment	26	18
Dementia	17	15
Combined groups	43	33

Comparison between imaging time points

Images were acquired as described in section 7.1 for both MIDAS and SUPeRB. Early and delayed planar cardiac MIBG images were processed to obtain HMR values as described previously, i.e using a 6cm circular cardiac region of interest and a 3 x 4 cm rectangular mediastinum region. Two operators processed the images: the author (GR) and Joseph Kane (JPMK) for MIDAS and the author and Jim Lloyd (JLL) for SUPeRB. The relationship between early and delayed imaging was assessed using linear regression, taking the mean HMR from the two operators for each image. The mean of two operators was taken as previous work on inter-operator repeatability section 7.1 showed variability was low. Bland-Altman plots were used to compare the difference between early and delayed images over the full HMR range more closely.

Comparison between mean early and mean delayed HMRs for LB and AD groups

The dementia and MCI groups were analysed separately and combined into larger Lewy body (LB) and Alzheimer’s disease (AD) groups. The mean HMR values and standard deviation for early and delayed images were calculated. Paired t-tests were used to determine whether a statistically significant difference between early and delayed HMR exists. The mean HMR for the two operators was used.

Accuracy of early and delayed planar imaging for LB and AD groups

The 29 healthy controls from SUPeRB with normal images were used to derive HMR cut-offs, as described in the previous chapter. The HMR cut-offs were used to assign each scan as normal or abnormal. Sensitivity, specificity and overall accuracy were calculated for each of the clinical groups described above. Since sensitivity and specificity values are expressed as a percentage, binomial confidence intervals for proportions were calculated using the “Exact Binomial and Poisson Confidence Intervals” online tool available at <http://statpages.info/confint.html>

7.3.3 Results

Comparison between imaging time points

The plots in Figure 70 show slightly worse agreement between HMRs at early and delayed time points than was seen between operators. This is also reflected in the goodness-of-fit parameters (Table 52),

with R^2 values of around 85% suggesting around 15% of the variance in delayed HMR is not predicted by early HMR. The goodness-of-fit is very similar for HMRs produced by a single operator (GR) as for both operators combined (Table 52; plots not shown). The Bland-Altman plots in Figure 71 show that for normal HMR values (mean early and delayed HMR greater than around 2.0) there is no systematic difference between the HMRs on early and delayed imaging. However abnormal scans (mean early and delayed HMR less than around 2.0) are much more likely to have a drop in HMR between early and delayed imaging. The HMR histograms in Figure 72 show a clearer separation of the data into two distinct normal and low uptake groups on delayed imaging.

Table 52: Goodness-of-fit parameters for linear regression analysis between early and delayed time points

Dataset	Operators	R^2 (% explained variance)	Standard error of regression (in units of HMR)
MIDAS	GR+JPMK mean	91%	0.21
	GR only	89%	0.23
SUPeRB	GR+JLJ mean	83%	0.25
	GR only	84%	0.24
MIDAS+SUPeRB	Mean	86%	0.24
	GR only	85%	0.24

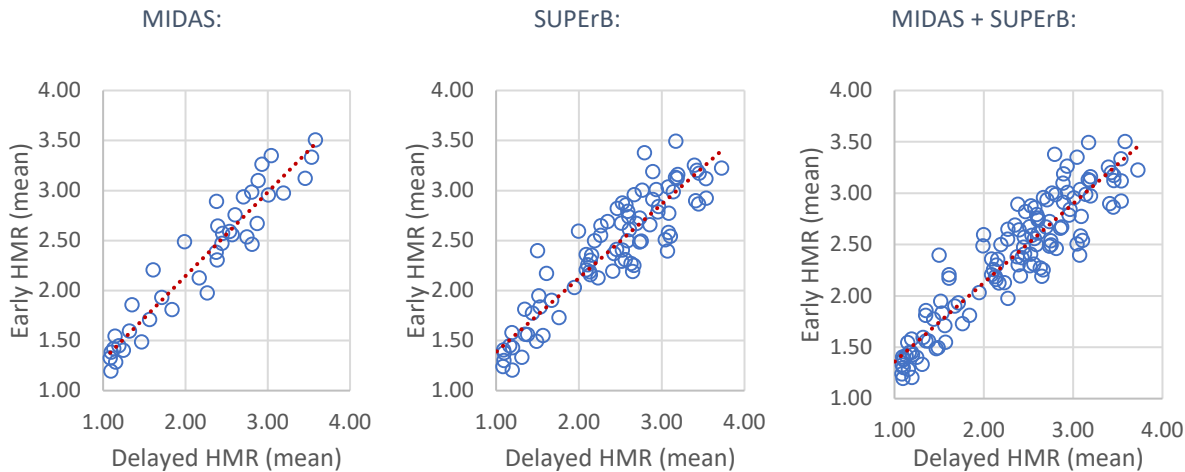


Figure 70: Correlation between early and delayed images (mean of both operator HMRs) for the MIDAS and SUPeRB datasets and for the two studies combined.

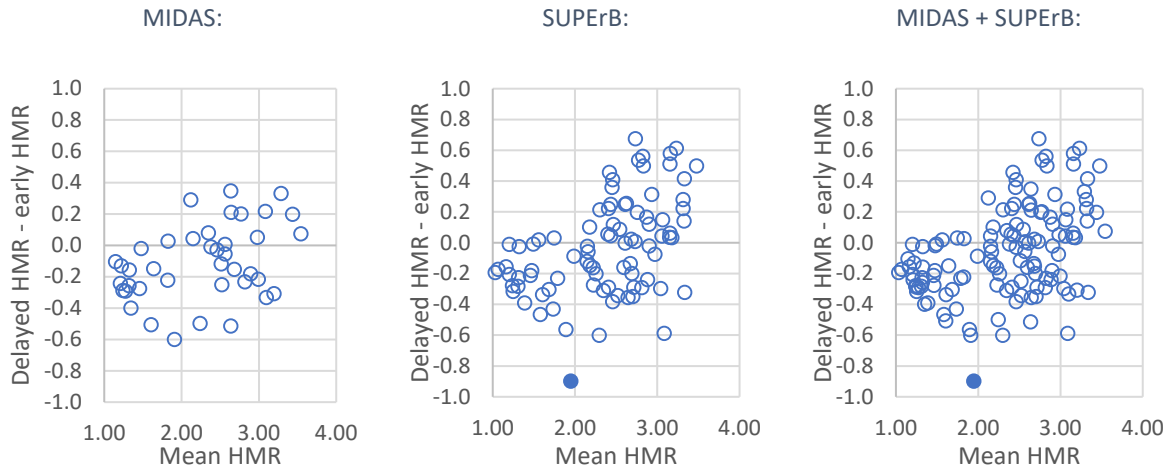


Figure 71: Bland-Altman plots of the difference between delayed and early HMR for MIDAS, SUPeRB and both datasets. The data point shown as a solid circle is one of the controls excluded from the main analysis as the early image was normal and delayed image abnormal.

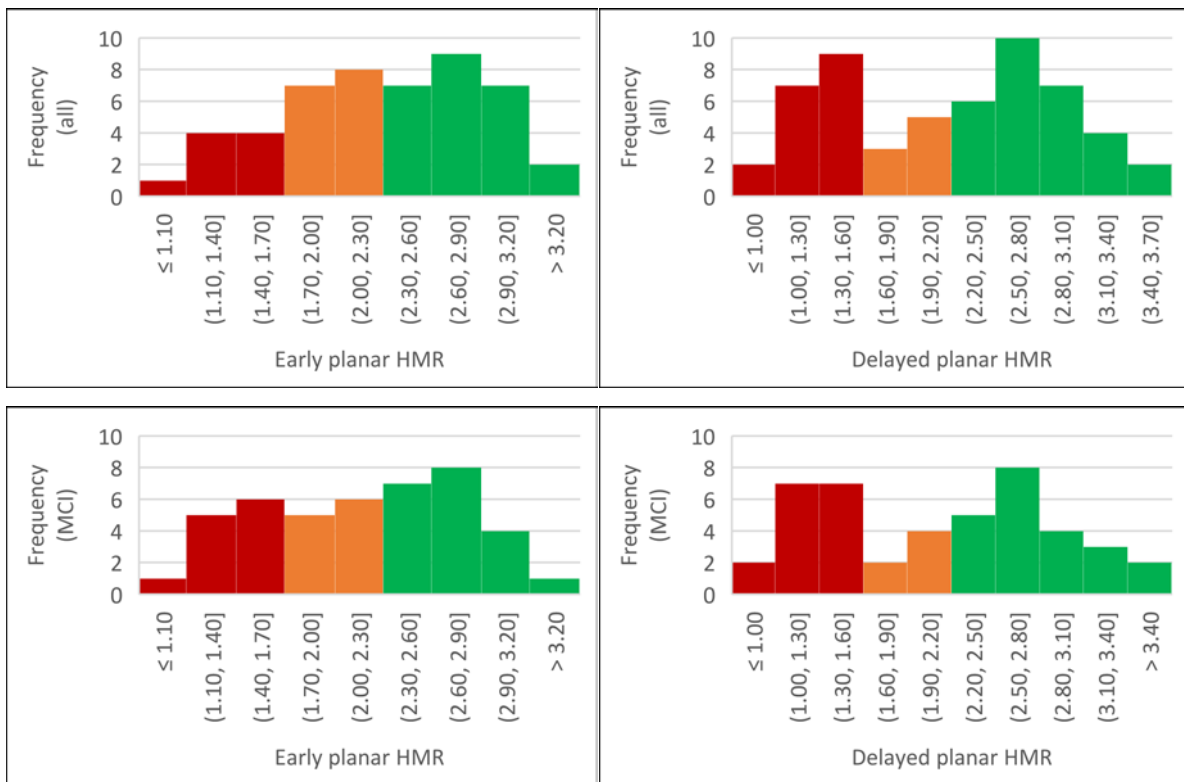


Figure 72: Histograms for early and delayed HMR for all subjects (top) and for MCI subjects (bottom). The red categories are below the normal cut-offs, orange borderline and green normal.

The numbers of cases included in each of the clinical groups is given in Table 53. Paired t-tests showed no significant difference in the group means between early and delayed imaging for controls or AD cases (Table 53). There was however a significant difference for probable DLB cases, with the early image HMR higher by 0.19 (16%) on average. There was no significant difference in mean early and delayed HMR for the probable MCI-LB cases from SUPeRB. The lack of difference in this dataset is to be expected, given the Bland-Altman plots above, as around half the MCI-LB cases have normal HMRS (see below).

Table 53: Difference between mean HMR on early and delayed images for the different clinical groups.

	n	Mean early HMR	Mean delayed HMR	Mean difference	Mean % difference	SD % difference	p-value (paired t- test)
Controls	29	2.71	2.78	0.07	2%	11%	0.238
MCI-AD	18	2.35	2.26	-0.10	-5%	14%	0.201
Prob MCI-LB	26	2.02	1.97	-0.05	-5%	14%	0.382
AD (MIDAS)	15	2.47	2.48	0.01	1%	7%	0.843
Prob DLB (MIDAS)	17	1.88	1.69	-0.19	-12%	11%	0.002
DLB (abnormal only)	12	1.47	1.23	-0.24	-16%	9%	<0.001
DLB+Prob MCI-LB	43	1.97	1.86	-0.10	-8%	13%	0.015

Accuracy of early and delayed planar imaging for LB and non-LB groups

The categorisation of scans into normal or abnormal groups is given in Table 54 for probable DLB and AD (MIDAS) and probable MCI-LB and MCI-AD (SUPeRB) with the sensitivities and specificities in Table 55. In all but three cases (3/76; 4%) the categorisations are the same on both early and delayed images (2 from MIDAS and 1 SUPeRB). One probable DLB case had a borderline-normal HMR on early imaging (2.15; cut-off 2.07), which fell to 1.61 on delayed imaging. One AD dementia case had a borderline-abnormal early HMR of 2.01, increasing to a clear normal 2.26 on delayed imaging. From the MCI dataset there was only one discrepancy between early and delayed results: an MCI-AD case with an early HMR of 2.17, dropping to 1.61 on delayed images. The images were reviewed and this does appear to be a true difference, rather than a collimator or processing error. As there are only 18 MCI-AD cases, this discrepancy is the cause of the 5% greater reported specificity for the diagnosis of MCI-LB on early images (83 vs 78%). If the datasets are grouped together, the specificity for diagnosing Lewy body disease is the same for early and delayed imaging at 82% (95% CI: 65 to 93%). The overall sensitivity is slightly higher for delayed imaging (60% vs 58%) - but this difference is due to just the one extra DLB case that fell to below the normal range on delayed imaging. Although the accuracies are

similar, the histograms in Figure 72 suggest there is better separation of abnormal and normal scans on delayed imaging.

Table 54: Early and delayed normal-abnormal categorisation for prob DLB, AD, prob MCI-LB and MCI-AD groups. Discrepancies between early and delayed results occur on the positive diagonal and are shown in red font.

Prob DLB		Early		AD		Early	
		normal > 2.10	abnormal < 2.10			normal > 2.10	abnormal < 2.10
Delayed	normal > 1.86	5	0	Delayed	normal > 1.86	12	1
	abnormal < 1.86	1	11		abnormal < 1.86	0	2

Prob MCI-LB		Early		MCI-AD		Early	
		normal > 2.10	abnormal < 2.10			normal > 2.10	abnormal < 2.10
Delayed	normal > 1.86	12	0	Delayed	normal > 1.86	14	0
	abnormal < 1.86	0	14		abnormal < 1.86	1	3

Table 55: Sensitivity, specificity and overall accuracy figures for the diagnosis of Lewy body disease on early and delayed cardiac MIBG imaging, with 95% confidence intervals calculated using the binomial distribution.

	Dementia		MCI		Combined data	
	Early	Delayed	Early	Delayed	Early	Delayed
Sensitivity	65% (38 to 86 %)	71% (44 to 90 %)	54% (33 to 73 %)	54% (33 to 73 %)	58% (42 to 73 %)	60% (44 to 75 %)
Specificity	80% (52 to 96 %)	87% (60 to 98 %)	83% (59 to 96 %)	78% (52 to 94 %)	82% (65 to 93 %)	82% (65 to 93 %)
Accuracy	72% (53 to 86 %)	78% (60 to 91 %)	66% (50 to 80 %)	64% (48 to 78 %)	69% (58 to 79 %)	71% (59 to 81 %)

7.3.4 Discussion

Bland-Altman plots of the pooled cardiac MIBG scans carried out for MIDAS and SUPeRb revealed a difference between early and delayed imaging for low HMR values only, with no apparent difference between early and delayed HMR for scans with normal cardiac uptake. Paired t-tests on well characterised subsets of the data confirmed that in healthy controls and AD there was on average no

difference between early and delayed HMR, whereas in probable Lewy body disease early HMR was significantly higher than delayed HMR, particularly for the DLB cases. It is likely that the early images contain a higher degree of non-specific uptake in the cardiac region, which washes out later and is not so apparent on delayed images. This reduction in non-specific uptake would have more effect when the level of specific uptake is low (or zero), which may explain why the DLB patients had lower HMRs on delayed images. The histograms of HMR distribution also suggested a clearer distinction between normal uptake and low uptake groups on delayed imaging than on early imaging, for both the combined dataset of dementia and MCI patients and the MCI dataset alone. While this is evidence in favour of delayed imaging in principle, a larger separation between LB and non-LB groups does not necessarily equate to superior diagnostic accuracy – the accuracy of early imaging may be equivalent in practice if the smaller separation between groups is more than adequate. We see in our dataset that the vast majority of scan results do not change between early and delayed, and that those clearly abnormal early on do not go on to become normal on delayed imaging. These results also suggest that the exact timing of the early and delayed imaging is not critical, which is useful to know in clinical practice as it can be difficult to schedule a patient for an exact timeslot.

Published papers from Japan [174, 232, 233] have shown no statistically significant difference in the area under the ROC curve between early and delayed imaging. However, this is a questionable statistical approach as the objective of these papers was to show that early imaging could be used instead of delayed. They use a null hypothesis of no difference but they have already said they want to show that early and delayed imaging give the same results. So while they showed that there was insufficient evidence that delayed imaging has better accuracy than early imaging, this does not prove that there is no difference in accuracy. In my opinion, one of the studies provides weak evidence, rather than no evidence, for delayed imaging being more accurate (AUC 0.871 early, 0.893 delayed; $p = 0.091$) [232]. These three studies were quite large with 453, 111 and 192 cases included. However, they were all done retrospectively, on clinical patients attending for diagnostic testing and as such the results are likely to have influenced the clinical groupings. A key strength of our work is that diagnoses were made as part of research studies and done entirely independently of cardiac MIBG imaging; however a limitation is the relatively small sample size, especially when only the patients with more certain diagnoses are included. Since only a small proportion of cases (3/76, 4%) changed their normal-abnormal designation between early and delayed imaging we can conclude that the accuracies of early and delayed imaging must be similar. However it is clear from reviewing the 95% confidence intervals that this study is not powered to detect clinically important difference in sensitivity or specificity between early and delayed imaging even with the MIDAS and SUPeR datasets combined. A power calculation (assuming 80% power and a 5% significance level) suggests that 8338 subjects per group would be needed to detect the 2% difference in accuracy seen here! The very small difference in

accuracy is not clinically important, especially in the context of other factors which might contribute to larger variations in the MIBG data such as obesity, cardiac disease and physiological fluctuations in uptake.

Taken together, the results in this chapter suggest that we should assume that delayed imaging is likely to be slightly more accurate than early imaging, but only by a small percentage - so the majority of the time the outcome on the early image will be the same. The question is then, what range of early HMR results can we accept as being definitively normal or abnormal (when requiring a binary result) and which range should be regarded as equivocal? I.e. what are the early HMR upper and lower cut-off values that result in delayed HMR values that are statistically likely to be above or below 1.86? The standard error that we found earlier from linear regression can be used to define a confidence interval, i.e. a range of values that the delayed HMR is likely to fall within for a given early HMR value. The least-squares best fit relationship between early and delayed HMR for all data is:

$$\textit{Delayed HMR} = (1.1124 * \textit{Early HMR}) - 0.3226$$

The standard error in delayed HMR from the linear fit is 0.286, so a 95% confidence interval, assuming a normal distribution, is $1.96 * SE$, which is ± 0.56 (in units of delayed HMR). This means that the delayed HMR is very likely to fall within ± 0.56 of the value calculated above. Rearranging, we find that the delayed cut-off value of 1.86 is given by an early HMR value of 2.02 ($0.7696 * 1.86 + 0.5887$).

$$\textit{Early HMR} = (0.7696 * \textit{Delayed HMR}) + 0.5887$$

But other delayed HMRs could also occur when the early HMR is 2.02: 95% of the time we would expect the delayed values to fall between 1.30 and 2.42 (1.86 ± 0.56). If we want to be reasonably confident of an early result being normal or abnormal, one approach is to work out what range of early HMRs could result in a delayed value of 1.86 with 95% confidence. The standard error in early HMR (from linear regression) is 0.238 so the 95% confidence interval is ± 0.467 . The early value most likely to give a delayed value of 1.86 is 2.02. 2.02 ± 0.467 results in a lower limit of 1.55 and upper limit of 2.47 for early HMR values. This means that an early HMR below 1.55 is very likely to be abnormal (<1.86) on delayed imaging and any value above 2.47 is likely to be normal. Applying this method in clinical practice, any patient with an early HMR between these values should return for delayed imaging to clarify. In our full dataset of 126 scans scanned with MELP collimators, 20 had early HMRs below 1.55, none of which were over 1.86 on delayed imaging. None of the 67 cases with early HMRs above 2.47 had abnormal delayed HMRs. We could therefore have omitted the delayed image in 69% of the cases and have still been confident of the results. In terms of the practicalities of this suggestion, since the HMR on the early image is very quick to calculate using our method presented in section 7.1, it would

be feasible to inform the patient whether they needed to return within 10 minutes of the early scan being completed.

We note that this method does not take into account the non-linearity seen over the HMR range (i.e. that early HMR is likely to be higher than delayed HMR when uptake is low), so one could argue that the confidence intervals would be best calculated separately for normal and abnormal HMRs. This would however increase the standard errors and therefore confidence intervals due to the smaller sample, making the equivocal range even wider. An alternative approach could be to assume that any case already below the delayed HMR cut-off of 1.86 on early imaging is very likely to remain below 1.86 on delayed imaging. There were 12 further cases with early HMR between 1.55 and 1.86 in our dataset and all were below 1.86 on delayed imaging. This means that in total 78% of the delayed scans (99/126) could have been omitted with this additional approach. Of the remaining 27 scans, there were five where the early and delayed result differed: three changed from normal on early images (>2.10) to abnormal on delayed (<1.86) and 2 changed from abnormal to normal.

7.3.5 Conclusion

Sensitivity and specificity values using HMR cut-offs derived from healthy controls are very similar for early and delayed imaging, showing that any difference in diagnostic accuracy between the time points is very small. However there is some suggestion from our data that the later imaging could separate Lewy body and non-Lewy body groups better than the early imaging; there would be some theoretical support for this view as non-specific uptake will wash out later.

Considering that much of the literature uses delayed imaging and there are evidence-based reasons to prefer it, we would recommend that imaging 3-4 hours after injection remains the reference standard. However we suggest that in clinical practice centres could omit the delayed image if the early results were unequivocal and have provided suggested ranges for this, valid for Siemens gamma cameras with medium energy collimators. In our dataset, 69% of cases would count as non-equivocal on early imaging, 78% if the lower limit of the equivocal range was increased to 1.86. Adopting this approach in clinical practice would remove the need for the delayed image in the majority of cases, with negligible potential impact on diagnostic accuracy.

7.4 Summary of Chapter 7

In section 7.1 we evaluated the accuracy and reproducibility of a simple method for calculating HMR on cardiac MIBG images, suitable for use in both routine clinical practice and research studies. This method gives similar accuracy to a bespoke cardiac region tailored to the patient and better reproducibility. The method has been adopted for HMR calculations in the SUPeRB study and is used throughout the thesis. We also confirmed that semi quantification is crucial for cardiac MIBG imaging, with visual raters tending to overcall subtle reductions in cardiac uptake, leading to reduced specificity. However, HMR analysis is of course only accurate if a suitable cut-off value is available, and currently this is a barrier to the technique being used outside research settings, at least in the UK.

The issue of normal thresholds is discussed in more detail in section 7.2, where we evaluate whether it is appropriate to apply HMR cut-off values from the Japanese literature to our UK population. It was seen that HMRs are significantly higher in Japan than in our study and could lead to misdiagnoses if applied in the UK. Potential population specific causes, such as obesity and heart disease, were discussed. The HMR values for controls from the large US ADMIRE-HF study [218] were compared with ours and shown to be lower, after correction for collimator differences. However HMR results for a cohort of 14 controls recruited and scanned in Manchester were similar to our results [216]. We presented a series of collimator-specific approximate HMR cut-offs based on control data from the SUPeRB study that could be used by centres not wishing to switch to medium energy collimators.

Cardiac MIBG imaging using the early phase only could be a good alternative to FP-CIT for people whose dementia symptoms make a lengthy procedure challenging. In our dataset we found that delayed images appear to separate patients with normal and abnormal uptake better than early images, but overall accuracy was very similar, with only 4% of subjects changing normal/abnormal categorisation between early and delayed. We suggest that, rather than switching to early imaging entirely, the delayed imaging could be omitted if the early image is non-equivocal, which was the case in 78% of scans in our combined MIDAS and SUPeRB dataset.

In summary, we have provided data, both this chapter and chapter 5, to enable us to propose an accurate and practical evidence-based cardiac MIBG protocol for image acquisition and analysis. This protocol could be used in clinical practice and in research studies requiring a binary normal/abnormal sympathetic innervation result. We suggest that patients should be administered 111 MBq and a planar anterior image of the chest acquired around 20 minutes later, using medium energy collimators. If the HMR, calculated using our 6 cm circle ROI method, is below 1.6 or above 2.5 the patient should be informed that the test is complete. If the HMR at 20 minutes is equivocal the patient should be asked to return for a repeat scan to be carried out 3-4 hours after injection. For research studies focussing on cardiac MIBG we recommend that delayed images are acquired for all subjects.

Chapter 8: Tomographic cardiac ^{123}I -MIBG analysis

In this chapter we explore more advanced methods for cardiac MIBG imaging using SPECT and SPECT-CT and assess whether these techniques are likely to improve the diagnosis of Lewy body disease compared to standard planar imaging.

In section 8.1 we use a phantom to demonstrate the effect of attenuation and scatter on apparent uptake with planar and tomographic methods. We investigate whether planar cardiac uptake results in control subjects are linked to patient size and explore whether using SPECT-CT to correct for attenuation and scatter can improve quantification in the phantom and controls. In section 8.2 we extend these methods to MCI patients from the SUPeRB study and compare the diagnostic accuracy of planar imaging with both SPECT (without attenuation and scatter correction) and SPECT-CT (with corrections).

In section 8.3 we outline methods for quantifying the absolute uptake of MIBG within the heart, and compare the accuracy of these methods to both planar and tomographic HMR methods. In section 8.4 we use SPECT imaging to examine the pattern of cardiac MIBG uptake within the myocardium, comparing controls with normal cognition to patients from the MIDAS study with either dementia with Lewy bodies or Alzheimer's disease. In section 8.5 we summarise these results and discuss whether there is a role for tomographic cardiac MIBG imaging in the diagnosis of Lewy body disease.

8.1 Planar cardiac MIBG HMRs depend on patient size – can SPECT-CT reduce dependence?

8.1.1 Introduction

Systematic equipment and processing differences in cardiac ^{123}I -MIBG HMRs between centres can be accounted for with planar phantom calibration, as described in sections 5.2 and 7.2. However, patient factors affecting HMR values, such as variation in the amount of attenuation and scatter due to patient size, cannot be taken into account with this method. There have been few publications commenting on whether patient factors may have a significant impact on diagnostic value of planar cardiac MIBG scintigraphy, and to our knowledge none that have explored whether SPECT with CT attenuation correction and scatter correction (CT ACSC) would be of benefit. Although several groups investigating the diagnostic accuracy of cardiac ^{123}I -MIBG for the diagnosis of Lewy body disease have acquired SPECT data alongside planar, often it is only the planar scans that have been used for analysis with the SPECT results not reported [55, 183, 236-238], suggesting the added value of SPECT is unclear. In 2016 Odagiri *et al.* reported that the accuracies of planar and SPECT-CT ^{123}I -MIBG cardiac scintigraphy were similar, but did not use the CT data for SPECT image corrections, only for localising the heart [239]. In their 2015 multicentre study, Yoshita *et al.* used SPECT for visual analysis but not for quantification of cardiac uptake [53].

Obesity has been linked to decreased apparent cardiac MIBG uptake by two previous studies [181, 228], and is thought to be due to increased attenuation and scatter in larger individuals, rather than to true differences in cardiac uptake. The largest study of healthy controls recruited to a cardiac MIBG research study (n=94) was that of Jacobsen *et al.* [218] carried out as part of the ADMIRE heart failure study [63]. The controls were aged between 29 to 82 years (mean 58.5 ± 10.6 years, with a wide BMI range of 20.4 to 44.3 kg/m² (mean 28.7 ± 5.1). The study reported a “modest” decrease in HMR with BMI, however this was a study into ageing and the incidental finding of a relationship between BMI and HMR was not discussed in detail. Pellegrino *et al.* reported lower HMRs in patients with BMI>30 (n=10) compared to those with BMI<30 (n=35) [228]. This was in heart failure patients and the results have not yet been replicated in healthy controls but do suggest that patient size may affect apparent cardiac MIBG uptake.

In this study we aimed to establish the degree to which planar cardiac ^{123}I -MIBG scintigraphy is affected by patient size and whether SPECT with CT ACSC can correct for this. We investigated this using an anthropomorphic chest phantom configured with different amounts of attenuation and a cohort of older adults with normal cognition with BMI between 22 and 38 kg/m².

8.1.2 Methods

Subjects

Twenty-nine volunteers aged over 60 years with normal cognition were recruited as part of the SUPeRB study, as described in section 7.2 and elsewhere in this thesis.

Image acquisition

Ten minute anterior planar images were acquired at 20 minutes and at 4 hours (± 30 minutes) after injection, as described in Chapter 7. SPECT-CT imaging was carried out immediately after the delayed planar image with the subject in the supine position with arms raised if possible. Only ten of the volunteers were able to be scanned with arms raised, the others kept their arms by their sides. SPECT images were acquired over 360 degrees in H mode (detectors oriented parallel to each other) to obtain 120 projections of 20 seconds using a non-circular autocontoured orbit. The energy window was $159 \text{ keV} \pm 7.5\%$, matrix size 64×64 and zoom factor 1.64. Following SPECT imaging, a low dose CT scan (130 kV with a quality reference mAs of 13) was acquired for attenuation and scatter correction.

Image reconstruction

All images were processed on a Hermes workstation (Hermes Medical Ltd, Stockholm Sweden). SPECT images were processed in Hermes Hybrid Recon Cardiology v1.1.2 using OSEM iterative reconstruction with 15 iterations and 3 subsets, with a Butterworth filter of 0.90 cm applied. Resolution recovery for the MELP collimator was applied, with parameters set for the Siemens Intevo gamma camera as provided by Hermes: hole diameter 0.294 cm; hole length 4.06 cm; detector intrinsic resolution at 140 keV 0.38 cm; radius of rotation off-set 4.76 cm. Two sets of images were reconstructed, both using resolution recovery: non-corrected (NC) and CT corrected for attenuation and scatter (ACSC). The scatter correction model is based on a Monte Carlo simulation of scatter within the patient [107, 123]. Full collimator and detector response modelling [125] is not included in this model.

Phantom image acquisition

Images were acquired with the RSD torso phantom with fillable heart, lung, liver and background compartments (Figure 73). The original heart insert supplied with the phantom did not give realistic-looking cardiac MIBG images so was substituted with a simple left ventricle insert manufactured in-house with fillable myocardial wall and cavity compartments. The phantom was filled and scanned with ^{123}I , using activity concentrations designed to mimic patient studies. The activities used are given in Table 56. The left ventricle insert was imaged alone on the scanner bed without any additional attenuation followed by three filled phantom scans with increasing amounts of attenuation: torso phantom, torso phantom with solid PMMA chest extension plate and torso phantom with chest plate and breasts (plastic shells filled with water) – see Figure 73.

The scan parameters were the same as those used for clinical research studies at our centre: MELP collimators, auto-contouring, 1.64 zoom, 64 x 64 matrix size. As with clinical scans, CT was acquired so that both attenuation and scatter corrected and non-corrected images could be reconstructed. Images were reconstructed using the clinical parameters given above.



Figure 73: Anthropomorphic torso phantom (centre), with in-house replacement LV insert (left) and additional PMMA chest plate and water filled breasts (right)

Table 56: activity concentrations used in the anthropomorphic torso phantom, selected to give count densities similar to healthy control images.

	Activity / MBq	kBq / ml	Concentration relative to background
LV dose (diluted into 200 ml, actual volume ~160 ml)	27.7	138.7	22
Background (approx. 8200 ml)	52.6	6.4	1
Liver (1000 ml)	78.6	78.6	12.2
Lungs (diluted in 1000 ml before filling)	50.1	50.1	7.8

Image analysis for cardiac uptake

The images of the torso phantom and the images of the older adult volunteers were analysed in the same way. Anterior planar images were analysed to obtain the heart to mediastinum ratio (HMR) using a 6 cm circular ROI placed over the left ventricle and 4 x 3 cm rectangular ROI between the lungs in the mediastinum, see [190]. The phantom was not moved from the scanner couch between acquisitions, allowing the same regions to be used for all three attenuation configurations.

The SPECT coronal slices were summed over all 64 slices to give images analogous to planar images. We obtained HMRs on these summed images in the same way as for the planar and we term these

“coronal SPECT HMRs”. The summed coronal SPECT-CT technique was included because it should be a simple and reproducible way of analysing attenuation and scatter corrected SPECT images, using planar analysis methods. The same regions of interest were used for planar and summed coronal SPECT HMRs, with no adjustment between the non-corrected and ACSC SPECT (see Figure 74 for example). This enables a direct assessment of the effects of attenuation and scatter correction to be made.

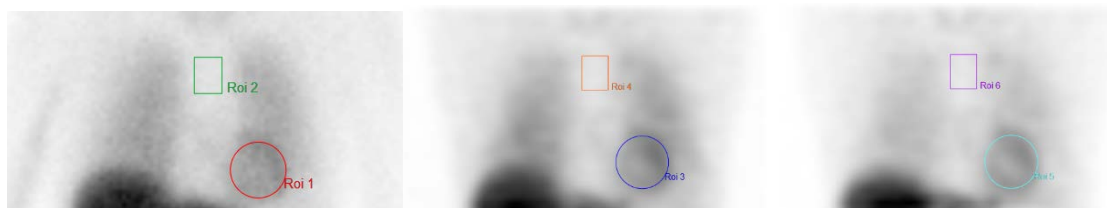


Figure 74: Planar image (left) with corresponding summed coronal SPECT slice images without CT ACSC (middle) and with corrections (right). Note the relative reduction in lung uptake and increased heart contrast on the ACSC image.

Volumetric HMRs on clinical and phantom SPECT images were calculated using 6cm diameter spherical VOIs placed over the left ventricle and cuboid VOIs over the mediastinum (Figure 75). The placement of VOIs was not modified between the ACSC and non-corrected datasets.

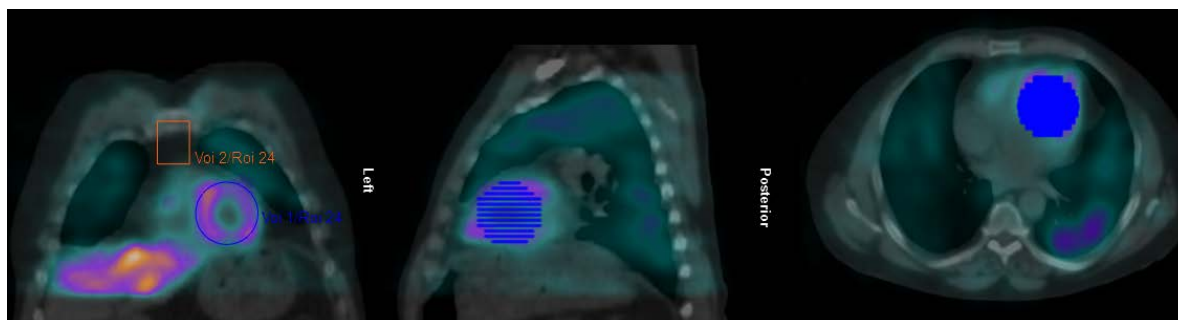


Figure 75: Example control SPECT-CT images showing coronal, sagittal and transverse slices centred on the left ventricle, with spherical cardiac VOI and rectangular mediastinal VOI shown.

Planar and SPECT HMR data for the controls were plotted against BMI and the slopes and standard errors calculated using linear regression. The standard errors of the slopes were used to calculate t-statistics to determine whether each relationship between HMR and BMI (for planar or SPECT) was significant at the $\alpha = 0.05$ level. To compare slopes, normalisation was necessary since average HMRs are much higher for SPECT than planar due to increased contrast. The SPECT slopes and the standard errors were scaled to be consistent with planar data by multiplying by the ratio between the mean

planar HMR and HMR in question. The differences between the normalised slopes and their combined standard errors were used to determine whether corrections for attenuation and scatter had a significant impact on reducing the dependence of HMR on BMI.

8.1.3 Results

Torso phantom

The appearance of the phantom on planar images was significantly affected by additional attenuation, whereas this was not the case with the attenuation and scatter corrected SPECT images (Figure 76). The HMR on planar imaging was reduced by 20% with the additional PMMA chest plate added and by 40% with the plate and water filled breasts added (Figure 76 and Figure 77). The summed coronal SPECT images showed an increase in HMR with attenuation, due to lower mediastinum counts – the appearance of the images was reasonably preserved. The SPECT count data, corrected for attenuation and scatter, and decay corrected, showed a much lower drop with additional attenuation.







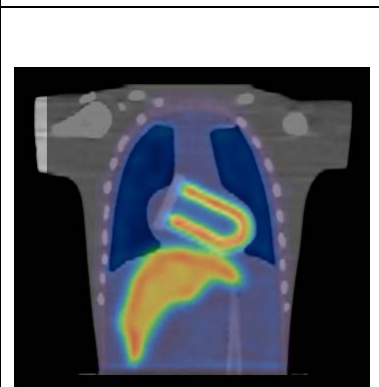
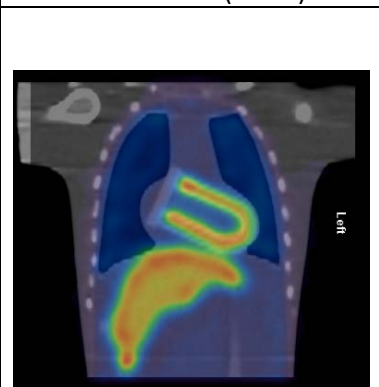
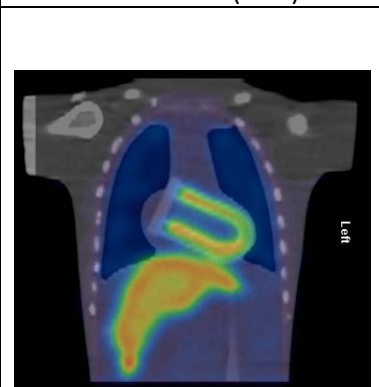
	Standard torso phantom	With chest plate attenuation	With chest plate and breast attenuation
Standard planar imaging			
	HMR = 3.99	HMR = 3.20 (-20%)	HMR 2.41 (-40%)
Summed coronal ACSC SPECT slices			
	HMR = 6.87	HMR = 7.84 (+14%)	HMR = 7.44 (+8%)
ACSC SPECT cardiac counts			
	2548	2510 (-2%)	2462 (-3%)

Figure 76: The effects of attenuation and scatter on planar and ACSC SPECT images of the torso phantom. The top row shows the planar images, middle row the ACSC summed coronal slices and the bottom row the ACSC SPECT-CT images. The planar and summed coronal SPECT HMRs are calculated using 6 cm circular cardiac ROIs; the ACSC SPECT mean cardiac counts using a 6 cm spherical VOI.

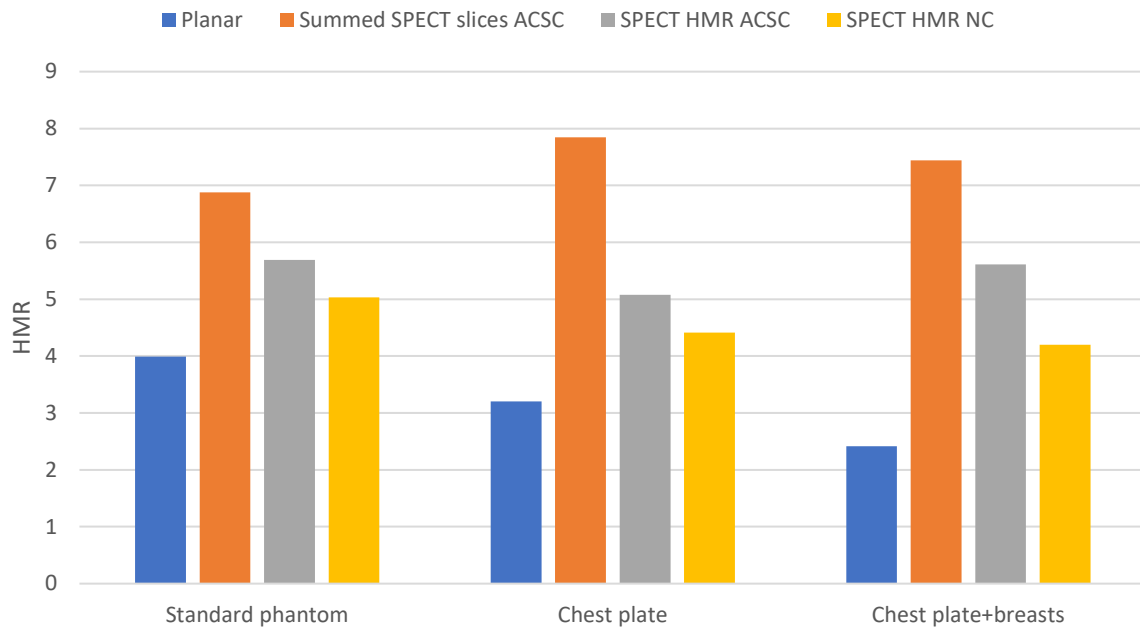


Figure 77: Phantom heart to mediastinum ratios for the three configurations, calculated from planar images, summed ACSC coronal SPECT images and volumetric HMRs from ACSC and non-corrected (NC) SPECT. Both the planar and non-corrected SPECT HMRs decrease with additional attenuating material.

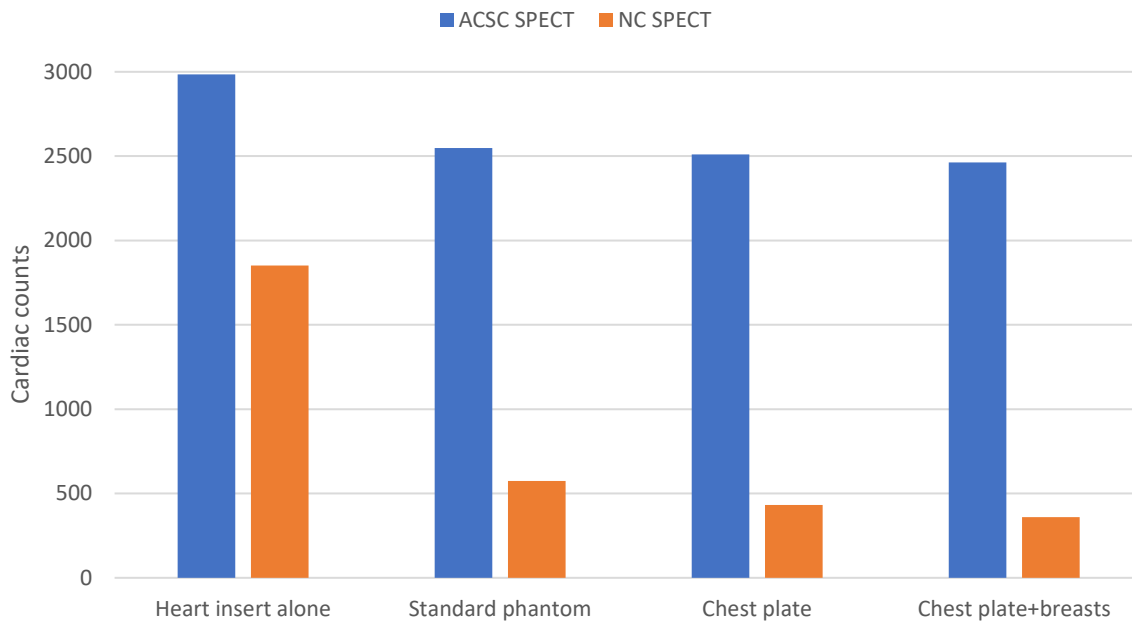


Figure 78: Counts in cardiac volume of interest with ACSC and non-corrected SPECT for the heart insert scanned alone and with increasing attenuation within the torso phantom. The drop in counts when the insert is scanned within the phantom is much less with ACSC applied.

Clinical results

On planar imaging a negative linear correlation between HMR and BMI is seen (Figure 79), with a slope that is significantly different from zero ($p < 0.01$), resulting in a 32% drop in HMR on average between BMI of 20 and 35 (Table 57). In comparison, the drop in HMR over this BMI range with attenuation and scatter corrected SPECT was 14% on average for the summed coronal slice method and 16% for the volumetric SPECT method (with 6cm diameter spherical VOI). However, comparison between normalised slope values (Table 58) shows no significant difference between planar and either summed coronal ACSC SPECT ($p = 0.11$) or volumetric ACSC SPECT ($p = 0.36$).

Figure 80 shows the plots of HMR or cardiac counts against BMI for the summed coronal SPECT images and SPECT volumetric HMR. These are displayed with the non-corrected data on the left and the ACSC data on the right for each method. The linear fits, and p values for slopes being different from zero for planar and SPECT imaging are given in Table 57. The p -values show that with the summed non-corrected SPECT coronal slices, there is a significant drop in HMR with BMI but with attenuation and scatter correction this is not statistically significant. However, when comparing the slopes for non-corrected and ACSC coronal HMR data directly (Table 58) there is no evidence of a statistically significant difference. SPECT volumetric HMR (spherical VOI) data show no significant relationship with BMI either with or without ACSC, and have the highest normalised standard errors, possibly due to greater uncertainty in cardiac and mediastinum volume of interest placement compared to the other 2D methods. The coefficients of variation of the HMR data (independent of any relationship with BMI) are also higher for the volumetric SPECT data than for the planar or summed coronal HMRs (Table 57).

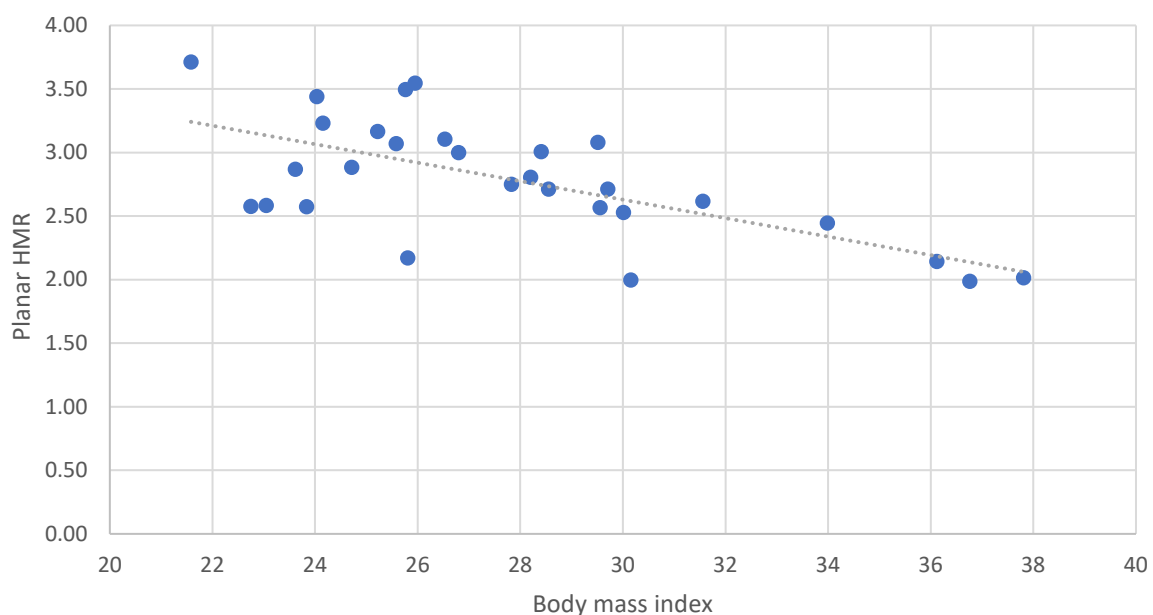


Figure 79: planar HMR plotted against BMI for the 29 older adults, with linear fit.

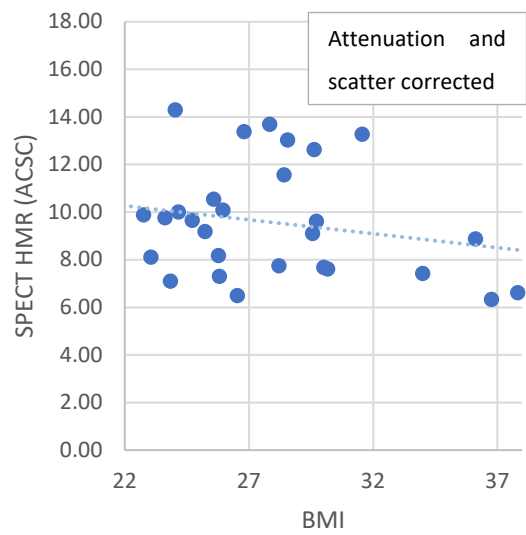
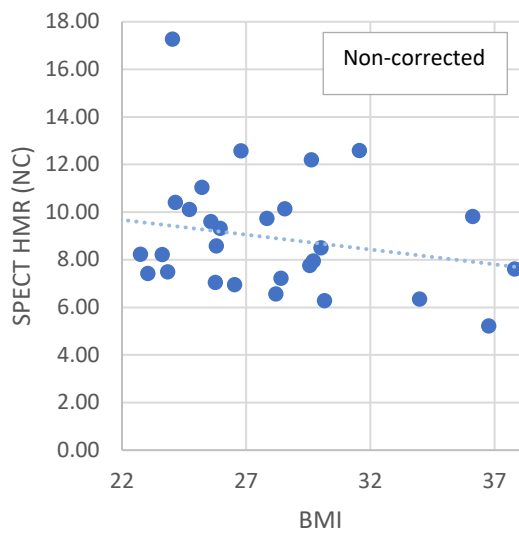
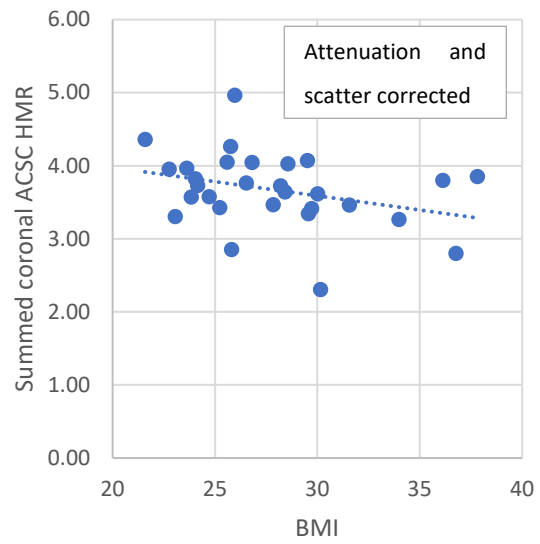
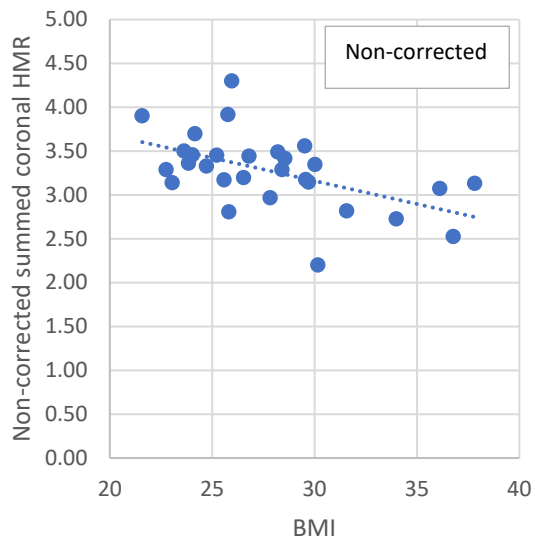


Figure 80: Summed coronal HMR with and without ACSC (top), SPECT volumetric HMR with and without ACSC (bottom).

Table 57: Linear regression between HMR and BMI for planar data and SPECT with and without attenuation and scatter correction.

Quantitative measurement	Planar	Summed SPECT coronal		Volumetric SPECT HMR	
	HMR	slices HMR			
CT ACSC?	n/a	Yes	No	Yes	No
Coefficient of variation	17%	14%	13%	25%	28%
Linear fit to BMI	-0.067*BMI + 4.61	-0.039*BMI + 4.76	-0.053*BMI + 4.74	-0.117*BMI + 12.84	-0.125*BMI + 12.42
R ²	0.44	0.11	0.28	0.05	0.05
Normalised slope	n/a	-0.028	-0.037	-0.034	-0.039
SE of slope	0.017	0.023	0.017	0.104	0.11
Normalised SE	n/a	0.017	0.015	0.03	0.034
p-value for non-zero slope	<0.01	0.09	<0.01	0.27	0.27
Predicted drop in value between BMI 20 and 35	3.36 to 2.27 (-32%)	3.95 to 3.39 (-14%)	3.69 to 2.90 (-21%)	10.5 to 8.74 (-16%)	9.93 to 8.06 (-19%)

Table 58: Comparison between normalised slopes for planar and SPECT HMR with BMI, and for attenuation and scatter corrected SPECT data vs no corrections.

Comparison	Difference in normalised slopes	Combined normalised standard error	T-statistic	P-value
Planar HMR vs ACSC summed coronal HMR	0.038	0.022	1.60	0.11
Planar HMR vs ACSC volumetric SPECT HMR	0.032	0.035	0.932	0.36
Summed coronal: ACSC vs no corrections	0.008	0.022	0.380	0.71
Volumetric SPECT: ACSC vs no corrections	0.005	0.046	0.104	0.92

8.1.4 Discussion

A strength of our study is the inclusion of both phantom and clinical data. The phantom data was acquired under controlled conditions allowing us to investigate the impact of increasing phantom size alone with no other sources of variation. The clinical data applies image reconstruction in more realistic conditions to determine whether the effect holds in practice.

Our results in the torso phantom show a clear advantage of SPECT-CT ¹²³I imaging with attenuation and scatter correction over planar imaging and over SPECT without CT ACSC. The visual appearance of the phantom reconstructed with the planar-equivalent summed ACSC SPECT images was greatly improved compared to the standard planar images; this is particularly apparent on the third configuration with both the PMMA chest plate and the water filled breast inserts added. The counts within the left ventricle insert showed a much greater reduction with increasing scatter and attenuation on the non-corrected images compared to those reconstructed with ACSC. Given that this was an anthropomorphic phantom with realistic dimensions and activity concentrations, the results can be assumed to hold in individual patients.

However, when examining our cohort of 29 older adults with normal cognition, the benefit of ACSC SPECT is not as apparent. Although on average, the relationship between BMI and cardiac uptake was less obvious with SPECT imaging than planar, there was no evidence that this was statistically significant. It is clear that there is a significant relationship between planar HMR and body mass index, so on average we would expect to see lower HMRs in larger people. However given the substantial variation between subjects (an R^2 value of 0.44 suggests that less than half the variation is related to BMI), it would not be practical to apply a linear correction to the HMR results based on BMI. This variation seems to be even greater with SPECT volumetric HMRs and there is furthermore no statistically significant negative slope between SPECT HMR and BMI, which we would expect for data without attenuation and scatter correction if this is the cause of the drop seen in planar imaging. The summed coronal slices do show a drop on the non-corrected images, suggesting that the summation of the SPECT slices reduces noise in the data. We recognise that BMI is more complex than the increases in size applied to our phantom study and there is likely to be substantial variation in the gamma ray interactions of two people of the same BMI. The inactive material used for the phantom chest plate and breasts is also not representative of active background tissue in patients. However, we cannot exclude the possibility that larger people may on average be more likely to have true reductions in cardiac uptake, for example due to underlying coronary artery disease, for which obesity is a risk factor.

Normal random variation in uptake between individuals may not be the only explanation for our findings. The ACSC summed coronal SPECT images do not appear noisier than the planar images but do suggest a drop in HMR with increasing BMI, albeit not statistically significant. A possible explanation is errors in the CT ACSC that are not seen in the phantom scans. Motion between SPECT and CT is a possibility, either due to volunteers moving slightly or internal mis-registration due to free breathing (e.g. diaphragm) or lack of cardiac gating. The images were examined for mis-registration during reconstruction, but this will not detect subtle shifts.

8.1.5 Conclusion

In an anthropomorphic torso phantom, increasing the size reduces cardiac uptake by 39% on planar images and 33% on non-corrected SPECT images but with attenuation and scatter correction applied the reduction is only 3%. Planar cardiac ^{123}I -MIBG images show a significant inverse relationship between cardiac uptake ratios and subject BMI. SPECT images with attenuation and scatter correction applied show no statistically significant drop in cardiac uptake with BMI, but there is considerable inter-subject variation and no statistically significant difference between the slopes of HMR against BMI for non-corrected and ACSC SPECT. It may be possible to reduce variation in future studies by using a SPECT-CT protocol that is more acceptable to older patients. This work suggests that ACSC SPECT does have the potential to improve cardiac ^{123}I -MIBG quantification, particularly in borderline planar cases but much larger studies would be needed to demonstrate this conclusively.

8.2 Does cardiac ^{123}I -MIBG SPECT-CT improve the diagnosis of Lewy body disease?

8.2.1 Introduction

Previously we saw that in a phantom, the planar HMR decreased with the addition of attenuating material. Corrections for attenuation and scatter, applied during SPECT reconstruction, reduced these effects. Cardiac MIBG images in controls showed an inverse relationship between planar HMR and BMI, indicating that patient size affects HMR.

In this section we explore whether SPECT with and without CT attenuation and scatter correction improves the diagnostic accuracy of cardiac MIBG imaging compared with the standard planar technique. Planar and SPECT control data is used to calculate normal cardiac uptake ranges for all the techniques considered. In this section, the term SPECT refers to reconstructions without attenuation and scatter correction, i.e. a technique available on any scanner. SPECT-CT refers to reconstructions with CT based attenuation and scatter correction. This reconstruction is less widely used in clinical practice as it requires use of a SPECT-CT scanner and a small additional CT radiation exposure.

SPECT data is available for both the MIDAS and SUPeRB subjects, but CT data for attenuation and scatter correction was only acquired for SUPeRB. In dementia and MCI, we therefore consider whether SPECT alone, without CT based image corrections, is of benefit over planar imaging. In MCI we also consider the benefit of SPECT-CT for relative quantification over both SPECT and planar imaging.

8.2.2 Methods

Subjects

The MIDAS and SUPeRB subjects included were the same as those in the previous chapters: 29 healthy older adults, 28 probable MCI-LB, 18 MCI-AD, 15 DLB (as 2/17 DLB did not have SPECT imaging) and 15 AD. In addition, the entire SUPeRB MCI dataset of 56 scans was used to compare the distribution of HMR values over the full range. The 29 control scans were used to produce normal cut-offs for SPECT imaging, analogous to the methods used with planar images in section 7.2. The MCI and dementia cases were used to calculate diagnostic accuracy for SPECT and the MCI alone for SPECT-CT. One probable MCI-LB case was scanned on the wrong collimator but was clearly normal on planar and SPECT. This case was included as normal for the accuracy calculations but excluded from quantitative analysis.

Image acquisition, reconstruction and processing

SPECT-CT images for the SUPeRB MCI dataset were acquired and processed as described for the controls in the previous section. Images were processed to obtain HMRs as described in the previous section, with the addition of a thresholded cardiac volume of interest for SPECT-CT (Figure 81), introduced to remove the left ventricular cavity and surrounding background from the VOI. The

hypothesis is that a VOI containing fewer non-specific counts may provide better separation between normal and abnormal images, and therefore have the potential to improve diagnostic accuracy. The thresholded VOI was generated by masking out all anatomy other than the left ventricle and including any voxels containing counts within 50% of the maximum voxel. The mean number of counts per voxel was then taken for comparison with the mediastinum reference VOI, as for the spherical cardiac VOI.

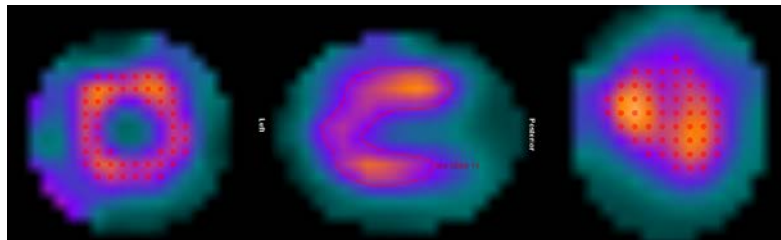


Figure 81: Slice data masked and reformatted into cardiac plane for thresholding, with 50% cardiac threshold VOI shown.

SPECT images for the MIDAS dataset were acquired and processed similarly, but without the additional attenuation and scatter corrected image reconstructions. Only the volumetric SPECT HMR method was therefore assessed for patients with dementia. The resultant set of processed images is summarised in Table 59 with the names they will be referred to throughout the rest of the section. As SPECT images were acquired after the delayed planar images only, not after the early planar images, the early planar HMR will not be used for comparison in this section. If insufficient cardiac uptake was visible for meaningful spherical VOI placement on SPECT images, the CT was used as a guide for SUPERB cases and a “best guess” based on lung and liver anatomy was made for the MIDAS cases. If there was insufficient uptake to apply the 50% threshold method, the images were classed as abnormal for the purposes of the sensitivity and specificity calculation.

Table 59: HMR methods used for each dataset in addition to the standard delayed planar HMR.

Name	Description	Controls	MCI	Dementia
Planar HMR	Delayed planar HMR	X	X	X
CSPECT HMR	Summed coronal SPECT HMR	X	X	
CSPECT-CT HMR	Summed coronal SPECT-CT HMR	X	X	
SPECT HMR	Volumetric SPECT HMR with spherical VOI	X	X	X
SPECT-CT HMR	Volumetric SPECT-CT HMR with spherical VOI	X	X	
SPECT-CT HMR _{50%}	Volumetric SPECT-CT HMR with 50% threshold cardiac VOI	X	X	

Analysis

The distribution of HMR values for the entire MCI dataset for each of the techniques was plotted using histograms, to reveal any differences in separating clusters of scans with different levels of cardiac uptake. The difference in mean HMR between MCI-LB and MCI-AD groups, and between DLB and AD groups was calculated for each method, and the standard error of the mean difference used to calculate confidence intervals for the difference. Scatter plots of SPECT HMR against planar HMR were produced to illustrate variation. Dot plots of HMR for each clinical group were produced for each HMR method.

The normal cut-off values for each method were derived from the 29 controls and taken as two standard deviations below the mean. This method of determining a normal range is used throughout the thesis and is also used in ¹²³I-FP-CIT commercial software. The coefficient of variation for each method was calculated from the control HMRs. The MCI results for each processing method were dichotomised as normal or abnormal according to the cut-offs and the resulting values used to calculate sensitivity, specificity and overall accuracy for each method. Contingency tables for each SPECT method were generated relative to delayed planar imaging. Percentage concordance between methods was calculated. For SPECT HMR, where MCI and dementia data was available, the datasets were assessed separately and as combined Lewy body and Alzheimer's disease groups.

8.2.3 Results

Normal cut-offs from control data

The mean HMR, standard deviation, coefficient of variation and normal-cut off calculated from the controls for each of the methods assessed is shown in Table 60.

Table 60: Control mean HMR, SD, coefficient of variation (CoV) and mean-2SD cut-off for the different processing methods used

	Mean	Standard deviation	CoV	Normal cut-off
Planar HMR	2.78	0.46	16.5%	1.86
CSPECT HMR	3.27	0.42	12.8%	2.43
CSPECT-CT HMR	3.67	0.51	13.9%	2.65
SPECT HMR	8.95	2.49	27.8%	3.96
SPECT-CT HMR	9.58	2.35	24.5%	4.87
SPECT-CT HMR _{50%}	10.69	2.46	23.0%	5.77

HMR distribution for each method

Figure 82 shows histograms of HMR distribution for the planar and SPECT HMR methods, using the 46 MCI cases. Categories below the normal cut-offs in .

Table 60 are displayed in red, above in green and borderline categories in orange. Delayed planar imaging shows reasonable separation between normal and abnormal uptake. The CSPECT method, where the SPECT slices are summed to give images similar to planar images, shows no clear separation. The distributions with CSPECT-CT and SPECT appear similar to planar. SPECT-CT HMR with both the spherical cardiac VOI and the threshold VOI show more distinct normal and abnormal uptake groups. The threshold VOI histogram is skewed towards zero because several scans had insufficient uptake for threshold analysis.

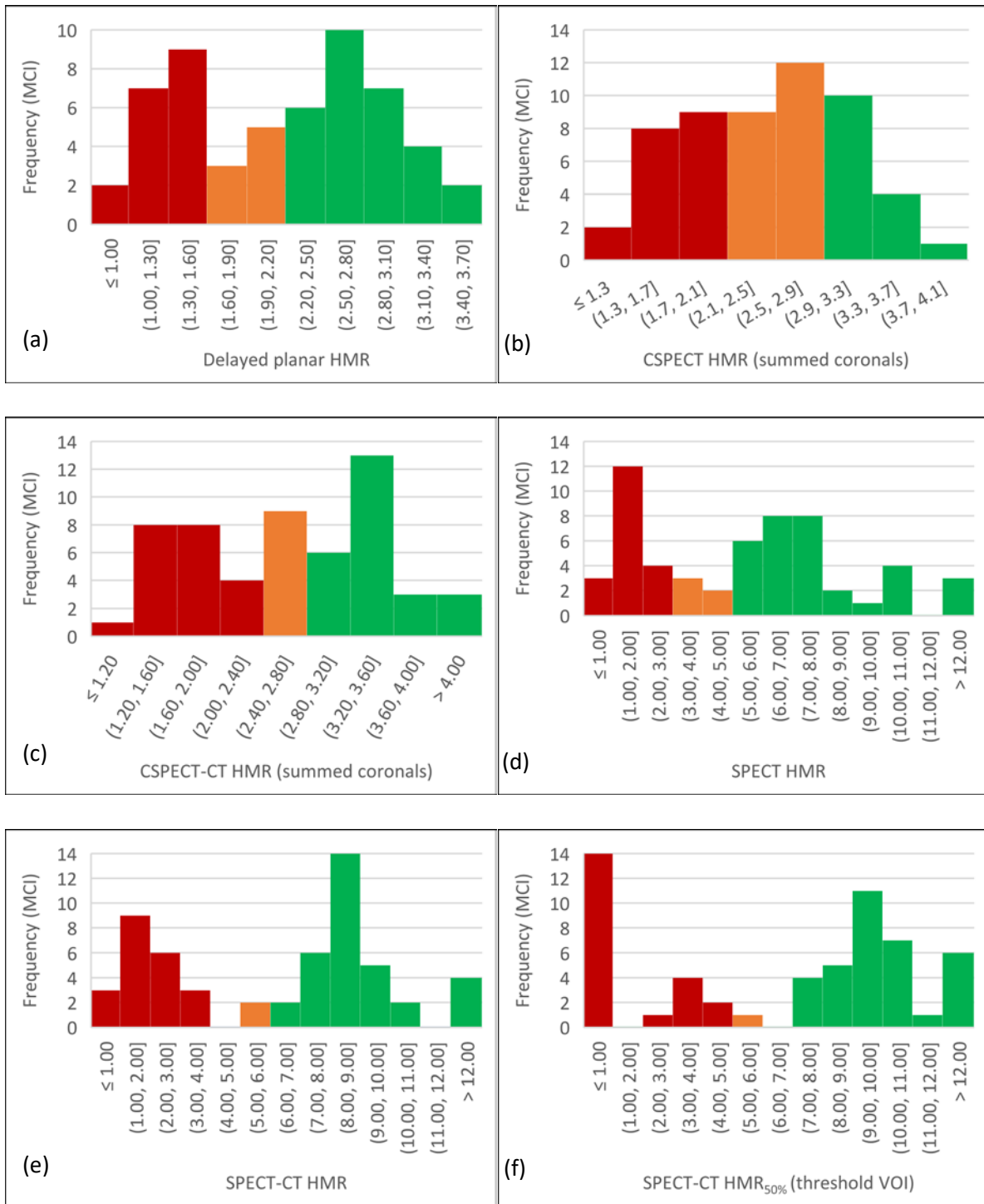


Figure 82: HMR histograms for MCI subjects in SUPeRB, showing the degree to which each imaging technique separates normal and low uptake. A: planar, B: summed coronal SPECT, C: summed coronal SPECT-CT, D: volumetric SPECT, E: volumetric SPECT-CT, F: SPECT-CT with cardiac threshold. Categories below the normal cut-offs in Table 60 are displayed in red, above in green and borderline categories in orange.

Difference between AD and LB groups for each method

Table 61 shows that there is a significant difference between the AD and DLB mean HMR for both planar and SPECT imaging, but not between the MCI-AD and MCI-LB groups. This is expected given the high proportion of MCI-LB scans designated as normal for all imaging methods (Figure 84). The difference between AD and DLB for MIDAS is much greater for SPECT than for planar, likely due to the increased SPECT contrast for high uptake scans. However, the dot plots for the MIDAS dataset in Figure 83 show that for both planar and SPECT HMR there is one AD case below the normal cut-off and 5 DLB cases above. These are the same patients, giving 100% concordance between planar and SPECT HMR for the MIDAS dataset.

Table 61: Differences between mean HMR for probable Lewy body disease and Alzheimer’s disease groups. In the MIDAS dataset mean HMR is significantly lower for DLB compared to AD for both planar and SPECT imaging. In the SUPeRB MCI dataset mean HMR appears lower in the Lewy body group but this is not statistically significant. The largest difference is seen for SPECT-CT HMR. SPECT-CT_{50%} is not included because many low uptake scans could not be analysed with this method.

		Lewy body disease		Alzheimer’s disease		Difference		
		Mean	Standard deviation	Mean	Standard deviation	Mean difference (AD-LB)	95% CI for mean difference	p-value (unpaired t-test)
Dementia (MIDAS)	Planar HMR	1.63	0.77	2.51	0.36	0.88	0.46 to 1.29	<0.001
	SPECT HMR	2.79	2.46	6.39	1.69	3.60	2.05 to 5.14	<0.001
Mild cognitive impairment (SUPeRB)	Planar HMR	1.98	0.84	2.26	0.56	0.21	-0.12 to 0.68	0.18
	CSPECT HMR	2.25	0.76	2.44	0.59	0.20	-0.21 to 0.58	0.36
	CSPECT-CT HMR	2.52	1.14	2.75	0.72	0.27	-0.30 to 0.77	0.40
	SPECT HMR	4.62	3.48	5.34	2.11	0.82	-0.89 to 2.34	0.38
	SPECT-CT HMR	5.28	3.81	6.60	2.43	0.92	-0.48 to 3.13	0.16

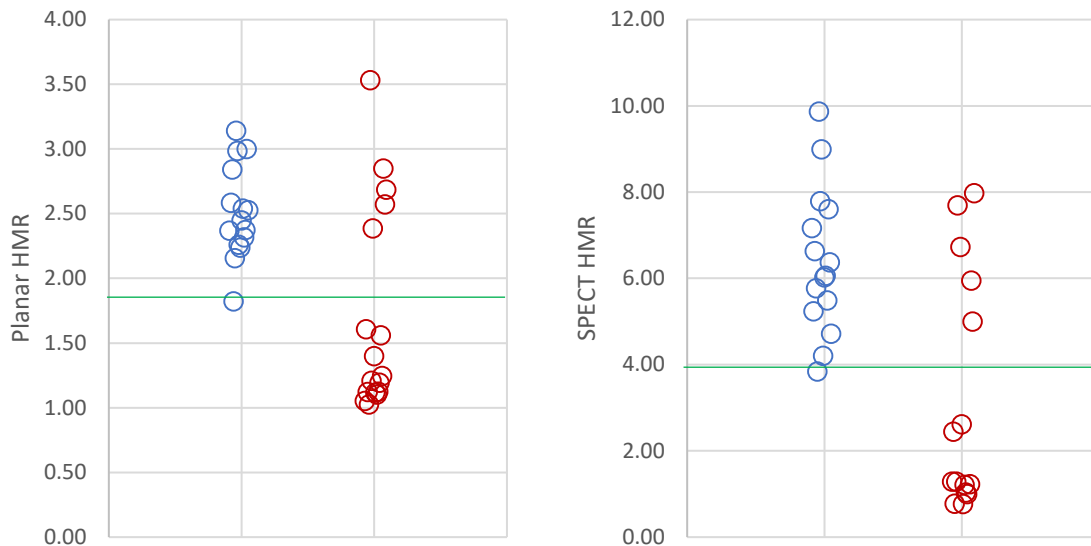


Figure 83: Dot plots for MIDAS for planar HMR (left) and SPECT HMR (right) showing no change in interpretation with SPECT. The blue circles are AD patients and the red circles DLB. The green line shows the normal cut-off – the borderline AD case is the same patient and is just below the normal cut-off for both planar and SPECT

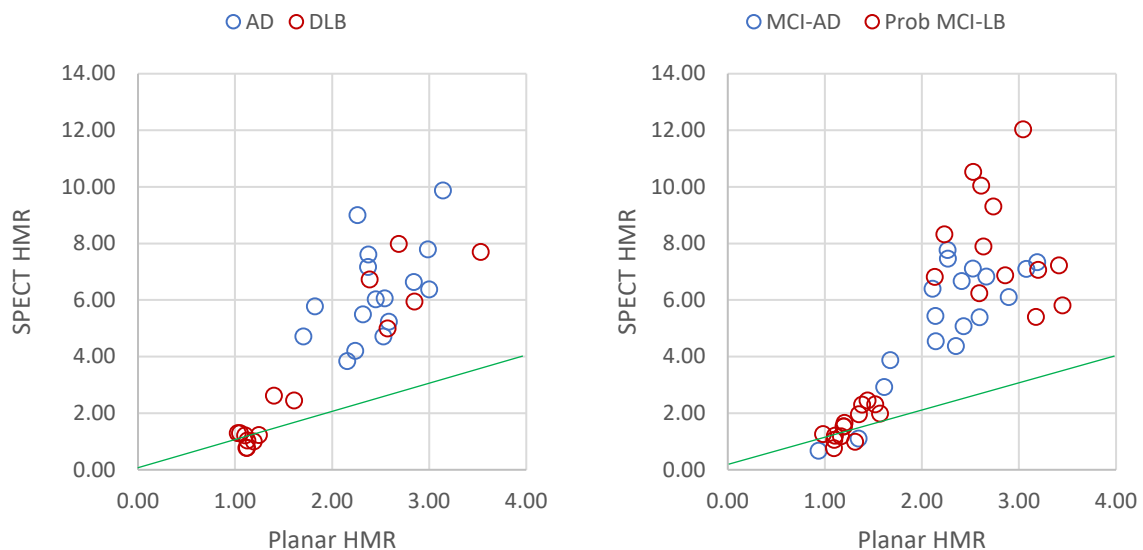


Figure 84: SPECT HMR plotted against delayed planar HMR for the MIDAS dementia dataset (left) and SUPeRB MCI (right). The green lines show where the data points would fall if planar and SPECT HMR were equivalent. In cases with cardiac uptake the SPECT HMR is much higher than the planar due to improved contrast, but when uptake is low there is little contrast with background for either method.

Comparison between HMR methods for MCI patients

Figure 85 shows that the CSPECT method provides little distinction between normal and abnormal scans for either the MCI-AD or MCI-LB dataset, which is in keeping with the histogram distributions in Figure 82. Visually there appears to be a better separation between normal and abnormal scans for SPECT-CT compared to SPECT and to planar. The SPECT-CT_{50%} threshold method appears similar to SPECT-CT, taking the scans with no visible uptake into account.

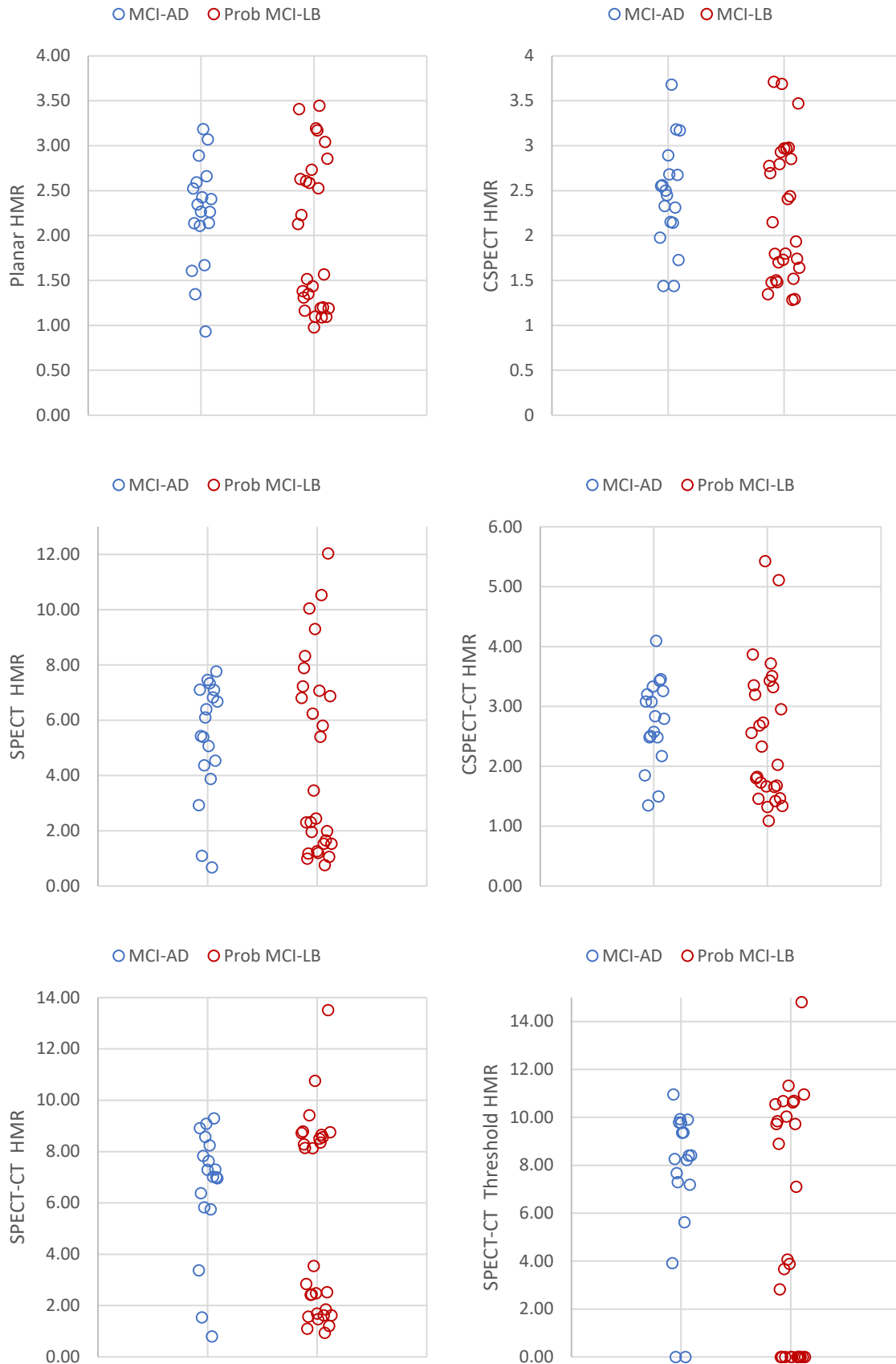


Figure 85: Dot plots for HMR for MCI-AD and MCI-LB groups. The MCI-LB group contains many normal uptake scans, and the MCI-AD some low uptake scans. However, the SPECT-CT HMR plot shows a clear division of the data into high and low uptake clusters, which is not as obvious for the other methods. The CSPECT methods show no distinction between groups.

Concordance between planar and SPECT techniques

Concordance between planar HMR and SPECT HMR was 100% for both the MIDAS and SUPeR datasets (Table 62). For the MCI dataset, concordance between planar HMR and SPECT-CT_{50%} HMR was also 100%. Concordance between planar HMR and summed coronal SPECT was slightly poorer (89%) both with and without attenuation and scatter correction, due to more cases being designated as abnormal with CSPECT and CSPECT-CT (Table 63). Concordance between planar and SPECT-CT HMR was 100% for the Lewy body group but 98% overall as one of the MCI-AD cases was below the planar cut-off but above the SPECT-CT cut-off.

Table 62: Contingency table showing perfect concordance between planar and SPECT HMR imaging for the MCI and dementia groups.

			MCI		Dementia	
			SPECT			
			> 3.96	< 3.96	> 3.96	< 3.96
LB	Planar	normal > 1.86	11	0	5	0
		abnormal < 1.86	0	15	0	10
AD	Planar	normal > 1.86	14	0	14	0
		abnormal < 1.86	0	4	0	1

Table 63: Contingency table showing concordance between planar and SPECT imaging in MCI for each SPECT method tested. Red figures indicate non-concordance. The CSPECT methods both designated four additional MCI-AD cases and one additional MCI-LB as abnormal. The only other disagreement was for a single MCI-AD case that was designated abnormal with planar HMR but was within the SPECT-CT normal range.

		CSPECT		CSPECT-CT		SPECT HMR		SPECT-CT HMR		SPECT-CT HMR _{50%}		
		>2.43	< 2.43	> 2.65	< 2.65	> 3.96	< 3.96	> 4.87	< 4.87	> 5.77	< 5.77	
MCI-LB	Planar	> 1.86	10	1	10	1	11	0	11	0	11	0
		< 1.86	0	15	0	15	0	15	0	15	0	15
MCI-AD	Planar	> 1.86	10	4	10	4	14	0	14	0	14	0
		< 1.86	0	4	0	4	0	4	1	3	0	4

Diagnostic accuracy of SPECT techniques compared with planar

The sensitivity, specificity and overall accuracy for distinguishing between MCI-LB and MCI-AD for each of the techniques assessed is given in Table 64. The specificities are lower for the summed coronal SPECT methods and higher for SPECT-CT due to the cases mentioned above. Overall there is little difference in diagnostic accuracy between any of the methods. Sensitivity is low for all methods due to the high proportion of normal scans within the MCI-LB group (as seen in the plots in Figure 85).

Table 64: Diagnostic accuracy with respect to clinical diagnosis for each of the image analysis methods assessed.

	Planar	CSPECT-CT	CSPECT	SPECT HMR	SPECT-CT HMR	SPECT-CT HMR _{50%}
Sensitivity	54% (15/28)	57% (16/28)	57% (16/28)	54% (15/28)	54% (15/28)	54% (15/28)
Specificity	78% (14/18)	56% (10/18)	56% (10/18)	78% (14/18)	83% (15/18)	78% (14/18)
Accuracy	63% (29/46)	57% (26/46)	57% (26/46)	63% (29/46)	65% (30/46)	63% (29/46)

8.2.4 Discussion

SPECT-CT HMR has theoretical advantages in improving the diagnostic accuracy of cardiac MIBG imaging as it potentially separates normal and abnormal uptake scans better than planar HMR due to improved contrast. Our results show a tantalizingly good separation of the MCI-LB cases into two groups; following up the existing cohort will therefore be very helpful to establish whether the potential improvement is borne out in practice. However, in our MCI dataset over half of the MCI-LB cases have scans that are normal, without any doubt. There are also MCI-AD cases that are clearly abnormal. This is reflected in the non-significant differences between mean HMR for the MCI dataset for all imaging methods. We therefore didn't see the improvement in accuracy with SPECT-CT in this particular dataset that we may have done if there had been more borderline images.

Summed coronal CSPECT-CT appears to be an unsuitable technique for calculating HMR, both considering the histograms that show no clear difference between normal and low uptake and the specificity results. This technique was included because it should be a simple and reproducible way of analysing attenuation and scatter corrected SPECT images. By summing the coronal slices, a "pseudo-planar" image is generated, which can be analysed using the reproducible planar HMR methods introduced in section 7.1. SPECT and SPECT-CT HMR is expected to be less repeatable as the operator needs to decide on the placement of the volumes in three dimensions. We found that the placement of the mediastinum region in particular was quite arbitrary and hypothesised that the simple geometry of the CSPECT-CT images could provide the advantages of attenuation and scatter correction, without the increase in variation.

The coefficient of variation is indeed lower for the CSPECT images than for SPECT, and is actually lower than that of the planar HMR. This could be due to the lower matrix size of the SPECT images, which makes the summed coronal images look smoother than the true planar images or possibly related to the use of resolution modelling to reduced noise. This can be seen in the example planar and summed coronal images displayed side by side in Figure 74 in section 8.1. The mean CSPECT HMR should be the same as the mean planar HMR as this is the same geometry and neither are corrected for scatter and attenuation; but in fact the CSPECT HMR is consistently slightly higher ($p < 0.001$).

The greater separation between high and low uptake scans seen on the histogram data for SPECT-CT compared to CSPECT-CT (Figure 82) suggests that the increase in image contrast with SPECT is a bigger factor than attenuation and scatter correction. However, attenuation and scatter correction does appear to increase separation for SPECT-CT HMR and led to one of the MCI-AD cases being designated as normal, whereas on planar imaging the patient had a low HMR. This HMR values for this patient were 1.70 on delayed planar imaging (normal cut-off 1.86) and 5.75 on SPECT-CT (normal cut-off 4.87). The value on SPECT without corrections was 3.88, just below the cut-off of 3.96. The planar and SPECT-

CT images in Figure 86 show a clear difference in appearance of the heart, with very little uptake visible on planar and uptake clearly visible on SPECT-CT. The BMI for this patient was 34.4, suggesting that the low HMR on the planar image may have been due to artefactual count reduction because of excess attenuation and scatter.

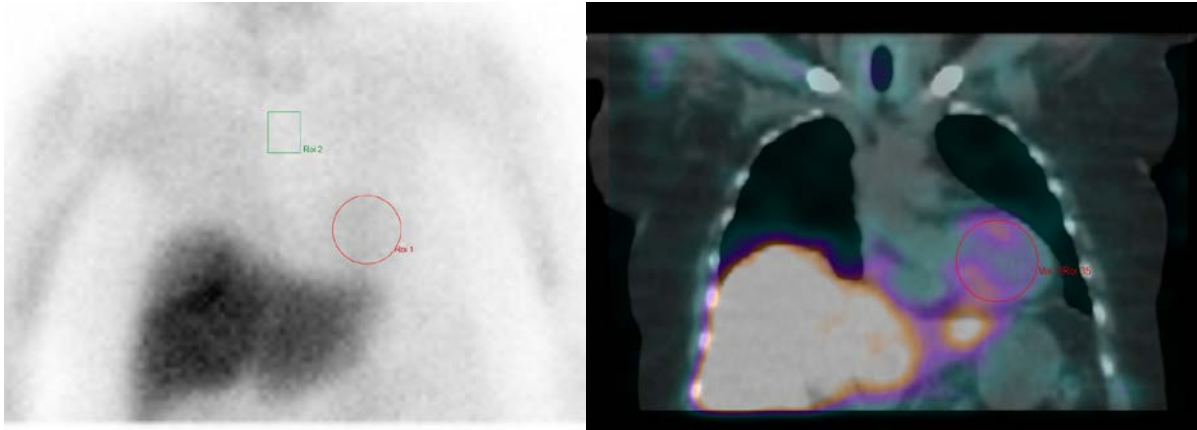


Figure 86: Planar image and SPECT-CT central slice image for the MCI-AD patient with low uptake on the planar scan and uptake within normal limits on SPECT-CT.

We expected to see more cases like the one above, given the BMI distribution in our dataset, with over half the patients classed as either overweight or obese (Figure 87) and given that several AD patients had unexpectedly low planar uptake. However, when we look at the impact of CT attenuation and scatter correction on HMR for the whole dataset, there does not appear to be any link with BMI (Figure 88). On average, the HMR values are higher for SPECT-CT than SPECT, including for slim people, but there is a lot of variation. This indicates that attenuation and scatter is not strongly dependent on BMI, which may help to explain why we did not see a significant difference in the slopes between SPECT HMR and BMI and SPECT-CT HMR and BMI for controls in the previous section. This is possibly because BMI is not a good surrogate for thickness of tissue overlying the heart.

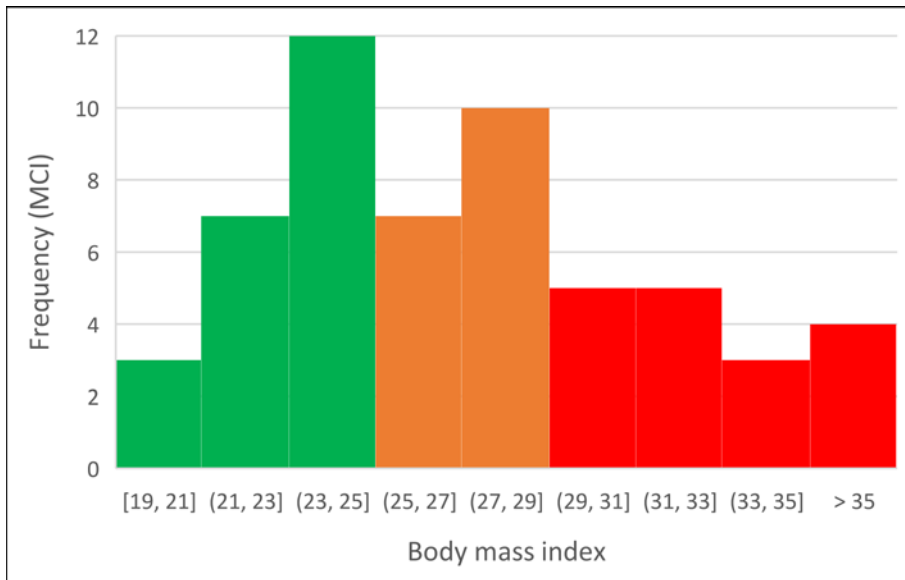


Figure 87: BMI distribution for the MCI dataset

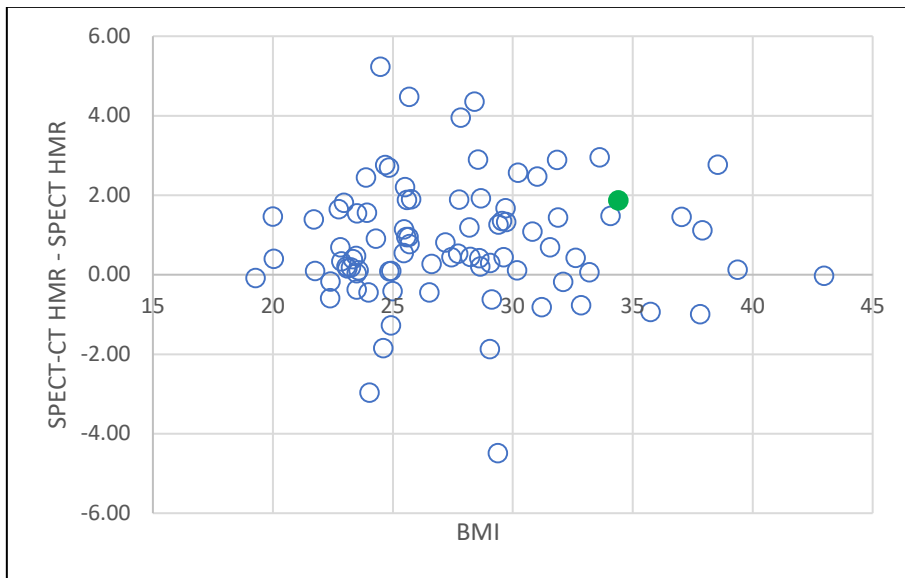


Figure 88: Bland-Altman plot showing difference in HMR for SPECT-CT and SPECT, plotted against BMI. The data point for the MCI-AD case discussed above is shown in green.

8.2.5 Conclusion

SPECT-CT improves the separation between normal and abnormal cardiac MIBG scans quite considerably, and may increase clinical confidence in scan interpretation. However, the increased separation translates into only a small improvement in MCI diagnostic accuracy because many MCI-LB scans are clear-cut normal cardiac MIBG cases, and some MCI-AD are clear-cut abnormal. In most cases planar imaging is more than sufficient to characterise the scans and SPECT-CT HMR doesn't add useful information. However, in clinical practice, SPECT-CT may be useful in borderline cases where the planar image result is unclear.

8.3 Absolute quantification of ^{123}I -MIBG cardiac uptake using SPECT-CT imaging

8.3.1 Introduction

Previous sections have used various relative quantification methods to assess cardiac sympathetic innervation. However, in theory a more direct absolute measure of specific uptake offers advantages and is the ultimate goal of nuclear medicine studies. Traditionally this has not been technically possible due to the very large uncertainty in scatter and attenuation effects. Now with SPECT-CT and advances in reconstruction techniques there is the potential to provide a far more accurate measure of true activity within a given organ and hence measure absolute uptake in kBq/ml, as was explored for striatal uptake of FP-CIT in section 6.2. In this section we report an investigation into the feasibility and accuracy of absolute quantification of cardiac MIBG in the diagnosis of Lewy body disease and compare this with previous relative methods. Absolute quantification for cardiac MIBG imaging has not previously been investigated in the literature.

In the previous section we showed that SPECT-CT relative quantification, via the volumetric heart-to-mediastinum ratio on attenuation and scatter corrected SPECT images, could separate normal and abnormal cardiac MIBG scans better than planar imaging or SPECT without ACSC. However a concern remained that these images could be subject to poor processing repeatability, due to the uncertainty in the placement of the mediastinum reference region. The coefficient of variation of the SPECT-CT HMR between subjects was substantial at 25%, compared to 17% for delayed planar HMR. By considering the absolute quantification of the activity concentration in the heart and mediastinum volumes of interest we will be able to show whether this was primarily due to variation in the cardiac VOI or mediastinum VOI.

If the variation in mediastinum region is substantial, it may be that absolute quantification of cardiac MIBG uptake could offer advantages over SPECT-CT HMR by providing a result less affected by noise and more in keeping with the actual sympathetic innervation of the heart. This has not been investigated before with CTACSC in the literature. However, as discussed for FP-CIT in section 6.2, absolute quantification has practical challenges that introduce potential sources of error and uncertainty and which may outweigh any advantages that the removal of dependence on a noisy reference region brings. These include the need to take height and weight measurements, accurately record the injected activity (including the residual activity in the syringe post-injection) and the time at which the activity was measured. There is also a need for accurate gamma camera calibration (to allow detected signal to be converted to radioactivity concentration), with regular QC checks on this. All of the above involves extra staff time so it would not be justified to implement this technique in routine clinical practice without having demonstrated a clinically important benefit. The scatter and attenuation correction method for torso studies requires CT imaging immediately after the SPECT,

increasing the radiation dose to the patient. This CT will always have at least some degree of mis-registration relative to the SPECT due to breathing, the point in the cardiac cycle at which the CT was acquired and voluntary patient movement, all of which will affect the accuracy of the reconstructed images. However this does also apply to the SPECT-CT HMR results, so is not a reason to prefer relative quantification (SPECT-CT HMR) over absolute cardiac uptake measurements. Finally, the use of a scaling factor to take patient size and therefore tracer volume of distribution (VoD) into account is a further source of variation as it is known that VoD does not scale linearly with patient weight, which is often used as a surrogate in PET imaging [240]. For example, an overweight patient will have a higher weight than a slim patient of the same height, but the tracer VoD will not be increased by the same proportion. However, some sort of adjustment for patient size is crucial, as smaller patients will have a lower tracer volume of distribution on average, with higher activity concentration in the heart (and in all other organs). An alternative surrogate is lean body mass, which scales better with tracer VoD [241, 242]. However, lean body mass cannot be measured directly, adding an additional source of uncertainty. LBM may be estimated from BMI and weight using the following sex-specific formulae[243], which have been validated for PET imaging [244]:

$$\text{LBM (males)} = 9,270 \times \text{weight} / (6,680 + 216 \times \text{BMI})$$

$$\text{LBM (females)} = 9,270 \times \text{weight} / (8,780 + 244 \times \text{BMI})$$

The objectives of this section are as follows:

- To assess whether patient weight or estimated lean body mass should be used for normalisation
- To calculate and compare the coefficients of variation for absolute cardiac uptake and absolute mediastinum (background) uptake
- To calculate normal cut-offs for absolute cardiac uptake and apply these to the cardiac uptake values in the MCI dataset
- To compare the diagnostic accuracy of absolute cardiac uptake with that of SPECT-CT HMR

8.3.2 Methods

Subjects

SPECT data is available for both the MIDAS and SUPeRB subjects, but CT data for attenuation and scatter correction was only acquired for SUPeRB. Attenuation and scatter correction are a pre-requisite for accurate activity quantification, so we are only able to assess the potential benefits of absolute quantification in the SUPeRB MCI cohort. The SPECT-CT scans of the controls with normal planar imaging are used to assess the variation in absolute cardiac uptake, and to establish a normal cut-off.

One control was excluded due to the injected activity not being recorded, so 28 control scans were used in this section. All the SUPeR MCI and control cases are included to assess the variation in mediastinum uptake, as this will be independent of cardiac innervation. The SUPeR probable MCI-LB and MCI-AD cases are used to assess diagnostic accuracy, as in section 7.2. As previously noted, one probable MCI-LB case was scanned on the wrong collimator but was clearly normal on planar and SPECT. This case was excluded from this section.

Absolute uptake calculation

In order to obtain a meaningful uptake measurement, the detected signal needs to be corrected for the amount of activity injected and the size of the patient (as a surrogate for VoD). In positron emission tomography (PET) this is usually achieved via the standardised uptake value (SUV), as applied to FP-CIT in section 6.2. The SUV requires the measurement of the activity concentration in a volume of interest, obtained from the count density in the volume via a gamma camera calibration factor, CF.

$$\text{Activity concentration (kBq/ml)} = \text{Count density} \times CF$$

The activity concentration is normalised by the weight or LBM and by the activity administered to the patient to give the SUV:

$$SUV = \frac{\text{Activity concentration (kBq/ml)}}{\text{Administered activity (MBq)/Patient weight (kg)}}$$

Cardiac MIBG SUVs may be calculated by Hermes Medical Hybrid Recon software, as we did for FP-CIT, but using a specific calibration factor for MELP collimators and torso geometry. This calibration factor has been obtained for MELP collimators. However, because we had already reconstructed the cardiac MIBG data with standard SPECT-CT reconstructions without SUV information to use for relative quantification in the previous section, we manually calculated a quantity analogous to SUV, the normalised count density, from the existing reconstructed data.

SUV can also be expressed as

$$SUV = \frac{\text{Count density} \times CF}{\text{Administered activity (MBq)/Patient weight (kg)}}$$

Assuming the calibration factor is unchanged between patients, SUV information is contained within a measurement of count density, normalised to injected activity and patient size. We define the normalised count density (CD_{norm}) as the raw count density (mean number of counts per voxel) (CD_{raw}) scaled to that of a standard sized person injected with 111 MBq ^{123}I -MIBG, referenced to the

scan time. We have defined CD_{norm} , which is simply a scaled version of SUV, as it can be calculated without requiring additional SUV reconstructions.

A standard sized person weighs 70 kg and has a lean body mass (LBM) of 55 kg. Either weight or LBM can be used to adjust the count density for patient size. The size scaling factor (SSF) is the patient weight or LBM divided by the reference weight or LBM. The activity scaling factor (ASF) is the decay corrected administered activity divided by the reference activity of 111 MBq.

The raw count density is multiplied by these scaling factors to give the normalised count density:

$$CD_{norm} = CD_{raw} \times ASF \times SS$$

$$CD_{norm_wt} = CD_{raw} \times Patient\ activity/111\ MBq \times weight/70\ kg$$

$$CD_{norm_LBM} = CD_{raw} \times Patient\ activity/111\ MBq \times LBM/55\ kg$$

For example, one of our controls (BMI 29.7; weight 105 kg; LBM 75kg) was administered 118 MBq ^{123}I -MIBG at 12:00, with a residual activity of 6 MBq left in the syringe. He was scanned 3.6 hours later, at which point the injected activity had decayed to 93 MBq, as the half-life of ^{123}I is 13.2 hours. The mean cardiac count density on the scan (CD_{raw}), sampled using a 6cm diameter spherical VOI, was 292 counts per voxel. The activity scaling factor (ASF) is $111/93 = 1.19$. The size scaling factor (SSF) is $105/70 = 1.50$ for weight scaling and $75/55 = 1.36$ for LBM scaling. The normalised count density is therefore $CD_{norm_wt} = 292 \times 1.19 \times 1.50 = 521$ for weight scaling and $CD_{norm_LBM} = 292 \times 1.19 \times 1.35 = 469$ for LBM scaling.

This benefit of this normalised count density approach is that the VOIs used for relative quantification can be reused, avoiding introducing an additional source of uncertainty. The method is valid because all our patients and controls were scanned using the same acquisition protocol on the same scanner. The scanner calibration factor that would have been required by Hybrid Recon software for SUV calculation is therefore just a scaling factor that is the same for all scans. However, if the acquisition time had varied between patients, for example, or if some patients had been scanned with different collimators, the SUV software would have been the simplest method to account for these differences.

Raw count density values were obtained for MCI patients and controls for the heart and mediastinum using the same spherical and threshold VOIs as in the previous section. Both the body weight and LBM normalisation methods were used, resulting in four CD_{norm} values for each subject (Table 65).

Table 65: Methods used to calculate normalised count density

		Size normalisation method	
		Weight	LBM
VOI method	Sphere	CD_{norm_wt}	CD_{norm_LBM}
	50% threshold	$CD_{norm_wt_50\%}$	$CD_{norm_LBM_50\%}$

Assessment of normalisation method

The raw and normalised count densities for the mediastinum were plotted against the size scaling parameter used for normalisation (i.e. weight or LBM). Mediastinum data for all MCI and control subjects were used. We expected to see a drop in CD_{raw} with increasing weight and LBM because the ^{123}I -MIBG is diluted over a larger volume of distribution. If the normalisation method is effective we should see no drop in CD_{norm} with patient size. Similar plots were produced for the cardiac count densities, but this was restricted to only the 28 controls with normal cardiac uptake on planar imaging.

The slopes of raw and normalised count densities against weight or LBM were calculated with 95% confidence intervals. The standard errors were used to calculate whether the slopes were significantly closer to zero with the normalisation methods applied.

The mean control CD_{norm} values and the standard deviations were used to calculate coefficients of variation (CoV) for each normalisation method for both the heart and the mediastinum. The CoV in this case is a measure of a combination of the variation between subjects and the uncertainty in VOI placement. The absolute quantification CoV results were compared with the CoV for SPECT-CT HMR of 25%.

Diagnostic accuracy of absolute cardiac uptake

Normal cut-off values for CD_{norm} were calculated from the control data, and defined as two standard deviations below the mean.

MCI scans were dichotomised as normal or abnormal to calculate sensitivity and specificity using the threshold described above. Sensitivity, specificity and accuracy was assessed for all four absolute uptake parameters listed in Table 65.

8.3.3 Results

Assessment of normalisation method

The plots of mediastinal uptake in Figure 89 confirm that count density decreases with patient size. The slopes are both significantly different from zero (Table 66). Both weight and LBM normalisation

have a statistically significant effect on the slopes, resulting in values that are not significantly different from zero. Visually the normalisation to LBM appears more effective but the histograms are similar for both methods (Figure 90). There is no cause identified for the obvious outlier – the image files have been checked and the correct collimator was used and the VOI placement was sensible. This was an MCI-LB patient with low cardiac activity so the cardiac counts are unhelpful in determining the cause.

The plots of the cardiac count density for the controls in Figure 91 again suggest that the LBM normalisation method may be slightly more effective than weight. This is supported by the statistical analysis (Table 66), with the weight normalisation showing no statistically significant change in slope and the residual slope still negative. The LBM normalisation method shows both a significant reduction in slope with normalisation applied and a resulting slope that is not significantly different from zero.

Table 66: Slopes of CD_{raw} and CD_{norm_wt} against weight and CD_{raw} and CD_{norm_LBM} against LBM, for cardiac and mediastinal count density. The 95% confidence intervals around the slope estimates are given as well as p-values for difference in slopes with and without normalisation applied.

	Cardiac counts against weight	Cardiac counts against LBM	Mediastinum counts against weight	Mediastinum counts against LBM
Slope (CD_{raw})	-5.02	-7.19	-0.46	-0.59
95% CI	-6.57 to -3.46	-9.21 to -5.17	-0.60 to -0.32	-0.79 to -0.40
Slope (CD_{norm})	-2.74	-2.00	-0.09	0.05
95% CI	-4.98 to -0.50	-4.88 to 0.87	-0.28 to 0.10	-0.21 to 0.31
Difference in slope with normalisation	2.27	5.18	0.37	0.64
p-value for difference	0.109	0.006	0.002	<0.001

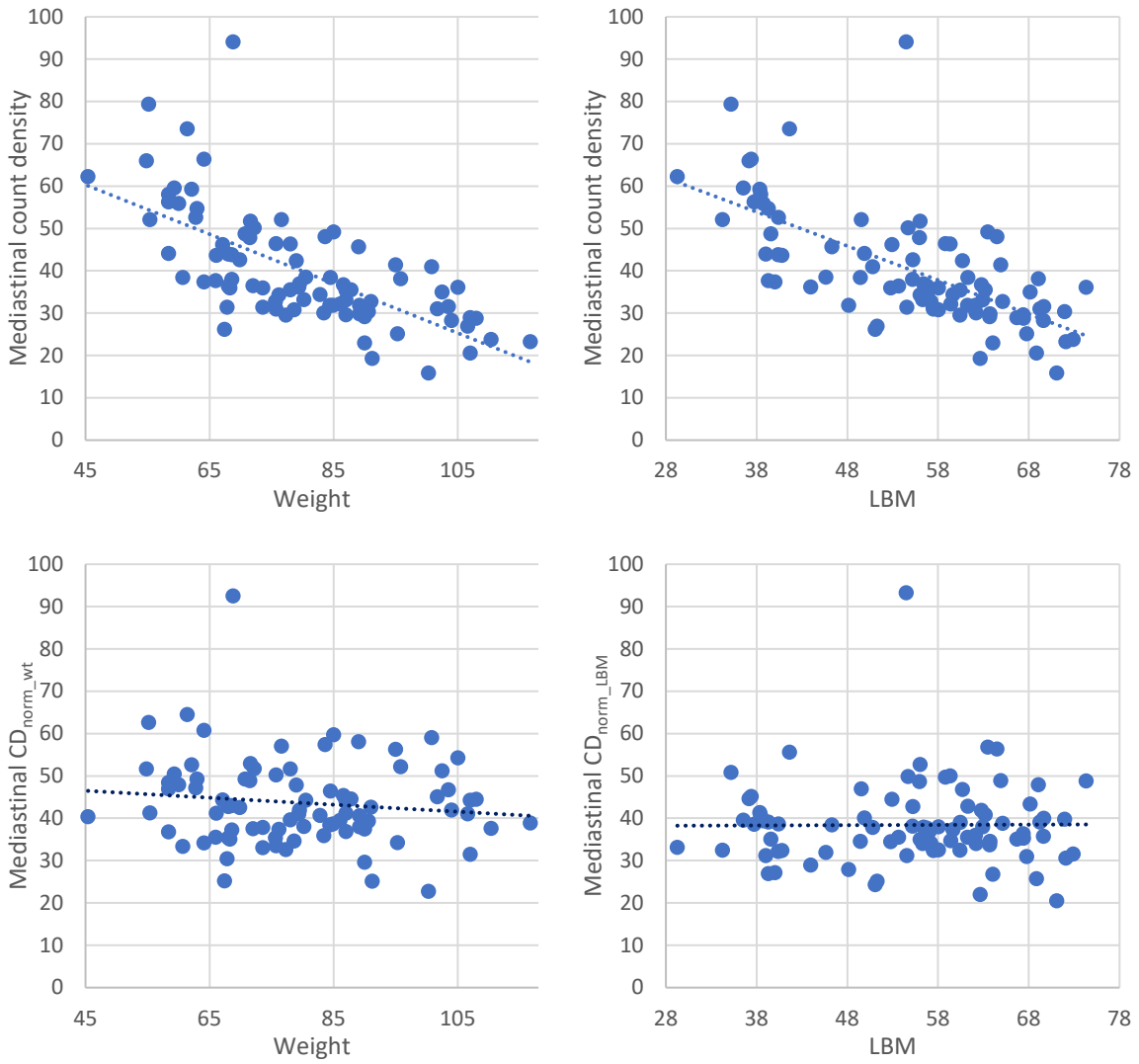


Figure 89: Top: Mediastinal count density plotted against weight (left) and against LBM (right). Bottom: Mediastinal CDnorm with weight normalisation (left) and LBM (right).

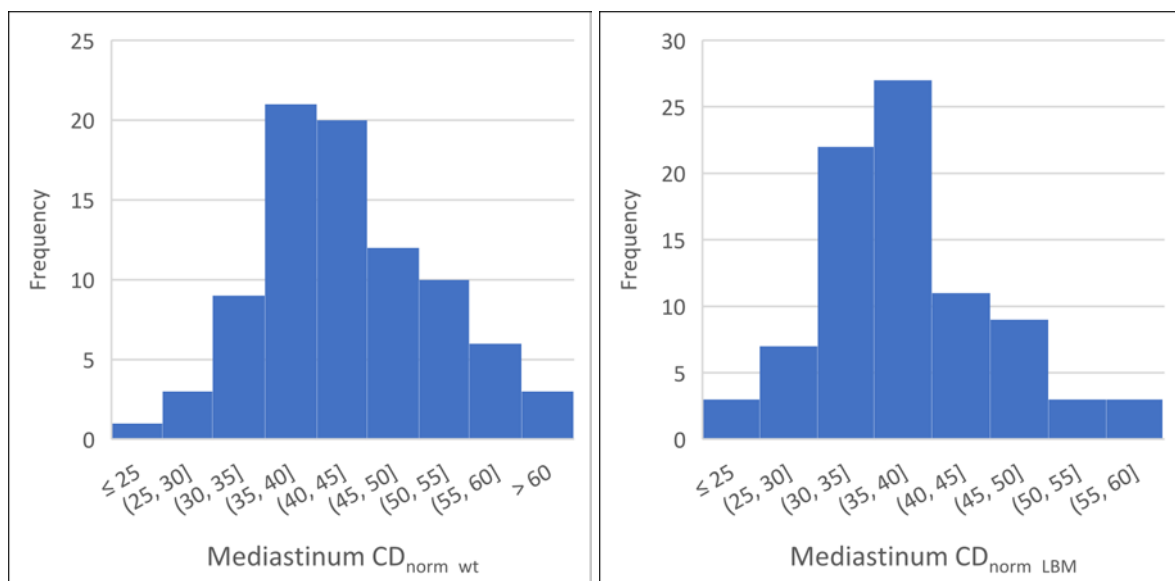


Figure 90: Mediastinum count density histogram with weight normalisation (left) and LBM (right).

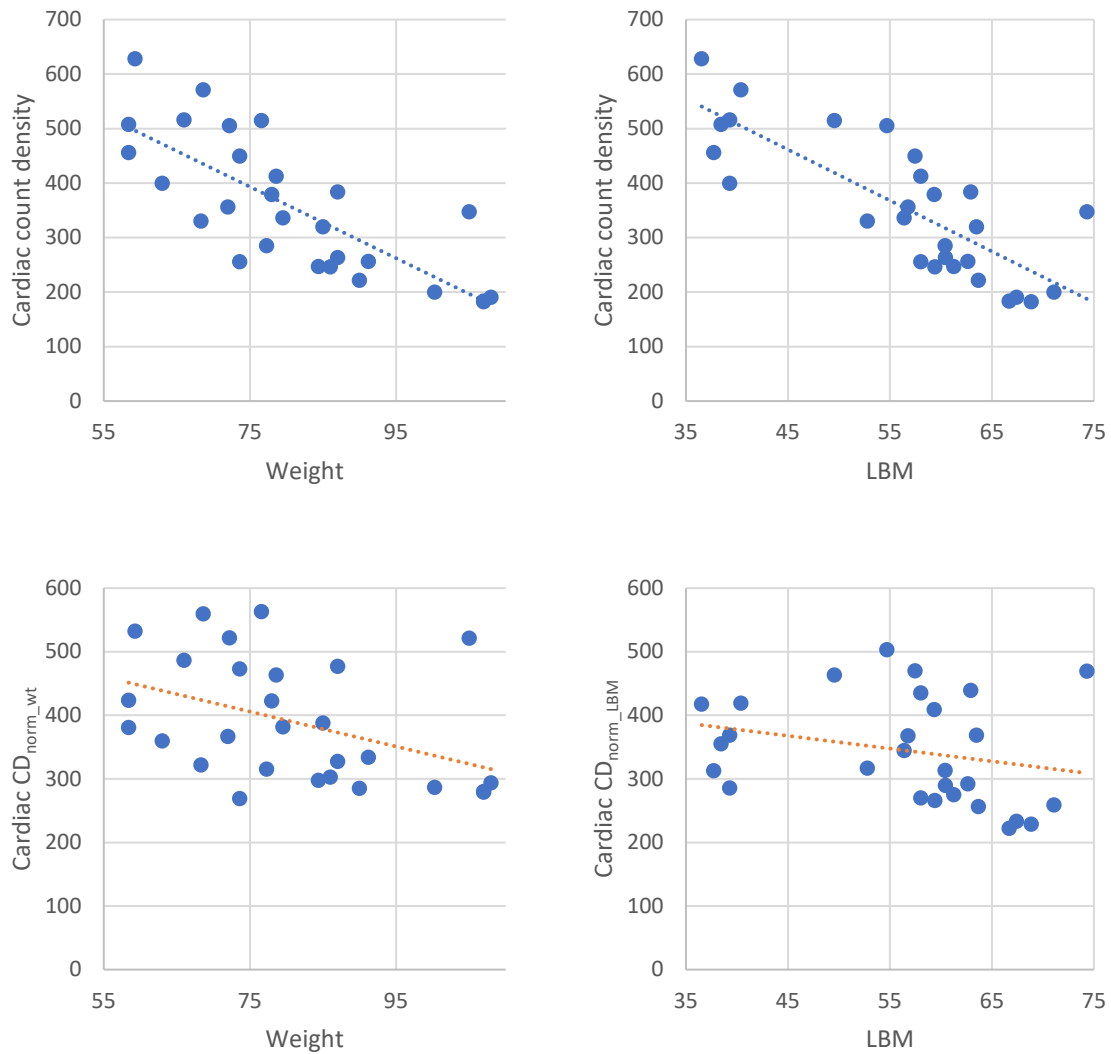


Figure 91: Top: Cardiac count density plotted against weight (left) and against LBM (right). Bottom: Cardiac CD_{norm} with weight normalisation (left) and LBM (right).

The coefficients of variation are shown in Table 67 and are similar for the mediastinum and heart VOIs, and in line with the SPECT-CT HMR CoV value of 25%, suggesting that uncertainty in mediastinum VOI placement is not contributing to the increased SPECT variation compared to planar, at least no more than the cardiac VOI.

Histograms of cardiac count density for the entire MCI group are shown in Figure 92, with SPECT-CT histograms from section 8.2 as a comparison. The separation between normal and abnormal scans seems similar for all cardiac CD_{norm} methods, and perhaps slightly less than for SPECT-CT HMR. However diagnostic accuracy is very similar to that of SPECT-CT HMR (Table 68).

Table 67: coefficients of variation in mediastinum uptake and cardiac uptake between controls subjects, showing similar variation for both VOIs and size normalisation methods.

Coefficient of variation			
	Mediastinum	Heart (sphere VOI)	Heart (50% threshold VOI)
No normalisation	30%	35%	32%
Activity only	31%	36%	33%
Activity and weight	22%	25%	21%
Activity and LBM	22%	24%	22%

Table 68: Diagnostic accuracy of the absolute quantification methods, with planar and SPECT-CT HMR values reproduced for comparison.

	Sensitivity	Specificity	Accuracy
Delayed planar	14/27 52%	14/18 78%	28/45 62%
SPECT-CT HMR	14/27 52%	15/18 83%	29/45 64%
CDnorm_wt	14/27 52%	15/18 83%	29/45 64%
CDnorm_wt_50%	14/27 52%	14/18 78%	28/45 62%
CDnorm_LBM	14/27 52%	14/18 78%	28/45 62%
CDnorm_LBM_50%	14/27 52%	14/18 78%	28/45 62%

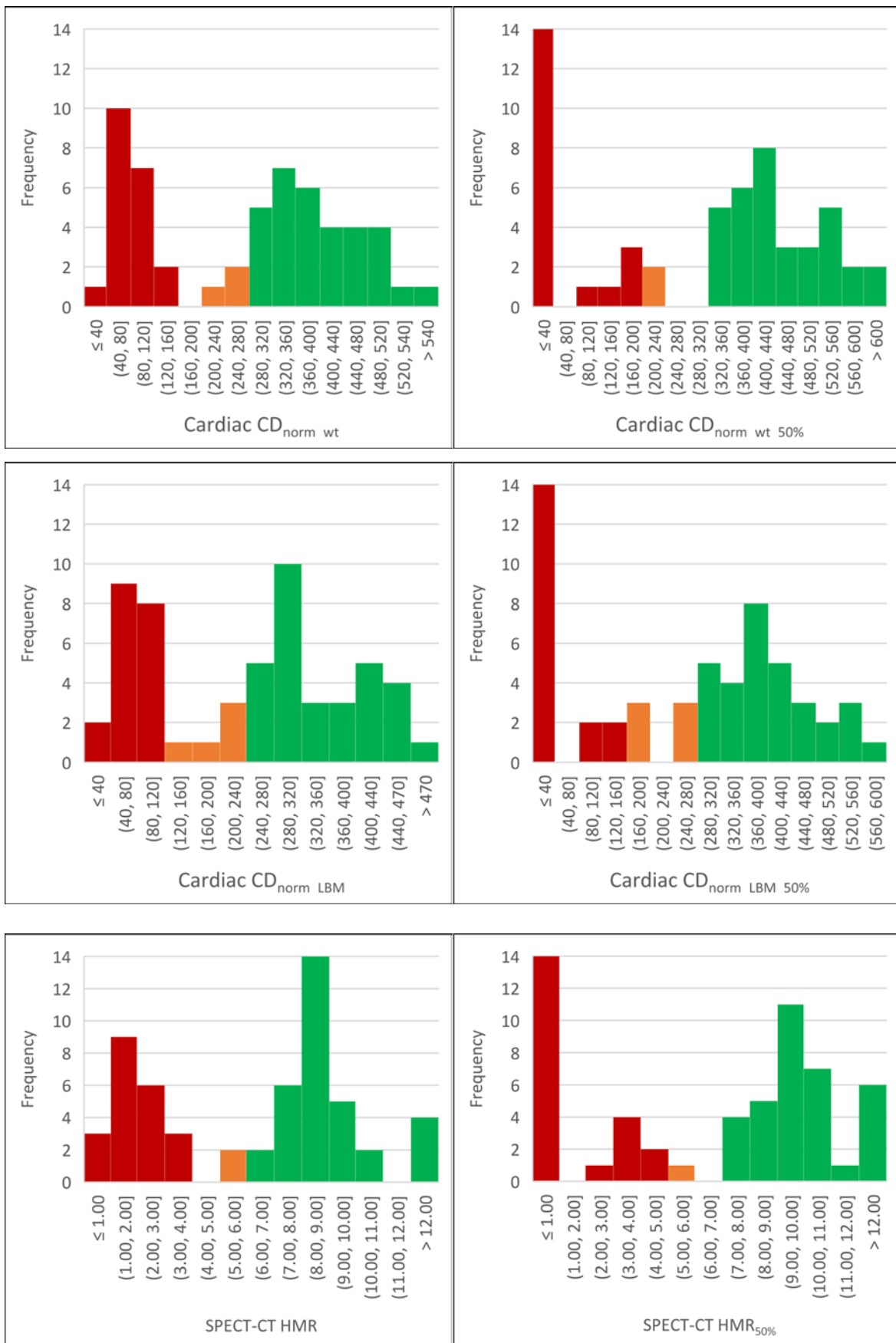


Figure 92: Histograms of normalised cardiac count density for the spherical mean counts and the 50% threshold mean. The top row shows weight normalisation and the middle row LBM. The SPECT-CT HMR histograms are shown in the bottom row for comparison. Red categories are below the control derived cut-offs, green above, and orange borderline.

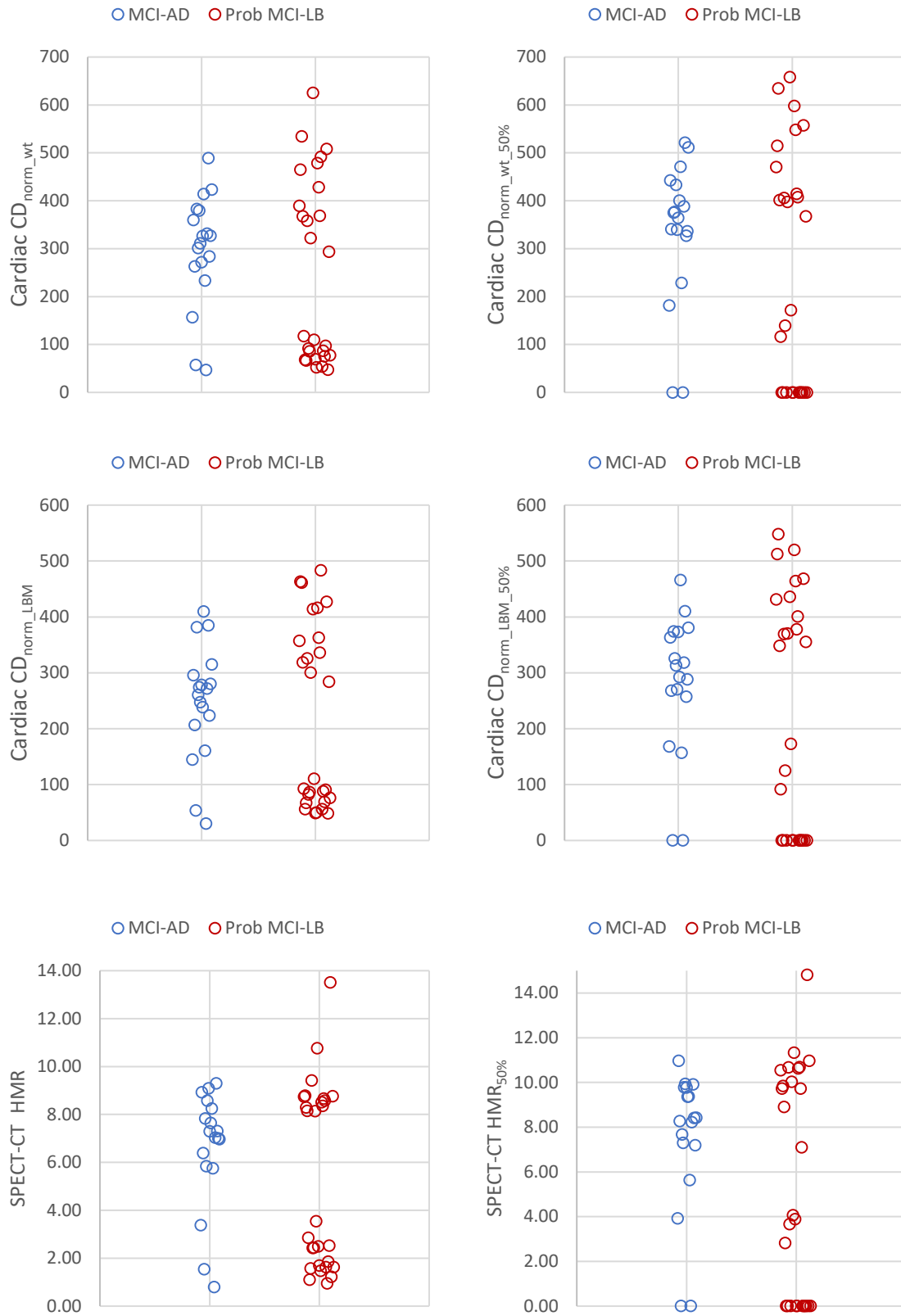


Figure 93: Plots of cardiac count density for MCI-AD and MCI-LB cases, with SPECT-CT HMR plots for comparison.

8.3.4 Discussion

VoD normalisation

Both the body weight and lean body mass normalisation methods corrected for the effect of patient size reasonably well, but the LBM method was the better of the two, with no residual dependence on size for either cardiac or mediastinum counts. We would therefore recommend this method over the weight based method. However, the normalisation method did not affect diagnostic accuracy, possibly due to the large separation between groups discussed in section 8.2. The accuracy with the 50% threshold VOI was the same for both weight and LBM normalisation and the accuracy with the spherical VOI was actually slightly better for the weight normalisation due to the MCI-AD patient discussed in section 8.2 being normal with weight normalisation and abnormal with LBM.

Comparison with SPECT-CT HMR

There was no difference in accuracy compared with SPECT-CT HMR. The histograms in Figure 92 and dot plots in Figure 93 suggest that the separation between normal and abnormal looking scans is slightly less than for SPECT-CT HMR, but the two methods are similar. The coefficients of variation of the normalised count density in the cardiac and mediastinum VOIs were both similar to that of SPECT-CT HMR. This demonstrates that the noise in the SPECT-CT HMR data is contributed approximately equally by the cardiac and mediastinum regions, so the hypothesis that absolute quantification may benefit from removal of the variable mediastinum region appears incorrect. Many of the factors that introduce uncertainty into absolute quantification, such as accurate activity measurements and correction for patient size, are cancelled out by use of relative quantification so SPECT-CT HMR is likely to be more reliable than cardiac CD_{norm} .

8.3.5 Conclusion

Based on this dataset, absolute quantification of cardiac MIBG uptake appears to offer no benefit over relative quantification with SPECT-CT HMR. The rationale for using absolute quantification over relative quantification so that uptake values are not dependent on a noisy reference region appears unfounded.

8.4 Regional cardiac ¹²³I-MIBG uptake in dementia and healthy ageing

8.4.1 Background

Cardiac MIBG studies are typically quantified on planar images using the heart-to-mediastinum ratio as a diagnostic indicator, as in previous sections. The HMR is an assessment of the overall cardiac uptake and does not take regional differences in uptake throughout the left ventricle into account. However, regional variation of uptake can be assessed using SPECT imaging.

Published studies exploring the use of SPECT cardiac MIBG imaging for diagnosing Lewy body disease have suggested that regional assessment of cardiac uptake could improve sensitivity as regional defects may occur before global loss is apparent [56, 177, 179, 239, 245, 246]. However, we have noted that regional count reductions are sometimes seen in probable AD subjects who have normal planar scan appearances, suggesting that discrete areas of reduced regional counts may be non-specific, e.g. due to unreported silent myocardial infarction, or artefactual, e.g. due to attenuation or the influence of extra-cardiac activity. It has been reported in the literature that heterogeneous cardiac MIBG count distribution can occur in healthy controls, with relative reduction most commonly seen in the inferior wall [188, 216, 246-251]. This may be due to true physiological variation in sympathetic nerve distribution or artefacts caused by the imaging process – either from the signal being attenuated by the body or by interference from non-cardiac activity, predominantly the liver [252-254]. Although several groups investigating the diagnostic accuracy of cardiac ¹²³I-MIBG for the diagnosis of Lewy body disease have acquired SPECT data alongside planar, often it is only the planar scans that have been used for analysis with the SPECT results not reported [55, 183, 236-238], suggesting that the added value of SPECT is unclear. In their 2015 multicentre study, Yoshita et al. used SPECT for visual analysis but it was not clear on what basis the SPECT images were deemed normal or abnormal [53].

We reviewed cardiac MIBG SPECT image data from older adults with normal cognition from the SUPeRB study and age matched DLB patients and Alzheimer's disease patients from the MIDAS study in order to gain a better understanding of cardiac MIBG distribution in these groups and therefore how to interpret patient scans. Our objective was to compare the distribution pattern of MIBG in the heart seen in these clinical groups. The underlying hope was that we may be able to distinguish confounding regional reductions in cardiac uptake due to cardiac disease from pathological ones due to Lewy body disease by looking at heterogeneity, thereby improving specificity. Sensitivity may also be improved if we were to find any characteristic patterns of reduction associated with Lewy body disease. The aim of this work was to carry out a pilot study to assess the patterns of regional uptake in controls, AD patients and DLB patients and determine whether this might have application in the diagnosis of Lewy body disease.

8.4.2 Methods

Subjects

Older people without cognitive impairment (controls):

Thirty-one healthy adults aged over 60 years were recruited as part of the Newcastle SUPeRB study. A detailed medical history was taken, which later enabled an estimate of cardiac risk to be made using the QRISK2 tool, recommended by NICE in Clinical Guideline 181 Cardiovascular disease: risk assessment and reduction [255]. QRISK2 uses the ratio of total cholesterol to high density lipoproteins as a risk factor, with a ratio of 4 used as default. These parameters were not part of our original study design so were not available for most of the volunteers. As the true ratios were not known, a ratio of 6, which is regarded as the upper limit by Heart UK, was input for those with high cholesterol disclosed in their medical history and the default of 4 used for everyone else. Two subjects had documented heart disease, making the QRISK2 score inappropriate - their risk scores were still calculated with this recorded as a limitation. QRISK2 scores in controls were assessed for any correlation with cardiac uptake using linear regression.

Older people with dementia:

Seventeen patients meeting the clinical criteria for probable dementia with Lewy bodies and 15 meeting criteria for dementia due to Alzheimer's disease were recruited from memory clinics, as described in Chapter 2 and our previous publications [83, 190]. Cardiac risk was not calculated for the dementia patients since their cardiac uptake was expected to be affected by Lewy body disease in many cases. However cardiac history was recorded to enable comparison with SPECT imaging results. We have previously showed no convincing evidence of lower cardiac MIBG uptake in patients with a previous MI, although this was a small subset [83].

Image acquisition

Planar and SPECT images were acquired as described in sections 7.1 and 8.1

Image processing

All images were processed on a Hermes workstation. Delayed anterior planar images were analysed to obtain the HMR as described in section 7.1. SPECT images were processed in Hermes Hybrid recon using filtered back projection (FBP) with a Butterworth filter of 0.38 cycles per cm cut-off and order 8

applied.⁵ The images were reoriented during reconstruction to align with the horizontal and vertical long axes of the myocardium, as is standard for myocardial perfusion SPECT, with as much extra cardiac activity (e.g. liver and lung) masked out as possible.

Image assessment

Quantitative Perfusion SPECT (QPS) software (Hermes Medical Ltd, Stockholm) was used to display the short axis, vertical long axis and horizontal long axis slice images in rows for visual assessment of relative uptake. Polar plots of relative uptake pattern were also reviewed. These plots display the average relative uptake in the apex, anterior, lateral, inferior and septal walls as a percentage of the maximum pixel value. The images and polar plots were reviewed blinded to the clinical groupings by a Consultant Clinical Scientist experienced in using the software for myocardial perfusion imaging (EJ). The overall degree of uniformity and the relative uptake for each of the left ventricle walls (anterior, inferior, septal, lateral and the apex) was recorded using the categories given in Table 69. The ratings categories were developed by GR, EJ and JLL by reviewing cardiac MIBG images not included in the current study.

Table 69: Rating categories for overall assessment of uptake pattern in the left ventricle and of uptake in each of the ventricular wall regions.

Overall rating categories	Uptake categories for each wall	Interference categories
<ul style="list-style-type: none"> • Not possible to score due to interference • Definitely uniform • Very slight heterogeneity • Patchy • Clear defect • Multiple clear defects • Other - see comment 	<ul style="list-style-type: none"> • Normal • Mildly reduced / patchy • Moderately reduced • Severely reduced • No uptake 	<ul style="list-style-type: none"> • No significant interference • Liver interference • Lung interference • Both liver and lung interference

8.4.3 Results

Controls with normal cognition

Of the 31 controls recruited, 2 had to be excluded due to their image appearances (as for the planar analysis in Chapter 7). The remaining 29 people included were aged between 62 and 94 years (mean

⁵ FBP was used as there was concern that resolution recovery used with the iterative OSEM method, when combined with MELP collimators and a non-circular autocontoured orbit may introduce artefactual non-uniformity, based on scans of the anthropomorphic torso phantom.

75.2). Most were male (22/29) and white (28/29). Body mass indices (BMI) ranged from 22 to 38 kg/m², with the mean value of 27.2 in the overweight category. Ten-year cardiac risk scores assessed with QRISK2 showed the group to be at high risk, as expected for older adults (Figure 94). Most subjects were at higher risk than expected for age due to the inclusion of one or more cardiac risk factors in the calculation. However, despite slight downward trends, there was no significant correlation between cardiac risk and either overall cardiac MIBG uptake as given by planar HMR (p=0.64) or minimum percentage relative uptake on SPECT (p = 0.08) (Figure 95).

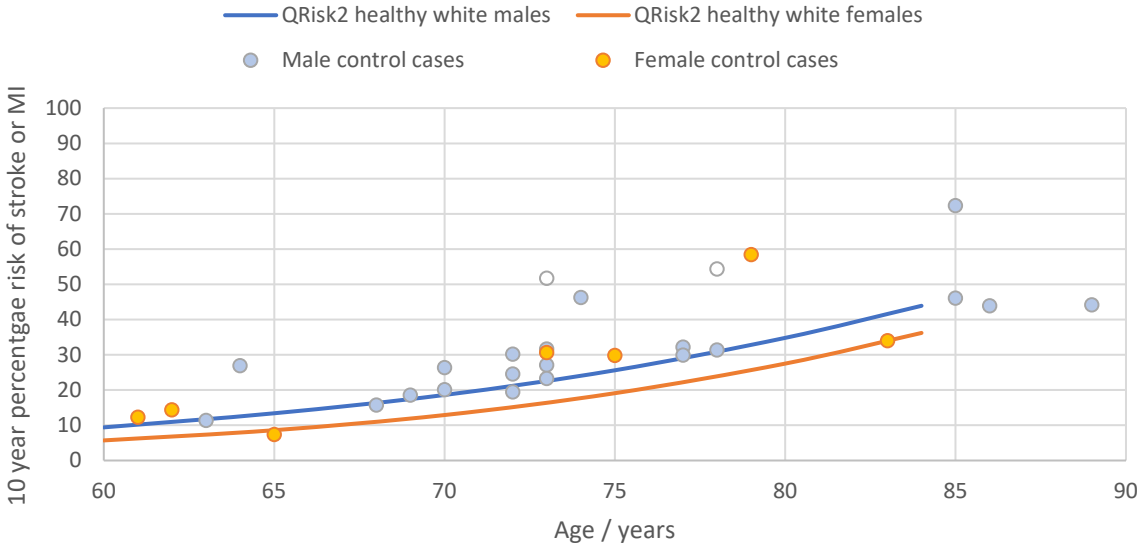


Figure 94: Plot of cardiac risk for our control subjects against subject age for males (blue data points) and females (orange data points). The two males with previous cardiac history are shown in open circles. The solid lines correspond to the mean QRISK2 database scores for white males and females with no cardiac risk factors. The database is only available up to 84 years.

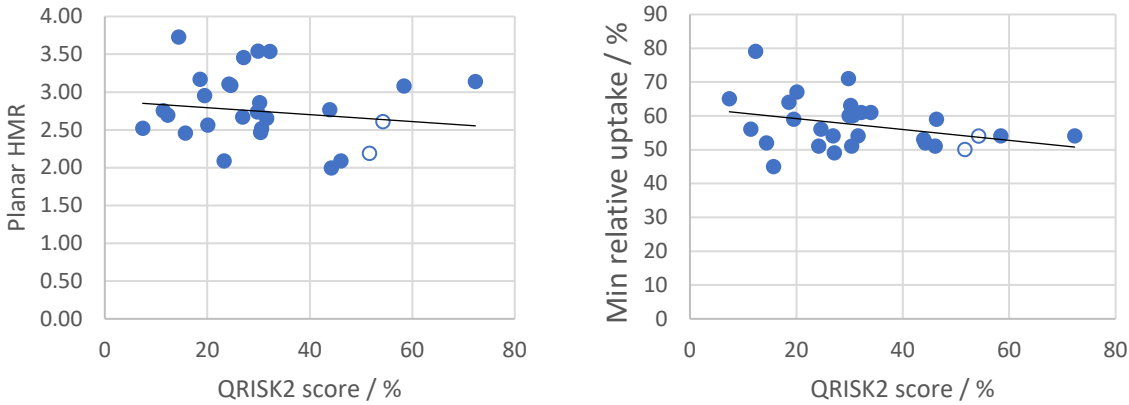


Figure 95: Plots of HMR (left) and minimum relative uptake (right) against cardiac risk. Controls with previous cardiac history are shown in open circles.

Two controls had abnormally low uptake on planar imaging (below the normal threshold of 2.2 suggested for medium energy collimators by the Japanese normal database of controls [188]) and no discernible uptake on SPECT imaging. These individuals had no significant medical history and were

not taking any medications suspected of interfering with cardiac MIBG uptake, including those without strong clinical evidence of an effect [256]. Of the remaining 29 controls, three were not possible to assess due to a high degree of lung interference. A further two with significant lung interference and two with liver interference were assessed but may be unreliable. Overall, 62% of the controls had some degree of interference; lung interference in 11/29 cases (38%) and liver interference in 7/29 (24%). Almost all controls assessed (25/26) displayed at least one area of clearly reduced relative uptake, most often in the inferior wall, which was rated as normal or mildly reduced in only 4/26 cases (15%). Some example images and ratings categories are shown in Figure 96. The results are summarised in Table 70 and Table 70 alongside those of the dementia patients.

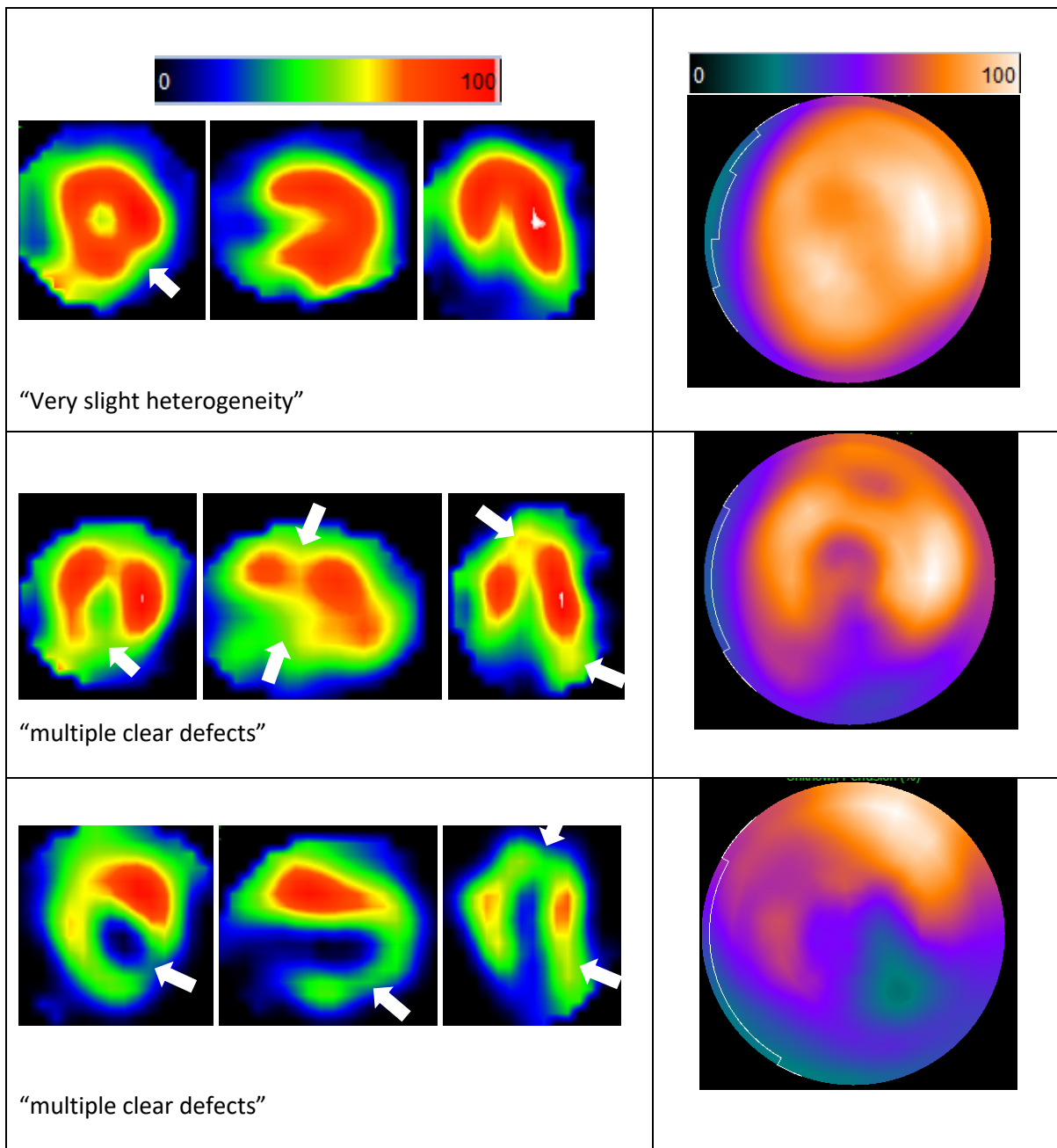


Figure 96: Top: QPS slice data and polar plot for the single control rated as within normal limits, showing almost perfectly uniform MIBG distribution, similar to that seen in normal Tc-99m myocardial perfusion images. Middle: Slice data and polar plot for a more typical control, with reduced relative MIBG uptake inferiorly and at the apex. Bottom: A very abnormal control but without prior known cardiac events. The sites of reduced relative count density are shown with white arrows.

People with dementia

Of the 15 people with DLB who completed both planar and SPECT imaging, ten had HMR values below 2.2. Of these, there was one with visible cardiac uptake on SPECT images, but it was too unclear for a meaningful SPECT reconstruction to be obtained. All five of the DLB subjects with normal planar cardiac uptake had uptake visible on SPECT, i.e. the normal HMRs were not caused by overlapping liver or lung. There were therefore only five DLB SPECT scans included for SPECT regional analysis, those with normal planar imaging. Of the 15 people with AD, three had HMR values below 2.2, however all 15 AD patients had sufficient cardiac uptake to reconstruct clear SPECT images.

The pattern of SPECT uptake was variable but overall similar to that seen in the controls for both AD and DLB (Table 70 and Table 71).

Table 70: Frequency of regions showing reduced relative uptake for control, AD and DLB groups.

	Apex	Anterior wall	Lateral wall	Inferior wall	Septal wall
Controls	6/26 (23%)	0	2/26 (8%)	22/26 (85%)	4/26 (15%)
AD	5/15 (33%)	0	2/15 (13%)	13/14 (93%)	4/15 (27%)
DLB	1/5 (20%)	0	0	4/5 (80%)	1/5 (20%)

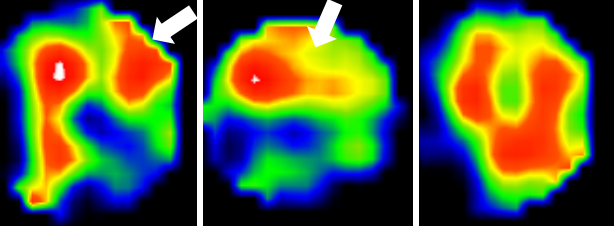
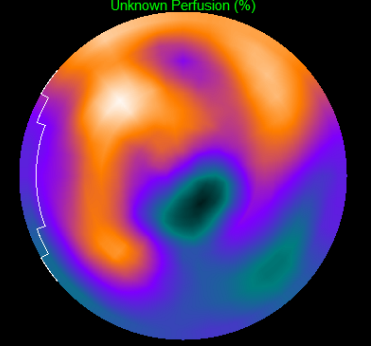
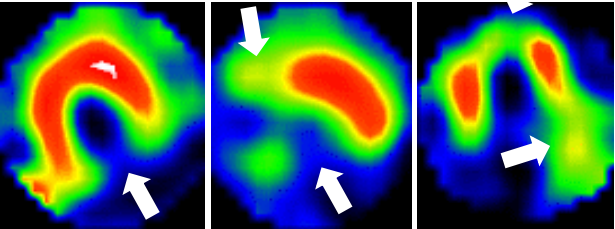
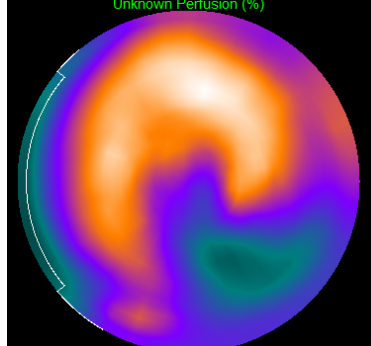
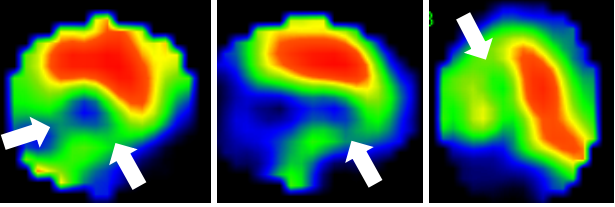
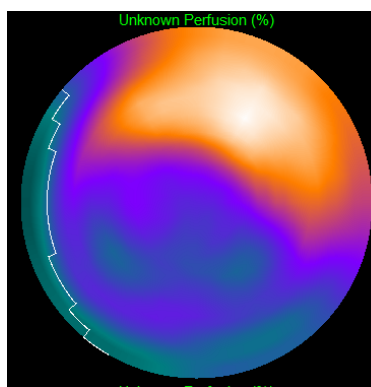
Table 71: Frequency of each overall rating category for control, AD and probable DLB groups.

	Control (n=29)	AD (n=15)	DLB (n=5)
Interference present?	18 (62%)	10 (67%)	2 (40%)
Not possible to score due to interference	7 (24%)	3 (20%)	1 (20%)
Definitely uniform	0	0	0
Very mild heterogeneity	1 (3%)	0	0
Patchy	0	1 (7%)	0
Clear defect	8 (28%)	3 (20%)	2 (40%)
Multiple clear defects	13 (45%)	8 (53%)	2 (40%)

Effect of myocardial infarction on planar and SPECT images

One control, three of the AD patients and two of the DLB patients had had a previous MI. Their images are displayed in Figure 97. The control had a borderline HMR of 2.19 (assuming a normal range >2.20) despite a large area of apparently reduced uptake at the apex and inferiorly (lung interference anteriorly means this should be interpreted with caution). Of the AD MI cases, one had a borderline HMR of 2.16 and two had normal uptake. Of the DLB MI cases one had normal HMR (2.85) and one had no visible cardiac uptake and so was one of the cases excluded from SPECT regional analysis.

The cases with previous MI showed defects of varying magnitude and extent on SPECT and variable planar HMRs, none of which were abnormally low. The number of cases is too low to establish a correlation between the extent of regional denervation and effect on global HMR but it seems likely that regional denervation would have to be widespread for planar HMR to be abnormally low.

	QPS slice data	Polar plot	Planar HMR
Control	 <p>“Not possible to score due to lung interference”</p>		2.19
Alzheimer's disease	 <p>“Multiple clear defects”</p>		2.99
Alzheimer's disease	 <p>“Multiple clear defects”</p>		2.16

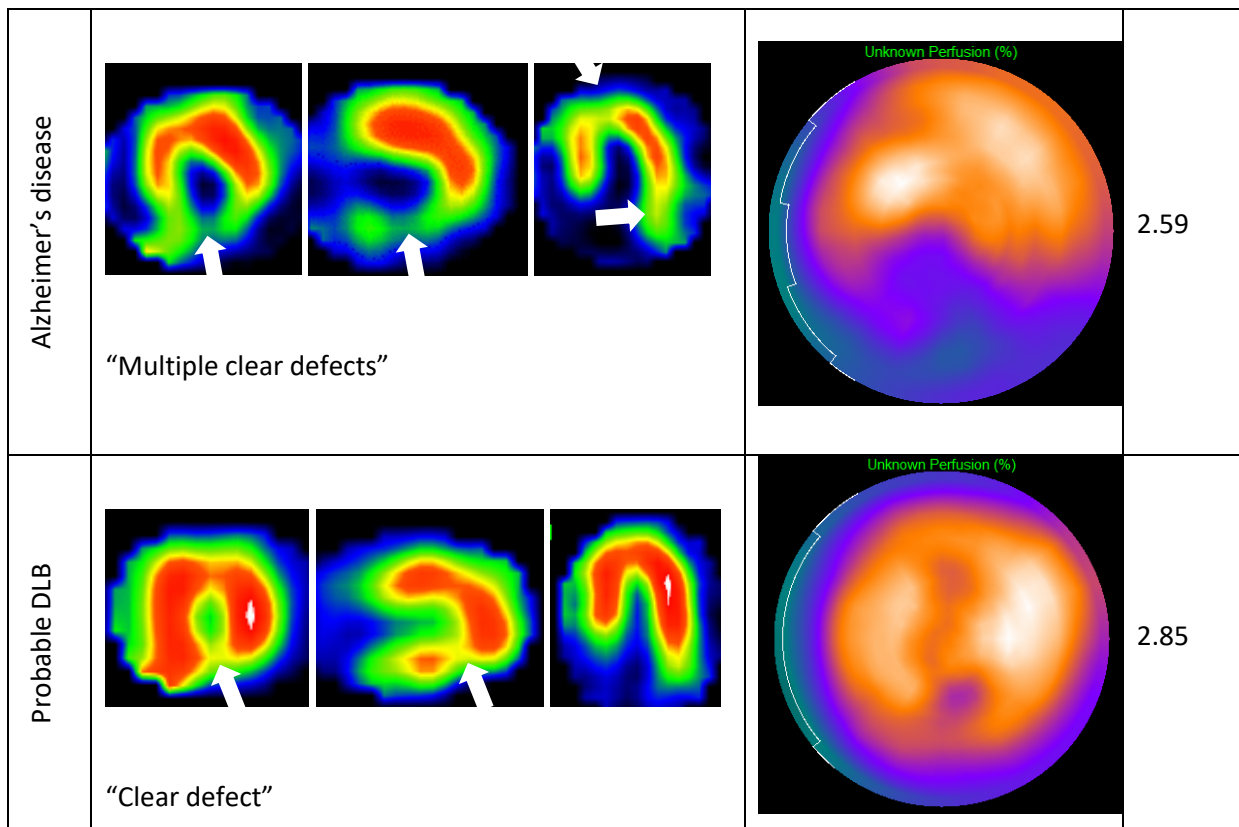


Figure 97: Slice data and polar plots for cases with previous MI, with corresponding planar HMRS. The sites of interference (top image) and relative reduced uptake are shown with arrows.

8.4.4 Discussion

The majority of the controls (25/26) and all the AD patients showed at least one region of reduced relative cardiac MIBG uptake, with 45% of controls and 53% of AD patients showing relative reduction in multiple regions. This suggests that the presence of areas of reduced uptake on SPECT should not be used to diagnose Lewy body disease, as some authors have suggested, as this is likely to be highly non-specific. Our controls were almost all at high risk of coronary events over the next 10 years, so cardiac disease as a cause for the relative reduction cannot be entirely excluded. However we note the lack of correlation between QRISK2 score and non-uniformity index suggesting the most at risk individuals are not at significantly greater risk of cardiac sympathetic denervation.

None of the control or dementia cases demonstrated reduced relative uptake in the anterior wall, which at first glance suggests that this region could be a useful target area in which to assess uptake in cases where the planar HMR is suspected of being affected by regional defects. However, 11 of the 29 control SPECT images were difficult to assess due to lung interference, three of whom were impossible to rate, making the anterior wall uptake unreliable.

The strengths of this study include the inclusion of older people thoroughly assessed to have normal cognition but who also had other morbidities associated with ageing, including cardiac risk factors, making them a representative sample of older adults. The patients with dementia had an expert panel

diagnostic review by three old-age psychiatrists, a method which has been validated against autopsy diagnosis and supported by regulatory authorities as a 'gold standard' [43, 207]. However, a limitation is the use of only one rater for the SPECT uniformity assessments. This does however reflect clinical practice in radionuclide image reporting and the results are convincing even with a single rater. A further limitation is the lack of detailed cardiac assessment and no myocardial perfusion imaging for correlation with cardiac MIBG. However, Simula et al. have shown that myocardial perfusion imaging does not correlate with cardiac MIBG SPECT in asymptomatic high coronary risk cases with mild coronary artery stenosis [257]. They found a link between reduced uptake on cardiac MIBG SPECT and degree of stenosis for the left anterior descending artery (but not the left circumflex or right coronary artery) and suggest that CAD could be the cause of regional inhomogeneities on cardiac MIBG SPECT previously reported as normal variants [257]. However, on our SPECT scans the areas showing reduced uptake are the inferior wall and/or apex for the majority of controls (21/26), which is against cardiac disease being the main cause as this would be expected to affect any of the cardiac territories. The pattern that we see suggests that either attenuation and scatter are causing this artefact, artefacts have been introduced in acquisition or reconstruction (e.g. due to high liver uptake), or there is a physiological reason for the apex and inferior wall to have reduced uptake.

The literature on regional sympathetic innervation in healthy controls is quite mixed. Most studies report a heterogeneous distribution within the left ventricle, although they do not agree on the sites of reduced uptake, perhaps related to small sample sizes and different age groups. Early work by Sisson and colleagues in the 1980s on 20 men aged 20-62 years showed a relative reduction at the apex only [258]. The European guidelines on ^{123}I -MIBG cardiac sympathetic imaging note that "normal cardiac MIBG distribution includes a relatively low uptake in the inferior wall, which is more pronounced in the elderly" referencing the 1993 study of Gill et al. [252] in seven young adults (29.4 ± 7.5 years) and eight middle aged adults (aged 53 ± 5.1 years – by no means elderly!). As well as the inferior reduction Gill et al. also reported septal reduction, with no difference between apex and base [252]. Tsuchimochi et al. studied 18 men and 11 women without cardiac diseases aged between 21 and 79 years (mean 42 ± 17 years) and reported that the inferior uptake of MIBG decreased with age in healthy controls, especially in men [246]. This is important in the context of our study, as all participants were over 60 years of age and most of our controls were men. Image quality was however quite limited compared to modern gamma camera images for both this study published in 1995 and those of Sisson [258] and Gill [252].

Our results are similar to those of a more recent study published by Asghar et al. in 2017 [216] who reported reduced inferior wall uptake in all 14 healthy controls studied, and suggested that these effects are not predominantly caused by imaging artefacts. However, all our control cases had normal anterior wall uptake, which was not the case in the Asghar study. Interestingly, in their 2012 ^{11}C

Hydroxyephedrine PET study, Wong et al. showed that the lateral wall not the inferior wall shows reduced regional innervation in healthy controls, with relative sparing of anterior and septal walls [259].

Post mortem tissue studies e.g. [260, 261] report non-uniform distribution of cardiac innervation, for example in the sinus node; but do not describe the relative distribution in the left ventricle so cannot be correlated with MIBG tracer uptake. However, in their post-mortem study on the distribution of autonomic nerves in the human heart, Kawano et al. showed a small but statistically significant reduction in the sympathetic innervation of the inferior wall of the left ventricle compared to the anterior wall [262]. The reduction is only around 20%, making this unlikely to explain the substantial differences seen in ^{123}I -MIBG uptake between the anterior and inferior wall in many controls.

It has been hypothesised that due to its tracer kinetics, ^{123}I -MIBG imaging is insensitive to substantial nerve losses in cardiac regions with mild to moderate levels of cardiac denervation and that regional nerve losses need to be relatively severe before a reduction in uptake becomes apparent [263]. If this is the case it makes the modest physiological differences seen by Kawano et al. [262] even less likely to explain the cardiac MIBG findings. We note that ^{11}C Hydroxyephedrine (HED) has similar kinetics to ^{123}I -MIBG, so this doesn't explain the difference in relative uptake patterns.

This leaves artefacts as a possible contributor to the apparent regional reductions in cardiac MIBG uptake. In a study using an anthropomorphic chest phantom with uniform myocardium activity, the count density within the myocardium was inhomogeneous with both medium and low energy collimators [253]. There was an overestimation of the measured count density in both the anterior and the septal wall and an underestimation of count density in the inferior wall [253], which is in line with typical control studies and which the authors assumed was due to attenuation. However, Gill considered attenuation artefacts an unlikely cause of the reduced inferior uptake in their control series, as myocardial perfusion imaging with thallium-201 did not show corresponding reductions [252]. We note that although we commonly see mildly reduced relative uptake in the inferior wall on clinical myocardial perfusion images obtained with $^{99\text{m}}\text{Tc}$ -tetrafosmin, these reductions are typically less obvious than those seen on the majority of our control ^{123}I -MIBG images. However both $^{99\text{m}}\text{Tc}$ and ^{201}Tl have different physical properties to ^{123}I , particularly in terms of high energy emissions and may not give the same apparent cardiac uptake patterns. This discrepancy could be potentially be resolved with the use of SPECT-CT imaging to correct for attenuation and scatter of the gamma rays from the MIBG tracer within the body. A recent phantom study by Shibutani et al. demonstrated an inferior artefact in a myocardium insert (scanned with a low-medium energy general purpose collimator), which resolved with Siemens Flash3D SPECT-CT reconstruction with ACSC [264].

Regarding our single control regarded as having uniform relative uptake, the following observations are of note. The planar HMR was 2.69, which is a typical value in line with the other controls (mean HMR 2.78 ± 0.46), rather than an outlier. Similarly, the BMI of this individual was unremarkable at 28.6 (mean 27.8 ± 4.2). However, the relative uptake of the heart is similar to that of the liver, whereas in most of the controls the uptake is much higher in the liver than the heart (Figure 98). In addition, the positioning of the dome of the liver is just inferior to the heart, which is not the case in many of the other controls. This suggests that high liver uptake may be contributing to the apparent reduced inferior wall uptake seen in most of the controls. One of the few studies to directly compare ^{11}C Hydroxyephedrine PET with ^{123}I -MIBG SPECT, in patients with heart failure and coronary heart disease, reported lower liver uptake on ^{11}C Hydroxyephedrine PET than ^{123}I -MIBG SPECT, with increased uptake in the inferior and septal regions [265], so this may explain the difference in relative distribution reported with the two tracers. A study comparing both tracers in ten healthy young men with normal BMI showed lower relative uptake in the inferior wall and septum with ^{123}I -MIBG SPECT compared with ^{11}C -HED [266]. This further suggests that liver uptake may be responsible for the inferior reduction, but we note that attenuation and scatter correction was used for the PET imaging but not the SPECT. To our knowledge, no other studies of ^{123}I -MIBG and ^{11}C -HED have been carried out in the same cohort of healthy controls and no studies at all with CT attenuation and scatter correction for ^{123}I -MIBG.

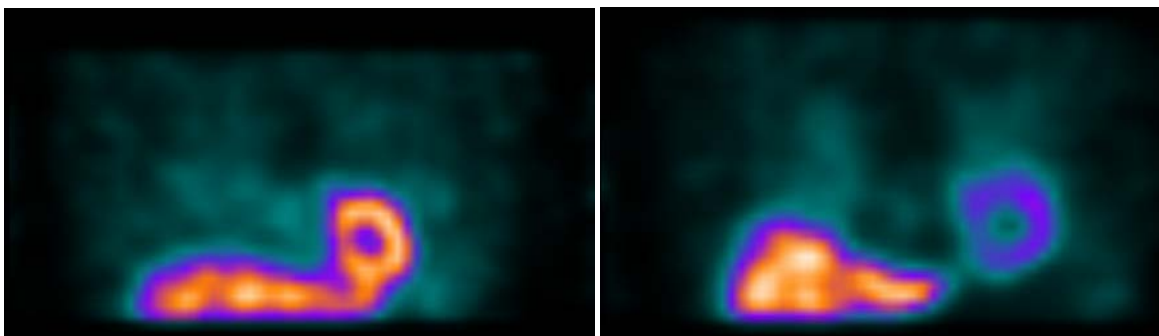


Figure 98: Left: coronal MIBG slice for the control subject with uniform relative cardiac uptake. Right: a more typical control with high liver uptake relative to the heart. Images are displayed scaled to the maximum voxel.

Limitations

As discussed above, a limitation of this study is the lack of definitive information on cardiac disease status of the participants. The QRISK2 score gives an idea of the likelihood of future events, but it is impossible to exclude undiagnosed asymptomatic CAD. The patients and controls were recruited to a dementia research study for which detailed cardiac assessment was not part of the study protocol; however an ECG and autonomic dysfunction tests were carried out. A further limitation of this study is the lack of data from younger subjects as only those over 60 years of age were recruited. SPECT-CT data for attenuation and scatter correction was not available for the patients so could not be assessed as part of this analysis.

8.4.5 Conclusion

Areas of relatively reduced cardiac ^{123}I -MIBG uptake on SPECT images are common, but not ubiquitous, in cognitively normal older adults and patients with dementia who have normal planar cardiac uptake. Uptake is often reduced in the inferior wall and sometimes in the septal wall, lateral wall and apex. Anterior wall uptake is often influenced by lung interference so appears unsuitable as a target region for uptake assessment, at least on images uncorrected for scatter and attenuation.

Reduced regional cardiac MIBG uptake in the inferior wall of the left ventricle is common in older adults and most likely related to attenuation or liver uptake. Reductions in other areas may also be artefactual or may be caused by mild underlying coronary artery disease. The pattern of reduction is in any case non-specific and therefore unlikely to aid in the diagnosis of Lewy body disease in typical clinical settings.

8.5 Summary of chapter 8

In section 8.1 we presented encouraging results in an anthropomorphic chest phantom, showing that increasing the amount of attenuating material had a notable impact on both the visual appearance and quantification of planar images, but that on SPECT-CT images the extra attenuation and scatter was compensated for accurately by the reconstruction algorithm. We saw a statistically significant drop in planar HMR with patient size, which is in keeping with the limited literature on this topic published to date [218, 228]. On attenuation and scatter corrected SPECT-CT imaging, there was no evidence of a significant relationship between HMR and BMI. However, when we looked at the difference between SPECT results (non-corrected) and SPECT-CT this was not significant, suggesting that in practice SPECT-CT ACSC may not be able to fully compensate for differences in patient size.

In the next section we showed that the impact on diagnostic accuracy in MCI of SPECT and SPECT-CT was small, with one MCI-AD patient misdiagnosed as abnormal with planar imaging having uptake results within normal limits on SPECT-CT. However, there was a considerably greater separation between normal uptake and low uptake scans with SPECT-CT than with either planar or non-corrected SPECT imaging, suggesting that SPECT-CT could have a role in cases that are borderline on planar imaging. We will review this when the SUPeRB cohort is complete and the full dataset has been processed. The implication for clinical practice is that HMR ratios can be influenced by the size and shape of the patient, and borderline results in obese patients could be the result of attenuation and scatter, rather than early changes in innervation.

In section 8.3 we presented a method for quantifying the absolute uptake within cardiac regions of interest on SPECT-CT images, which has not been applied cardiac MIBG imaging before. We found that cardiac signal should be normalised to lean body mass rather than patient weight. We compared the accuracy of absolute quantification, using our controls to give normal cut-offs, with that of relative quantification with SPECT-CT and found no differences. The distribution of the absolute quantification results, seen on dot plots and histograms, was similar to that of the SPECT-CT HMR results, suggesting that absolute quantification is not likely to improve the diagnosis of Lewy body disease. In section 8.4 we saw that reduced regional cardiac MIBG uptake is non-specific and therefore unlikely to aid in the diagnosis of Lewy body disease in typical clinical settings.

To summarise, the work presented in this chapter provides evidence that SPECT-CT may be beneficial in research studies, where a more precise estimate of cardiac sympathetic innervation is desirable. SPECT-CT may also have a limited role in clinical practice, for example it may be indicated if early and delayed planar imaging has been carried out and the results are equivocal. We would not recommend that SPECT-CT is routinely used in clinical practice as this would more than double the effective radiation dose to an individual with only a small chance of affecting the outcome of the test.

Chapter 9: FP-CIT and MIBG uptake in healthy ageing

Published literature on the relationship between striatal FP-CIT uptake and age shows clear evidence that younger adults have higher uptake than older adults. However, the mathematical model for the decline with age and whether this continues into old age is not clear. Literature on cardiac MIBG uptake and age is mixed, with some studies demonstrating no significant change in uptake with age and some showing normal uptake is lower in old age. In this chapter we investigate whether there is evidence of a relationship between uptake and age in older adults without evidence of Lewy body disease between the ages of 60 and 92 and discuss the implications for the diagnosis of Lewy body disease in this age group.

Section 9.1 explores whether age-correction is needed for FP-CIT quantification in older adults and is adapted from our 2019 publication [267]. Section 9.2 applies similar methodology to planar and tomographic cardiac MIBG uptake measurements, using the SUPeRB control group of older adults with normal cognition. An overall chapter summary is given in section 9.3

9.1 Do ^{123}I -FP-CIT specific binding ratios decline with age in older adults?

9.1.1 Introduction

Studies in healthy controls over a wide range of ages have demonstrated that striatal uptake levels decrease with age, confirmed by Karrer *et al.* in their 2017 meta-analysis [200]. It is therefore important to interpret SBRs by comparing with age-matched healthy controls and Z scores are commonly age-adjusted for this purpose. However, the exact form of the relationship between SBR and age is not clear, leading commercial software packages to implement different age correction models. Patients referred for FP-CIT scans for dementia diagnosis are typically over 60 years of age and often much older. It is therefore vital to ensure that age-correction is appropriate for this age group.

Some dopamine transporter availability studies report a linear decline in SBR with age, for example the ENC-DAT study [152]. This is a large European initiative led by the European Association of Nuclear Medicine (EANM), which included FP-CIT scans of 139 adults aged between 20 and 83 years and found slopes (absolute change in SBR per year) of -0.015 (95% confidence interval: -0.021 to -0.009) for men and -0.018 (-0.025 to -0.011), which they state corresponds to a 4 to 7% decline in SBR per decade depending on the striatal region assessed [152]. Prior to this, a large single-centre study of 186 adults found a loss of 46% over 18-88 years, on average 6.6% per decade, also applying a linear model [268]. A recent publication studying 256 healthy individuals in Japan aged between 30 and 83 [269] demonstrated a linear fit for both males and females, reporting an average decline of 6.3% per decade. However, a study using a Tc-99m ligand similar to FP-CIT [270] concluded that an exponential or bi-linear function of SBR with age is a better model, with the decline tailing off in older age. The meta-analysis by Karrer *et al.* [200], which combined data on a variety of PET and SPECT dopamine transporter ligands, concluded that while there is some suggestive evidence for steeper decline in young adults, a linear relationship describes the decrease in function with age as well as an exponential function (coefficient of determination (R^2) values of 38.6% for linear fit and 37.1% for log-linear fit). However, most subjects were not scanned with ^{123}I -FP-CIT, with only 5 such studies included within the 33 studies on dopamine transporter availability. In addition, the majority of the 289 ^{123}I -FP-CIT scans included appear to come from two studies published by the ENC-DAT team [142, 271] suggesting that double counting may have occurred. The relationship in older subjects alone was not studied, as this was not the focus of the meta-analysis, so it was not demonstrated that a significant age-dependency exists in this age group.

Two commercial software packages were used in the current study: DaTQUANT and BRASS. The DaTQUANT normal database consists of 119 healthy control subjects (aged 30 - 84 years) taken from the Parkinson's Progression Marker Initiative study (PPMI <https://www.ppmi-info.org> [272]). A bi-

linear model with different slopes for individuals under 55 and over 68, with a transition region in between is implemented. This was generated by the DaTQUANT developers by splitting the control SBRs into two clusters with different linear fits, since it was noticed that the SBR does not change as rapidly in older age (Personal communication, GE Healthcare). This was done separately for the left and right striata and for each sub-region resulting in small differences between regions and sides. The slopes implemented for the whole striatum (i.e. absolute change in SBR per year) are -0.025 for both striata under 55 years of age and -0.012 and -0.007 for the left and right striata respectively over 68 years (Personal communication, GE Healthcare). The DaTQUANT SBR intercepts (i.e. projected SBR at age zero) are 3.5 (left striatum) and 3.4 (right) for the younger cluster and 2.7 (left) and 2.3 (right) for the older cluster. These slopes and SBR intercept values apply to images reconstructed with OSEM iterative reconstruction without attenuation correction, and correspond to decreases in SBR of 0.25 (10%) between 40 and 50 years, 0.12 (7%) between 70 and 80 years for the left striatum and 0.07 (4%) between 70 and 80 years for the right striatum. A bi-linear model can also be implemented within BRASS, but we used the default model, which is a linear fit to 103 of the controls from the ENC-DAT study aged between 21 and 83 years, used as the healthy control database in BRASS.

Our experience with both software packages has been that SBRs for the oldest adults can be quite low and the scans can have abnormal visual appearances, yet the age-corrected Z scores can be within the normal range. This raises the possibility that the age correction required for older people may be overestimated. Published data showing SBR against age for the healthy controls from the ENC-DAT study [152], which applied a single linear fit to all ages, shows little evidence of a clear continuing decrease in SBR with age in adults over 60 years old, substantial variation between subjects and relatively few data points for subjects over 75 years. The DaTQUANT database also shows no convincing decline in SBR with age for subjects over 60 (SBR data accessed within the software).

We recruited healthy older adults and older adults with cognitive impairment but no signs of Lewy body disease to determine whether there is evidence of age-dependency of SBR in adults over 60 years of age. We also used a separate independent test group of individuals over 60 with and without Lewy body disease to test our findings concerning age correction.

9.1.2 Methods

Subject selection

Healthy older adults, adults with mild cognitive impairment (MCI) and adults with dementia were taken from the SUPeRb, LewyPro and MIDAS studies. The healthy older adults and the MCI-AD individuals from one of the studies were combined to give a large group of controls (i.e. older people without evidence of Lewy body disease). An independent group of individuals with MCI and individuals with dementia was used to test the effects of age correction. For both the control and test groups, MRI

brain scans and CT scans (where available) were reviewed by a radiologist to exclude other causes of cognitive impairment.

Control group: MCI-AD and healthy older adults

Twenty-six individuals meeting the NIA-AA clinical criteria for MCI due to Alzheimer's disease (MCI-AD) [273] were recruited from memory clinics, specialist dementia services, elderly medicine clinics and neurology clinics in the North East of England. Potential subjects had been diagnosed with MCI in these services and were approached to take part in the Newcastle University LewyPro study [37]. Diagnoses were confirmed as MCI-AD by an expert consensus clinical panel of old-age psychiatrists. As Lewy body disease may be misdiagnosed as MCI-AD, we excluded any MCI-AD individual with an abnormal FP-CIT scan from our control group. An abnormal FP-CIT scan was defined by an abnormal consensus visual rating, given by a panel of five experts, as described in section 6.5. Three of the 26 MCI-AD individuals (aged 72, 76 and 79) were thus excluded because their FP-CIT scans were rated as abnormal.

Thirty-one healthy controls were taken from the SUPeRB study. Incidental Lewy body disease is common in older people [274] so we excluded any healthy older adults with an abnormal rated FP-CIT scan from our control group to ensure that our data reflected age-related changes in adults without any evidence of Lewy body disease. Two of the 31 healthy older adults (aged 63 and 73) were thus excluded because their FP-CIT scans had been rated as abnormal by the expert panel (see section 6.1).

Test group: individuals with dementia and MCI

Sixteen individuals over 60 years of age with probable DLB, 17 with probable AD, 20 with MCI due to probable Lewy body disease (MCI-LB) and 18 with MCI due to probable AD (MCI-AD) were included to study the effect of age correction on FP-CIT results in older patients. Those with dementia were recruited as part of the MIDAS study [83]. Those with MCI were recruited as part of an ongoing Newcastle University follow up study to the LewyPro study [37]. Note the MCI-AD individuals included here are not included in the above control group. Baseline diagnoses were used for these groups as one year follow up diagnoses were not available.

Image processing and SBR calculation

FP-CIT images were acquired as described previously in chapters 4 and 6. SBRs were calculated with BRASS and DaTQUANT. Age corrected Z scores and Z scores without age-correction were recorded from BRASS. It is not possible to turn off age-correction when using the DaTQUANT normal database so only age-corrected Z scores were recorded from DaTQUANT. Sub-regions such as the caudate, putamen and posterior putamen can be evaluated by both packages but we chose to focus on whole striatum results to reduce the effects of random noise due to small, possibly mis-registered volumes of interest in our data.

Analysis

A linear mixed effects model [275] (IBM SPSS v.24) was used to test for age dependency and any effect of gender, side (left vs right striatum) and dataset (MCI-AD vs adults with no cognitive impairment). Age, gender, side and dataset were input as fixed effects in the model and study ID input as a random effect. A mixed effects approach was used so that data from both striata could be included to give 104 SBR measurements; the random effect of study ID accounts for the fact that the SBR in the left and right striatum of an individual are likely to be correlated. The model was run for DaTQUANT and BRASS data separately, with both SBR and corrected Z score used as the dependent variable. P-values of less than 0.05 were taken to indicate effects significantly different from zero. The rate of change in SBR and corrected Z score with age and 95% confidence limits were recorded.

9.1.3 Results

Subject characteristics

The age and sex of the individuals included and proportion of women in each group is given in Table 72.

Table 72: Ages and gender balance of the individuals with MCI due to AD, the healthy older adults, the two groups taken together as a control group and the test individuals with DLB, AD, MCI-LB and MCI-AD.

		Number included	Mean age (SD)	Proportion female
Controls	Probable MCI-AD	23	77.7 (8.3)	0.61 (14/23)
	Healthy older adults	29	74.5 (7.6)	0.28 (8/29)
	Total controls	52	75.9 (8.0)	0.42 (22/52)
Test group	Probable DLB	16	77.2 (7.5)	0.06 (1/16)
	Probable AD	17	77.2 (6.5)	0.24 (4/17)
	Probable MCI-LB	20	73.6 (6.8)	0.10 (2/20)
	Probable MCI-AD (test group)	18	73.8 (6.9)	0.44 (8/18)
	Total test group	71	n/a	n/a

SBR age-dependency in the control group

The mixed effects model showed that the effect of age, gender, striatum side or dataset (MCI-AD vs healthy adults) on SBR was not significant for either DaTQUANT or BRASS (Table 73). The only

significant effect on age-corrected Z score was striatum side, which applied to DaTQUANT only, likely due to different slopes implemented in the software for right and left.

Table 73: Estimate of fixed effects from SPSS mixed effects model for DaTQUANT and BRASS for dependent variables SBR and Z score.

Co-variate		Specific binding ratio (SBR)		SBR Z score	
		DaTQUANT	BRASS	DaTQUANT	BRASS
Age	Estimate	-0.041	-0.066	0.170	0.015
	T statistic	-0.947	-1.048	1.328	1.142
	95% CI	-0.013 to +0.005	-0.019 to +0.006	-0.009 to 0.427	-0.011 to +0.041
	p-value	0.348	0.300	0.190	0.259
Gender	Estimate	0.136	0.149	0.396	0.309
	T statistic	1.868	1.419	1.865	1.443
	95% CI	-0.010 to +0.283	-0.062 to +0.361	-0.310 to +0.824	-0.121 to +0.739
	p-value	0.068	0.162	0.068	0.155
Control dataset	Estimate	0.090	0.205	0.251	0.430
	T statistic	1.238	1.951	1.185	2.018
	95% CI	-0.056 to +0.237	-0.006 to +0.416	-0.175 to +0.678	+0.001 to +0.859
	p-value	0.222	0.057	0.242	0.049
Striatum side	Estimate	0.006	0.022	0.159	0.002
	T statistic	0.429	1.152	3.372	0.049
	95% CI	-0.022 to +0.034	-0.016 to +0.060	+0.064 to +0.254	-0.077 to +0.0810
	p-value	0.669	0.255	0.001	0.961

Plots of SBR against age for the 52 controls are shown in Figure 99. Visually, there is no clear evidence of age-dependency in this age group.

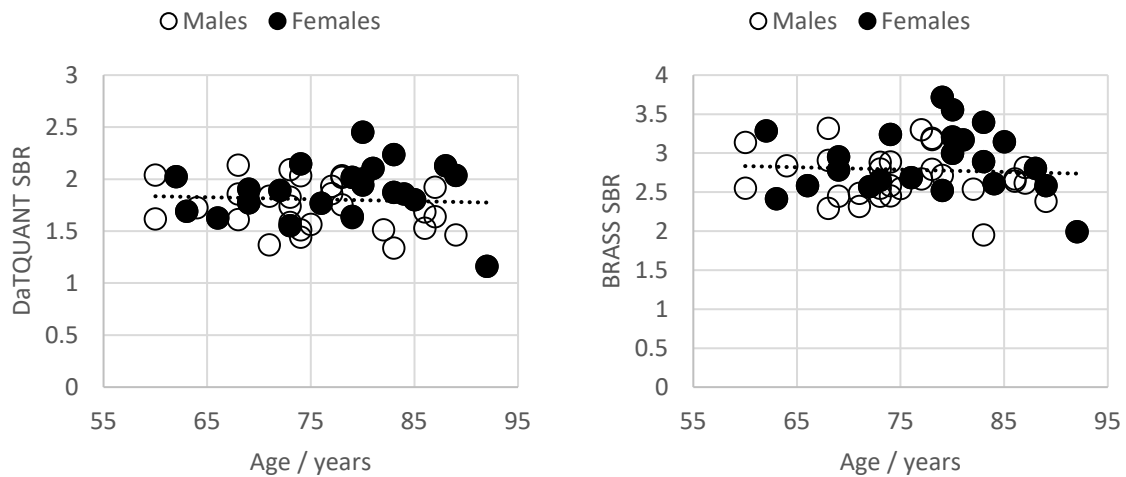


Figure 99: Mean striatum SBR values plotted against age for DaTQUANT (left) and BRASS (right).

The slopes of the linear fits to the DaTQUANT and BRASS SBR data as a function of age for controls from the mixed effects model are given in Table 74. The SBR slopes are negative but not significantly different from zero. The corrected Z score slopes are larger and positive, however these are also not significantly different from zero.

Table 74: Rate of change in SBR and age corrected Z score per year for Newcastle control subjects over 60 years of age calculated with DaTQUANT and BRASS

	SBR			Age-corrected Z score		
	Slope	95% confidence interval	p-value	Slope	95% confidence interval	p-value
DaTQUANT	-0.004	-0.013 to 0.005	0.35	0.017	-0.009 to 0.043	0.19
BRASS	-0.007	-0.019 to 0.006	0.30	0.015	-0.011 to 0.041	0.26

Z score results for our controls with age correction applied are shown plotted against age in Figure 100 (a) and (b) for BRASS and DaTQUANT. Both appear to show an increase in corrected Z score with age, although the mixed effects model above showed that this was not statistically significant. Control Z scores with the age correction function switched off in BRASS are plotted in Figure 100 (c). These Z scores result from comparing to the mean SBR taken over the entire ENC-DAT database age range. It is not possible to switch off age correction in DaTQUANT when using the factory installed PPMI database so we have not compared Z scores with and without age correction for DaTQUANT.

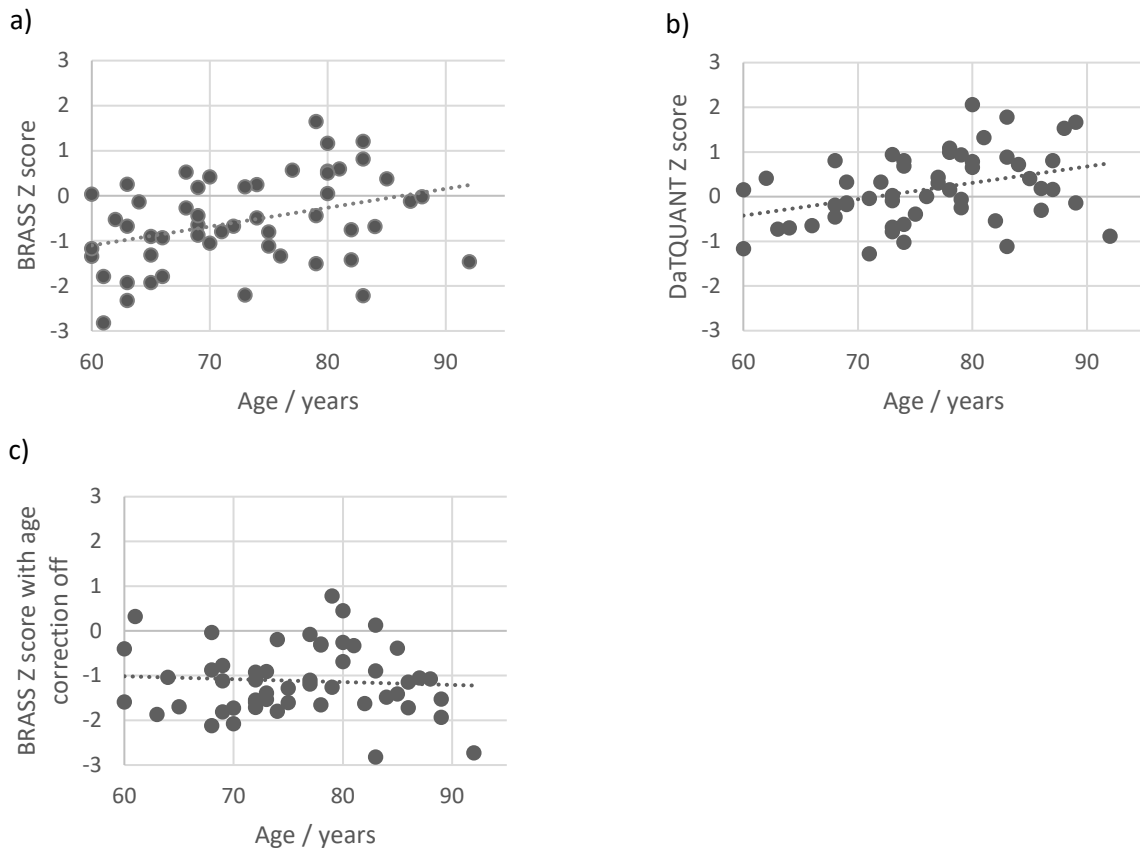


Figure 100: (a) BRASS age corrected Z scores for controls, (b) DaTQUANT age corrected Z scores and (c) BRASS Z scores without age correction. The average values of the right and left striatum SBRs are used.

Age correction in the test groups

Figure 101 shows BRASS Z scores with and without age correction for probable DLB, AD, MCI-LB and MCI-AD (the test group). The lowest of the scores for the right and left striata have been taken, rather than one side or an average as this reflects clinical practice. The number of images correctly classified based on BRASS Z score values with and without age correction is shown in Table 75.

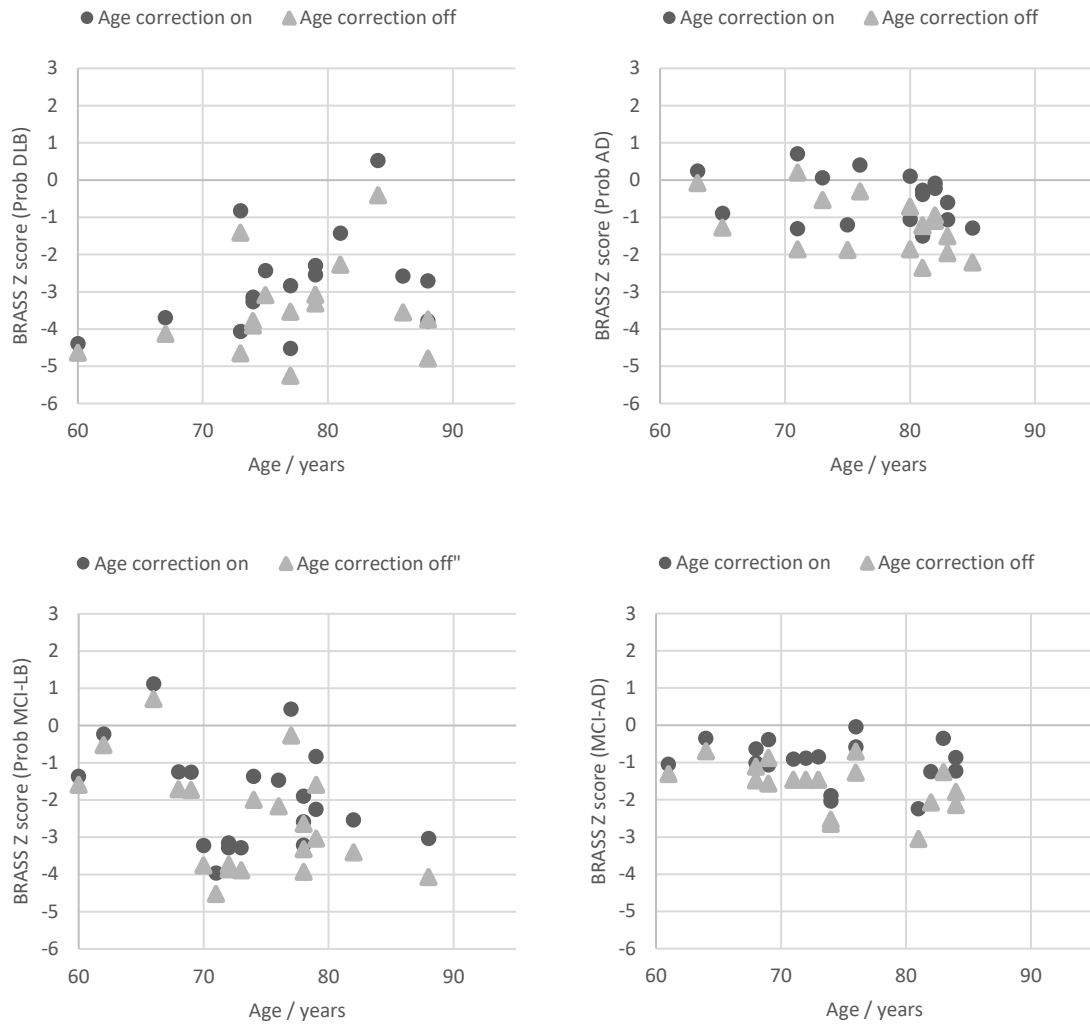


Figure 101: BRASS Z scores (lowest of left and right striatum) with and without age correction for probable DLB (a) probable AD (b), probable MCI-LB (c) and probable MCI-AD (d).

Table 75: Number of individuals with BRASS age corrected and non-corrected Z scores above and below our threshold of $-2SD$.

	BRASS age correction on			BRASS age correction off		
	Z below -2	Z above -2	% with expected Z score	Z below -2	Z above -2	% with expected Z score
Controls	4	48	92%	4	48	92%
Prob DLB	13	3	81%	14	2	88%
Prob AD	0	17	100%	2	15	88%
Prob MCI-LB	10	10	50%	13	7	65%
Prob MCI-AD	2	16	89%	5	13	72%

9.1.4 Discussion

Our results show that any decrease in the SBR of older adults with age appears to be very small, which suggests that a single linear fit over the entire adult age range may be inappropriate for older adults. The 95% confidence intervals for our linear fits (absolute SBR change per year) are -0.013 to +0.005 with DaTQUANT and -0.019 to +0.006 with BRASS. These correspond to percentage changes in SBR per decade in the control group of -7 to +3% with DaTQUANT and -7 to +2% with BRASS, taken over the range 70-80 years. The confidence intervals encompass zero, and the linear mixed effects model showed no significant difference from zero in any measured slope, so our results do not support age-dependency in this age range. The current default SBR decreases modelled in the software correspond to approximately -5% between 70 and 80 years for BRASS and -4 to -7% for DaTQUANT, depending on striatal region. The slope between age and standardised uptake measurement (absolute change per year) reported in the 2017 meta-analysis was -0.032 [200], which was stated to correspond to a -8.9% change per decade. This is an average over all decades so the percentage change per decade would be smaller for younger adults and greater for older adults, giving values much greater than our small estimates. We note that the 2007 study by Koch *et al.* concluding that the decline in dopamine transporter availability tails off in older age [270] was not included in the meta-analysis. This study reported an average decline of 9.5% per decade under 48 years of age and 6.3 % on average over 48 years. This is evidence against a linear model, where the percentage effect would be expected to increase in magnitude with age. As was the case with the meta-analysis [200], Koch *et al.* [270] also reported similar R^2 values for logarithmic and linear regression, suggesting that an exponential relationship is equally plausible. Yamamoto *et al.* [276] studied 30 healthy controls aged 50 to 86 years and reported a linear relationship between SBR and age (parameters not specified). However, they found that the percentage change in SBR per decade was lowest (2.4%) for those over 70 years, which is not in keeping with a linear model.

We cannot exclude a small decrease in SBR with age, or indeed an increase, based on our dataset of 52 controls. However we note that the databases used by the manufacturers included similar numbers of participants over 60 years of age to our study (68 PPMI participants for DaTQUANT and around 55 ENC-DAT participants for BRASS). To our knowledge, no publications reporting a decrease in SBR with age have demonstrated that a statistically significant decrease also exists for older adults (>60 years), including the meta-analysis. A strength of our study lies in the focus on the age group most relevant to clinical practice.

Reviewing DaTQUANT database participants aged over 60 years within the software shows results similar to our controls, with no clear continuing decline in SBR in this age group. Visual review of published ENC-DAT control data also shows no definite decline in SBR with age beyond 60 years of age.

However, the largest published dataset is the Japanese multicentre database [269], which does appear to show a continuing decline in SBR in older age. The SBR calculation method used in this study was the Southampton method [277], which unlike the commercial software does not attempt automated registration to a template. The Southampton method places large VOIs around each striatum to reduce partial volume effects and calculates SBRs based on an assumed striatal volume which is either kept the same for all patients or fixed values assigned depending on gender [144]. As the authors discussed, any loss of striatal volume with age will cause SBR to be underestimated with this method, possibly exaggerating the rate of decline with age [269]. Although it is possible to calculate SBR values for sub-striatal regions such as the caudate and putamen separately, we chose to consider only the whole striatum for this study. Since the caudate and putamen can usually not be fully distinguished on an FP-CIT scan, both BRASS and DaTQUANT have to use a somewhat arbitrary dividing line between them. In our experience, even when the putamen shows clearly reduced uptake visually it may have a normal SBR, due to counts from the caudate nucleus. In addition, with smaller regions, uncertainty due to lower counts is greater. The limitation of this is that the whole striatum analysis will be less sensitive to any losses occurring only in individual regions. It is known that putaminal uptake is often affected first in Lewy body disease and Nobili *et al.* [142] showed that for the ENC-DAT normal database the putamen to caudate ratio dropped with age. However, their generalised linear model analysis showed no statistically significant difference between regions, so we conclude that any such effect is small and unlikely to explain our findings.

As with all such studies recruiting healthy controls, it is possible that the sample recruited is misrepresentative of the general population. This issue is likely to increase with age as a smaller proportion of older people may be able and willing to undergo the travel, tests and scans required to participate. We do not know whether older adults with above average general health are likely to have above average dopaminergic function but we cannot exclude this as a source of bias both for our study and the databases used in BRASS and DaTQUANT. Although we achieved a good gender balance within our control group, consisting of 44% women and 56% men, this was not the case for the MCI and dementia groups, predominantly due to the higher prevalence of Lewy body disease in males.

For the oldest adults (over 80 years of age) a drop in BRASS Z score of around one standard deviation is seen with age correction switched off. In our DLB dataset this would have changed the interpretation of the result from normal to abnormal in one case (from $Z = -1.42$ with age correction to -2.27 without). In our MCI-LB dataset there are more borderline results and interpretation would have changed from normal to abnormal in three of the 20 cases. However, as our AD and MCI-AD Z scores are slightly lower than expected with age correction applied, removing the age correction shifts 2 AD and three MCI-AD cases below Z scores of -2 . We note that turning age correction off in BRASS means that images are compared against the whole database, (mean age of 52), which exaggerates the effect

somewhat as it is clear that an age-dependency exists in younger adults. Ideally, we would compare our data against a subset of older database subjects only. Unfortunately it is not possible to recalculate a mean SBR using only the older ENC-DAT subjects as this data is not published or available within the software – only the mean and standard deviation values for the whole dataset are used within BRASS. It is also not possible to adjust the DaTQUANT database to include only subjects over 60 as the original scans are required for creating a new database. At present, the only way of comparing SBRs against an older database without age correction is to set up a local control database.

9.1.5 Conclusion

We found that the rate of change in SBR with age for control subjects over 60 was not significantly different from zero, suggesting that any decline in SBR with age in healthy older adults is likely to be lower than previously thought. Both BRASS and DaTQUANT apply slopes that are greater than our estimates, leading to potentially large differences between the software age-adjusted SBR and individual values for the oldest patients. With BRASS, Z scores are around 1SD lower with age correction switched off for subjects over 80 years of age. Use of current age correction methods in older people being assessed for DLB and MCI-LB could contribute to false negative reporting and misdiagnosis and therefore caution is suggested in clinical practice and research applications.

9.2 Does cardiac ¹²³I-MIBG uptake decrease with age in healthy older adults?

9.2.1 Introduction

So far in this thesis we have assumed that cardiac uptake, as measured by either planar or SPECT methods, does not depend on age. This was the conclusion of a relatively large study of 94 healthy controls aged between 29 and 82 enrolled in the US ADMIRE-HF study [181] but other publications in the literature are conflicting. The ADMIRE-HF study, published by Jacobson et al., used both planar and SPECT HMR measurements and found no statistically significant relationship between age and HMR for any of the methods [181]. Prior to this, a smaller Spanish study of 38 patients aged between 16 to 75 and enrolled in a cancer drug trial had reported a significant decrease in planar HMR with age ($r = -0.6264$; $p < 0.001$) [278]. However as Jacobson et al. pointed out [181], a relatively small number of patients were used to span a wide age range, with only 6 patients aged between 40 and 60 and 10 aged over 60. Tsuchimochi found no correlation between HMR and age, but included only 29 subjects covering an age range of 21 to 79 years [279]. Sakata et al. studied 300 adults in Japan with normal coronary angiography results, aged between 40 and 79 years [280]. They found a significant increase in MIBG washout rate with age on SPECT images, but unfortunately did not perform planar imaging or calculate SPECT HMR. Unfortunately it is therefore impossible to compare their results, which suggest a change with age, with those of the ADMIRE-HF study, which do not. Rengo et al. used multivariate analysis in a group of 180 middle aged and elderly heart failure patients in Italy to show that age was an independent predictor of cardiac MIBG uptake [281], with lower uptake in older patients.

The most recent evidence on cardiac MIBG and ageing comes from the Japanese Society of Nuclear Medicine working group normal database and is presented by Nakajima et al. [221]. The database consists of 62 patients aged between 20 and 84 years, mean 57 ± 19 years. A small, yet significant effect of ageing was seen for delayed HMR (a decline of approximately 0.2 in units of HMR between the ages of 50 and 80). Interestingly the decrease was only significant after the data, acquired with a range of collimators, were corrected to give results consistent with medium energy collimators. The subjects included are patients with normal scans, rather than healthy controls and only those with a low likelihood of cardiac disease were included.

Interestingly, in the ADMIRE-HR control study it was noted that activity-normalised cardiac uptake measurements (counts per pixel/voxel per MBq administered) actually increased with age, but so did mediastinal counts, producing no significant effect overall. This was the case for both planar and SPECT imaging, with both cardiac and mediastinal regions showing a statistically significant correlation. With SPECT the HMR also appeared to increase slightly with age but this wasn't statistically significant. This study did not include CT data for scatter and attenuation correction for the SPECT images and used

LEHR collimators [181], which may be unsuitable for measuring absolute uptake in torso studies, although a point spread function based correction for septal penetration was made.

As both ACSC and non-corrected SPECT data can be reconstructed for SUPeRB study participants, including both relative and absolute quantification results, the SUPeRB control dataset is suitable for assessing these effects further. The use of medium energy collimators means that the results should be less dependent on source to detector distance than those of Jacobson et al. Although there are only 31 subjects, the age range of interest (60 years +) is relatively narrow. In addition, the entire dataset of SUPeRB subjects can be used to assess any variation in mediastinal region uptake with age, which will of course affect HMR. If cardiac uptake does decrease with age then correcting for this may give improved accuracy in diagnosis of DLB. The aim of this study was therefore to measure the change in cardiac uptake with age in cognitively normal control subjects.

9.2.2 Methods

Subjects and imaging

Cognitively normal controls from the SUPeRB study, with normal visual planar imaging appearances are included in this section; 2 of the 31 controls were excluded. One further control was excluded as the administered activity was not recorded, so 28 individuals were included in total for cardiac uptake assessment. We obtained HMR from early and delayed planar imaging as described in chapter 7, SPECT and SPECT-CT HMRs as detailed in section 8.1 and 8.2 and normalised cardiac and mediastinal count density as described in section 8.3. The 6 cm diameter spherical VOI results were used for the SPECT images, rather than the threshold results as the two were shown to give similar results in section 8.2. Lean body mass normalisation was used for the absolute quantification results as we showed this to be more effective than weight normalization in section 8.3.

In addition, we studied the mediastinum VOI uptake on SPECT imaging in all 83 SUPeRB subjects (MCI plus controls), since we assume that the mediastinum uptake is not affected by the Lewy body disease process and this allows a larger sample to be studied. The absolute quantification results from section 8.3 were used to assess the relationship between mediastinal uptake and age.

Analysis

A similar method to that used for striatal uptake in the previous section was adopted. Linear mixed effects models [275] (IBM SPSS v.24) were used to test for dependence of the different cardiac and mediastinal uptake measures on various covariates. Age, gender, BMI, weight and lean body mass were input as fixed effects in the model. Age was combined with weight, BMI and LBM to check for possible interactions. With planar HMR, the mixed effects approach allowed with study ID to be input as a random effect so that data from both early and delayed planar imaging could be grouped to give

56 HMR measurements; the random effect of study ID accounts for the fact that the early and delayed planar HMRs of an individual are likely to be correlated. P-values of less than 0.05 were taken to indicate effects significantly different from zero. Scatter plots of uptake against age were produced for early HMR, delayed HMR, SPECT HMR, SPECT-CT HMR in order to visualise the relationships.

9.2.3 Results

Plots of cardiac uptake against age are shown in Figure 102 for early and delayed planar HMR, SPECT HMR, SPECT-CT HMR and normalised cardiac count density measured on SPECT and SPECT-CT images. All except early HMR show a trend towards lower cardiac uptake with age, but this is only statistically significant for SPECT-CT HMR. The results are presented in more detail below.

Planar imaging

The results of the multivariate analysis are given in Table 76 for planar (early and delayed combined with mixed effect model). The planar HMR results show weak evidence of relationships with both age and measures of patient size (p values between 0.07 and 0.11), however with age and size variables combined the relationship becomes significant, with a p-value for LBM*age of 0.09, suggesting an interaction between these parameters. Gender is important for both planar cardiac and planar mediastinum counts, but appears to completely cancel out with planar HMR. Only the data for early and delayed planar HMR combined is shown, but delayed planar HMR alone did not show any statistically significant covariates.

Table 76: HMR and count density results of linear mixed effects model for planar images in normal controls (early and delayed data combined in model). P values for covariates with a significant effect on uptake are shown in bold italics.

	Planar HMR (early and delayed combined)		Planar cardiac count density (activity normalised)		Planar mediastinal count density (activity normalised)	
	Wald Chi- Square	P value	Wald Chi- Square	P value	Wald Chi- Square	P value
Weight	3.242	.072	8.614	.003	3.109	.078
BMI	2.587	.108	7.145	.008	2.338	.126
LBM	3.227	.072	10.015	.002	4.274	.039
Gender	.330	.566	4.904	.027	9.277	.002
Age	2.567	.109	.294	.588	1.041	.308
Time point	.684	.408	.025	.873	.157	.692
LBM * Age	6.792	.009	5.571	.018	.208	.648
BMI * Age	3.197	.074	4.387	.036	.447	.504
Weight * Age	5.156	.023	6.241	.012	.733	.392

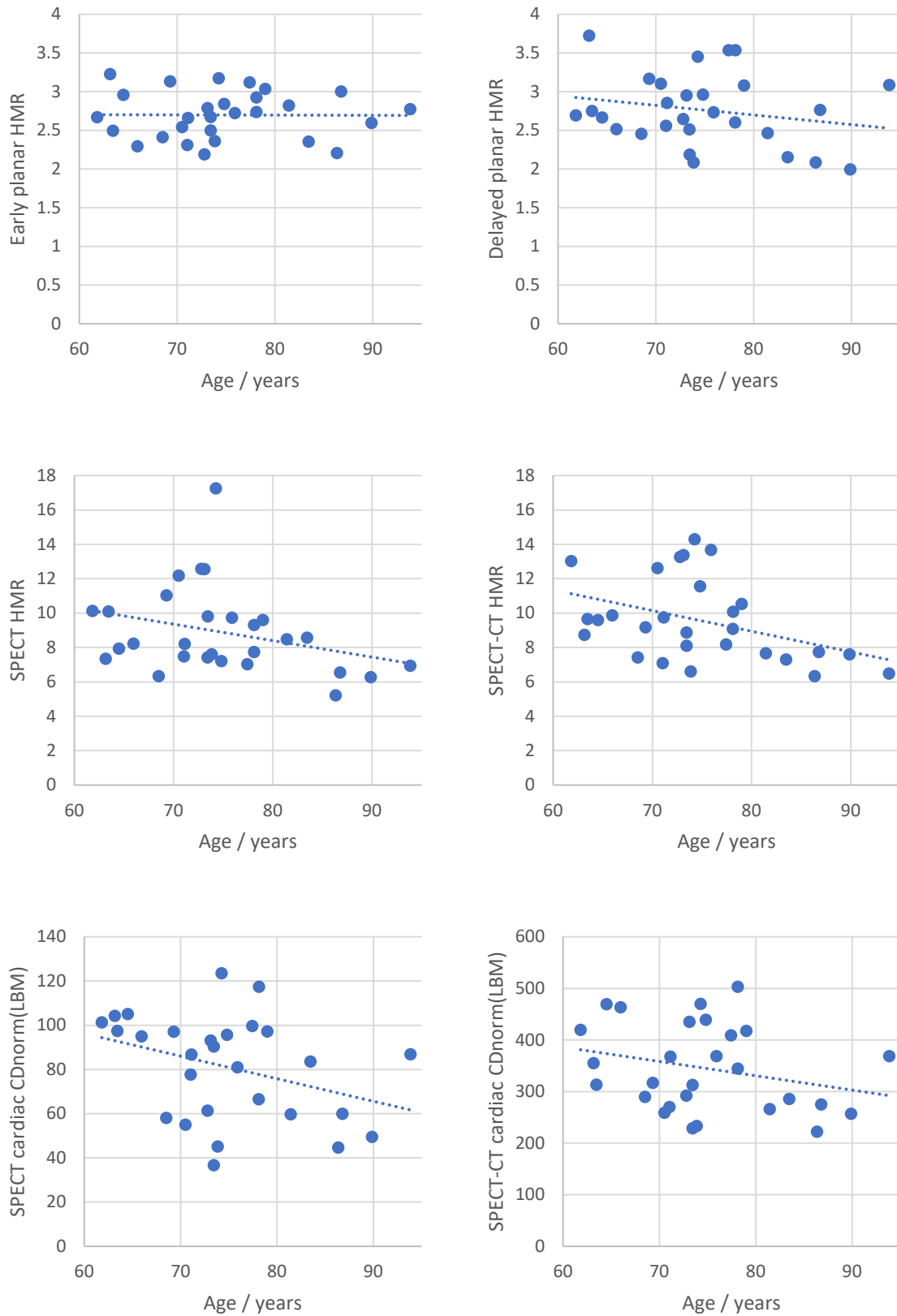


Figure 102: Cardiac uptake in cognitively normal controls plotted against age for early and delayed planar HMR, SPECT HMR, SPECT-CT HMR and normalised cardiac count density measured SPECT and SPECT-CT images

SPECT and SPECT-CT

SPECT imaging without attenuation and scatter correction results show no significant covariates for either HMR or normalised count density (Table 77). For SPECT-CT HMR the only significant variable is age (p=0.04) with all other covariates showing p values >0.5. SPECT-CT cardiac count density is not dependent on age (p=0.26) indicating that it is variation in the mediastinum counts which have a major contribution to the HMR age relationship. SPECT-CT HMR is related to patient size, even after LBM normalisation.

Table 77: Results of linear mixed effects model for SPECT image uptake measurements in controls. P values for covariates with a significant effect on uptake are shown in bold italic font.

	SPECT HMR		SPECT-CT HMR		SPECT cardiac count density		SPECT-CT cardiac count density	
	X ²	P	X ²	P	X ²	P	X ²	P
Weight	.206	.320	.044	.704	.883	.348	7.390	.007
BMI	.608	.650	.050	.834	.954	.329	5.484	.019
LBM	.254	.435	.366	.823	.518	.472	5.519	.019
Gender	.004	.614	4.359	.545	.349	.555	.468	.494
Age	1.242	.949	.193	.037	.349	.555	1.254	.263
BMI * Age	.498	.265	.185	.661	1.034	.309	5.908	.015
Weight * Age	.171	.480	.335	.667	1.558	.212	9.520	.002
LBM * Age	.273	.679	.357	.563	1.521	.217	9.689	.002

Mediastinum uptake (plotted for both MCI and control subjects in Figure 103) appears on SPECT and SPECT-CT images appears to show a slight increase with age even with activity and LBM normalization, but there is substantial variation between subjects and the effect does not come close to being statistically significant. Plots of LBM normalised SPECT and SPECT-CT mediastinum uptake are shown in Figure 103. None of the covariates included in the linear mixed effects model had a significant effect on mediastinal uptake (Table 78).

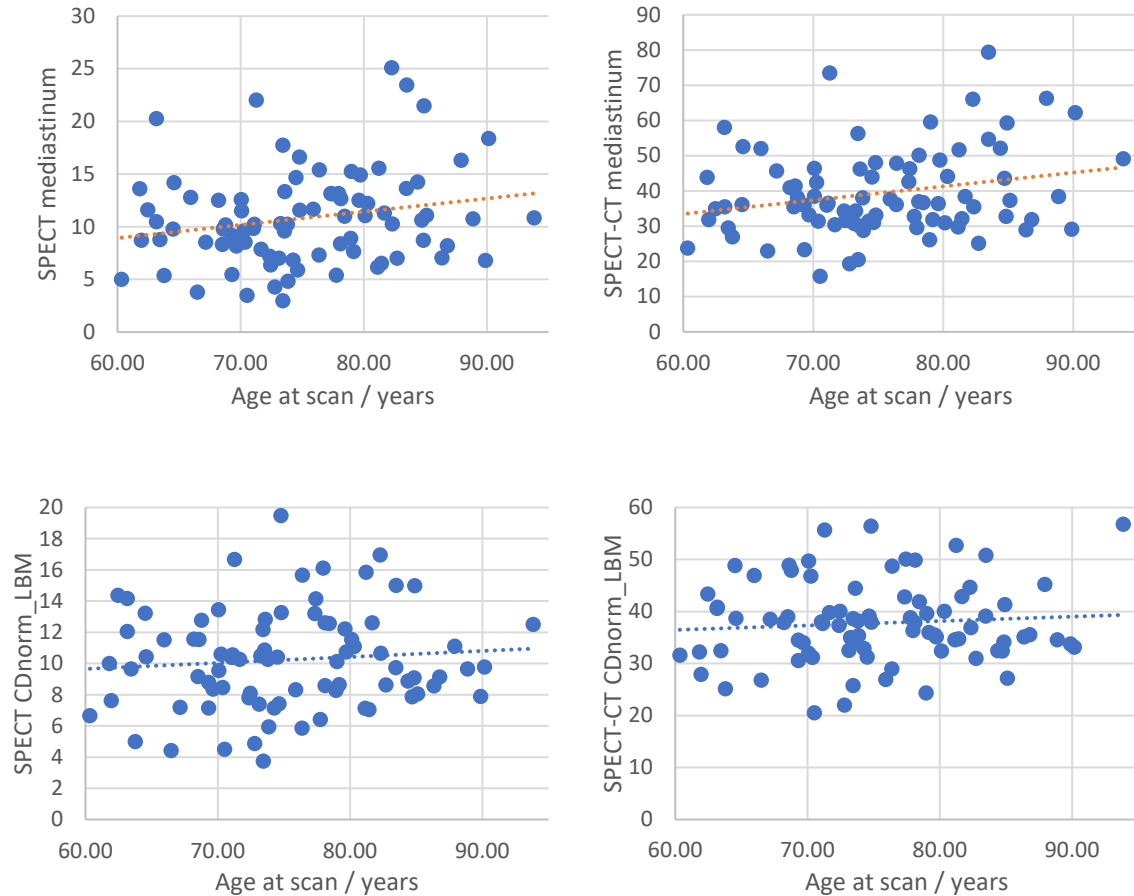


Figure 103: Top: activity-normalised mediastinal count density for all subjects on SPECT and SPECT-CT images. Bottom: activity and lean body mass normalised mediastinal count density on SPECT and SPECT-CT images.

Table 78: Results of linear mixed effects model for mediastinum uptake, showing no variation with age or any of the covariates tested.

	SPECT mediastinal count density (activity normalised)		SPECT-CT mediastinal count density (activity normalised)		SPECT mediastinal count density (activity and LBM normalised)		SPECT –CT mediastinal count density (activity and LBM normalised)	
	Chi- Square	P value	Chi- Square	P value	Chi- Square	P value	Chi- Square	P value
Age	.615	.433	.001	.972	.179	.672	.077	.782
Weight	.001	.975	.560	.454	.002	.964	.278	.598
BMI	.011	.916	1.050	.306	.031	.861	.347	.556
Gender	.085	.770	.016	.900	.828	.363	.154	.694
LBM	.000	.998	.322	.571	.003	.956	.188	.664
Age*Weight	.023	.879	.248	.618	.005	.944	.225	.635
Age*BMI	.062	.804	.471	.493	.096	.757	.160	.689
Age*LBM	.105	.746	.018	.894	.058	.810	.267	.605

9.2.4 Discussion

Our planar results are in keeping with the findings of the ADMIRE-HF study, with both our dataset and the ADMIRE-HF controls showing no evidence of a relationship between early HMR and age, and both showing a suggestion of a decline in delayed planar HMR with age, which is not statistically significant. It is possible that with a larger sample size or wider age range we would have seen a significant effect, especially given the positive result with SPECT-CT HMR and the small significant decline in delayed HMR with age seen in the Japanese database, also with medium energy collimators [221].

In our planar HMR dataset, the combination of age with LBM had a significant effect on HMR, even though these covariates separately did not show strong evidence for a relationship (

Table 76). This suggests that there is an interaction between LBM and age, although we note that there is not an obvious linear correlation between them.

Our tomographic results are quite different from those in the ADMIRE-HF study, where a statistically significant increase in SPECT cardiac and mediastinal count density was seen with age and a trend towards higher SPECT HMR [181]. Our results show a trend towards lower SPECT HMR with age, which becomes statistically significant with attenuation and scatter correction applied. The SPECT mediastinal count density without LBM correction, which is the value most similar to that measured by Jacobson et al., does show a suggestion of increasing uptake with age, but this is removed when corrected for patient size with LBM normalisation. The cardiac count density does appear to decrease with age, but the slope is not significant. However, the slope of SPECT-CT HMR against age is significant, even though there is no corresponding increase in mediastinum count density with age to increase the effect. This suggests that the relationship between cardiac count density and age may be masked by inter-subject variation, and taking the ratio to obtain HMR may cancel out some of this variation between patients. This would also support the use of SPECT-CT HMR over absolute quantification for cardiac MIBG imaging.

On delayed planar imaging and non-corrected SPECT imaging, HMR does appear to decline with age, although this is not statistically significant. Our results with SPECT-CT indicate that this is a real effect and that the planar relationship is likely masked by variation between subjects due to other factors such as physiological variation and patient size. The effect of ageing on delayed planar HMR appears to account for a non-trivial amount of the variation seen between subjects, an average decrease of around 0.4 HMR points in the range 60-90 years, compared to a total variation of around ± 1 HMR point ($SD=0.46$). Accounting for age has the potential to improve the diagnosis of Lewy body disease, by reducing the variation between normal subjects. A larger sample of normal SPECT-CT scans would be required to characterize the relationship between HMR and age more comprehensively, which could be obtained from the MCI-AD dataset after follow up to confirm the subjects are unlikely to have Lewy

body disease. This would be similar to our approach in the previous section on FP-CIT uptake and ageing, where MCI-AD cases from the LewyPro study with several years clinical follow up were used as controls.

9.2.5 Conclusion

In cognitively normal controls, our results show a significant decline in HMR with age on attenuation and scatter corrected SPECT-CT images. This suggests that correcting for age could be important for improving the diagnosis of Lewy body disease, including on delayed planar images. Further work with a larger dataset is indicated to characterize the relationship and the correction required.

9.3 Summary of chapter 9

In section 9.1 we showed that visually there appeared to be no decline in striatal SBR with age in people over 60 years of age without Lewy body disease. On inspection of the control data used by the software manufacturers there also appeared to be no clear drop with age in this age group. Our mixed methods analysis showed no dependence on age. Although these results are convincing, our sample size was too small to be able to exclude a small decrease, or indeed increase, in uptake with age. Our 95% confidence intervals overlapped with the slopes used by BRASS and DaTQUANT even though our estimated slope was smaller. We speculate that an exponential relationship between striatal uptake and age may be a better model of the biological effect, as this would account for both the large effects seen in young adults and the negligible effect in old age. Further work with larger samples over a wider age range would be necessary to fully characterise the relationship between striatal uptake and age.

In section 9.2 we showed that unlike striatal FP-CIT uptake, HMR results do appear to decrease in old age, with the results obtained on scatter and attenuation corrected SPECT-CT images of cognitively normal older adults showing a statistically significant inverse linear relationship with age. One possible reason for the discrepancy between cardiac sympathetic innervation and dopaminergic innervation in older adults may be the interaction with cardiac disease, which would of course not affect the striata. The majority of our controls did not have a history of cardiac problems, but the risk of asymptomatic cardiac disease increases with age and it is possible that this could account for the observed decline in cardiac MIBG uptake. Nevertheless, the results suggest that, on average, a correction for age could reduce the variation in cardiac MIBG uptake between subjects – regardless of the physiological reasons for the variation.

The strengths of the work presented in this chapter include the recruitment of elderly subjects with normal cognition, representing the ages of people typically presenting for memory and cognition assessment in the NHS. This makes the results more relevant to clinical practice than studies covering the whole adult age range, where typically few adults over the age of 70 are included. A further strength is the prospective recruitment of volunteers, rather than the retrospective selection of patients with normal imaging from clinical databases. The main limitation is the sample sizes, which are rather too small to definitively confirm the relationships in question.

Chapter 10: Discussion and conclusions

This thesis focussed on the optimisation and quantification of two scintigraphic biomarkers for the diagnosis of DLB: ^{123}I -FP-CIT (DaTSCAN™) and cardiac ^{123}I -MIBG. Both scans have technical limitations that suggest that, whilst very accurate in distinguishing between Lewy body disease and other conditions at the dementia stage, their performance may not be ideal for detecting early neurodegenerative change with current methods used in clinical practice. In this thesis these issues were addressed using a range of novel state-of-the-art approaches. Throughout the thesis these novel techniques were compared with the standard techniques to evaluate their potential benefit. The application of both biomarkers was studied in advanced anthropomorphic phantom models, cohorts of highly characterised dementia and MCI subjects and older adults with normal cognition recruited as age matched controls.

This chapter presents an overview of the main findings presented in the thesis and the novel contributions to the field. The use of both biomarkers in clinical practice and research is then discussed in light of the findings and the limitations of the work and future directions are summarised.

10.1 Overview of main findings and novel contributions

10.1.1 FP-CIT summary

Software comparison

We showed that ^{123}I -FP-CIT specific binding ratios (SBRs) calculated by BRASS and DaTQUANT software packages are systematically different, although give similar diagnostic accuracy. These have not been compared directly in the literature. The SBRs are affected by the reconstruction method used and by factors such as image registration that affect the SBR calculation. The associated Z scores output by the software also depend on the normal database used for comparison and the method used for age-correction. Quantification results obtained with different methods are therefore not directly comparable and a consistent reconstruction method and software package should be used. For the SUPeRB study the BRASS software package was chosen based on technical advantages over DaTQUANT.

Factors affecting BRASS quantification

Phantom experiments and reprocessing of scans from the LewyPro study showed that pixel size, repeat scanning and reorientation during processing had small random effects on BRASS quantification of less than 10%.

Visual rating and BRASS quantification

The reconstruction method required for comparison with the BRASS normal database, denoted RRACSC_Uniform in this thesis, was adopted for both the standard quantification method and for visual assessment. The visual rating of FP-CIT scans by a consensus panel of raters was compared to BRASS quantification using the standard RRACSC_Uniform reconstruction method applied in SUPeRB, selecting only probable MCI-AD and MCI-LB cases for analysis. The sensitivity of visual assessment of FP-CIT scans by raters with access to BRASS quantification was 48% (95% CI: 29 to 68%) and specificity 88% (62 to 98%). The sensitivity and specificity of BRASS quantification alone was 44% (25 to 65%) and 100% (79 to 100%). There were three cases where there was discordance between the panel and BRASS quantification but the overall accuracy of the visual ratings method and quantification alone was very similar. Long term follow up, including repeat FP-CIT scans, is ongoing and will help to clarify the status of the considerable number of cases assumed to be early Lewy body disease but with normal FP-CIT scans, as well as that of the cases where there was discordance between the panel and BRASS.

Advanced FP-CIT reconstruction methods

Standard BRASS SBR quantification results were compared with absolute quantification with SUV-SPECT. Preliminary results suggest that SUV-SPECT using either the mean striatal SUV or the SUV_{peak}

separates normal and abnormal subjects less well than semi-quantification, with more variation seen among normal subjects. We recommend that SBRs continue to be used for quantification rather than SUV. This is in line with the results of recent publications on this topic from Japan [196, 197]. However, the scaling of the images to the absolute value rather than relative to the maximum voxel for visual rating could be a more robust way of detecting subtle balanced loss and further investigation is warranted as this has not been investigated in the literature.

It was found that advanced FP-CIT reconstruction methods using SPECT-CT and collimator modelling affected specific binding ratio values, but the overall interpretation of the scans with respect to the SBRs of the age-matched SUPeRb control database reconstructed with the same method was not affected. It was therefore not recommended to use SPECT-CT for calculating SBRs in FP-CIT in clinical practice. The lack of improved diagnostic accuracy with CT attenuation and scatter correction is in keeping with the literature on this topic [102, 162, 198], but there have been no published studies applying full collimator modelling to FP-CIT.

Test-retest variation

The repeatability of visual rating and BRASS quantification was assessed using baseline and follow up scans in cognitively normal controls. Two of the 20 controls with repeat visual ratings were assessed as abnormal at baseline but normal at one year follow up and overall concordance was 75% for our modified and extended Benamer ratings scale and 90% for binary abnormal-normal categorisation. SBR results on several of the baseline and repeat scans differed by more than 20% and the mean of the absolute test-retest difference was 12%, similar to the coefficient of variation among subjects at baseline and in keeping with previous work at Newcastle [204], but higher than early work on FP-CIT repeatability [201]. The causes of the intra and inter-subject variation are unknown, but the results suggest that a drop in SBR of less than around 20% in repeat scans should not be regarded as significant.

Striatal uptake of FP-CIT in healthy ageing

The scan data for controls with normal cognition was used to explore ageing effects. Multivariate analysis using linear mixed effects models was performed to determine whether there is an association between uptake and ageing between 60 and 90 years of age. We found that SBR does not continue to decrease with age in older adults, as was previously assumed to be the case, indicating potential misdiagnosis if linear age correction is applied. This novel finding was published in the *Annals of Nuclear Medicine* journal [267] and presented at international conferences. Linear age corrections are currently applied in BRASS and DaTQUANT, although a meta-analysis found that an exponential model fits the data equally well [200].

In clinical practice this potential problem is difficult to avoid as it requires the use of a local age-matched database for older adults rather than the default age-corrected software databases. Switching off age correction is not possible with some commercial software packages, and would be inappropriate for a database made up of a wide range of ages as there is clear evidence of a relationship between SBR and age in younger adults. As turning off age correction in commercial software is unhelpful and determining a local age-matched normal range is not a practical solution for most clinical centres, we investigated the practical clinical implications of using quantification software with the default age correction applied. To do this we compared the accuracy of BRASS quantification in the SUPeR study using our local control database with the accuracy of BRASS using the ENC-DAT database with age correction. There was no change in clinical interpretation of the scans, indicating that the linear age correction method used in BRASS, did not lead to misdiagnosis in this dataset. It is likely therefore that the effect of age correction is not large. Nevertheless, possible overcorrection in subjects over 80 years should be borne in mind, particularly in borderline cases. In research studies aiming to identify subtle changes, it may be more important to use a more appropriate model.

10.1.2 Cardiac MIBG summary

Preliminary work

Anthropomorphic phantom scans established that, as expected from the literature, medium energy collimators give better image contrast than low energy due to reduced septal penetration of high energy gamma rays and confirmed their use for the SUPeR study. A method for normalising HMR results obtained with different acquisition settings using simple phantom scans was presented. We found that an administered activity of 111 MBq with planar scan duration of ten minutes (the protocol used in Japan [185]) provides adequate counting statistics for heart to mediastinum ratio (HMR) calculations to be accurate to within 2%, suggesting that the recommendation of 370 MBq in UK ARSAC guidelines is excessive. Phantom scans showed that large changes in count density due to increasing scan time or administered MIBG activity did not affect HMR.

Processing methods for planar HMR calculation

Using data from the MIDAS study we found that different methods of planar MIBG analysis are operator dependent and give systematically different results. We also explored the use of visual rating and found that semi-quantification gave higher specificity, provided an appropriate normal HMR cut-off is used. We recommended the use of a 6 cm circular cardiac ROI for calculating HMR as this method reduced inter-operator variation when compared to a region tailored to the patient, without changing accuracy and is in line with the method used in the Japanese literature [173, 219]. These results have been published in Nuclear Medicine Communications [190] and disseminated to the local nuclear medicine community at regional meetings.

Normal HMR cut-off values

We found that previously published HMR cut-off values in Japan and the USA may not be valid in a UK population and proposed suitable alternatives based on the SUPeRB control dataset for both early and delayed images. These cut-offs are 2.10 for early imaging and 1.86 for delayed imaging, suitable for use with medium energy collimators and Siemens cameras only.

This work was accepted for publication in the Journal of Nuclear Cardiology in August 2019 and will be published with an accompanying editorial piece by Prof Nakajima of Kanazawa University, Japan with whom we collaborated to compare Japanese and UK data directly. Previous publications have addressed the need to account for technical factors in HMR thresholds [165, 167, 168] – this work shows that population differences also affect interpretation.

Early vs delayed imaging

We acquired early and delayed images in both the MIDAS and SUPeRB studies and found that HMR values obtained on the standard delayed images gave marginally superior accuracy to early images. Greater separation in uptake between normal and abnormal scans on delayed images both in the SUPeRB and MIDAS studies was found, in keeping with that reported in the literature [235]. In research, there are benefits in the maintaining the standard delayed method.

In our series 78% of cases gave a clear non-equivocal result on early imaging and the early and delayed result differed in only 4% of cases. This finding suggests that in clinical practice an alternative protocol could be adopted in which only an equivocal early HMR result would lead to patients having the delayed scan. This protocol could save time in the majority of cases so would be more convenient for patients and save NHS staff and scanner time.

We brought all the planar cardiac MIBG data together to propose an accurate and practical evidence-based cardiac MIBG protocol for image acquisition and analysis. This protocol could be used in clinical practice within the NHS and in research studies requiring a binary normal/abnormal sympathetic innervation result. We suggested that patients should be administered 111 MBq and a planar anterior image of the chest acquired around 20 minutes later, using medium energy collimators. If the HMR, calculated using our 6 cm circle ROI method, is below 1.6 or above 2.5 the patient should be informed that the test is complete. If the HMR at 20 minutes is equivocal the patient should be asked to return for a repeat scan to be carried out 3-4 hours after injection.

Effect of patient size on cardiac MIBG uptake results

The effect of patient size on apparent uptake was assessed with an anthropomorphic torso phantom and using the SUPeRB controls. This was performed for both planar and SPECT acquisitions; for the latter the effect of using advanced image reconstruction methods with CT attenuation and scatter

corrections compared to conventional processing was explored. In the phantom, increasing the size reduced apparent cardiac uptake by 39% on planar images and 33% on non-corrected SPECT images but with SPECT-CT images with attenuation and scatter correction applied the reduction was only 3%. For the control subjects, planar cardiac ^{123}I -MIBG images show a significant inverse relationship between cardiac uptake ratios and subject BMI. SPECT images with CT attenuation and scatter correction applied showed no statistically significant drop in cardiac uptake with BMI, suggesting that the corrections are effective in patients. However, there was considerable inter-subject variation, which may explain why there is no statistically significant difference between the slopes of HMR against BMI for non-corrected and ACSC SPECT. This work suggests that ACSC SPECT has the potential to improve cardiac ^{123}I -MIBG quantification, particularly in borderline planar cases. This has not been explored in the literature and larger studies would be needed to confirm our findings.

This work is currently under peer review for publication in Biomedical Physics and Engineering express.

Use of SPECT-CT for quantification of cardiac MIBG uptake

The diagnostic accuracy and the separation of normal and abnormal scans into clusters was compared for planar, SPECT-CT relative quantification and SPECT-CT absolute quantification. We saw that the separation between normal and abnormal cardiac MIBG scans on planar imaging is around 1.5 HMR points, increasing to around 5 points with SPECT and around 7 points with SPECT-CT. This is a novel finding as SPECT with CT ACSC has not previously been used in the literature for cardiac MIBG imaging. However this very promising finding translates into only a small improvement in MCI diagnostic accuracy because in most cases planar imaging is sufficient to characterise the scans as normal or abnormal. Application in routine clinical practice may be limited, but results here suggest potential value in research for the measurement of serial change over time and correlation of uptake with clinical symptoms, for example.

Preliminary exploration of cardiac ^{123}I -MIBG regional uptake patterns

The uniformity pattern of MIBG uptake within the myocardium was assessed for controls and dementia patients, as a feasibility study to see if this might be useful for diagnosing Lewy body disease in patients with global uptake within the normal range. Studies in the literature had been mixed, with some suggesting that assessment of regional reductions improves diagnosis [56, 177, 179, 239, 245, 246] and some that this is not specific [188, 216, 246-251].

We found that areas of relatively reduced cardiac ^{123}I -MIBG uptake on SPECT images are very common in cognitively normal older adults and patients with dementia who have normal planar cardiac uptake. Reduced regional cardiac MIBG uptake in the inferior wall of the left ventricle is the most common defect and likely related to attenuation or liver uptake. Reductions in other areas may also be

artefactual or may be caused by mild underlying coronary artery disease. The pattern of reduction is in any case non-specific and therefore unlikely to aid in the diagnosis of Lewy body disease in typical clinical settings. This work is currently under review with the Journal of Nuclear Cardiology.

Cardiac MIBG uptake in healthy ageing

In cognitively normal controls, we saw a small but significant decline in HMR with age on attenuation and scatter corrected SPECT-CT images, of around 0.5 HMR points on SPECT-CT per decade (around 5%). On planar and non-corrected SPECT imaging HMR also appears to decline with age but this is not statistically significant. This was in contrast to FP-CIT, where no evidence of any relationship between striatal uptake and age was seen in our cohort of individuals without Lewy body disease over 60 years of age. Estimates in the literature for the relationship between FP-CIT uptake and age are around 7% per decade and these are accounted for in quantification software packages. Previous literature on the relationship between cardiac MIBG and age is mixed, with a large US study showing no relationship and the Japanese normal patient database showing slightly lower mean HMR in over 60s compared to under 60s. Neither study used SPECT-CT however. Although the effect of ageing on sympathetic innervation appears small compared to the overall variation seen between subjects, correction could reduce this variation and help to improve diagnostic accuracy.

10.2 Comparison between FP-CIT and cardiac MIBG

A common finding throughout this thesis is that many individuals with probable MCI-LB diagnosed based on clinical features have clear normal image appearances and quantification results on FP-CIT and/or cardiac MIBG scans, with the specific reconstruction and processing method making little difference to the binary interpretation of the scans. In cardiac MIBG we saw a greater separation of scan data into two distinct groups with SPECT-CT than with planar imaging or with FP-CIT, but this did not translate into greater sensitivity with respect to the clinical diagnoses. If Lewy body disease is not present in the relevant tissue the imaging technique will always have limited sensitivity, even with technical improvements. Longer term follow up of these patients and eventual autopsy will help to clarify the utility of both scans, and it will be of particular interest to see whether probable MCI-LB patients with both normal FP-CIT and MIBG scans do go on to develop DLB.

We know that Lewy body disease is found throughout the nervous system [81] and does not appear to have a consistent pattern of spread [1] so it is not surprising that some scans in probable MCI-LB cases are normal, particularly as this is early stage disease. Our results add to the body of evidence contradicting the Braak hypothesis that the progression of Lewy body pathology follows “an ascending course with little interindividual variation” [282], for example, some MCI-LB patients had normal cardiac MIBG and abnormal FP-CIT. It has been shown in an autopsy study at Newcastle that 10% of people diagnosed with DLB who meet pathologic criteria for Lewy body disease had normal FP-CIT imaging in life [48] suggesting that it is possible for the striatum to be spared even in advanced disease.

The concordance between planar MIBG HMR and visual rating of FP-CIT scans with access to quantification results was around 80%: 82% for probable MCI-LB, categorised using clinical symptoms only, 67% for MCI-AD and 88% for the controls. Two controls were rated abnormal on each of the scans, different individuals for each scan. The overall diagnostic accuracy results presented in this thesis are similar for both techniques, showing that sometimes the FP-CIT result is more in line with the clinical picture and sometimes the cardiac MIBG result. The diagnoses with respect to clinical symptom count only (0=AD, 2+=MCI-LB) allow us to compare the scans directly. The sensitivity with respect to clinical symptoms is 59% for both scans. The specificity is 81% for FP-CIT and 85% for MIBG (the difference due to just one patient with no clinical features of Lewy body disease having a normal MIBG scan and abnormal FP-CIT). Our results from SUPeRb so far suggest that around half of people with MCI-LB develop pathology affecting both dopaminergic function and sympathetic function at an early stage, around a third have no signs of significant disease in either the heart or substantia nigra and the rest have either only the dopaminergic nerves affected early on or signs of cardiac denervation only, around one in 12 (8%) for each case. The level of concordance is similar to that reported by Tiraboschi et al. who compared cardiac MIBG and FP-CIT DLB and AD in the only other prospective

study in the literature [67] and found the results to be concordant in 51 of the 59 patients (86%). The numbers are too low to come to firm conclusions, especially without autopsy data but the relatively high concordance in SUPeR suggests it would require a very large study to detect any difference in diagnostic accuracy with certainty.

There were no controls or MCI-AD cases with both biomarkers abnormal, suggesting that the use of both may be justified in situations where high specificity is required, for example in selecting people with AD *without* evidence of additional Lewy body disease. The combined sensitivity would be reduced but only from 59% to 50%, so it may be feasible to use both biomarkers to select MCI-LB cases with more certainty for clinical trials of new disease modifying therapies.

Currently lack of knowledge on how to interpret cardiac MIBG scans is a significant barrier to their use in routine clinical practice in the UK, despite the inclusion of cardiac MIBG as an alternative to FP-CIT in the 2018 NICE dementia guidelines. This could be addressed with UK wide standardisation of imaging protocols and scanner calibration (see section 10.4 below).

10.3 Strengths and limitations

Clinical diagnoses

A key strength of this work is the use of well-characterised dementia and MCI cohorts and age-matched controls with normal cognition recruited into research studies. These subjects underwent a wide range of cognitive and neuropsychology tests, more comprehensive than those carried out in a clinical setting, and their results were reviewed by a panel of expert old-age psychiatrists with particular interest and expertise in prodromal diagnosis before coming to a consensus diagnosis. MCI-LB subjects with only one core clinical feature of Lewy body disease or positive biomarker were excluded from analysis, as were MCI-AD cases with one positive biomarker⁶. We are therefore as certain as possible that the diagnoses are accurate but this can only be confirmed by future neuropathology at autopsy, included as part of the study. With the MCI studies, such autopsy confirmation is not likely to be received for several years for the majority of patients and will therefore not be fully representative of the pathology present at recruitment. A limitation is that tests for detecting Alzheimer's disease pathology, such as amyloid or tau PET scanning, were not performed so we are not able to tell on a biomarker level, whether the profile of the MCI-LB cases is likely to have a contribution from AD pathology. This however, is in keeping with the NIA-AA framework [24], in which a clinical diagnosis of AD is made via the exclusion of other causes rather than via confirmation of AD pathology.

We used the 2017 international DLB consortium criteria for DLB diagnosis [1] to categorise the MCI-LB cases into possible and probable categories, applying the same criteria at the MCI stage as used at the dementia stage, including the use of MIBG and FP-CIT results for assigning possible and probable groups. The possible MCI-LB cases were then excluded from analysis within this thesis. In the FP-CIT section MIBG was used as a biomarker and in the MIBG section FP-CIT was used. Flow charts showing the decision making process based on core clinical symptoms at one year and baseline biomarker results were given in the clinical background chapter in section 2.3.3.

There was non-concordance between cardiac MIBG and FP-CIT results in the SUPeRb study in around 20% of cases, leading to slight differences in the diagnostic groupings in the FP-CIT and cardiac MIBG sections of this thesis. An alternative approach would have been to use only the core clinical features, resulting in the same clinical groups for both the FP-CIT and MIBG work, making it easier to compare the scans. However, many MCI-LB participants had only one core clinical feature, even with one or both scans abnormal, and would have been excluded with this approach. In this thesis, only the baseline biomarkers were used for diagnosis, even though follow-up FP-CIT scans were available in the

⁶ No MCI-AD cases had both FP-CIT and MIBG abnormal but if so they would have been excluded rather than moved into the probable MCI-LB category as this requires at least one core clinical feature of Lewy body disease.

majority of cases included. This was because we have not acquired repeat cardiac MIBG scans and because baseline biomarker results are more in keeping with clinical practice, where a patient's diagnosis will be revised following changes in symptoms but it is less common to request a repeat FP-CIT scan.

Quantification work

Strengths of the work on FP-CIT and cardiac MIBG quantification presented in this thesis include the following advantages not generally available in routine clinical practice:

- Access to two of the most widely used FP-CIT quantification packages, BRASS and DaTQUANT (typically NHS departments will have one or the other or an in-house method)
- Access to an advanced FP-CIT reconstruction method not yet released by the developer for clinical use
- Visual FP-CIT results from a consensus ratings panel of five experts, rather than a single clinical reporter
- Use of ENC-DAT FP-CIT normal database (purchased for Fellowship)
- Data from controls from the SUPeRB study with normal cognition
- Use of well-characterised MCI and dementia research cases with follow up (as discussed above)
- Use of sophisticated anthropomorphic phantoms (purchased for Fellowship)
- Calibration of gamma camera for planar cardiac MIBG imaging with help from colleagues in Japan and the Netherlands. This has not been done in any other UK centre

A limitation of the work on advanced FP-CIT analysis methods was the lack of a suitable method for SUV-SPECT analysis for the FP-CIT data that would produce results that can be compared to SBR methods. If the manufacturer of BRASS (Hermes Medical Ltd) were to agree to make changes to the SBR calculation method in future releases of BRASS, such that the results could be used for absolute quantification, this would be ideal. The evidence presented in this thesis and in a limited number of small published studies suggests that SUV-SPECT is slightly less accurate than SBR, but this could not be concluded with as much certainty as we would have liked.

Time factors were a major limitation in all the work involving phantom scans. The availability of the gamma camera was limited to weekday evenings after clinical scans had been completed. Lengthy scans were required to mimic patient count rates so only a few scans could be acquired in each session, as often complex manipulations of activity distribution in the phantom were required between each scan. Weekend work was not thought to be feasible as ^{123}I can only be delivered on weekday mornings and has a half-life of 13.2 hours making it at least twice as expensive to order sufficient activity, however in hindsight this would probably have been worthwhile. Much of the phantom work would

have benefitted from access to the equipment for whole day rather than sessions of a few hours as multiple scans of the exact same activity distribution could have been achieved.

Although the sample sizes of the control and MCI groups were reasonable for an imaging study they remain too small for us to be able to detect small effects with certainty, for example the lack of decline in FP-CIT uptake with age in older adults and the relationship between SPECT-CT cardiac uptake and patient size. Many of the results in this thesis should therefore be regarded as preliminary findings, requiring larger multi-centre studies for confirmation.

10.4 Future directions

10.4.1 Ongoing work with SUPeRB study

As the SUPeRB study is ongoing, with recruitment only finishing in August 2019, only a subset of the data was considered in this thesis (59 MCI subjects recruited up until January 2018). This was so that one year follow up diagnoses could be used to determine the clinical groups to be used in the thesis and to allow time for analysis. I will be starting a part time post-doctoral position at Newcastle University following the fellowship to enable some of the work presented here to be extended to the full SUPeRB dataset of around 100 MCI subjects. We plan to publish much of the work presented in this thesis after re-analysis with the complete dataset.

10.4.2 UK wide calibration of gamma cameras for cardiac MIBG imaging

The results of the MIDAS pilot study into the use of cardiac MIBG imaging for dementia diagnosis in a UK setting were encouraging and we have been approached several times by interested clinicians and clinical scientists in other UK centres for advice on implementing a cardiac MIBG service. The results in section 7.2 on HMR thresholds suggest that for other UK centres would be better off applying our results than other values in the literature from other countries; however to do so would require calibration of the gamma cameras, using similar methods as those presented in section 5.2. We are therefore planning to apply for funding for a UK wide calibration project. In this project we would manufacture similar planar phantoms in house, constructed to give known activity concentration ratios and take or send them to different centres. We would use the results to convert our results to HMR cut-off values appropriate for other centres to use and develop national guidelines.

10.4.3 Simulation work using Monte Carlo modelling

Simulating scans using Monte Carlo methods enables investigation of the effects of patient factors in a more flexible and controlled manner than can be obtained with physical anthropomorphic phantoms, overcoming difficulties of gold standards with patient studies. In addition to the large amount of work presented here I also carried out some preliminary work on developing a Monte Carlo model of our gamma camera (see Appendix G) which will form the basis for future work.

10.4.4 Multicentre clinical study

We plan to apply for funding in future to conduct a large multi-centre study into the use of FP-CIT and cardiac MIBG in the diagnosis of Lewy body disease in the UK. This will enable our findings to be replicated in different settings and will increase the power of the study, clarifying some of the findings in this thesis that appear promising but are hampered by small clinical groups.

10.5 Concluding remarks

This thesis presented an in-depth and comprehensive study of technical aspects of nuclear medicine biomarker quantification using ^{123}I labelled radiopharmaceuticals for the diagnosis of Lewy body disease. This provides a solid foundation for clinical and research application of these techniques in both early and established disease. Proposals are made for further work to advance these studies via a larger multicentre study and phantom calibration.

Appendix A: Key technical and practical differences between DaTQUANT and BRASS

	DaTQUANT	BRASS	
Technical considerations	Image reconstruction	Accepts raw and transaxial data - but internal reconstruction of raw data recommended to ensure consistency with normal database. Iterative reconstruction (OSEM) or FBP with or without uniform attenuation correction available. Scatter correction and CTAC not supported.	Accepts reconstructed transaxial data only, which must have been reconstructed within Hermes using recommended settings (OSEM with resolution recovery, scatter and attenuation correction) for normal database comparison to be valid.
	Normal database	116 healthy control studies from 20 centres scanned as part of Parkinson's Progression Markers Initiative. 31 to 84 years old (mean 60.5).	103 healthy control studies from European Association of Nuclear Medicine ENC-DAT database (10 different centres, mean age 51 years)
	Image registration	Deformable registration of subject image to fit normal template	Deformable registration of <i>normal template image to subject image</i> . Two hemispheres registered separately.
	Adjustment of regions	Whole striatum VOIs can be moved individually but sub-regions (caudate etc...) cannot be adjusted. The occipital background region cannot be adjusted.	The entire subject image can be shifted, rotated or scaled to fit the template. The template is fixed so no individual regions can be adjusted.
	Camera correction	Not used	SBR results can be adjusted for differences between cameras. Calibration scans using a striatal phantom are required.
	Z scores and age correction	Calculated with respect to SBR interpolated from normal controls. Non-linear fit used for interpolation. $Z = (SBR - \text{control mean for age})/SD$ Same SD used for all values of SBR.	Calculated as a linear function of age. $((SBR + ((Age - 51.5631) * 0.0138)) - 3.3334) / 0.5004$
Practical considerations	Saving results	Results can be saved as pdf and added to a .csv file containing all studies processed – however .csv file does not store Z scores	Results can be saved within Hermes database and exported as an individual .csv file. Not possible to process in batches or save multiple results to a file
	Display of results	Subject SBRs are displayed with age-matched control means and SDs and Z scores. A plot of normal control SBRs against age is displayed with the subject result highlighted.	Subject SBRs are displayed with Z scores. No plot.
	Access to normal database results	No access to PPMI scans Results are available indirectly by reading results from plots of SBR vs age.	No access to any control data
	Local normal databases	Easy to create local databases and swap between them	Possible to create and implement local databases with support from Hermes
	Published work	None as of September 2016	Development work published and used in several published studies
	Cost and accessibility	Much less expensive than BRASS (<£1000 for single computer licence) and available as a standalone package for any PC.	Only available on Hermes workstations. Much more expensive than DaTQUANT (>£10,000). However can be made available on multiple workstations.

Appendix B: Factors affecting FP-CIT Z scores

This appendix details possible reasons for the differences seen between BRASS and DaTQUANT Z scores in Chapter 4.

Camera correction

The DaTQUANT PPMI normal database SBR values were used to determine whether there is a significant difference between controls acquired with different cameras and therefore whether lack of calibration might explain the difference between our local normal patients and the PPMI controls. The SBR data for PPMI controls is not freely available in the literature or from GE Healthcare but was instead obtained by recording SBR results from the normal database plot within DaTQUANT. The camera used to scan each PPMI control was recorded (if given) and results were divided into three groups: Siemens (i.e. same manufacturer as the Newcastle cameras), other manufacturers and unknown (Figure 104). The subject ages ranged from 30 to 85. A comparison between the Siemens results (59 subjects) and the results for other known cameras (40 subjects) showed that Siemens SBRs were on average lower than other manufacturers by 0.15 ($p=0.02$). Sixteen subjects with unknown camera were excluded. However, for older subjects only (over 60) the difference was 0.09 and was not statistically significant. Furthermore, since the local normal patients were scanned with Siemens cameras, the SBR results with DaTQUANT would be expected to be lower than average rather than higher. It is therefore highly unlikely that a lack of camera correction accounts for the high Z scores for the local normal patients with DaTQUANT.

The ENC-DAT database used by BRASS is not available within the software and publications have not supplied the full dataset with the camera used for each subject. This makes it difficult to comment on whether there is a significant difference between cameras. The original work on phantom calibration for the database provided a range of different calibration factors for each camera model [158], (although the SBRs were not compared to test for significant differences). However, a recent publication by members of the ENC-DAT group [283] suggested that correcting for camera differences has very little impact on diagnostic accuracy. However this study was limited to clinical patients with three year's follow up, rather than research subjects and therefore it is likely that the original FP-CIT scan report influenced the diagnosis. It is also difficult to reconcile this lack of camera effect with the +15% SBR correction factor we are applying to make our results compatible with the normal database. Without this correction our normal patients would have low SBRs on BRASS compared to the ENC-DAT database.



Figure 104: DaTQUANT SBRs for PPMI normal database for Siemens, other manufacturer and unknown scanners. Visually there does not appear to be any systematic difference between Siemens camera results and other cameras for subjects over 60 years and this is supported by a non-significant t-test result.

Age correction differences

By default, BRASS applies a Z score age correction formula that is linear across all ages, resulting in a large adjustment of Z scores upwards for older subjects. The ENC-DAT healthy control database used by BRASS is not available within the software for inspection but the data has been published by Varrone et al.[111], where plots of SBR against age are given (SBRs calculated with an older version of BRASS without resolution recovery so not directly comparable to our local results). The linear age correction formula in BRASS is derived from this work (Figure 105). Note that although Varrone et al. showed a small but significant difference between SBRs for men and women (higher for women) [111], gender is not taken into account within BRASS (nor within DaTQUANT).

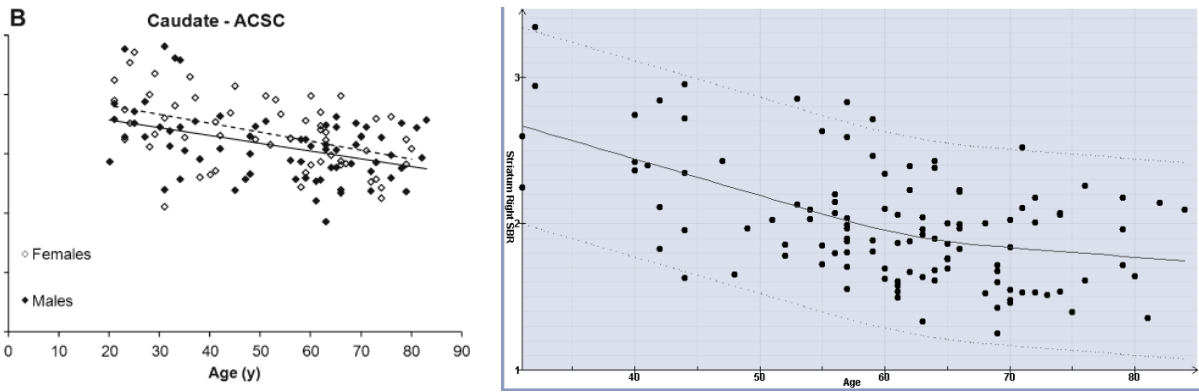


Figure 105: ENC-DAT (left) and PPMI SBR data (right) for healthy controls as a function of age. The ENC-DAT plot is taken from [111] showing a linear fit with age and the PPMI plot is a screenshot from the DaTQUANT program showing a non-linear fit that drops off less steeply with age. Visually, despite the different fits displayed, the distribution of data points is similar and there does not seem to be a big difference in the relationship between SBR and age.

The age correction in DaTQUANT works quite differently. A non-linear fit of SBR as a function of age is performed and the interpolated SBR for the patient is used for calculating patient Z scores. In the PPMI dataset the decline in SBR with age occurs at a lower rate for subjects over 60 than for younger subjects. However this does not account for the higher Z scores with DaTQUANT at all. The opposite effect would be expected since DaTQUANT takes account of SBRs being higher for older subjects than would be expected from a linear relationship between SBR and age. The need for age correction in older adults is discussed in much more detail in section 9.1, and in our publication based on this work [267].

Appendix C: Effect of scan reorientation and zoom on FP-CIT quantification with BRASS

Introduction

Clinical ^{123}I -FP-CIT scans are usually manually reorientated in three planes during image reconstruction to correct for variation in patient positioning and give a standard output for visual reporting. The transverse slices are adjusted for head rotation, the coronal slices for sideways head tilt and the sagittal slices are adjusted for “up-down” tilt. Because the patient is scanned with their head back in a headrest, the sagittal slices often need rotating forward to bring the brain in line with the plane of the anterior and posterior commissure (AC-PC line) as shown in Figure 106. It is straightforward to rotate the transverse and coronal planes so that the striata are symmetrical but the AC-PC plane is more difficult to judge on an FP-CIT scan as anatomical features are less obvious.

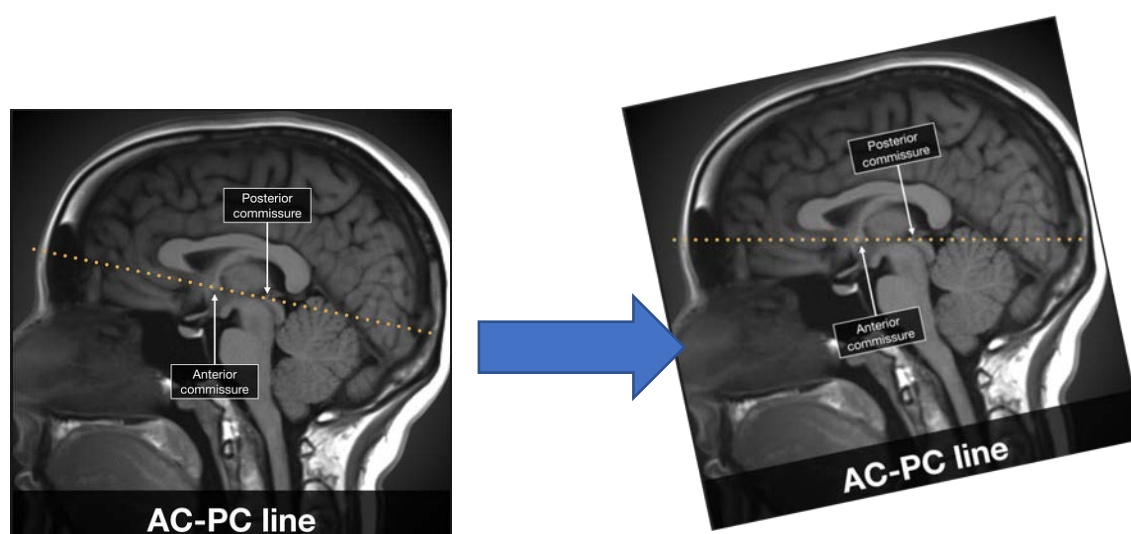


Figure 106: Illustration of the AC-PC line orientation, taken from Radiopaedia.org (<https://radiopaedia.org/articles/ac-pc-line-1>)

The BRASS software application (Hermes Medical Solutions Ltd) automatically registers the FP-CIT scan to a standard template (as illustrated in Figure 107), so manual reorientation during reconstruction is not necessary - as long as the images are only intended to be used for quantification and not for visual assessment. In fact, Hermes warn the user not to adjust the images prior to processing with BRASS: “In order to obtain the most accurate and reproducible results when reconstructing studies which will be evaluated using the HybridViewer BRASS application for DaTScan with the EARL database, the following guidance should be followed:

- The ENCDAT reconstruction protocol provided by Hermes should be used
- The Chang attenuation outlines should not be adjusted
- The reconstructed images should not be aligned manually
- The reconstructed images should not be zoomed” [284]

This is because rotation and zooming require interpolation, which reduces image resolution [284]. BRASS was validated at the RVI using a set of local normal patient scans with clinical follow up that were retrieved and reconstructed as recommended by Hermes. Following the above guidance, these were not reorientated during reconstruction.

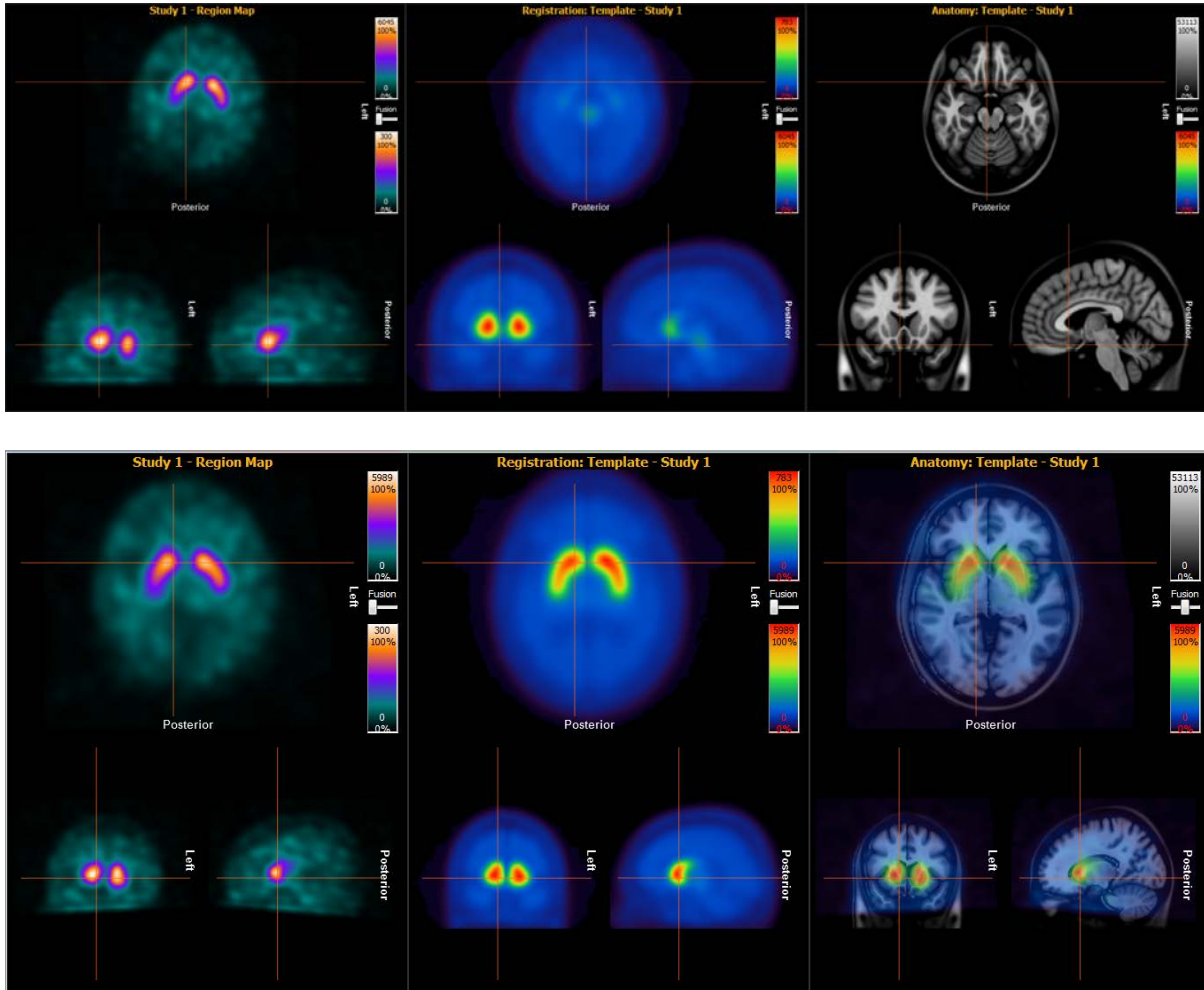


Figure 107: Top: Example study input into BRASS before registration (left). The normal FP-CIT template is shown in the centre and the MRI anatomical template for reference on the right. Bottom: The images after registration to the template.

Previously, visual reporting of FP-CIT scans at the Newcastle Royal Victoria Infirmary (RVI) was done using data reconstructed with the filtered back projection algorithm without attenuation and scatter correction (FBP). BRASS (v2.6) quantification requires a different reconstruction method, which uses the OSEM iterative algorithm with corrections for resolution, attenuation and scatter (OSEM_RRACSC). Given that the visual appearance of the OSEM_RRACSC images is superior to the FBP_NC images in terms of image noise, resolution and contrast we decided to implement the former for the SUPeRB research study. It was decided that a gradual switch to clinical reporting with OSEM_RRACSC would be made within NuTH, with the non-corrected FBP images also generated at first. These would also act as a quality control step to ensure no artefacts were introduced by applying resolution modelling, attenuation or scatter correction.

To use the OSEM_RRACSC images for visual reporting they need to be reorientated, which as discussed above has not been done for quantification. In order to make the image display as close as possible to the previous processing it is also necessary to add an additional zoom of 200% (2x) in the reconstruction. Finally, because the OSEM_RRACSC algorithm takes several minutes to complete on the full field of view raw projection images it is beneficial to crop the raw images both superiorly and inferiorly so that just the brain is included and not the space above the head or the jaw and salivary glands (Figure 108). Usually over half the image can be removed in this way without affecting the appearance of the reconstructed images.

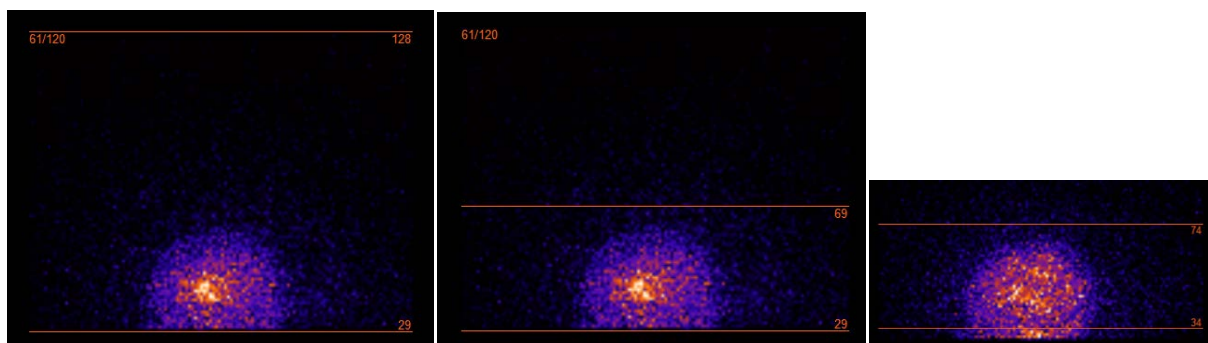


Figure 108: Raw FP-CIT projection data showing full field of view (left), superior cropping (middle) and cropping of the salivary glands inferiorly (right – different patient).

It would be much simpler to use the same reconstructed images for reporting and quantification with BRASS and theoretically these changes should not make much difference to the automated registration. However a conference talk from a colleague at the Royal Free hospital (British Nuclear Medicine Society 2013) suggested adjusting scans for tilt did affect BRASS quantification. This work was done using an older version of BRASS rather than the current Hybrid Viewer version (v2.6). We were unable to find any literature published on this topic (Spring 2017).

The purpose of this piece of work was to assess whether reorientation, image cropping and zoom have a significant effect on BRASS quantification results and therefore whether it is necessary to process a separate reconstruction using the full raw dataset with original orientation and zoom for use with BRASS.

Method

Thirty FP-CIT scans⁷ taken from the LewyPro study were reconstructed using the OSEM_RRACSC method within Hermes HybridRecon (Neurology) in the following four ways:

1. Using the full field of view raw dataset with no reorientation or zoom.
2. Using the full field of view raw dataset and applying reorientation as required in all three planes – angles of rotation recorded.
3. Cropping the raw dataset - no reorientation or zoom.
4. Using the same cropped dataset with no reorientation as above but with an additional 2x zoom.

Figure 109 shows the reconstructed images before and after reorientation for an example scan. The 120 resulting scans were processed with BRASS and the striatal to background ratios with camera correction for the left and right striatum and left and right caudate and putamen recorded. The Z-scores, which compare the SBR results to the Hermes normal database, were also recorded for each of the six regions.

The differences in SBR values between each of the following reconstruction methods were calculated:

- Datasets 1 and 2 (effect of *reorientation* on SBR values)
- Datasets 1 and 3 (effect of *cropping* raw projection dataset)
- Datasets 3 and 4 (effect of *2x zoom*)

The mean difference in SBR for each region was recorded. Paired t-tests were used to test for any significant differences in SBR values.

In clinical practice age corrected Z-scores are often used to judge whether an SBR value is below the normal range expected for a healthy individual of the same age, with a Z-score of -2 taken to indicate a significantly reduced SBR. We therefore used Z-score discrepancies to assess whether any differences in SBR value between reconstruction methods were clinically significant. The frequency of a mis-match between the Z-scores of the two reconstructions in question was recorded. For example, if the Z-score for one reconstruction was lower than -2 (i.e. abnormal quantification, more than 2 SD from the normal mean) and for the other it was higher, or vice versa, then it was considered a clinically significant mismatch. E.g. if the right Z-score was -1.9 for one recon and -2.1 for the other it would be clinically significant, assuming -2SD is used as threshold. As there are 30 scans used in this sample, each with six regions, the total number of Z-score matches / mismatched for each reconstruction method being compared is 120.

⁷ The first thirty scans were selected from the LewyPro dataset of 74 scans as the reprocessing required is time consuming and thirty was considered an adequate sample.

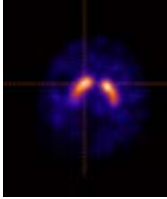
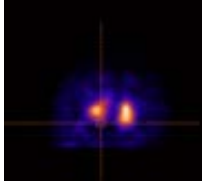
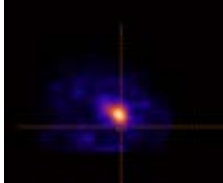
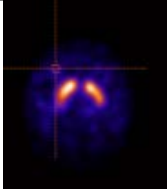
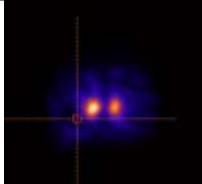
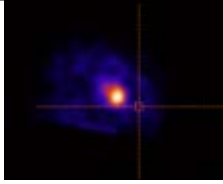
	Transverse plane	Coronal plane	Sagittal plane
Original images without rotation			
Images after rotation			
Angle rotated	+ 8	+ 8	- 10

Figure 109: Example images displayed in HybridRecon before and after reorientation in all three planes, with the angles of rotation applied for this study. This is LewyP001, used as an example throughout this document.

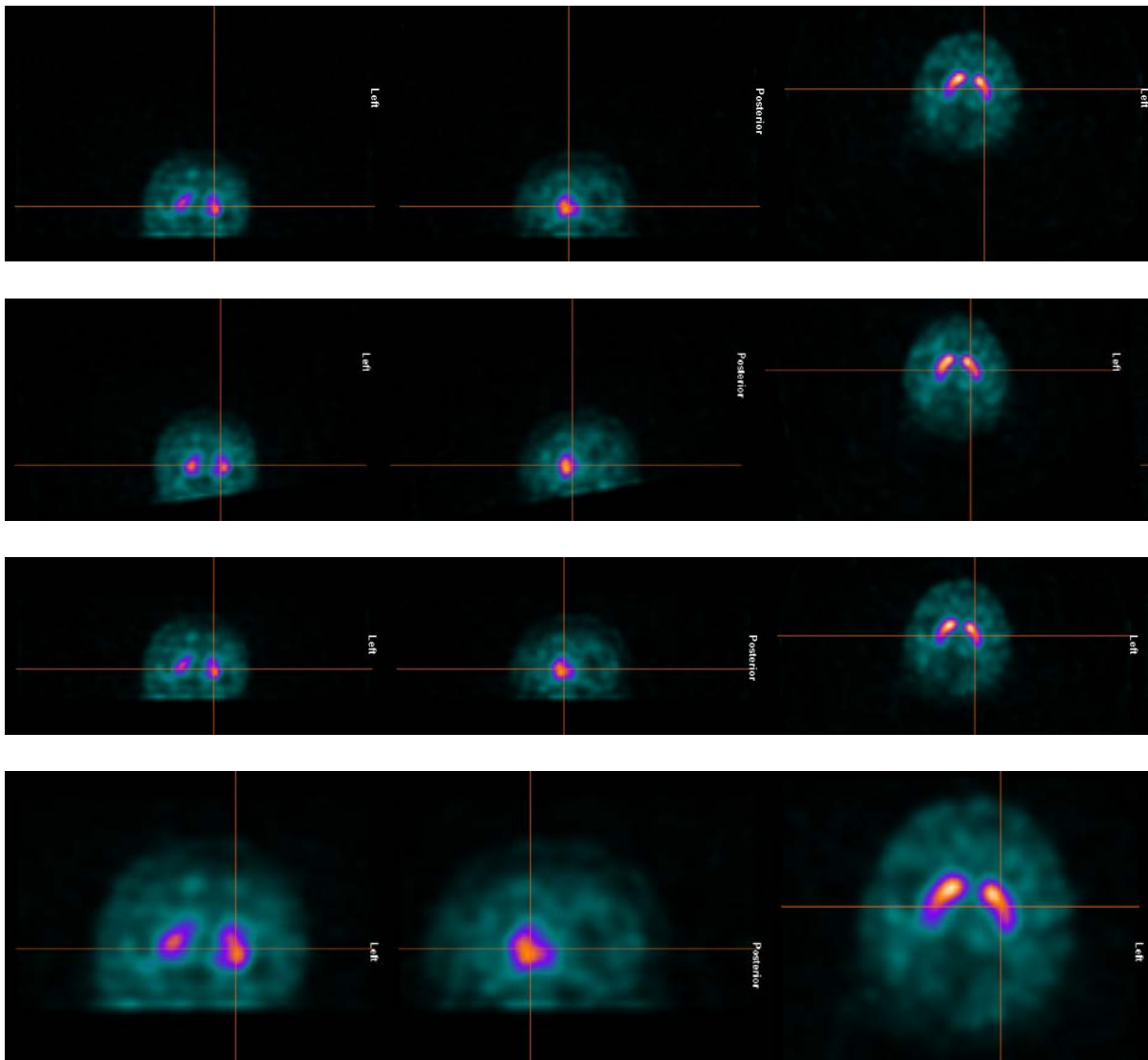


Figure 110: Example image reconstructed using the four methods discussed in the text. Top to bottom: 1) Original images with no reorientation or cropping; 2) reorientated images; 3) cropped images, 4) cropped and zoomed images.

Results and discussion

The mean striatal to background (SBR) values for the 30 scans reconstructed without reorientation, cropping or zoom is shown at the top of Table 79. The mean changes in SBR values given by reorientation, cropping and zoom are shown in bold. The p-value result from the paired t-test to check for statistical significance is shown in the middle row. P values of less than 0.05 indicate that it is very unlikely that the difference seen is due to chance. Finally, the number of mismatched Z-scores (out of a possible 30 for each region) is given in the bottom row.

		Right striatum	Left striatum	Right caudate	Left caudate	Right putamen	Left putamen
Mean SBR values for original images		2.50	2.53	2.64	2.72	2.33	2.31
Reorientation (Dataset 2 – Dataset 1)	Mean SBR difference	-0.07	-0.07	-0.05	-0.07	-0.08	-0.08
	p-value	<i>0.002</i>	<i>0.002</i>	0.123	<i>0.013</i>	<i>0.008</i>	<i>0.011</i>
	Z-score matches	29	30	30	29	30	30
Cropping (Dataset 3 – Dataset 1)	Mean SBR difference	0.01	0.01	-0.01	0.01	0.03	0.00
	p-value	0.613	0.378	0.609	0.591	0.275	0.983
	Z-score matches	30	30	30	29	30	30
Zoom (Dataset 4 – Dataset 3)	Mean SBR difference	0.02	0.02	0.03	0.02	0.01	0.03
	p-value	<i>0.010</i>	<i>0.010</i>	0.111	0.518	0.566	0.379
	Z-score matches	30	30	30	30	30	30

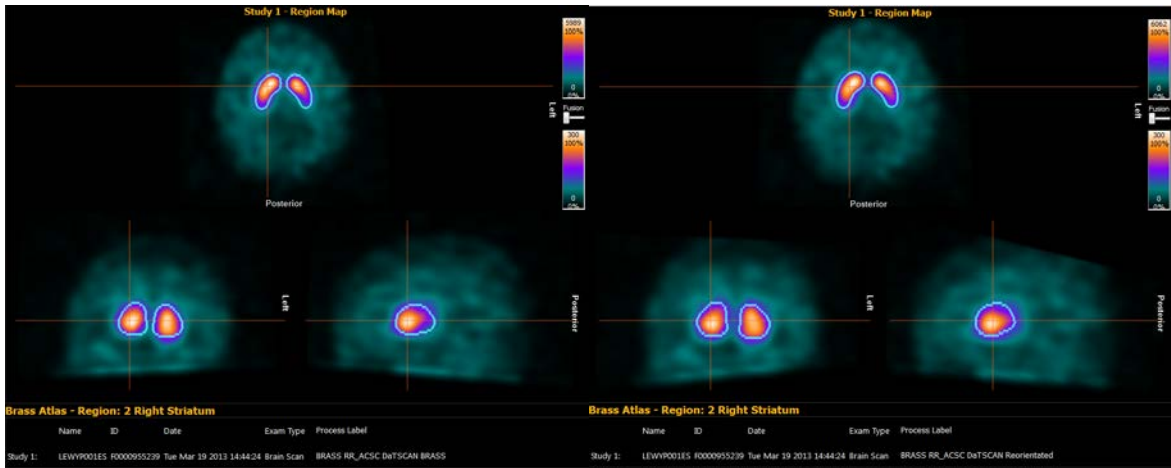
Table 79: Differences in mean SBR values for the reconstruction methods compared. For each effect (reorientation, cropping and zoom) the mean change in SBR values is shown in bold on the top row, the p-value for the paired t-tests in the middle row and the number of Z-score matches (out of 30) in the bottom row.

Effect of reorientation

The average angle of rotation was -1.8 degrees in the transverse plane, -1.5 degrees in the coronal plane and -8.51 in the sagittal plane. On average the SBR values and Z-scores for all regions were lower for the reoriented / adjusted reconstructions than the original ones. Using paired t-tests to test for significance only the right caudate showed no significant difference (mean SBR difference = -0.05,

p = 0.12). For all other regions the average SBR differences were -0.07 or -0.08, p<0.05. However, these mean differences are quite small, less than 5% drop in SBR.

A Z-score mismatch between the original orientation and reorientated images occurred just twice in the whole dataset of 180 Z-scores (6 x 30), each for different patients. So of the 30 patients, two had one Z-score of a possible six that was not consistent between the two reconstructions. None had more than one Z-score mis-match. Example BRASS output for original and reorientated images is shown in Figure 111.



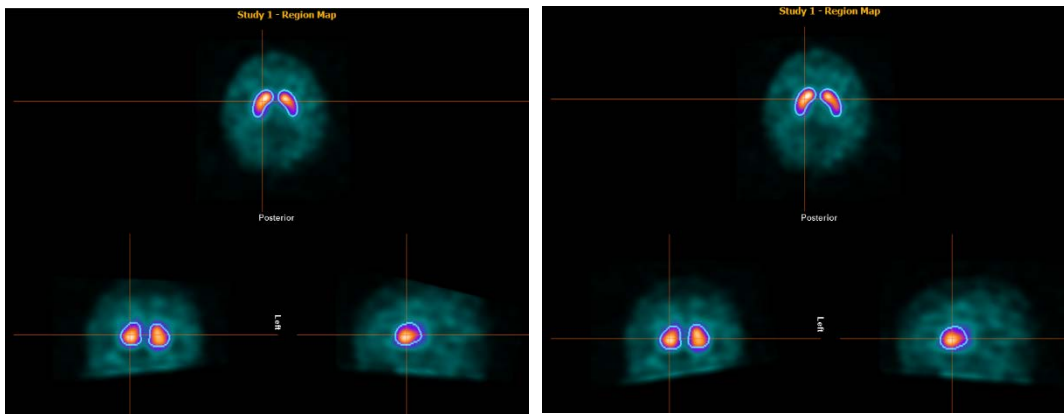
Region Name	Specific Ratio (Z)
Occipital Ctx.	0.00 (nan)
Right Striatum	4.02 (2.18)
Left Striatum	3.86 (1.91)

Region Name	Specific Ratio (Z)
Occipital Ctx.	0.00 (nan)
Right Striatum	4.05 (2.25)
Left Striatum	3.89 (1.97)

Figure 111: Example BRASS output for original orientation scan (left) and reorientated images (right). This is the same example subject as shown above. Note that on average the SBR values were lower for the reorientated images; here they are slightly higher.

Effect of cropping the projection data

The root mean squared change in SBR values resulting from cropping the raw images prior to reconstruction was very small and not statistically significant. One Z-score (of a possible 120) showed a mismatch between the full field of view and cropped reconstructions. Example BRASS output for the same example patient used throughout is shown in Figure 112.



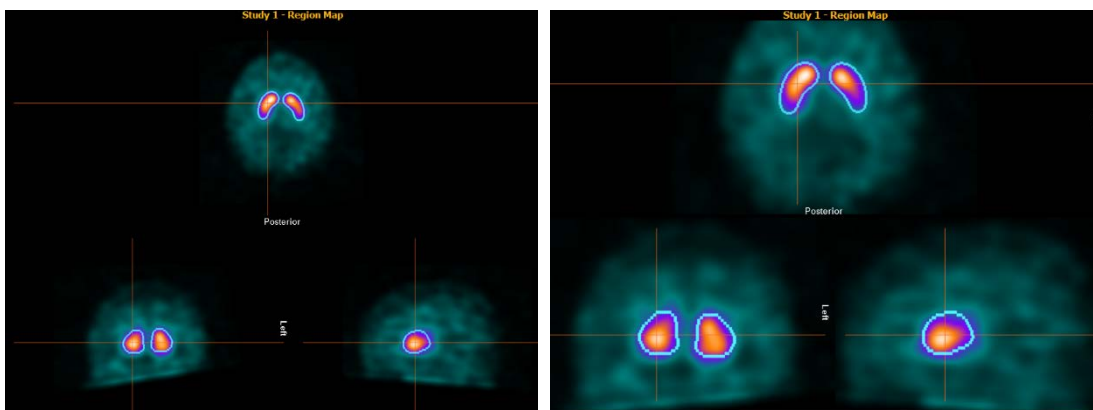
Region Name	Specific Ratio (Z)
Occipital Ctx.	0.00 (nan)
Right Caudate	4.22 (2.27)
Left Caudate	4.05 (1.86)
Right Putamen	3.86 (2.07)
Left Putamen	3.71 (1.94)

Region Name	Specific Ratio (Z)
Occipital Ctx.	0.00 (nan)
Right Caudate	4.26 (2.34)
Left Caudate	4.23 (2.22)
Right Putamen	3.93 (2.20)
Left Putamen	3.50 (1.52)

Figure 112: BRASS output for the example patient without cropping (left) and with cropping (right).

Effect of 2x zoom

On average all the regions (and Z-scores) were very slightly higher for the zoomed reconstructed images than the original ones, but the difference was even smaller than for reorientation. The mean SBR differences were between +0.01 and +0.03, which is approximately a 1% increase on average. Paired t-tests indicated that the whole striatum SBR increases were statistically significant but the individual caudate and putamen increases were not. This is likely to be because there is more random variation between results for the small regions so such a small increase would be difficult to demonstrate. There were no instances of Z-score mismatch between original scale and zoomed images. Example BRASS output for original and 2x zoomed images is shown in Figure 113.



Region Name	Specific Ratio (Z)	Region Name	Specific Ratio (Z)
Occipital Ctx.	0.00 (nan)	Occipital Ctx.	0.00 (nan)
Right Caudate	4.26 (2.34)	Right Caudate	4.41 (2.63)
Left Caudate	4.23 (2.22)	Left Caudate	4.29 (2.34)
Right Putamen	3.93 (2.20)	Right Putamen	3.72 (1.78)
Left Putamen	3.50 (1.52)	Left Putamen	3.45 (1.41)

Figure 113: BRASS output for example original scale and 2x zoomed images.

Conclusion

Cropping the dataset to include only the brain has no significant effect on the results. Reorientation and zoom both have a small statistically significant effect on the BRASS quantification results but these are both clinically insignificant and furthermore act in opposite directions. We can therefore proceed to include the OSEM_RRACSC reconstruction used in research studies alongside the FBP reconstruction used clinically and eventually move to using the same images for reporting and quantification in clinical practice.

Appendix D: The effect of pixel size on the quantification of FP-CIT images

Introduction

¹²³I-FP-CIT quantification with BRASS software (Hermes Medical Ltd, Stockholm) is used extensively throughout this thesis. Hermes states that a FP-CIT image pixel size of between 2 and 3 mm should be used for quantification with BRASS, as this is used in the ENC-DAT project from which the BRASS normal database is taken [111] and is recommended by the European Association of Nuclear Medicine [113] in order to achieve a pixel size “one-third to one-half of the expected resolution”. However the current pixel size used at the RVI for both NHS scans and ongoing University studies is 3.9mm, which is recommended for quantification with DaTQUANT. This value was set by GE Healthcare, the manufacturer of both FP-CIT and DaTQUANT in accordance with the US Society of Nuclear Medicine guidelines, which state that “experimental studies with a striatal phantom suggest that optimal images are obtained when the selected matrix size and zoom factors give a pixel size of 3.5–4.5 mm” [112]. Other than the society guidelines, we could not find any published data to support either approach to pixel size. Using a larger or smaller pixel size than recommended could potentially affect how well the BRASS and DaTQUANT automated registration algorithms can register the scan to the template defining the striatal regions, which could affect the results.

The pixel size could be reduced to fulfil BRASS requirements by increasing the zoom applied when acquiring the images, as this means the same number of pixels would cover a smaller area. However a zoom of x1.23 is already applied in our standard clinical FP-CIT protocol. Increasing the zoom would make it more difficult to set up the patient within the field of view correctly, so is not a change to implement without evidence that it would be beneficial. Of course, reducing the pixel size in line with BRASS recommendations would mean a move away from the pixel size recommended for DaTQUANT.

The purpose of this work is to use an anthropomorphic striatal phantom to acquire images with the current 3.9mm pixel size used in clinical practice and with pixel sizes within the recommended 2-3 mm range and to use this data to determine whether there is a difference in BRASS and DaTQUANT results with pixel size. This information will help us decide whether reducing the pixel size is necessary for the SUPeR study, which we would prefer not to do. It is likely that there will be normal random variation between subsequent scans even under the same conditions so repeated scans of the phantom scanned with the same zoom will be used to quantify this.

Objectives

- To investigate the effect of pixel size on FP-CIT quantification with BRASS and DaTQUANT using phantom images.
- To acquire phantom images with SPECT-CT with a range of striatal to background activity ratios and known activity concentrations.

Methods

Image acquisition

The striatal phantom (RSD Radiology Support Devices, Long Beach, CA, USA) was filled with ^{123}I activity concentrations intended to give normal striatal to background ratios on quantification ($\sim 15:1$ and $\sim 10:1$). Figure 17 in section 4.1 shows the phantom being filled and assembled. The exact activity ratios were later verified using samples of solution analysed in a gamma counter. The phantom was scanned three times using the current RVI imaging protocol which uses a zoom of 1.23. The phantom was repositioned between each repeat acquisition to mimic repeat scans of a patient and allow an estimation of random variation between repeat scans to be made. Without repositioning, the phantom was then scanned with the three zooms available on the scanner calculated to give pixel sizes between 2 and 3 mm (1.78x, 2x and 2.29x). This resulted in four scans taken with the phantom in the exact same position with pixel sizes of 3.9 mm, 2.7 mm, 2.4 mm and 2.1 mm. The phantom was deliberately not moved between these acquisitions so that the effect of zoom could be separated from the effect of differences in positioning. Note however that there will still be a random element in the imaging due to the stochastic nature of radionuclide imaging with limited counts.

The phantom background activity concentration was then approximately doubled to give striatal to background ratios intended to give borderline results on quantification ($\sim 7:1$ and $\sim 5:1$). Again, the exact ratios were later verified using samples analysed with a gamma counter. The phantom was scanned using the current zoom of 1.23, with the time per projection reduced from 25 seconds to 12 seconds to compensate for the increased activity. No adjustment for decay was made. Without repositioning the phantom, the images were repeated using a zoom of 2. There was insufficient time available on the scanner to scan with all three increased zooms so the middle value was chosen.

Finally, the background activity concentration in the phantom was approximately doubled again to give striatal to background ratios of $\sim 3.5:1$ and $2.5:1$ and the phantom scanned with zooms of 1.23 and 2, again halving the time per projection to compensate for the increased activity. This resulted in scan data for six different striatal to background ratios ranging from very normal to very abnormal, with at least two different pixel sizes for each. All acquisitions included a CT scan for attenuation and scatter correction.

Determining activity ratios and absolute activities

The absolute activities in the striata and background compartments (and striatal to background activity concentration ratios) were determined using 1 ml sample solutions from each compartment counted in a gamma counter alongside a standard solution with known activity concentration.

Two solutions of approximately 40 kBq/ml and 50 kBq/ml were prepared and used to fill the striata. The volume solution in each striatum was measured by weighing the container of solution before and after filling. The volume of the background compartment was estimated by measuring the amount of distilled water required to fill it. This volume information was used to calculate the activities needed to give the different striatal to background ratios required.

After each acquisition, before increasing the background activity, sample counting tubes were filled with background solution samples. The two striatal concentrations were unchanged throughout so only needed to be sampled once after all acquisitions were completed. Three samples were taken from each compartment and the sample tubes were weighed before and after the solution was added so that an exact volume could be determined for each sample.

A standard ^{123}I solution of known activity concentration was prepared using a syringe of ^{123}I measured as 5.23 MBq in the radionuclide calibrator at reference time of 11:00. This was diluted to exactly 1 litre in a calibrated flask to give a known concentration of 5.23 kBq / ml. Three samples of approximately 1 ml were taken from this to count alongside the phantom samples so that the absolute activities in the phantom could be determined. Again the sample tubes were weighed to determine the exact volume of each sample.

The samples from the three background concentrations, two striatal concentrations and the standard solution were counted in the gamma counter for 120 seconds each, alongside 1ml samples of water used to correct for environmental background counts. The timestamp for each measurement was used to decay correct the counts to the reference time of 11:00. The counts were also normalised to the volume of the solution in each tube to give the counts per ml. The mean counts per ml measured from the three standard solution samples was used to determine the activity concentration for each of the samples from the phantom.

Image processing

BRASS:

Each image was first reconstructed using Hermes HybridRecon(Neurology). An iterative ordered subsets expectation maximum (OSEM) algorithm was applied with resolution modelling for the specific collimator used, Monte Carlo derived scatter correction and attenuation correction. Both uniform (geometrical) based attenuation and scatter correction and CT data driven methods were used. The resultant images were loaded into BRASS and the SBRs with camera correction were recorded for the right and left whole striata, caudate and putamen.

DaTQUANT:

The raw data for each image was exported from a Hermes workstation and loaded onto a desktop PC with DaTQUANT installed. A dicom directory was set up to import the raw data into DaTQUANT. The images were reconstructed within DaTQUANT using the built in GE Xeleris OSEM algorithm but without correcting for attenuation or scatter⁸. The SBRs were recorded for the right and left whole striata, caudate and putamen. The Hermes uniform AC and CTAC reconstructed images were also exported and analysed with DaTQUANT to allow direct comparison between the two software packages with the same input data.

Analysis

The range (max-min) of SBR values for each region over the three repeated normal uptake scans was calculated and expressed as a percentage of the average SBR. The percentage variation for each region was then averaged to give a mean SBR variation with repositioning for a normal uptake scan. This was done for BRASS and DaTQUANT results for all reconstructions tested. The variation between the four normal uptake scans with different zooms was calculated in the same way. The percentage variation for repositioning was compared with the percentage variation with zoom for a normal uptake scan.

The percentage differences in SBR values for zooms of 1.23 and 2.0 were compared for the six striatal to background activity ratios given from the normal, borderline and abnormal uptake phantom scans.

Results

The activity concentrations and absolute amount of ¹²³I activity in the left and right striatal compartments and background section of the phantom are shown in Table 80.

Table 80: Activity concentrations measured for each phantom compartment with reference to a standard solution.

	Activity concentration / kBq/ml	Compartment volume / mls	Total activities at reference time / MBq
Right striatum	58.5	10.4	0.61
Left striatum	38.5	10.6	0.41
Background 1	4.0	1295	5.1
Background 2	8.8	1295	11.5
Background 3	19.0	1295	24.6

⁸ Scatter correction and CTAC are unavailable within DaTQUANT and we found the DaTQUANT uniform attenuation correction to be unreliable using the default thresholding values.

For quantification, striatal to background ratios are defined as (striatal counts – background counts)/background counts. This means that the SBRs measured from the samples are the ratio of striatal and background activity concentration measurements, minus 1. These measured values are given in Table 81 with the SBR results from BRASS and DaTQUANT alongside. The SBRs are underestimated by both software packages, suggesting that the partial volume effect is large. However, it is the change in SBR with pixel size that is assessed here, rather than the absolute error.

Table 81: Activity ratios and measured SBRs shown alongside BRASS and DaTQUANT SBR results for the standard 1.23x zoom.

		Measured activity concentration ratio	Measured SBR (ratio – 1)	BRASS SBR (Chang AC)	BRASS SBR (CTAC)	DaTQUANT SBR (NC)	DaTQUANT SBR (Hermes Chang AC)	DaTQUANT SBR (Hermes CTAC)
Normal uptake	Right striatum	14.7	13.7	5.65	6.06	5.17	4.58	5.95
	Left striatum	9.7	8.7	3.25	3.44	2.98	2.63	3.51
Borderline uptake	Right striatum	6.6	5.6	2.32	2.33	1.95	1.69	2.36
	Left striatum	4.4	3.4	1.28	1.21	1.04	0.92	1.23
Low uptake	Right striatum	3.1	2.1	1.05	0.78	0.63	0.59	0.87
	Left striatum	2.0	1.0	0.52	0.27	0.28	0.24	0.49

Images of the phantom with normal striatal to background ratio for each pixel size / zoom are given in Figure 114, with the images from BRASS and DaTQUANT with striatal regions overlaid shown below.

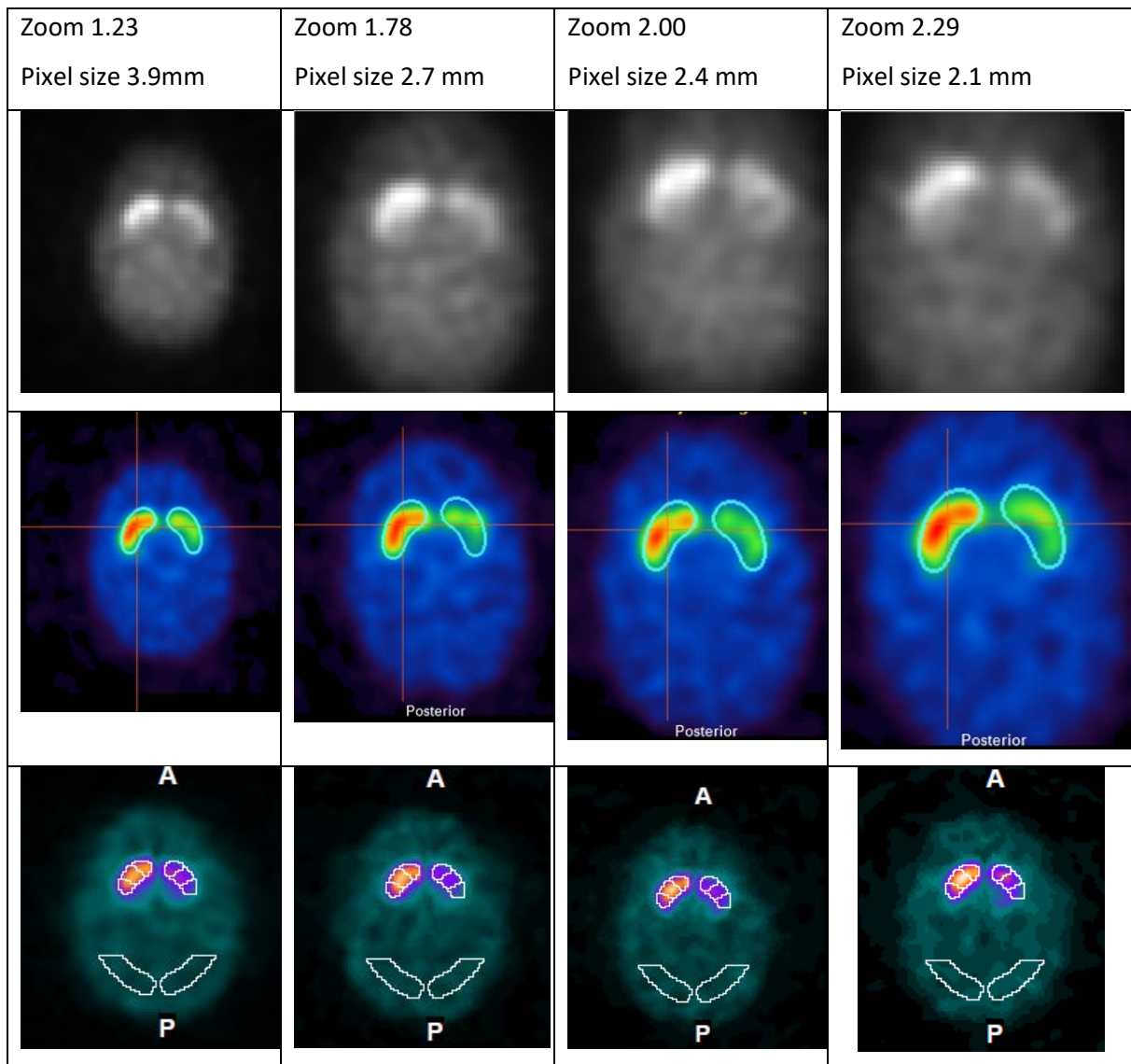


Figure 114: Top: Phantom images with normal filling ratios with increasing zoom, illustrating the images becoming less “pixelated” at higher zoom. Middle: BRASS images for each zoom with striatal contours (occipital background not shown). BRASS registers the template to the subject image so the image size displayed increases with zoom. Bottom: DaTQUANT images for each zoom with caudate, putamen and background contours shown. DaTQUANT registers the subject image to a template, so the images are resized to match the template.

The percentage variation in SBR values over three repeat scans, averaged over all regions, was 7.0% and 7.5% with the BRASS uniform AC and CTAC reconstructions respectively (Table 82). The percentage variation with the four pixel sizes / zooms was 8.3% and 7.2% respectively. With scans reconstructed in DaTQUANT (no AC) the percentage variation for the repeat scans was 8.9% and was 13.6% for the scans with different pixel zooms. Plots of SBR value against zoom factor for BRASS and DaTQUANT (Figure 115) suggest that there is no systematic variation for BRASS but SBRs appear lower for higher zooms with DaTQUANT. The results for zooms of 1.23 and 2.0 for normal uptake, borderline and low uptake are given in Table 83. There is generally much more variation between the scans with different zooms with lower uptake. Table 82 and Table 83 also show results for the same reconstructed data (Hermes uniform AC and Hermes CTAC) processed with BRASS and DaTQUANT. There is more variation in the DaTQUANT results than the BRASS results particularly at low uptake.

Table 82: The variation in quantification results seen with repeat scanning with the same zoom (top rows) and with increasing zoom (bottom rows).

		BRASS Uniform AC	BRASS CTAC	DaTQUANT NC	DaTQUANT Hermes Uniform AC	DaTQUANT Hermes CTAC
Variation in quantification with repeat scanning	Average SBR difference (all regions)	0.28	0.35	0.32	0.25	0.29
	Percentage difference	7.0%	7.5%	8.9%	6.0%	8.3%
Variation in quantification with zoom	Average SBR difference (all regions)	0.29	0.33	0.47	0.52	0.50
	Percentage difference	8.3%	7.2%	13.6%	12.3%	13.3%

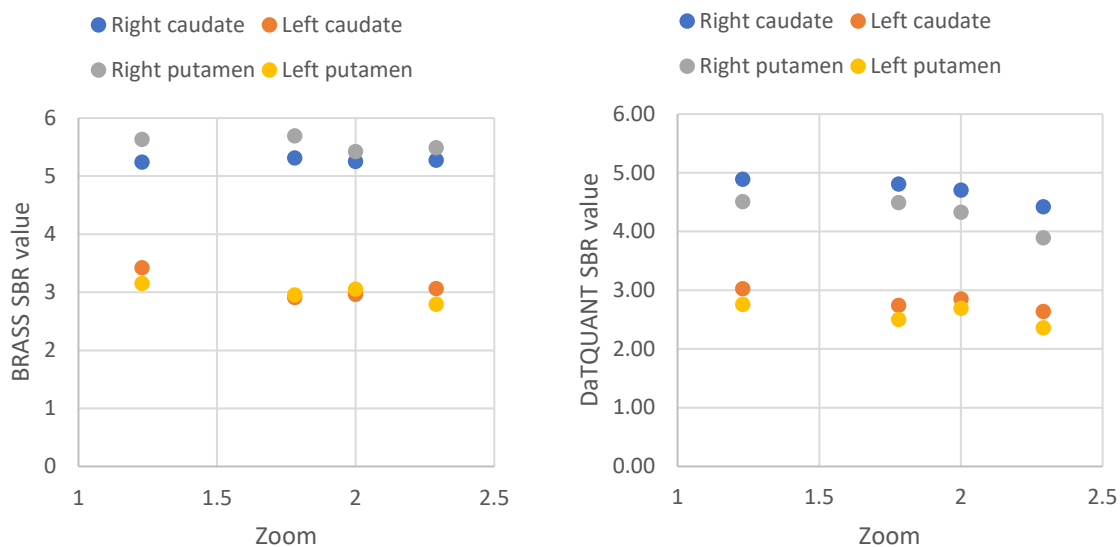


Figure 115: Plots of SBR against zoom factor for BRASS (with uniform AC) and DaTQUANT (no AC).

Table 83: Variation in SBR values between zooms of 1.23 and 2.0 for normal, borderline and low uptake ratios

			BRASS Uniform ACSC	BRASS CT ACSC	DaTQUANT NC	DaTQUANT Hermes Uniform ACSC	DaTQUANT Hermes CT ACSC
Normal uptake ratios	Average difference (all regions)	SBR (all regions)	0.20	0.05	0.15	0.18	0.26
	Percentage difference		5.5%	1.1%	4.0%	4.7%	6.9%
Borderline uptake ratios	Average difference (all regions)	SBR (all regions)	0.13	0.09	0.07	0.23	0.19
	Percentage difference		6.9%	4.4%	5.5%	13.0%	11.7%
Low uptake ratios	Average difference (all regions)	SBR (all regions)	0.07	0.08	0.05	0.17	0.14
	Percentage difference		8.0%	14.7%	9.6%	19.1%	23.6%

Discussion

The data here show that changing the zoom has a similar effect on quantification results with BRASS as repeating the scan with the phantom repositioned. This suggests that the variation is random and could be due to slight changes in the input image. There is therefore no definite evidence to support changing the zoom to give a pixel size closer to that used with the Hermes normal database in BRASS. Changing the zoom just for research studies would be confusing for staff acquiring and processing the data, and changing the zoom for all studies would mean making changes to the way clinical images are presented for reporting, since the slice thickness would also change.

For DaTQUANT these results show that changing the zoom does appear to have a greater effect on the SBR results than repositioning, and the plot in Figure 115 suggests that SBR results decrease with zoom, or increase with pixel size. The same reconstructed input image data analysed with DaTQUANT and BRASS also showed more variation with zoom with DaTQUANT than with BRASS. Since the current zoom factor of 1.23 is recommended by GE Healthcare it would be sensible to stick with this factor given that it seems to be more important for DaTQUANT than BRASS.

DaTQUANT and BRASS register the images in different ways, which is one reason why there might be the discrepancies between them for the same input data. DaTQUANT “warps” the study image to fit the normal template image, whereas BRASS adjusts the normal template to the study image, and does a separate registration for each hemisphere of the brain.

This amount of data in this study is limited due to unavoidable time constraints. Each acquisition took around 40 minutes in total and all were carried out in the same imaging session so that the same phantom striatal activity could be used. Ideally repeat scans with the same zoom would have been carried out for all striatal activity ratios, not just the normal uptake scan. If time permitted it would also have been better to have taken images for all zooms each time the phantom was repositioned, rather than only the standard 1.23 zoom. It would also have been useful to collect data for repeat scans with the same parameters without any movement of the phantom to explore the effect of statistical variation in image counts on the quantification results as it is impossible to separate this from the effects of zoom. The phantom was filled with activity designed to give count rates similar to patient studies. If higher activities had been used, the acquisition time could have been reduced, but we wanted the count rates to match clinical scanning as closely as possible.

Conclusion

Although the results are limited and insufficient for statistical analysis, the results obtained support keeping the current zoom factor of 1.23 for both BRASS and DaTQUANT processing.

Appendix E: Cardiac MIBG imaging protocol details from the literature

Paper evaluating MIBG for DLB vs AD or other dementias (date order)	Country; setting	Subjects included	Activity / MBq;	Image duration	Post-injection delay	HMR normal cut-off; origin	Camera and collimator	Sensitivity and specificity
Yoshita et al. PLOS 2015 Mar 20;10(3) [53]	Japan; multicentre	87 DLB; 46 AD	111 MBq;	not stated	20-30 mins & 3-4 hours	2.10 (both time points); ROC analysis	Mixed (multicentre) results scaled to MEGP	Sens 69% Spec: 89% (both time points)
Lebasnier et al. Eur J Nucl Med Mol Imaging. 2015 Jan 28. [56]	France; single centre	45 with Lewy body disease; 19 without	<60 kg 148 MBq, 60–90 kg 222 MBq, >90 kg 296 MBq;	10 minutes	4 hours	1.47; ROC analysis	Siemens Symbia T2; LEHR	Sens: 82% Spec: 84%
Slaets et al. Int J Geriatr Psychiatry. 2014 Nov 3. [59]	Belgium; single centre	20 possible DLB at baseline with 6+ months follow up	74 MBq	not stated	4 hours	1.68; Yoshita et al. 2006	Phillips XCT or GE Varicam; LEHR	Sens: 100% (16/16) Spec: 75% (3/4)
Inui et al. Ann Nucl Med. 2014 Oct;28(8):796-804. [236]	Japan; single centre	35 prob DLB; 34 prob AD	111 MBq	6 mins	4 hours	1.7; ROC analysis	ADAC Vertex-Plus, HITACHI; LEGP	Sens: 97%; Spec: 100%
Sakamoto et al. Ann Nucl Med. 2014 Apr;28(3):203-11. [285]	Japan; single centre	24 prob DLB; 76 without DLB inc PD	111 MBq	5 mins	15 mins	2.2; ROC analysis in independent group	Millennium VG, GE; MEGP	Sens: 83% Spec: 85%
Fukui et al. Dement Geriatr Cogn Dis Extra. 2013 May 18;3(1):148-60 [286]	Japan; single centre	53 DLB (MIBG not done in AD)	Not stated	Not stated	20 mins & 4 hours	Not clear	Not stated	Not assessed

Paper evaluating MIBG for DLB vs AD or other dementias (date order)	Country; setting	Subjects included	Activity / MBq;	Image duration	Post-injection delay	HMR normal cut-off; origin	Camera and collimator	Sensitivity and specificity
Treglia et al. J Neuroimaging. 2014 Mar-Apr;24(2):149-54. [68]	Italy; single centre	20 DLB; 11 other dementia; 20 control subjects	111 MBq	5 mins	15 mins & 4 hours	1.55; 2SD below control mean	“Imaged according to EANM guidelines” Control mean 1.89; probably low energy	Sens: 90% Spec: 91%
Camacho et al. Q J Nucl Med Mol Imaging. 2011 Aug;55(4):476-83. [287]	Spain; single centre	28 DLB	Not in abstract; article not accessible			1.56; not accessible	Not accessible; HMR cut-off in abstract suggests low energy	Sens: 82% (23/28)
Novellino et al. Neurobiology of Aging. 2010 Nov;31(11):1903-11. [184]	Italy; single centre	9 DLB; 6 FTD; 16 controls	111 MBq	Not stated	10 mins & 4 hours	3SD below control mean. 1.91 ±0.17 = 1.40 at 10 mins; 1.99 ±0.19 = 1.42 at 4 hours	Axis, Picker; collimator not stated	Sens: 100% (9/9) Spec: 100% (6/6) (both time points)
Estorch et al. Eur J Nucl Med Mol Imaging. 2008 Sep;35(9):1636-41. [54]	Spain; single centre	19 DLB; 12 AD; 13 other non LBD	370 MBq	Not stated	15 mins & 4 hours	1.56 (local database over 60 years); 1.36 ROC analysis on 4 hour images	GE Millenium Hawkeye V3; LEGP	Sens 94% Spec: 96% (using ROC cut-off)
Wada-Isoe et al. J Neurol Sci. 2007 Sep 15;260(1-2):33-7. [237]	Japan; single centre	20 DLB; 32 AD; 29 controls	111 MBq	Not stated	30 mins & 3.5 hours	1.81 (30 mins); 1.82 (3.5 hours);	Not given	30 mins: Sens 85%; Spec 100% 3.5 hours:

Paper evaluating MIBG for DLB vs AD or other dementias (date order)	Country; setting	Subjects included	Activity / MBq;	Image duration	Post-injection delay	HMR normal cut-off; origin	Camera and collimator	Sensitivity and specificity
						ROC analysis		Sens 100%; Spec 91%
Hanyu et al. Dement Geriatr Cogn Disord. 2006;22(5-6):379-84. [55]	Japan; single centre	32 DLB; 9 PDD; 40 AD	111 MBq	5 mins	4 hours	1.73 2SD below control mean 2.37 ± 0.32	PRISM 2000VP, Picker; LEHR	Sens 95% Spec 87%
Yoshita et al. Neurology. 2006 Jun 27;66(12):1850-4. [58]	Japan; multi centre	37 DLB; 42 AD; 10 controls	111 MBq	5 mins	20 mins & 3 hours	1.61 (20 mins), 1.68 (3 hours); ROC analysis	PRISM 2000XP, Shimadzu or 9300A, Toshiba LEHR	20 mins:Sens 97% Spec 100% 3 hours:Sens 100% Spec 100%
Oide et al. Intern Med. 2003 Aug;42(8):686-90.	Japan; single centre	10 DLB, 10 AD	148 MBq	2 mins	15 mins & 4 hours	Not stated	(ZLC-7500, Siemens); LEHR	Not stated but around 1.8 from plot
Yoshita et al. J Neurol Neurosurg Psychiatry. 2001 Nov;71(5):583-8.	Japan; single centre	14 DLB; 14 AD	111 MBq	2 mins	20 mins & 3 hours	Not stated	GCA9300A/HG, Toshiba; LEHR	Not stated but around 1.7 from plot
Watanabe et al. J Neurol Neurosurg Psychiatry. 2001 Jun;70(6):781-3.	Japan; single centre	11 DLB; 10 AD; 11 controls	111 MBq	5 mins	30 mins & 4 hours	Not stated	Not stated	Not stated but around 1.8 from plots

Appendix F: Consistency of FP-CIT processing for semi-quantification

This appendix explores whether careful reprocessing of baseline and repeat FP-CIT scans by the same operator gives more consistent quantification results.

Three subjects were noted to have very different baseline and repeat SBR results with BRASS. The orientation of the reconstructed images, including the extent of the reconstructed data inferiorly, differed between the baseline and repeat images. An example is shown in Figure 116. However, careful reprocessing of the scans, using the CT data as a guide, did not result in more consistent SBR results (Table 84). The changes in SBR between baseline and repeat for the three scans with original processing were -23%, +43% and +32%; after reprocessing the changes were -25%, +43% and +27%. The rest of the scans were therefore not reprocessed and the original results used.

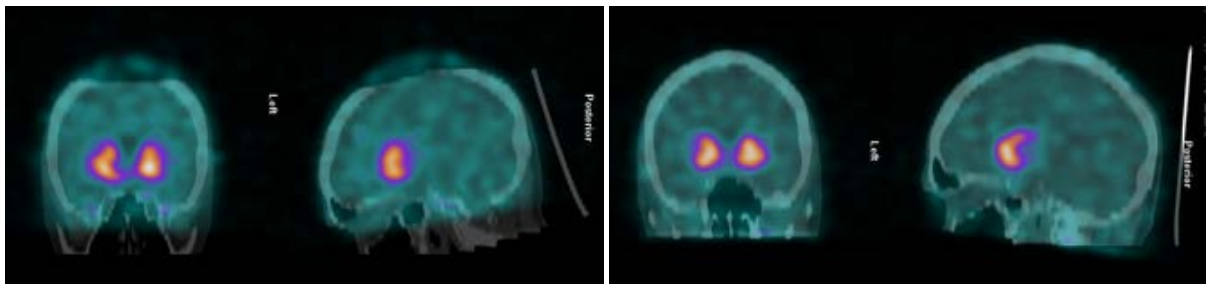


Figure 116: Baseline (left) and repeat (right) coronal and sagittal views showing quite different inferior scan extent and different reorientation on the sagittal views. The CT data is shown to aid visualisation, but was not used in the reconstruction.

Table 84: Baseline and repeat SBRs and Z scores for the regions with greatest change for the three controls noted to have different results. The differences with both the original processing and following careful reprocessing by the same operator are shown.

		Example 1	Example 2	Example 3
Original independent processing	Region with greatest change	Right caudate	Right caudate	Left caudate
	Baseline SBR (Z score)	3.06 (-0.06)	2.69 (-1.21)	2.93 (-0.54)
	Repeat SBR (Z score)	2.36 (-1.37)	3.84 (1.05)	3.86 (1.31)
	% SBR change (Z score change)	-23% (-1.31)	+43% (+2.26)	+32% (+1.85)
Re-processing baseline and repeat	Region with greatest change	Right caudate	Left caudate	Left caudate
	Baseline SBR (Z score)	3.13 (0.08)	2.77 (-0.92)	3.08 (-0.24)
	Repeat SBR (Z score)	2.36 (-1.36)	3.96 (1.37)	3.91 (1.40)
	% SBR change (Z score change)	-25% (-1.44)	+43% (+2.29)	+27% (+1.64)

Appendix G: Simulation work using Monte Carlo modelling

This Appendix details preliminary work on developing a Monte Carlo model of the Siemens Intevo gamma camera, which will form the basis for future work.

Monte Carlo methods are statistical techniques used to model repetitive processes containing a random element, such as the interaction of radiation with matter [288]. They enable investigation of the effects of patient factors in a more flexible and controlled manner than can be obtained with physical anthropomorphic phantoms, and overcome difficulties of gold standards with patient studies.

A typical calculation uses a probability distribution and a random number generator to determine an outcome for an event – this is then repeated many times for subsequent events to build up an overall representation of the system being modelled [289]. To use an example relevant to nuclear medicine, we might wish to model the interaction of gamma rays with material of a certain density. In a Monte Carlo simulation of a nuclear medicine scan, for each gamma ray photon emitted from the patient a random number is generated and this determines the outcome for that individual photon. The random numbers are assigned to the outcomes according to the interaction cross-section (probability) of each outcome (absorption, scatter, no interaction...) occurring – so the most likely will have the most numbers assigned to it. This means that if a high number of interactions is simulated the results will be a close match to the predefined probability distribution, but no two simulations will give identical results.

In nuclear medicine, using a Monte Carlo method to simulate a scan has significant advantages over physical experiments [289]. It is more straightforward to change parameters and run a new simulation than it is to modify a physical set up. Data from Monte Carlo simulations could be used to investigate the effect of differences in size and shape of internal organs on semi-quantification, which would be impossible to achieve with human studies and difficult with physical phantom models. In addition, with simulations it is possible to study the underlying physical processes, such as the amount of attenuation and scatter. This is not possible with a physical experiment as we can't separate out the primary and scattered components with certainty.

I have completed some key tasks related to Monte Carlo modelling during the Fellowship – completion of a basic gamma camera model (Figure 117), acquisition of energy spectrum data from point sources and phantoms with which to test the model, export of raw simulated data and reconstruction within Hermes HybridRecon (Figure 118) and the purchase of the commercial XCAT anatomical CT dataset with which I can create activity distributions in MATLAB (examples shown in Figure 119).

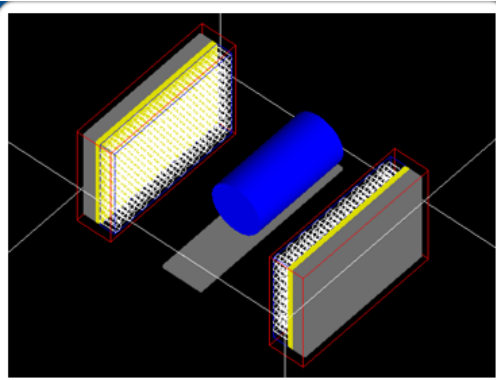


Figure 117: Visualisation of model of dual headed gamma camera showing collimators, NaI crystal and backscatter component, with uniform phantom on couch

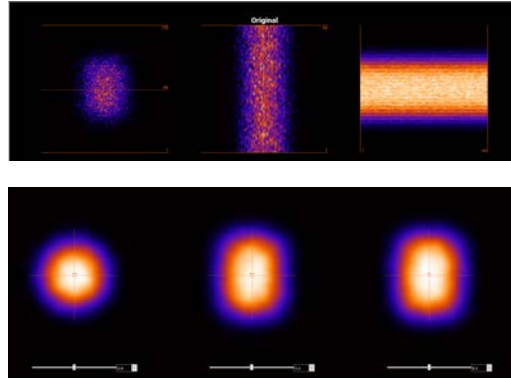


Figure 118: Top: raw simulation data and sinogram in Hermes Hybrid Recon program. Bottom: axial, coronal and sagittal planes through reconstructed images of simulated phantom.

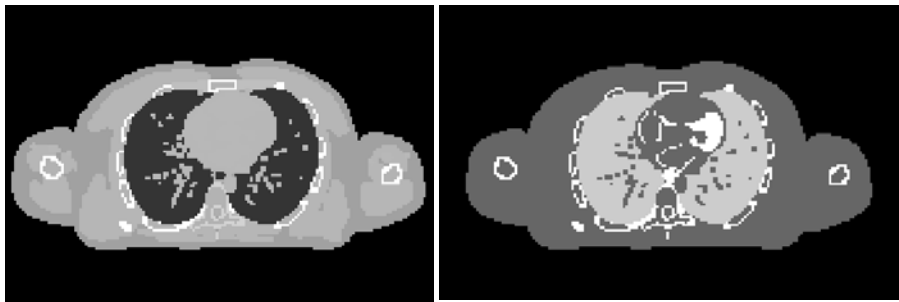


Figure 119: Slice data generated using XCAT virtual phantom showing CT Hounsfield units (left) and activity distribution in bone, lung, heart and background (right).

References

1. 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.
2. Orimo, S., [Clinical and pathological study on early diagnosis of Parkinson's disease and dementia with Lewy bodies]. *Rinsho Shinkeigaku*, 2008. **48**(11): p. 831-4.
3. Orimo, S., et al., *Cardiac sympathetic denervation precedes neuronal loss in the sympathetic ganglia in Lewy body disease*. *Acta Neuropathol*, 2005. **109**(6): p. 583-8.
4. McKeel, D.W., et al., *Dementia: An Atlas of Investigation and Diagnosis*. 2007, Oxford: Clinical Publishing.
5. Alzheimer's Society, *Turning Up the Volume: unheard voices of people with dementia*, Alzheimer's Society, Editor. 2017.
6. WHO, *Dementia Factsheet*, W.H. Organisation, Editor. 2016.
7. Patel, V., *Deaths registered in England and Wales (Series DR): 2015. Registered deaths by age, sex, selected underlying causes of death and the leading causes of death for both males and females.*, O.f.N. Statistics, Editor. 2016, Office for National Statistics.
8. Martin Prince, G.-C., Maëlen Guerchet, Yu-Tzu Wu, Matthew Prina, Anders Wimo, *World Alzheimer Report 2015. The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends*. 2015, Alzheimer's Disease International.
9. McCulloch, D.H.Y., *Factsheet 400: What is dementia*, A.s. Society, Editor. 2017.
10. Warner, J.P., R. Butler, and S. Gupta, *Dementia*. *BMJ Clin Evid*, 2010. **2010**.
11. Perri, R., et al., *Word list and story recall elicit different patterns of memory deficit in patients with Alzheimer's disease, frontotemporal dementia, subcortical ischemic vascular disease, and Lewy body dementia*. *J Alzheimers Dis*, 2013. **37**(1): p. 99-107.
12. 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.
13. Kane, J.P.M., et al., *Clinical prevalence of Lewy body dementia*. *Alzheimers Res Ther*, 2018. **10**(1): p. 19.
14. 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.
15. 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.
16. Galvin, J.E., et al., *Lewy body dementia: the caregiver experience of clinical care*. *Parkinsonism Relat Disord*, 2010. **16**(6): p. 388-92.
17. 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.
18. Schneider, J.A., et al., *Cognitive impairment, decline and fluctuations in older community-dwelling subjects with Lewy bodies*. *Brain*, 2012. **135**(Pt 10): p. 3005-14.
19. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium*. *Neurology*, 2005. **65**(12): p. 1863-72.
20. Aarsland D, P.R., Larsen JP, McKeith IG, O'Brien JT, Perry EK, Burn D, Ballard CG, *Neuroleptic sensitivity in Parkinson's disease and parkinsonian dementias*. *J Clin Psychiatry* 2005. **66**: p. 633-637.
21. Sorbi S, H.J., Erkinjuntti T, Fladby T, Gainotti G, Gurvit H, Nacmias B, Pasquier F, Popescu BO, Rektorova I, Religa D, Rusina R, Rossor M, Schmidt R, Stefanova E, Warren JD, Scheltens P, *EFNS Scientist Panel on Dementia and Cognitive Neurology. EFNS-ENS Guidelines on the diagnosis and management of disorders associated with dementia*. *Eur J Neurol.*, 2012. **19**(9): p. 1159-1179.
22. McKeith IG, R.E., Askew K, Naidu A, Allan L, Barnett N, Lett D, Mosimann UP, Burn D, O'Brien JT, *More severe functional impairment in dementia with lewy bodies than Alzheimer disease is*

- related to extrapyramidal motor dysfunction. *The American Journal of Geriatric Psychiatry* 2006. **14**(7): p. 582-588.
23. Walker, Z., et al., *Comparison of cognitive decline between dementia with Lewy bodies and Alzheimer's disease: a cohort study*. *BMJ Open*, 2012. **2**: p. e000380.
 24. Guy M. McKhann, D.S.K., Howard Chertkow, Bradley T. Hyman, Clifford R. Jack Jr., Claudia H. Kawas, William E. Klunk, Walter J. Koroshetz, Jennifer J. Manly, Richard Mayeux, Richard C. Mohs, John C. Morris, Martin N. Rossor, Philip Scheltens, Maria C. Carrillo, Bill Thies, Sandra Weintraub, Creighton H. Phelps, *The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease*. *Alzheimer's & Dementia*, 2011. **7**: p. 263-269.
 25. Walker, L., et al., *Neuropathologically mixed Alzheimer's and Lewy body disease: burden of pathological protein aggregates differs between clinical phenotypes*. *Acta Neuropathol*, 2015. **129**(5): p. 729-48.
 26. McKeith, I., et al., *Revisiting DLB Diagnosis: A Consideration of Prodromal DLB and of the Diagnostic Overlap With Alzheimer Disease*. *J Geriatr Psychiatry Neurol*, 2016. **29**(5): p. 249-53.
 27. Fujishiro, H., et al., *Prodromal dementia with Lewy bodies*. *Geriatr Gerontol Int*, 2015. **15**(7): p. 817-26.
 28. Donaghy, P.C., J.T. O'Brien, and A.J. Thomas, *Prodromal dementia with Lewy bodies*. *Psychol Med*, 2015. **45**(2): p. 259-68.
 29. Belden, C.M., et al., *Clinical characterization of mild cognitive impairment as a prodrome to dementia with Lewy bodies*. *Am J Alzheimers Dis Other Demen*, 2015. **30**(2): p. 173-7.
 30. Welsh-Bohmer, K.A., *Defining "prodromal" Alzheimer's disease, frontotemporal dementia, and Lewy body dementia: are we there yet?* *Neuropsychol Rev*, 2008. **18**(1): p. 70-2.
 31. Fujishiro, H., et al., *Dementia with Lewy bodies: early diagnostic challenges*. *Psychogeriatrics*, 2013. **13**(2): p. 128-38.
 32. Iranzo, A., et al., *Neuropathology of prodromal Lewy body disease*. *Mov Disord*, 2014. **29**(3): p. 410-5.
 33. Chiba, Y., et al., *Retrospective survey of prodromal symptoms in dementia with Lewy bodies: comparison with Alzheimer's disease*. *Dement Geriatr Cogn Disord*, 2012. **33**(4): p. 273-81.
 34. Donaghy PC, O'Brien JT, and T. AJ, *Prodromal dementia with Lewy bodies*. *Psychol Med*, 2015. **45**(2): p. 259-268.
 35. Sonni, I., et al., *Clinical validity of presynaptic dopaminergic imaging with (123)I-ioflupane and noradrenergic imaging with (123)I-MIBG in the differential diagnosis between Alzheimer's disease and dementia with Lewy bodies in the context of a structured 5-phase development framework*. *Neurobiol Aging*, 2017. **52**: p. 228-242.
 36. Sakakibara, R., et al., *MIBG Myocardial Scintigraphy Identifies Premotor PD/DLB During a Negative DAT Scan Period: Second Report*. *Mov Disord Clin Pract*, 2019. **6**(1): p. 46-50.
 37. Thomas, A.J., et al., *Diagnostic accuracy of dopaminergic imaging in prodromal dementia with Lewy bodies*. *Psychol Med*, 2018: p. 1-7.
 38. Fujishiro, H., et al., *Early diagnosis of Lewy body disease in patients with late-onset psychiatric disorders using clinical history of rapid eye movement sleep behavior disorder and [(123) I]-metaiodobenzylguanidine cardiac scintigraphy*. *Psychiatry Clin Neurosci*, 2018. **72**(6): p. 423-434.
 39. Iranzo, A., et al., *Dopamine transporter imaging deficit predicts early transition to synucleinopathy in idiopathic rapid eye movement sleep behavior disorder*. *Ann Neurol*, 2017. **82**(3): p. 419-428.
 40. Kasanuki, K., et al., *123I-FP-CIT SPECT findings and its clinical relevance in prodromal dementia with Lewy bodies*. *Eur J Nucl Med Mol Imaging*, 2017. **44**(3): p. 358-365.
 41. Tatsch, K. and G. Poepperl, *Nigrostriatal dopamine terminal imaging with dopamine transporter SPECT: an update*. *J Nucl Med*, 2013. **54**(8): p. 1331-8.
 42. Piggott, M.A., et al., *Striatal dopaminergic markers in dementia with Lewy bodies, Alzheimer's and Parkinson's diseases: rostrocaudal distribution*. *Brain*, 1999. **122 (Pt 8)**: p. 1449-68.

43. 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.
44. O'Brien, J.T., et al., *Dopamine transporter loss visualized with FP-CIT SPECT in the differential diagnosis of dementia with Lewy bodies*. Arch Neurol, 2004. **61**(6): p. 919-25.
45. O'Brien, J.T., et al., *Diagnostic accuracy of 123I-FP-CIT SPECT in possible dementia with Lewy bodies*. Br J Psychiatry, 2009. **194**(1): p. 34-9.
46. Walker, R.W. and Z. Walker, *Dopamine transporter single photon emission computerized tomography in the diagnosis of dementia with Lewy bodies*. Mov Disord, 2009. **24 Suppl 2**: p. S754-9.
47. Siepel, F.J., et al., *Loss of Dopamine Transporter Binding and Clinical Symptoms in Dementia With Lewy Bodies*. Mov Disord, 2016. **31**(1): p. 118-25.
48. 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.
49. O'Brien, J.T., et al., *Is ioflupane 123 injection diagnostically effective in patients with movement disorders and dementia? Pooled analysis of four clinical trials*. BMJ Open, 2014. **4**(7): p. e005122.
50. Pink, J., et al., *Dementia: assessment, management and support: summary of updated NICE guidance*. Bmj, 2018. **361**: p. k2438.
51. Chung, E.J. and S.J. Kim, *(123)I-Metaiodobenzylguanidine Myocardial Scintigraphy in Lewy Body-Related Disorders: A Literature Review*. J Mov Disord, 2015. **8**(2): p. 55-66.
52. Treglia, G., et al., *(1)(2)(3)I-MIBG cardiac scintigraphy in Lewy body-related disorders*. Mov Disord, 2011. **26**(10): p. 1949-5; author reply 1950.
53. 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.
54. 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.
55. 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.
56. Lebasnier, A., et al., *Potential diagnostic value of regional myocardial adrenergic imaging using (123)I-MIBG SPECT to identify patients with Lewy body diseases*. Eur J Nucl Med Mol Imaging, 2015. **42**(7): p. 1043-51.
57. Nakajima, K., et al., *Iodine-123-MIBG sympathetic imaging in Lewy-body diseases and related movement disorders*. Q J Nucl Med Mol Imaging, 2008. **52**(4): p. 378-87.
58. Yoshita, M., et al., *Value of 123I-MIBG radioactivity in the differential diagnosis of DLB from AD*. Neurology, 2006. **66**(12): p. 1850-4.
59. Slaets, S., et al., *Diagnostic value of MIBG cardiac scintigraphy for differential dementia diagnosis*. Int J Geriatr Psychiatry, 2015. **30**(8): p. 864-9.
60. 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.
61. Treglia, G., et al., *Diagnostic performance of iodine-123-metaiodobenzylguanidine scintigraphy in differential diagnosis between Parkinson's disease and multiple-system atrophy: a systematic review and a meta-analysis*. Clin Neurol Neurosurg, 2011. **113**(10): p. 823-9.
62. 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.
63. Jacobson, A.F., et al., *Myocardial iodine-123 meta-iodobenzylguanidine imaging and cardiac events in heart failure. Results of the prospective ADMIRE-HF (AdreView Myocardial Imaging for Risk Evaluation in Heart Failure) study*. J Am Coll Cardiol, 2010. **55**(20): p. 2212-21.

64. Verberne, H.J., et al., *Prognostic value of myocardial 123I-metaiodobenzylguanidine (MIBG) parameters in patients with heart failure: a systematic review*. Eur Heart J, 2008. **29**(9): p. 1147-59.
65. Travin, M.I., et al., *How do we establish cardiac sympathetic nervous system imaging with (123)I-MIBG in clinical practice? Perspectives and lessons from Japan and the US*. J Nucl Cardiol, 2018.
66. Orimo, S., [The clinical significance of MIBG myocardial scintigraphy in Parkinson disease]. Brain Nerve, 2012. **64**(4): p. 403-12.
67. 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.
68. 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.
69. Bostrom, F., et al., *Cerebrospinal fluid total tau is associated with shorter survival in dementia with Lewy bodies*. Dement Geriatr Cogn Disord, 2009. **28**(4): p. 314-9.
70. Shimizu, S., et al., *Utility of the combination of DAT SPECT and MIBG myocardial scintigraphy in differentiating dementia with Lewy bodies from Alzheimer's disease*. Eur J Nucl Med Mol Imaging, 2016. **43**(1): p. 184-92.
71. Orimo, S., et al., *Degeneration of cardiac sympathetic nerve begins in the early disease process of Parkinson's disease*. Brain Pathol, 2007. **17**(1): p. 24-30.
72. Zaccai, J., et al., *Patterns and stages of alpha-synucleinopathy: Relevance in a population-based cohort*. Neurology, 2008. **70**(13): p. 1042-8.
73. Walker, Z., et al., *Evolution of clinical features in possible DLB depending on FP-CIT SPECT result*. Neurology, 2016. **87**(10): p. 1045-51.
74. Walker, Z., et al., *Clinical usefulness of dopamine transporter SPECT imaging with 123I-FP-CIT in patients with possible dementia with Lewy bodies: randomised study*. Br J Psychiatry, 2015. **206**(2): p. 145-52.
75. Siepel, F.J., et al., *(123I)FP-CIT SPECT in suspected dementia with Lewy bodies: a longitudinal case study*. BMJ Open, 2013. **3**(4).
76. van der Zande, J.J., et al., *[(123)I]FP-CIT SPECT scans initially rated as normal became abnormal over time in patients with probable dementia with Lewy bodies*. Eur J Nucl Med Mol Imaging, 2016. **43**(6): p. 1060-6.
77. Durcan, R., et al., *Imaging in prodromal dementia with Lewy bodies: Where do we stand?* Int J Geriatr Psychiatry, 2019. **34**(5): p. 635-646.
78. Walker, Z., et al., *Dementia with Lewy bodies: a comparison of clinical diagnosis, FP-CIT single photon emission computed tomography imaging and autopsy*. J Neurol Neurosurg Psychiatry, 2007. **78**(11): p. 1176-81.
79. 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.
80. Jung, Y., et al., *Clinicopathological and (123)I-FP-CIT SPECT correlations in patients with dementia*. Ann Clin Transl Neurol, 2018. **5**(3): p. 376-381.
81. Beach, T.G., et al., *Multi-organ distribution of phosphorylated alpha-synuclein histopathology in subjects with Lewy body disorders*. Acta Neuropathol, 2010. **119**(6): p. 689-702.
82. Takahashi, M., et al., *Quantitative correlation between cardiac MIBG uptake and remaining axons in the cardiac sympathetic nerve in Lewy body disease*. J Neurol Neurosurg Psychiatry, 2015. **86**(9): p. 939-44.
83. 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.
84. Donaghy, P.C., et al., *Neuropsychiatric symptoms and cognitive profile in mild cognitive impairment with Lewy bodies*. Psychol Med, 2018. **48**(14): p. 2384-2390.

85. Gillen, R., et al., *CT-based attenuation and scatter correction compared with uniform attenuation correction in brain perfusion SPECT imaging for dementia*. *Phys Med Biol*, 2015. **60**(17): p. 6775-87.
86. Bailey, D.L. and K.P. Willowson, *An Evidence-Based Review of Quantitative SPECT Imaging and Potential Clinical Applications*. *Journal of Nuclear Medicine*, 2013. **54**(1): p. 83-89.
87. Bailey, D.L. and K.P. Willowson, *Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality*. *European Journal of Nuclear Medicine and Molecular Imaging*, 2014. **41**: p. S17-S25.
88. Dickson, J., J. Ross, and S. Voo, *Quantitative SPECT: the time is now*. *EJNMMI Phys*, 2019. **6**(1): p. 4.
89. Willowson, K., D.L. Bailey, and C. Baldock, *Quantitative SPECT reconstruction using CT-derived corrections*. *Phys Med Biol*, 2008. **53**(12): p. 3099-112.
90. IAEA, *Nuclear medicine physics : a handbook for students and teachers*. 2014: International Atomic Energy Authority.
91. IAEA, *Quantitative nuclear medicine imaging : concepts, requirements and methods*. 2014, Vienna: International Atomic Energy Agency.
92. Hutton, B.F., I. Buvat, and F.J. Beekman, *Review and current status of SPECT scatter correction*. *Physics in Medicine and Biology*, 2011. **56**(14): p. R85-R112.
93. Frey, E.C., J.L. Humm, and M. Ljungberg, *Accuracy and Precision of Radioactivity Quantification in Nuclear Medicine Images*. *Seminars in Nuclear Medicine*, 2012. **42**(3): p. 208-218.
94. Ljungberg, M., *Absolute Quantitation of SPECT Studies*. *Semin Nucl Med*, 2018. **48**(4): p. 348-358.
95. Gilliland, D.R., et al., *Quantitative SPECT reconstruction of iodine-123 data*. *J Nucl Med*, 1991. **32**(3): p. 527-33.
96. Hudson, H.M. and R.S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data*. *IEEE Trans Med Imaging*, 1994. **13**(4): p. 601-9.
97. Koch, W., et al., *Is iterative reconstruction an alternative to filtered backprojection in routine processing of dopamine transporter SPECT studies?* *J Nucl Med*, 2005. **46**(11): p. 1804-11.
98. Dickson, J.C., et al., *The impact of reconstruction method on the quantification of DaTSCAN images*. *Eur J Nucl Med Mol Imaging*, 2010. **37**(1): p. 23-35.
99. Hutton, B.F., *The origins of SPECT and SPECT/CT*. *Eur J Nucl Med Mol Imaging*, 2014. **41 Suppl 1**: p. S3-16.
100. Brown, S., et al., *Investigation of the relationship between linear attenuation coefficients and CT Hounsfield units using radionuclides for SPECT*. *Appl Radiat Isot*, 2008. **66**(9): p. 1206-12.
101. Fleming, J.S., *Technique for the use of standard outlines for attenuation correction and quantification in SPECT*. *Nucl Med Commun*, 1990. **11**(10): p. 685-96.
102. Lange, C., et al., *CT-based attenuation correction in I-123-ioflupane SPECT*. *PLoS One*, 2014. **9**(9): p. e108328.
103. Papanastasiou, E., E. Moralidis, and A. Siountas, *The effect of scatter correction on planar and tomographic semiquantitative (123)I cardiac imaging. A phantom study*. *Hell J Nucl Med*, 2017. **20**(2): p. 154-159.
104. Lagerburg, V., et al., *A comparison of different energy window subtraction methods to correct for scatter and downscatter in I-123 SPECT imaging*. *Nuclear Medicine Communications*, 2012. **33**(7): p. 708-718.
105. Inoue, Y., et al., *Acquisition protocols and correction methods for estimation of the heart-to-mediastinum ratio in 123I-metaiodobenzylguanidine cardiac sympathetic imaging*. *J Nucl Med*, 2013. **54**(5): p. 707-13.
106. Bexelius, T. and A. Sohlberg, *Implementation of GPU accelerated SPECT reconstruction with Monte Carlo-based scatter correction*. *Ann Nucl Med*, 2018. **32**(5): p. 337-347.
107. Sohlberg, A., H. Watabe, and H. Iida, *Acceleration of Monte Carlo-based scatter compensation for cardiac SPECT*. *Phys Med Biol*, 2008. **53**(14): p. N277-85.
108. Shcherbinin, S., et al., *Accuracy of quantitative reconstructions in SPECT/CT imaging*. *Phys Med Biol*, 2008. **53**(17): p. 4595-604.

109. Robinson, A.P., et al., *The influence of triple energy window scatter correction on activity quantification for Lu-177 molecular radiotherapy*. *Physics in Medicine and Biology*, 2016. **61**(14): p. 5107-5127.
110. Cherry, S., G.L. Sorensen, and M.E. Phelps, *Physics in nuclear medicine*. 4th ed. 2012: Elsevier.
111. Varrone, A., et al., *European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis*. *Eur J Nucl Med Mol Imaging*, 2013. **40**(2): p. 213-27.
112. Djang, D.S., et al., *SNM practice guideline for dopamine transporter imaging with 123I-ioflupane SPECT 1.0*. *J Nucl Med*, 2012. **53**(1): p. 154-63.
113. Darcourt, J., et al., *EANM procedure guidelines for brain neurotransmission SPECT using (123)I-labelled dopamine transporter ligands, version 2*. *Eur J Nucl Med Mol Imaging*, 2010. **37**(2): p. 443-50.
114. Macey, D.J., et al., *Comparison of low- and medium-energy collimators for SPECT imaging with iodine-123-labeled antibodies*. *J Nucl Med*, 1986. **27**(9): p. 1467-74.
115. Dobbeleir, A.A., A.S. Hambye, and P.R. Franken, *Influence of high-energy photons on the spectrum of iodine-123 with low- and medium-energy collimators: consequences for imaging with 123I-labelled compounds in clinical practice*. *Eur J Nucl Med*, 1999. **26**(6): p. 655-8.
116. Fletcher, A.M., et al., *I-123 MIBG cardiac uptake measurements: limitations of collimator choice and scatter correction in the clinical context*. *Nucl Med Commun*, 2010. **31**(7): p. 629-36.
117. Inoue, Y., et al., *Effect of collimator choice on quantitative assessment of cardiac iodine 123 MIBG uptake*. *J Nucl Cardiol*, 2003. **10**(6): p. 623-32.
118. Inoue, Y., et al., *Septal penetration in iodine-123 metaiodobenzylguanidine cardiac sympathetic imaging using a medium-energy collimator*. *J Nucl Cardiol*, 2014. **21**(1): p. 71-7.
119. Inoue, H., et al., *Acquisition protocols and correction methods for estimation of the heart-to-mediastinum ratio in 123I-metaiodobenzylguanidine cardiac sympathetic imaging*. *J Nucl Med*, 2013. **54**(5): p. 12.
120. 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.
121. Inoue Y, et al., *Septal penetration in iodine-123 metaiodobenzylguanidine cardiac sympathetic imaging using a medium-energy collimator*. *J Nucl Cardiol*, 2014. **21**(1): p. 7.
122. Sohlberg, A., H. Watabe, and H. Iida, *Three-dimensional SPECT reconstruction with transmission-dependent scatter correction*. *Ann Nucl Med*, 2008. **22**(7): p. 549-56.
123. Kangasmaa, T., et al., *Half-time myocardial perfusion SPECT imaging with attenuation and Monte Carlo-based scatter correction*. *European Journal of Nuclear Medicine and Molecular Imaging*, 2011. **38**: p. S204-S204.
124. Kangasmaa, T.S., C. Constable, and A.O. Sohlberg, *Evaluation of quantitative 123I and 131I SPECT with Monte Carlo-based down-scatter compensation*. *Nucl Med Commun*, 2018. **39**(12): p. 1097-1102.
125. Sohlberg, A.O. and M.T. Kajaste, *Fast Monte Carlo-simulator with full collimator and detector response modelling for SPECT*. *Annals of Nuclear Medicine*, 2012. **26**(1): p. 92-98.
126. Kangasmaa, T.S., et al., *Multicenter evaluation of single-photon emission computed tomography quantification with third-party reconstruction software*. *Nucl Med Commun*, 2016. **37**(9): p. 983-7.
127. Brendle, C., et al., *Is the standard uptake value (SUV) appropriate for quantification in clinical PET imaging? - Variability induced by different SUV measurements and varying reconstruction methods*. *Eur J Radiol*, 2015. **84**(1): p. 158-162.
128. Mercieca, S., et al., *Comparison of SUVmax and SUVpeak based segmentation to determine primary lung tumour volume on FDG PET-CT correlated with pathology data*. *Radiother Oncol*, 2018. **129**(2): p. 227-233.
129. Akamatsu, G., et al., *Influence of Statistical Fluctuation on Reproducibility and Accuracy of SUVmax and SUVpeak: A Phantom Study*. *J Nucl Med Technol*, 2015. **43**(3): p. 222-6.

130. Laffon, E., et al., *SUVpeak Performance in Lung Cancer: Comparison to Average SUV from the 40 Hottest Voxels*. J Nucl Med, 2016. **57**(1): p. 85-8.
131. Laffon, E., et al., *Variability of average SUV from several hottest voxels is lower than that of SUVmax and SUVpeak*. Eur Radiol, 2014. **24**(8): p. 1964-70.
132. McCleery, J., et al., *Dopamine transporter imaging for the diagnosis of dementia with Lewy bodies*. Cochrane Database Syst Rev, 2015. **1**: p. Cd010633.
133. DC, C., et al., *Dementia with Lewy bodies versus Alzheimer's disease: role of dopamine transporter imaging*. Mov Disord, 2003. **18**(Suppl 7): p. S34-8.
134. Walker Z, et al., *Striatal dopamine transporter in dementia with Lewy bodies and Parkinson disease: a comparison*. Neurology, 2004. **62**(9): p. 5.
135. Pencharz, D.R., et al., *Automated quantification with BRASS reduces equivocal reporting of DaTSCAN (123I-FP-CIT) SPECT studies*. Nucl Med Rev Cent East Eur, 2014. **17**(2): p. 65-9.
136. Soderlund, T.A., et al., *Value of semiquantitative analysis for clinical reporting of 123I-2-beta-carbomethoxy-3beta-(4-iodophenyl)-N-(3-fluoropropyl)nortropine SPECT studies*. J Nucl Med, 2013. **54**(5): p. 714-22.
137. Booij, J., et al., *Diagnostic Performance of the Visual Reading of (123)I-Ioflupane SPECT Images With or Without Quantification in Patients With Movement Disorders or Dementia*. J Nucl Med, 2017. **58**(11): p. 1821-1826.
138. Nicastro, N., et al., *Added Value of Combined Semi-Quantitative and Visual [123I]FP-CIT SPECT Analyses for the Diagnosis of Dementia With Lewy Bodies*. Clin Nucl Med, 2017. **42**(2): p. e96-e102.
139. Werner, R.A., et al., *Visual and Semiquantitative Accuracy in Clinical Baseline 123I-Ioflupane SPECT/CT Imaging*. Clin Nucl Med, 2019. **44**(1): p. 1-3.
140. Koch W, et al., *Clinical testing of an optimized software solution for an automated, observer independent evaluation of dopamine transporter SPECT studies*. J Nucl Med, 2005. **46**(7): p. 10.
141. Calvini, P., et al., *The basal ganglia matching tools package for striatal uptake semi-quantification: description and validation*. Eur J Nucl Med Mol Imaging, 2007. **34**(8): p. 1240-53.
142. Nobili, F., et al., *Automatic semi-quantification of [123I]FP-CIT SPECT scans in healthy volunteers using BasGan version 2: results from the ENC-DAT database*. Eur J Nucl Med Mol Imaging, 2013. **40**(4): p. 565-73.
143. Dickson, J.C., et al., *The impact of reconstruction and scanner characterisation on the diagnostic capability of a normal database for [(123)I]FP-CIT SPECT imaging*. EJNMMI Res, 2017. **7**(1): p. 10.
144. Tossici-Bolt L, et al., *Quantification of [123I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio*. Eur J Nucl Med Mol Imaging, 2006. **33**(12).
145. Iwabuchi, Y., et al., *Impact of a combination of quantitative indices representing uptake intensity, shape, and asymmetry in DAT SPECT using machine learning: comparison of different volume of interest settings*. EJNMMI Res, 2019. **9**(1): p. 7.
146. Buchert, R., et al., *Diagnostic performance of the specific uptake size index for semi-quantitative analysis of I-123-FP-CIT SPECT: harmonized multi-center research setting versus typical clinical single-camera setting*. EJNMMI Res, 2019. **9**(1): p. 37.
147. Crespo, C., et al., *Quantification of dopaminergic neurotransmission SPECT studies with 123I-labelled radioligands. A comparison between different imaging systems and data acquisition protocols using Monte Carlo simulation*. Eur J Nucl Med Mol Imaging, 2008. **35**(7): p. 1334-42.
148. Gallego, J., et al., *Validation of semi-quantitative methods for DAT SPECT: influence of anatomical variability and partial volume effect*. Physics in Medicine and Biology, 2015. **60**(15): p. 5925-5938.
149. Furuta, A., et al., *Impact of quantitative index derived from 123I-FP-CIT-SPECT on reconstruction with correction methods evaluated using a 3D-striatum digital brain phantom*. Radiol Phys Technol, 2018.
150. Furuta, A., H. Onishi, and K. Nakamoto, *Development of Realistic Striatal Digital Brain (SDB) Phantom for (123)I-FP-CIT SPECT and Effect on Ventricle in the Brain for Semi-quantitative*

- Index of Specific Binding Ratio*. Nihon Hoshasen Gijutsu Gakkai Zasshi, 2017. **73**(10): p. 1018-1027.
151. Furuta, A., H. Onishi, and H. Amijima, *Quantitation of specific binding ratio in (123)I-FP-CIT SPECT: accurate processing strategy for cerebral ventricular enlargement with use of 3D-striatal digital brain phantom*. Radiol Phys Technol, 2018. **11**(2): p. 219-227.
 152. Varrone A, et al., *European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis*. Eur J Nucl Med Mol Imaging, 2013. **40**(2): p. 15.
 153. Lavalaye J, et al., *Effect of age and gender on dopamine transporter imaging with [123I]FP-CIT SPET in healthy volunteers*. Eur J Nucl Med Mol Imaging, 2000. **27**(7): p. 3.
 154. Morton RJ, et al., *Variation of DaTSCAN quantification between different gamma camera types*. Nucl Med Commun, 2005. **26**(12).
 155. Soderlund TA, et al., *Value of semiquantitative analysis for clinical reporting of 123I-2-Î²-carbomethoxy-3Î²-(4-iodophenyl)-N-(3-fluoropropyl)nortropane SPECT studies*. J Nucl Med, 2013. **54**(5): p. 9.
 156. Dickson JC, et al., *The impact of reconstruction method on the quantification of DaTSCAN images*. Eur J Nucl Med Mol Imaging, 2010. **37**(1): p. 13.
 157. Kita, A., et al., *Standardization of the specific binding ratio in [123I]FP-CIT SPECT: study by striatum phantom*. Nucl Med Commun, 2019. **40**(5): p. 484-490.
 158. Tossici-Bolt, L., et al., *Calibration of gamma camera systems for a multicentre European I-123-FP-CIT SPECT normal database*. European Journal of Nuclear Medicine and Molecular Imaging, 2011. **38**(8): p. 1529-1540.
 159. Brogley, J.E., *DaTQUANT: The Future of Diagnosing Parkinson Disease*. J Nucl Med Technol, 2019. **47**(1): p. 21-26.
 160. Tossici-Bolt, L., et al., *[[123I]FP-CIT ENC-DAT normal database: the impact of the reconstruction and quantification methods*. EJNMMI Phys, 2017. **4**(1): p. 8.
 161. Lapa, C., et al., *Influence of CT-based attenuation correction on dopamine transporter SPECT with [[123I]FP-CIT*. Am J Nucl Med Mol Imaging, 2015. **5**(3): p. 278-86.
 162. Warwick, J.M., et al., *The Role of CT-Based Attenuation Correction and Collimator Blurring Correction in Striatal Spect Quantification*. Int J Mol Imaging, 2011. **2011**: p. 195037.
 163. Maebatake, A., et al., *An anthropomorphic phantom study of brain dopamine transporter SPECT images obtained using different SPECT/CT devices and collimators*. J Nucl Med Technol, 2015. **43**(1): p. 41-6.
 164. Yokoyama, K., et al., *Computed-tomography-guided anatomic standardization for quantitative assessment of dopamine transporter SPECT*. Eur J Nucl Med Mol Imaging, 2017. **44**(3): p. 366-372.
 165. 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.
 166. 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.
 167. Verschure, D.O., et al., *A European myocardial 123I-MIBG cross-calibration phantom study*. J Nucl Cardiol, 2017.
 168. 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.
 169. 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.
 170. 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.

171. 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.
172. Klene, C., et al., *Influence of ROI definition on the heart-to-mediastinum ratio in planar 123I-MIBG imaging*. J Nucl Cardiol, 2016.
173. 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.
174. 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.
175. Dimitriu-Leen AC, Scholte AJ, and J. AF., *123I-MIBG SPECT for Evaluation of Patients with Heart Failure*. J Nucl Med, 2015. **56**(Suppl 4): p. 25S - 30S.
176. van der Veen BJ, et al., *Assessment of global cardiac I-123 MIBG uptake and washout using volumetric quantification of SPECT acquisitions*. J Nucl Cardiol, 2012. **19**(4): p. 11.
177. Chen J, et al., *Quantitative I-123 MIBG SPECT in differentiating abnormal and normal MIBG myocardial uptake*. J Nucl Cardiol, 2012. **19**(1): p. 8.
178. Alvi, R., et al., *Quantification and Determination of Normal (123)I-Meta Iodobenzylguanidine Heart-to-Mediastinum Ratio (HMR) from Cardiac SPECT/CT and Correlation with Planar HMR*. J Nucl Med, 2018. **59**(4): p. 652-658.
179. Oh JK, et al., *Comparison of I-123 MIBG planar imaging and SPECT for the detection of decreased heart uptake in Parkinson disease*. J Neural Transm (Vienna), 2015.
180. Somsen GA, et al., *Cardiac 123I-MIBG uptake is affected by variable uptake in reference regions: implications for interpretation in clinical studies*. Nucl Med Commun, 1996. **17**: p. 5.
181. Jacobson AF, et al., *Impact of age on myocardial uptake of 123I-mIBG in older adult subjects without coronary heart disease*. J Nucl Cardiol, 2013. **20**: p. 9.
182. Inoue Y, et al., *Validity of the mediastinum as a reference region to evaluate cardiac accumulation of iodine-123 metaiodobenzylguanidine*. Ann Nucl Med, 2015. **29**(7): p. 8.
183. Oide, T., et al., *Usefulness of [123I]metaiodobenzylguanidine ([123I]MIBG) myocardial scintigraphy in differentiating between Alzheimer's disease and dementia with Lewy bodies*. Intern Med, 2003. **42**(8): p. 686-90.
184. Novellino, F., et al., *Myocardial (123)I-MIBG scintigraphy for differentiation of Lewy bodies disease from FTD*. Neurobiol Aging, 2010. **31**(11): p. 1903-11.
185. Nakajima, K., et al., *Cardiac sympathetic nervous system imaging with (123)I-meta-iodobenzylguanidine: Perspectives from Japan and Europe*. J Nucl Cardiol, 2017. **24**(3): p. 952-960.
186. Nakajima, K., et al., *Multicenter cross-calibration of I-123 metaiodobenzylguanidine heart-to-mediastinum ratios to overcome camera-collimator variations*. J Nucl Cardiol, 2014. **21**(5): p. 970-8.
187. Nakajima, K., et al., *Correction of iodine-123-labeled meta-iodobenzylguanidine uptake with multi-window methods for standardization of the heart-to-mediastinum ratio*. J Nucl Cardiol, 2007. **14**(6): p. 843-51.
188. Nakajima, K., et al., *Normal values and standardization of parameters in nuclear cardiology: Japanese Society of Nuclear Medicine working group database*. Ann Nucl Med, 2016. **30**(3): p. 188-99.
189. Matsuo, S., et al., *Characterization of Japanese standards for myocardial sympathetic and metabolic imaging in comparison with perfusion imaging*. Ann Nucl Med, 2009. **23**(6): p. 517-22.
190. Roberts, G., et al., *A comparison of visual and semiquantitative analysis methods for planar cardiac 123I-MIBG scintigraphy in dementia with Lewy bodies*. Nucl Med Commun, 2019.
191. Nakajima, K. and M. Yamada, *(123)I-Meta-iodobenzylguanidine Sympathetic Imaging: Standardization and Application to Neurological Diseases*. Chonnam Med J, 2016. **52**(3): p. 145-50.

192. van der Veen, L., A. Scholte, and M. Stokkel, *Mathematical methods to determine quantitative parameters of myocardial 123I-MIBG studies: a review of the literature*. Nucl Med Commun, 2010. **31**(7): p. 617-28.
193. Verberne, H.J., et al., *Variations in 123I-metaiodobenzylguanidine (MIBG) late heart mediastinal ratios in chronic heart failure: a need for standardisation and validation*. Eur J Nucl Med Mol Imaging, 2008. **35**(3): p. 547-53.
194. Benamer, T.S., et al., *Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group*. Mov Disord, 2000. **15**(3): p. 503-10.
195. Lloyd, J.J., et al., *A new visual rating scale for loflupane imaging in Lewy body disease*. Neuroimage Clin, 2018. **20**: p. 823-829.
196. Wakabayashi, Y., et al., *Semi-quantitative dopamine transporter standardized uptake value in comparison with conventional specific binding ratio in [123I] FP-CIT single-photon emission computed tomography (DaTscan)*. Neurol Sci, 2018. **39**(8): p. 1401-1407.
197. Toriihara, A., et al., *Semiquantitative analysis using standardized uptake value in (123)I-FP-CIT SPECT/CT*. Clin Imaging, 2018. **52**: p. 57-61.
198. Akahoshi, M., et al., *Attenuation and scatter correction in I-123 FP-CIT SPECT do not affect the clinical diagnosis of dopaminergic system neurodegeneration*. Medicine (Baltimore), 2017. **96**(45): p. e8484.
199. Tossici-Bolt, L., et al., *Calibration of gamma camera systems for a multicentre European (1)(2)(3)I-FP-CIT SPECT normal database*. Eur J Nucl Med Mol Imaging, 2011. **38**(8): p. 1529-40.
200. Karrer, T.M., et al., *Reduced dopamine receptors and transporters but not synthesis capacity in normal aging adults: a meta-analysis*. Neurobiol Aging, 2017. **57**: p. 36-46.
201. Booij, J., et al., *Imaging of dopamine transporters with iodine-123-FP-CIT SPECT in healthy controls and patients with Parkinson's disease*. J Nucl Med, 1998. **39**(11): p. 1879-84.
202. Matsuoka, K., et al., *Test-retest reproducibility of extrastriatal binding with (123)I-FP-CIT SPECT in healthy male subjects*. Psychiatry Res Neuroimaging, 2016. **258**: p. 10-15.
203. Tsuchida, T., et al., *Reproducibility of dopamine transporter density measured with 123I-FPCIT SPECT in normal control and Parkinson's disease patients*. Ann Nucl Med, 2004. **18**(7): p. 609-16.
204. Colloby, S.J., et al., *Progression of dopaminergic degeneration in dementia with Lewy bodies and Parkinson's disease with and without dementia assessed using 123I-FP-CIT SPECT*. Eur J Nucl Med Mol Imaging, 2005. **32**(10): p. 1176-85.
205. Jakobson Mo, S., et al., *Pre- and postsynaptic dopamine SPECT in idiopathic Parkinsonian diseases: a follow-up study*. Biomed Res Int, 2013. **2013**: p. 143532.
206. Jakobson Mo, S., et al., *(1)(2)(3)I-FP-Cit and 123I-IBZM SPECT uptake in a prospective normal material analysed with two different semiquantitative image evaluation tools*. Nucl Med Commun, 2013. **34**(10): p. 978-89.
207. 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.
208. 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.
209. 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.
210. Lee, K., et al., *Characteristics and regulation of 123I-MIBG transport in cultured pulmonary endothelial cells*. J Nucl Med, 2006. **47**: p. 437-42.
211. 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.
212. 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.

213. Miller, J.W., et al., *Effect of l-dopa on plasma homocysteine in PD patients: Relationship to B-vitamin status*. *Neurology*, 2003. **60**(7).
214. Waqar, F. and M.C. Gerson, *Cardiac sympathetic imaging in the diagnosis of cardiac autonomic neuropathy in pre-diabetes*. *J Nucl Cardiol*, 2015. **22**(6): p. 1269-72.
215. Asghar, O., et al., *Individuals with impaired glucose tolerance demonstrate normal cardiac sympathetic innervation using I-123 mIBG scintigraphy*. *J Nucl Cardiol*, 2015. **22**(6): p. 1262-8.
216. Asghar, O., et al., *Iodine-123 metaiodobenzylguanidine scintigraphy for the assessment of cardiac sympathetic innervation and the relationship with cardiac autonomic function in healthy adults using standardized methods*. *Nucl Med Commun*, 2017. **38**(1): p. 44-50.
217. Orimo, S., et al., *(123)I-meta-iodobenzylguanidine (MIBG) cardiac scintigraphy in alpha-synucleinopathies*. *Ageing Res Rev*, 2016. **30**: p. 122-33.
218. Jacobson, A.F., et al., *Impact of age on myocardial uptake of (1)(2)(3)I-mIBG in older adult subjects without coronary heart disease*. *J Nucl Cardiol*, 2013. **20**(3): p. 406-14.
219. Nakajima, K., et al., *The time has come to standardize (123)I-MIBG heart-to-mediastinum ratios including planar and SPECT methods*. *Eur J Nucl Med Mol Imaging*, 2016. **43**(2): p. 386-8.
220. Jacobson, A.F., et al., *123I-mIBG scintigraphy to predict risk for adverse cardiac outcomes in heart failure patients: design of two prospective multicenter international trials*. *J Nucl Cardiol*, 2009. **16**(1): p. 113-21.
221. Nakajima, K., et al., *Is (123)I-metaiodobenzylguanidine heart-to-mediastinum ratio dependent on age? From Japanese Society of Nuclear Medicine normal database*. *Ann Nucl Med*, 2018. **32**(3): p. 175-181.
222. Nakata, T., et al., *Regional cardiac sympathetic nerve dysfunction and the diagnostic efficacy of metaiodobenzylguanidine tomography in stable coronary artery disease*. *Am J Cardiol*, 1996. **78**(3): p. 292-7.
223. Travin, M.I., *Potential for adrenergic imaging to serve as a unique tool for guidance of patient management during and after an acute ischemic event*. *J Nucl Cardiol*, 2018. **25**(2): p. 581-585.
224. Fallavollita, J.A. and J.M. Canty, Jr., *Dysinnervated but viable myocardium in ischemic heart disease*. *J Nucl Cardiol*, 2010. **17**(6): p. 1107-15.
225. Hinton, W., et al., *Incidence and prevalence of cardiovascular disease in English primary care: a cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC)*. *BMJ Open*, 2018. **8**(8): p. e020282.
226. Digital, N., *Quality and Outcomes Framework, Achievement, prevalence and exceptions data - 2017-18*. Quality and Outcomes Framework (QOF), 2018.
227. Kitamura, A., et al., *Trends in the incidence of coronary heart disease and stroke and their risk factors in Japan, 1964 to 2003: the Akita-Osaka study*. *J Am Coll Cardiol*, 2008. **52**(1): p. 71-9.
228. Pellegrino, T., et al., *Impact of obesity and acquisition protocol on (123)I-metaiodobenzylguanidine indexes of cardiac sympathetic innervation*. *Quant Imaging Med Surg*, 2015. **5**(6): p. 822-8.
229. Bradbury, K.E., et al., *Association between physical activity and body fat percentage, with adjustment for BMI: a large cross-sectional analysis of UK Biobank*. *BMJ Open*, 2017. **7**(3): p. e011843.
230. Cui, R., et al., *Body mass index and mortality from cardiovascular disease among Japanese men and women: the JACC study*. *Stroke*, 2005. **36**(7): p. 1377-82.
231. Rascol, O. and L. Schelosky, *123I-metaiodobenzylguanidine scintigraphy in Parkinson's disease and related disorders*. *Mov Disord*, 2009. **24 Suppl 2**: p. S732-41.
232. 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.
233. 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.
234. 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.

235. 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.
236. Inui, Y., et al., *Comparison of (123)I-MIBG myocardial scintigraphy, brain perfusion SPECT, and voxel-based MRI morphometry for distinguishing between dementia with Lewy bodies and Alzheimer's disease*. Ann Nucl Med, 2014. **28**(8): p. 796-804.
237. Wada-Isoe, K., et al., *Diagnostic markers for diagnosing dementia with Lewy bodies: CSF and MIBG cardiac scintigraphy study*. J Neurol Sci, 2007. **260**(1-2): p. 33-7.
238. Watanabe, H., et al., *Cardiac (123)I-meta-iodobenzylguanidine (MIBG) uptake in dementia with Lewy bodies: comparison with Alzheimer's disease*. J Neurol Neurosurg Psychiatry, 2001. **70**(6): p. 781-3.
239. Odagiri, H., et al., *On the Utility of MIBG SPECT/CT in Evaluating Cardiac Sympathetic Dysfunction in Lewy Body Diseases*. PLoS One, 2016. **11**(4): p. e0152746.
240. Tsutsui, Y., et al., *Multicentre analysis of PET SUV using vendor-neutral software: the Japanese Harmonization Technology (J-Hart) study*. EJNMMI Res, 2018. **8**(1): p. 83.
241. Sun, M., et al., *Prognostic value of pretreatment PET/CT lean body mass-corrected parameters in patients with hepatocellular carcinoma*. Nucl Med Commun, 2018. **39**(6): p. 564-571.
242. Decazes, P., et al., *A Method to Improve the Semiquantification of 18F-FDG Uptake: Reliability of the Estimated Lean Body Mass Using the Conventional, Low-Dose CT from PET/CT*. J Nucl Med, 2016. **57**(5): p. 753-8.
243. Janmahasatian, S., et al., *Quantification of lean bodyweight*. Clin Pharmacokinet, 2005. **44**(10): p. 1051-65.
244. Tahari, A.K., et al., *Optimum lean body formulation for correction of standardized uptake value in PET imaging*. J Nucl Med, 2014. **55**(9): p. 1481-4.
245. Frederic Courbon, et al., *Cardiac MIBG Scintigraphy Is a Sensitive Tool for Detecting Cardiac Sympathetic Denervation in Parkinson's Disease*. Movement Disorders, 2003. **18**(8): p. 890-897.
246. Shinsaku Tsuchimochi, et al., *Age and Gender Differences in Normal Myocardial Adrenergic Neuronal Function Evaluated by Iodine-123-MIBG Imaging*. J Nucl Med, 1995. **36**: p. 969-974.
247. Takakazu Morozumi, et al., *Myocardial Iodine-123 -Metaiodobenzylguanidine Images and Autonomic Nerve Activity in Normal Subjects*. J Nucl Med, 1997. **38**: p. 49-52.
248. Momose, M., et al., *How heterogeneous is the cardiac autonomic innervation?* Basic Res Cardiol, 2001. **96**: p. 539-546.
249. Kazuyuki Sakata, et al., *Physiological Fluctuation of the Human Left Ventricle Sympathetic Nervous System Assessed by Iodine- 123-MIBG*. J Nucl Med, 1996. **39**: p. 1667-1671.
250. Somsen, G.A., et al., *Normal values and within-subject variability of cardiac I-123 MIBG scintigraphy in healthy individuals: implications for clinical studies*. J Nucl Cardiol, 2004. **11**(2): p. 126-33.
251. Shimizu, M., et al., *Heterogeneity of cardiac sympathetic nerve activity and systolic dysfunction in patients with hypertrophic cardiomyopathy*. J Nucl Med, 2002. **43**(1): p. 15-20.
252. Gill, J.S., et al., *Heterogeneity of the human myocardial sympathetic innervation: in vivo demonstration by iodine 123-labeled meta-iodobenzylguanidine scintigraphy*. Am Heart J, 1993. **126**(2): p. 390-8.
253. Verberne, H.J., et al., *Influence of collimator choice and simulated clinical conditions on 123I-MIBG heart/mediastinum ratios: a phantom study*. Eur J Nucl Med Mol Imaging, 2005. **32**(9): p. 1100-7.
254. Verschure, D.O., et al., *(1)(2)(3)I-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. 279-85.
255. NICE, *Clinical Guideline 181: Cardiovascular disease: risk assessment and reduction, including lipid modification*, N.I.f.H.a.C. Excellence, Editor. 2014.
256. Jacobson, A.F. and M.I. Travin, *Impact of medications on mIBG uptake, with specific attention to the heart: Comprehensive review of the literature*. J Nucl Cardiol, 2015. **22**(5): p. 980-93.

257. Simula, S., et al., *Cardiac adrenergic innervation is affected in asymptomatic subjects with very early stage of coronary artery disease*. J Nucl Med, 2002. **43**(1): p. 1-7.
258. Sisson, J.C., et al., *Metaiodobenzylguanidine to map scintigraphically the adrenergic nervous system in man*. J Nucl Med, 1987. **28**(10): p. 1625-36.
259. Wong, K.K., et al., *Pattern of cardiac sympathetic denervation in idiopathic Parkinson disease studied with ¹¹C hydroxyephedrine PET*. Radiology, 2012. **265**(1): p. 240-7.
260. Crick, S.J., et al., *Innervation of the human cardiac conduction system. A quantitative immunohistochemical and histochemical study*. Circulation, 1994. **89**(4): p. 1697-708.
261. Kimura, K., M. Ieda, and K. Fukuda, *Development, maturation, and transdifferentiation of cardiac sympathetic nerves*. Circ Res, 2012. **110**(2): p. 325-36.
262. Kawano, H., R. Okada, and K. Yano, *Histological study on the distribution of autonomic nerves in the human heart*. Heart Vessels, 2003. **18**(1): p. 32-9.
263. Raffel, D.M., et al., *Radiolabeled phenethylguanidines: novel imaging agents for cardiac sympathetic neurons and adrenergic tumors*. J Med Chem, 2007. **50**(9): p. 2078-88.
264. Shibutani, T., et al., *Characteristics of iodine-123 IQ-SPECT/CT imaging compared with conventional SPECT/CT*. Ann Nucl Med, 2019. **33**(2): p. 103-111.
265. Matsunari, I., et al., *Iodine-123 metaiodobenzylguanidine imaging and carbon-11 hydroxyephedrine positron emission tomography compared in patients with left ventricular dysfunction*. Circ Cardiovasc Imaging, 2010. **3**(5): p. 595-603.
266. Yoshinaga, K., et al., *Prone-position acquisition of myocardial (123)I-metaiodobenzylguanidine (MIBG) SPECT reveals regional uptake similar to that found using (11)C-hydroxyephedrine PET/CT*. Ann Nucl Med, 2014. **28**(8): p. 761-9.
267. Roberts, G., et al., *(123)I-FP-CIT striatal binding ratios do not decrease significantly with age in older adults*. Ann Nucl Med, 2019. **33**(6): p. 434-443.
268. van Dyck, C.H., et al., *Age-related decline in dopamine transporters: analysis of striatal subregions, nonlinear effects, and hemispheric asymmetries*. Am J Geriatr Psychiatry, 2002. **10**(1): p. 36-43.
269. Matsuda, H., et al., *Japanese multicenter database of healthy controls for [(123)I]FP-CIT SPECT*. Eur J Nucl Med Mol Imaging, 2018. **45**(8): p. 1405-1416.
270. Koch, W., et al., *Extended studies of the striatal uptake of ^{99m}Tc-NC100697 in healthy volunteers*. J Nucl Med, 2007. **48**(1): p. 27-34.
271. Koch, W., et al., *Extrastriatal binding of [(1)(2)(3)I]FP-CIT in the thalamus and pons: gender and age dependencies assessed in a European multicentre database of healthy controls*. Eur J Nucl Med Mol Imaging, 2014. **41**(10): p. 1938-46.
272. *The Parkinson Progression Marker Initiative (PPMI)*. Prog Neurobiol, 2011. **95**(4): p. 629-35.
273. Albert, M.S., et al., *The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease*. Alzheimers Dement, 2011. **7**(3): p. 270-9.
274. Markesbery, W.R., et al., *Lewy body pathology in normal elderly subjects*. J Neuropathol Exp Neurol, 2009. **68**(7): p. 816-22.
275. Pinheiro, J.C. and D.M. Bates, *Mixed-Effects Models in S and S-PLUS*. 2000: Springer Science & Business Media. 528.
276. Yamamoto, H., et al., *Age-related effects and gender differences in Japanese healthy controls for [(123)I] FP-CIT SPECT*. Ann Nucl Med, 2017. **31**(5): p. 407-412.
277. Tossici-Bolt, L., et al., *Quantification of [123I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio*. Eur J Nucl Med Mol Imaging, 2006. **33**(12): p. 1491-9.
278. Estorch, M., et al., *Myocardial iodine-labeled metaiodobenzylguanidine 123 uptake relates to age*. J Nucl Cardiol, 1995. **2**(2 Pt 1): p. 126-32.
279. Tsuchimochi, S., et al., *Age and gender differences in normal myocardial adrenergic neuronal function evaluated by iodine-123-MIBG imaging*. J Nucl Med, 1995. **36**(6): p. 969-74.
280. Sakata, K., et al., *Physiological fluctuation of the human left ventricle sympathetic nervous system assessed by iodine-123-MIBG*. J Nucl Med, 1998. **39**(10): p. 1667-71.

281. Rengo, G., et al., *Impact of aging on cardiac sympathetic innervation measured by (123)I-MIBG imaging in patients with systolic heart failure*. Eur J Nucl Med Mol Imaging, 2016. **43**(13): p. 2392-2400.
282. Braak, H., et al., *Staging of brain pathology related to sporadic Parkinson's disease*. Neurobiol Aging, 2003. **24**(2): p. 197-211.
283. Dickson JC, T.-B.L., Sera T, Booij J, Ziebell M, Morbelli S,, B.T. Assenbaum-Nan S, Pagani M, Kapucu OL, Hesse S, Van Laere K, Darcourt, and V.A. J, Tatsch K, *The impact of reconstruction and scanner characterisation on the diagnostic capability of a normal database for [(123)I]FP-CIT SPECT imaging*. Ejnmmi Research, 2017. **7**(1): p. 10.
284. HermesMedicalSolutions, *Hybrid Recon Neurology Handbook version 3.0*, H.M. Solutions, Editor. 2018.
285. Sakamoto, F., et al., *Diagnosis of dementia with Lewy bodies: diagnostic performance of combined (1)(2)(3)I-IMP brain perfusion SPECT and (1)(2)(3)I-MIBG myocardial scintigraphy*. Ann Nucl Med, 2014. **28**(3): p. 203-11.
286. Fukui, T., et al., *Prevalence and clinical implication of microbleeds in dementia with lewy bodies in comparison with microbleeds in Alzheimer's disease*. Dement Geriatr Cogn Dis Extra, 2013. **3**(1): p. 148-60.
287. Camacho, V., et al., *Cardiac sympathetic impairment parallels nigrostriatal degeneration in Probable Dementia with Lewy Bodies*. Q J Nucl Med Mol Imaging, 2011. **55**(4): p. 476-83.
288. Buvat, I. and I. Castiglioni, *Monte Carlo simulations in SPET and PET*. Q J Nucl Med Mol Imaging, 2002. **46**: p. 48-61.
289. Ljungberg, M., S.E. Strand, and M.A. King, *Monte Carlo Calculations in Nuclear Medicine*. 2 ed. 2013: CRC Press.