



University
of Glasgow

Wallace, Hugh James (2023) *Advancing quantitative imaging of neuroblastoma*. PhD thesis.

<https://theses.gla.ac.uk/83630/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Advancing Quantitative Imaging of Neuroblastoma

Hugh James Wallace

MSc Clinical Physics

MSci Physics with Astrophysics

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy,

School of Medicine, College of Medical, Veterinary & Life Sciences

University of Glasgow, February 2023

Abstract

Neuroblastoma is an embryonal cancer that exhibits marked heterogeneity of presentation and prognosis, ranging in outcomes from amongst the poorest in paediatric oncology to spontaneous remission. Scintigraphy using ^{123}I -meta-Iodobenzylguanidine (MIBG) provides unique functional information and is commonly used to manage Neuroblastoma alongside MRI. This thesis advances the imaging of neuroblastoma through SPECT/MR integration and quantitative SPECT/CT optimisation.

A simulation study of Dixon Magnetic Resonance Attenuation Correction (MRAC) was undertaken using $^{99\text{m}}\text{Tc}$ -MDP SPECT/CT studies. These simulations showed that SPECT MRAC could achieve superior performance to PET/MR in cases without significant lung coverage. Observer studies were also undertaken to evaluate a novel ^{123}I -MIBG SPECT/MR fusion dataset and the impact of this technique on semi-quantitative scoring of Neuroblastoma. The observers identified multiple clinically significant findings when using SPECT/MR versus planar scintigraphy.

The ^{123}I imaging performance of Siemens parallel hole collimators was assessed using a custom precision sensitivity phantom. The contributions from photopeak and septal penetration were spectrally decomposed and the medium energy collimator shown to be optimal. Existing methodologies for optimisation were found to be inadequate for quantitative SPECT. A measurement-dependent methodology was proposed and used to optimise reconstruction parameters for ^{123}I -MIBG SPECT/CT.

Optimised parameters were applied to wholebody SPECT/CT scans of Neuroblastoma and patterns and ranges of uptake in the liver and brain were evaluated. Right lobe of liver was shown to provide a significantly lower and more statistically consistent physiological reference than the left lobe. Standardised Uptake Values (SUVs) scaled to lean body mass were found to be superior to body weight scaled SUVs and absolute activity concentrations. Normal ranges and limits of variation were recommended as quality control measures for quantitative scans of Neuroblastoma.

Lay Summary

The primary goal of this thesis is to advance imaging of the cancer Neuroblastoma through research of quantitative methods for Nuclear Medicine imaging, i.e. imaging methods from which clinically relevant numerical values can be derived. In Nuclear Medicine (a discipline that uses chemically labelled radioisotopes, typically gamma-ray emitters, for imaging and measurement) the primary radiopharmaceutical used in Neuroblastoma is ^{123}I -meta-Iodobenzylguanidine (MIBG) which is a noradrenaline analogue. Neuroblastoma cancer cells exhibit a preponderance of noradrenaline transporters and MIBG exploits these to specifically image these cells.

At the outset of the project, effectively all Nuclear Medicine departments used the technique of planar scintigraphy, where images akin to plain-film x-rays are acquired from in-front and behind a patient lying on the scanner table. These are generally not suitable for quantitative measurements due to a number of phenomena including gamma-ray absorption and scatter within the patient. However 3-dimensional (3D) images known as Single Photon Emission Computed Tomography (SPECT) can be generated by rotating the gamma-ray detectors around the patient. This is typically a slow process but allows for multiple corrections to be applied to the images to achieve a much closer representation of the true radiopharmaceutical uptake within the patient. When combined with an anatomical imaging technique such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) this can give clinicians a much better understanding of the location, intensity and nature of a site of uptake.

The primary correction required to visualise the true uptake is attenuation correction (AC); a correction for the gamma-rays lost to absorption and scattering within the patient. This is most readily achieved through an attenuation map derived from a spatially matched CT scan, as CT is essentially a map of the density of tissues within the patient. While combined SPECT/CT scanners were commercially available at the outset of this work, the Royal Hospital for Sick Children (RHSC) in Glasgow had only a standalone SPECT scanner. However, a recent service development had

permitted a novel same-day SPECT and MRI service for Neuroblastoma and therefore this work initially focused on combining SPECT and MRI for the purposes of Magnetic Resonance Attenuation Correction (MRAC).

Chapter 2 is a simulation of a basic technique for achieving MRAC. The simulated MRI technique produces separate images of the fat-based (e.g. white fat) and water-based soft tissues (e.g. muscles, organs) within the patient which are then assigned a value of attenuation representative of the average attenuation of that tissue type. This technique produces essentially no signal from bone or lung, so these tissues must be filled-in using image processing. The simplest such step suggested in the literature involves the filling of any thin holes (assumed to be bone) with attenuation equivalent to water-based soft tissue. The identification of lung requires more sophisticated steps and it was hypothesised that SPECT would not require these corrections, due to being less prone to attenuation effects than other modalities.

The simulation study reverse engineered MRI attenuation maps from CT scans acquired on hybrid SPECT/CT scanners so that the effect of the simulated attenuation map on the SPECT scan could be compared to the true AC supplied by the original CT. To allow for assessment of the relative error generated from soft tissues, bone and lung; changes to each of these regions were also applied separately. The study found that the simulation produced an average error of 12.5-15.1% in cases with lung tissue present and 8.2-11.0% in cases without lung tissue. Soft tissue averaging was found to contribute $\pm 0.4\%$, while replacement of bone with soft tissue contributed 2.4-3.6%. The remaining error was initially thought to be the result of simulating lung tissue as zero attenuation but was, surprisingly, found to contribute 6.0-8.3% even in cases with no lung tissue. This led to the discovery of a systematic error in the image processing related to the lowest CT value modelled as fat-based tissue. The simulation study concluded that best-case uptake errors achievable were 2.2-2.7% in cases with no lung tissues and 6.5-6.9% in cases with lung tissues. These figures were considered comparable or even superior to the attenuation errors achievable in other modalities and therefore even this simple MRAC technique showed promise for clinical utility.

A method of semi-quantitative image assessment was explored in Chapter 3. The

technique of image scoring allows observers to assign each patient image a numerical value, representative of the total disease burden. The Curie and SIOOPEN scores used in this work have been shown to be linked to patient prognosis in large clinical trials. The work consists of two single observer studies with 15 and 17 patients included in each. In both studies the observers were blinded to patient identity and scored all the planar nuclear medicine images in a single session and in a random order. In a separate session, the observers scored the combination of planar images and positionally matched SPECT and MRI images. The differences between the planar-only and planar with SPECT/MR were calculated and statistically assessed.

The first observer study made use of a local radiologist at the Royal Hospital for Children (RHC) in Glasgow and this found statistically significant differences in Curie and SIOOPEN scores of +1.0 and +1.1 score units respectively. Multiple clinically significant findings including better visualisation of uptake within the spine, abdomen and head were also identified. The second observer study involved an internationally recognised expert from another centre and tightened the criteria for inclusion to focus on a single time-point (diagnosis). This study was unable to find a statistically significant result for either score but identified many similar clinical significant findings. Multiple limitations of the semi-quantitative methodology were identified, most significantly the variability in observer scoring technique. The need for an objective, truly quantitative imaging technique was reinforced and this became the focus of the next part of the project.

As the work of Chapter 3 concluded, SPECT/CT became available locally at RHC and the project focus shifted from SPECT and MRI integration to quantitative imaging of Neuroblastoma using SPECT/CT. The goal of this work was to generate images that accurately reflect the concentration of gamma-ray emitters at each point within a patient and allow for numerical measurements of uptake in different lesions and structures.

The first step of this sub-project was to characterise the performance of each of the lead collimators used to image gamma-rays on the SPECT/CT system. Using a custom test object the sensitivity of each collimator was assessed globally and at a range of gamma-ray energies. These data showed that the Medium Energy Low

Penetration (MELP) collimator had only an 8.6% contribution from septal penetration (i.e. erroneous events caused by gamma-rays passing straight through the collimator) compared to the 42.6% and 53.4% measured for the typically recommended Low Energy All Purpose (LEAP) and Low Energy High Resolution (LEHR) collimators. Using a novel analytical technique, the MELP collimator was shown to have a higher relative sensitivity to true, useful gamma-ray events than the low energy collimators and was selected for further development of quantitative imaging.

Commercial SPECT test objects were then used to make measurements of the accuracy of uptake quantification in a range of fillable spheres, imaged with SPECT/CT using the MELP collimators. The total number of iterations and size of spatial filter used by the SPECT reconstruction algorithm were tested, as were a number of different measurement techniques. The maximum uptake and 1ml peak uptake were recommended as the optimal measurements and these yielded an optimal set of reconstruction parameters of 80 total iterations and a 15mm filter. These parameters avoided errors due to over-correction of uptake in larger spheres and balanced this with accurate uptake measurement in a large background region, equivalent to the liver.

Finally, these optimised parameters were used to reconstruct 50 scans from 13 patients and the distribution of liver and brain uptake was assessed in each. The goal of this work was to suggest an optimal methodology for scaling patient measurements and to establish normal ranges for uptake. Brain uptake was shown to be a cold background with very low uptake values and was best expressed in absolute units (kBq/ml). The uptake in left and right lobes of liver was shown to be systematically different with the right lobe of liver showing a more normal distribution across the set of patient scans. Scaling by lean body mass was shown to provide the most consistent measurements in the right lobe of liver and a normal range of 0.5-2.1 standard uptake values scaled to lean body mass (SUV/LBM) was calculated. A 22mm spherical region was found to be suitable for use in patients as young as 4 months of age. A variability limit of 0.2SUV/LBM in the right lobe of liver was recommended for as a quality control measure for quantitative work. This work could be used to move the Neuroblastoma community towards a set of common standards for quantitative disease burden assessment.

Table of Contents

Abstract	2
Lay Summary	3
Table of Contents	7
List of Tables	10
List of Figures	11
Acknowledgements	14
Author's Declaration	15
Definitions/Abbreviations	16
1 Introduction & Literature Review	19
1.1 Neuroblastoma	19
1.2 Scintigraphic Imaging	28
1.3 Neuroblastoma Imaging	32
1.4 Research Direction	34
2 A Simulation of MRI Attenuation Correction	35
2.1 Introduction	35
2.2 Methods and Materials	38
2.3 Results	40
2.4 Discussion	47
2.5 Conclusion	50
2.6 Reflection	50
3 Impact of Fused SPECT-MRI on Neuroblastoma Scores	51
3.1 Introduction	51
3.2 Experiment 1: Blinded Clinical Audit	54

3.2.1	Introduction	54
3.2.2	Methods and Materials	55
3.2.3	Results	58
3.2.4	Discussion	63
3.2.5	Conclusion	67
3.3	Experiment 2: Blinded Observer Study	67
3.3.1	Introduction	67
3.3.2	Methods and Materials	68
3.3.3	Results	70
3.3.4	Discussion	78
3.3.5	Conclusion	84
3.4	Reflection	84
4	Wholebody Quantitative SPECT-CT for Neuroblastoma	86
4.1	Introduction	86
4.2	Collimator Selection	95
4.2.1	Introduction	95
4.2.2	Methods and Materials	96
4.2.3	Results	101
4.2.4	Discussion	108
4.2.5	Conclusion	111
4.3	Phantom Evaluation	112
4.3.1	Introduction	112
4.3.2	Methods and Materials	112
4.3.3	Results	116
4.3.4	Discussion	127
4.3.5	Conclusion	132
4.4	Clinical Evaluation	133
4.4.1	Introduction	133
4.4.2	Methods and Materials	133
4.4.3	Results	138
4.4.4	Discussion	152
4.4.5	Conclusion	160
4.5	Reflection	161

5	Conclusions & Future Work	162
5.1	Conclusions	162
5.2	Future Work	167
	Bibliography	170

List of Tables

2.1	Soft tissue region ranges for 2σ , 3σ and 4σ models	42
2.2	Least-squares fitted gradients grouped model and lung coverage	47
3.1	SIOPEN and Curie total scores for the local observer	58
3.2	SIOPEN and Curie total scores for the expert observer	71
3.3	Grouped SIOPEN and Curie score differences	74
3.4	Inter-observer comparison of cases scored by both observers	76
4.1	Sensitivity measurements and model parameters for each collimator	103
4.2	Iteration limits for each filter size	123
4.3	Absolute activity concentrations and relative recovery coefficients for quantitatively viable reconstructions	130
4.4	Supporting information and ranges of clinically relevant parameters	139
4.5	Weighted mean uptake measured in left liver, right liver and brain	153

List of Figures

1.1	Electron microscopy of neural crest cells in an embryo	20
1.2	Migration of neural crest cells into adult tissues and organs	20
1.3	Examples of histology of neuroblastoma subtypes	21
1.4	Stratification of neuroblastoma histology under the INPC	24
1.5	Prognostic effects of histology, differentiation, MKI and patient age	24
1.6	Prognostic effects of favourable/unfavourable histology on each INSS stage	25
1.7	INRG consensus pre-treatment classification schema	26
1.8	Neuroblastoma treatment schedule for SIOPEN High Risk Neu- roblastoma 1 Clinical Trial	27
1.9	Schematic of a gamma camera detector	30
1.10	Operational flow of an OSEM algorithm	31
1.11	Planar wholebody scintigraphy compared to wholebody MRI	33
2.1	Example of pseudo-CT generation workflow	41
2.2	HU histogram of low-dose CT scan voxel values	42
2.3	Comparison of pseudo-CTs of the thorax generated with 3σ soft tissue regions	43
2.4	Visual comparison of reconstructed SPECT slices using each of five attenuation correction maps	44
2.5	Bland-Altman style voxel plots with significant lung volume coverage	45
2.6	Bland-Altman style voxel plots with no lung volume coverage	46
3.1	Anatomical regions used in Curie and SIOPEN scoring systems	52
3.2	Example planar and SPECT/MR images as presented to the ob- server during scoring	57
3.3	Histograms of local clinical observer scoring 15 cases at a range of clinical time points.	59

3.4	Bland-Altman plots of local clinical observer scoring 15 cases at a range of clinical time points.	60
3.5	Key finding from case 11.	61
3.6	Key finding from case 3.	62
3.7	Key finding from case 13.	63
3.8	Key finding from case 14.	64
3.9	Histograms of expert observer scoring 15 cases of high-risk neuroblastoma at diagnosis.	72
3.10	Bland-Altman plots of expert observer scoring 15 cases of high-risk neuroblastoma at diagnosis.	73
3.11	Clinically significant finding in case 7	75
3.12	Clinically significant finding in case 9	77
3.13	Clinically significant finding in case 10	78
3.14	Clinically significant findings in case 10 (continued)	79
4.1	Longitudinal tumour and liver uptake in a single case from diagnosis to week 73 of treatment	94
4.2	Schematic of PMMA sensitivity phantom	98
4.3	System sensitivity versus detector-phantom distance for each collimator	102
4.4	Septal penetration derived from sensitivity model versus detector-phantom distance for each collimator	105
4.5	Sensitivity spectra acquired with phantom positioned at 50, 150 and 450mm distant from each collimator	106
4.6	Fitted sensitivity spectra model parameters for parallel-hole collimators	107
4.7	Photographs of SPECT phantoms	113
4.8	Visualisation of VOI analysis	115
4.9	NEMA contrast recovery and background variability	117
4.10	1ml peak values and boxplots of absolute activity concentrations for 37mm and 28mm spheres	119
4.11	1ml peak values and boxplots of absolute activity concentrations for 22mm and 17mm spheres	120

4.12	Boxplots of absolute activity concentrations for 13mm and 10mm spheres	121
4.13	1ml peak values and boxplots of absolute activity concentrations for 37mm and 28mm background VOIs	124
4.14	1ml peak values and boxplots of absolute activity concentrations for 22mm and 17mm background VOIs	125
4.15	Boxplots of absolute activity concentrations for 13mm and 10mm background VOIs	126
4.16	Placement of 22mm VOIs in right liver, left liver and brain in a patient with grossly abnormal anatomy	136
4.17	Patient 12 - the lone instance of liver surgery within the dataset .	140
4.18	Boxplots of absolute activity concentration in right liver, left liver and brain	142
4.19	Boxplots of SUV/BW scaled uptake in right liver, left liver and brain	143
4.20	Boxplots of SUV/LBM scaled uptake in right liver, left liver and brain	144
4.21	Scatter plots of mean uptake vs uptake time for right liver, left liver and brain for each unit scale	148
4.22	Scatter plots of mean uptake vs decay corrected activity for right liver, left liver and brain for each unit scale	149
4.23	Histograms of standard deviations of uptake for right liver, left liver and brain for each unit scale	150
4.24	Scatter plots of left liver vs right liver and left-right paired difference plots for each measurement scale	151
4.25	Comparison of balanced uptake and diffuse left lobe uptake patterns	155
4.26	Comparison of left lobe uptake patterns with and without compression by a central mass	157

Acknowledgements

This research would not have been possible without valuable contributions from Dr Lorenzo Biassoni; Consultant Nuclear Medicine Physician, Great Ormond Street Hospital; and Dr Harjeet Kaur; Consultant Paediatric Radiologist, Royal Hospital for Children. They collaborated as clinical observers and provided valuable insights into the imaging of neuroblastoma. I would also like to extend my deepest appreciation to my supervisors; Dr Michael Bradnam, Dr Alison Bolster and Dr John Foster; I could not have undertaken this journey without your patience, feedback and support.

Many thanks should also go to the Paediatric Radiology and Oncology teams who inspired and encouraged me. I especially wish to thank Dr Andrew Watt, Dr Greg Irwin and Dr Katharine Orr; who reported the diagnostic images and championed the hybrid imaging of neuroblastoma. I also wish to deeply thank Dr Dermot Murphy, Dr Jairam Sastry and Dr Milind Ronghe; who managed the neuroblastoma patients and strongly supported the research of quantitative imaging techniques.

I wish to acknowledge the staff of the Mechanical Workshop, Department of Clinical Physics & Bioengineering; who constructed an imaging phantom to my specification.

I am very grateful to my close colleagues at the Royal Hospital for Children; especially Mrs Shona Morris, Mrs Nicola Disbrowe, Mrs Helen Paterson and Dr Pauline Hall-Barrientos; for their dedication to patient care, expertise in imaging and unwavering moral support. Further thanks to the Nuclear Medicine and Medical Physics teams at Glasgow Royal Infirmary and University Hospital Crosshouse; with special mention to Dr Gillian Ainslie-McLaren, Mr Jim McGarvie and Dr Mani Manivannan for their support and encouragement.

Lastly, I would be remiss in not thanking my friends and family; especially my partner Steph and my parents Mary and James. Their continued belief in me, even when I doubted myself, has sustained me through to completion of this thesis.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Hugh James Wallace

February 2023

Definitions/Abbreviations

2D 2-dimensional

3D 3-dimensional

AC attenuation correction

ALK anaplastic lymphoma kinase

ANOVA Analysis of Variance

BNMS British Nuclear Medicine Society

BSA body surface area

BW body weight

COG Children's Oncology Group

CT Computed Tomography

CZT Cadmium-Zinc-Telluride

DCPB Department of Clinical Physics and Bioengineering

DICOM Digital Imaging and Communications in Medicine

DNA deoxyribonucleic acid

EANM European Association of Nuclear Medicine

EFS event free survival

ESOPNM European Symposium of Paediatric Nuclear Medicine

FH favourable histology

FDG fluorodeoxyglucose

FOV field of view

FWHM Full Width at Half Maximum

GA General Anaesthetic

GCAT Gamma Camera Assessment Team

GRI Glasgow Royal Infirmary

GN ganglioneuroma

GNB ganglioneuroblastoma

GNBn nodular ganglioneuroblastoma

HU Hounsfield Unit

IEC International Electrotechnical Commission
INPC International Neuroblastoma Pathology Classification
INSS International Neuroblastoma Staging System
INRG International Neuroblastoma Risk Group
INRGSS International Neuroblastoma Risk Group Staging System
IQR interquartile range
LBM lean body mass
LEAP Low Energy All Purpose
LEHR Low Energy High Resolution
MDP methylene diphosphonate
MDU Medical Devices Unit
MELP Medium Energy Low Penetration
MIBG meta-Iodobenzylguanidine
MKI mitosis-karyorrhexis index
MR Magnetic Resonance
MRAC Magnetic Resonance Attenuation Correction
MRI Magnetic Resonance Imaging
NEMA National Electrical Manufacturers Association
NPL National Physical Laboratory
OSEM Ordered Subsets Expectation Maximisation
OS Overall Survival
PCC Pearson Correlation Coefficient
PERCIST Positron Emission tomography Response Criteria in Solid Tumours
PET Positron Emission Tomography
PET/CT Positron Emission Tomography/Computed Tomography
PET/MR Positron Emission Tomography/Magnetic Resonance
PHA Pulse Height Analyser
PMTs photomultiplier tubes
PMMA poly-methyl methacrylate
RECIST Response Evaluation Criteria in Solid Tumours
RHC Royal Hospital for Children
RHSC Royal Hospital for Sick Children
ROC receiver-operator characteristic
SIOPEN International Society of Paediatric Oncology Neuroblastoma Europe

SPECT Single Photon Emission Computed Tomography
SPECT/CT Single Photon Emission Computed Tomography/Computed Tomography
SPECT/MR Single Photon Emission Computed Tomography/Magnetic Resonance
STIR Short Tau Inversion Recovery
SUL Standardised Uptake Value scaled to lean body mass
SUV Standardised Uptake Value
SUV/BW standard uptake values scaled to body weight
SUV/LBM standard uptake values scaled to lean body mass
TSE Turbo Spin Echo
TLCRR Tumour-to-Liver Count-Rate Ratio
UH unfavourable histology
UK United Kingdom
UTE ultra-short echo time
USA United States of America
VOI volume of interest
WHO World Health Organisation

Chapter 1

Introduction & Literature Review

1.1 Neuroblastoma

Neuroblastoma is a cancer of the primitive nerve cells that form within the neural crest, a transient embryonal structure unique to vertebrates that migrates throughout the foetus and contributes to the healthy development of many different systems and organs (Simões-Costa and Bronner 2013). First described in chicken embryos in 1868 by Wilhelm His (Richardson and Keuck 2022), the cells of the neural crest typically arise in week 4 of human embryonal development as shown in Figure 1.1. After forming within the crest, these multipotent cells then disseminate throughout the embryo and mature into a range of final forms dependent on their final landing niche. The cells of the neural crest are therefore most properly thought of as four distinct populations which are organised by position; cranial, vagal, trunk and sacral. Under normal development, each of these cell groups stream to their intended destinations and contribute to a unique range of tissues as shown in Figure 1.2.

These primitive and multipotent origins give rise to the key characteristics of the disease. For example, the migratory nature of the origin cells to some extent dictates the sites of malignancy; the locations listed in Figure 1.2 include the most common sites for primary neuroblastic masses (adrenal medulla and paraspinal sympathetic chain) as well as some of the most common metastatic sites (nodes around the major vessels and within the mediastinum, liver, skin, bones of the face and skull). However Neuroblastoma is also known to spread extensively using the lymphatic and circulatory systems, and extensive bone and bone marrow deposits in the axial and peripheral skeleton are common at the point of diagnosis (Kushner 2004).

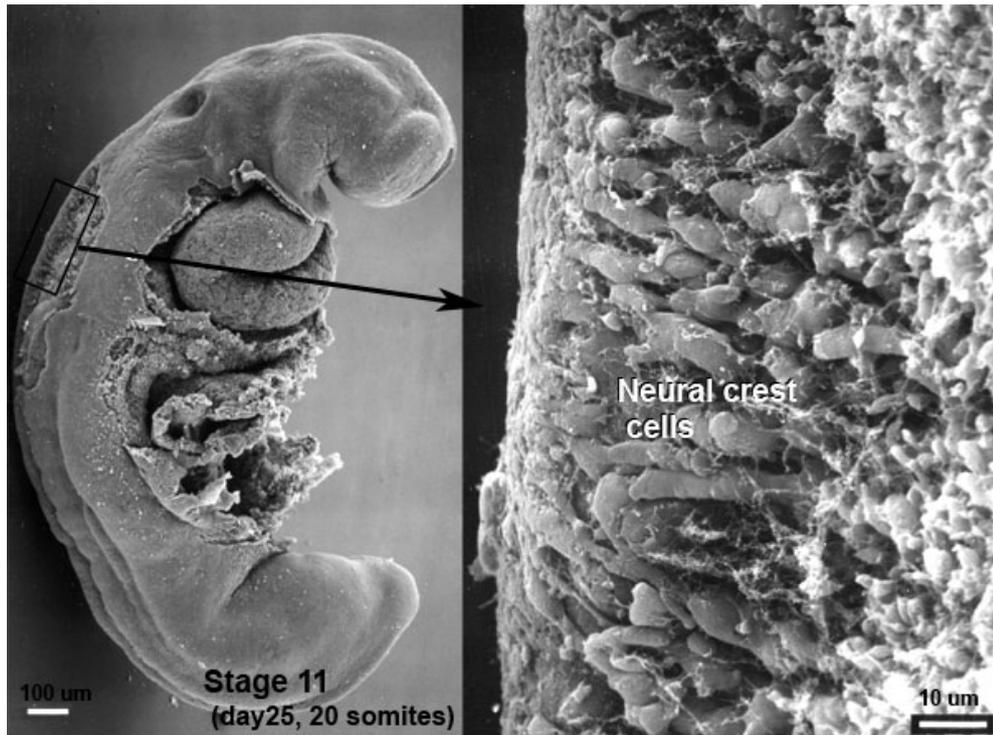


Figure 1.1: Scanning electron microscopy of an embryo at Carnegie stage 11 (23-26 days gestation) showing 10 micron scale images of neural crest cells in situ. ©Dr Mark Hill 2023, UNSW Embryology ISBN: 978 0 7334 2609 4. Retrieved from https://embryology.med.unsw.edu.au/embryology/index.php/File:Stage11_sem21.jpg and re-used with permission.

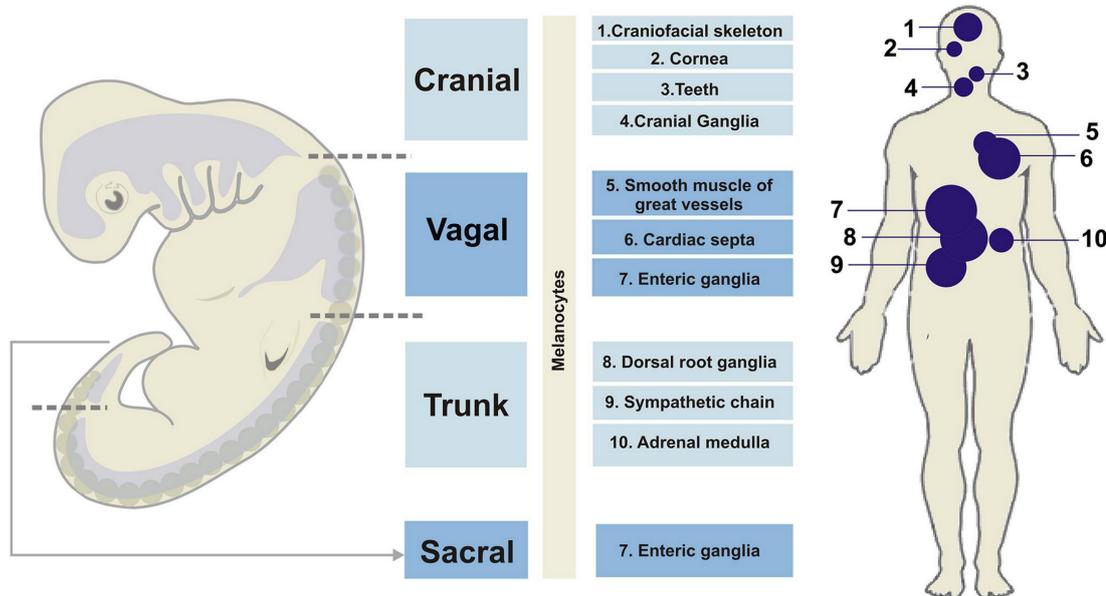


Figure 1.2: Migration of neural crest cells into adult tissues and organs. Re-use from Simões-Costa and Bronner (2013) granted under Creative Commons License (Attribution-NonCommercial 3.0 Unported) as described at <https://creativecommons.org/licenses/by-nc/3.0/legalcode>.

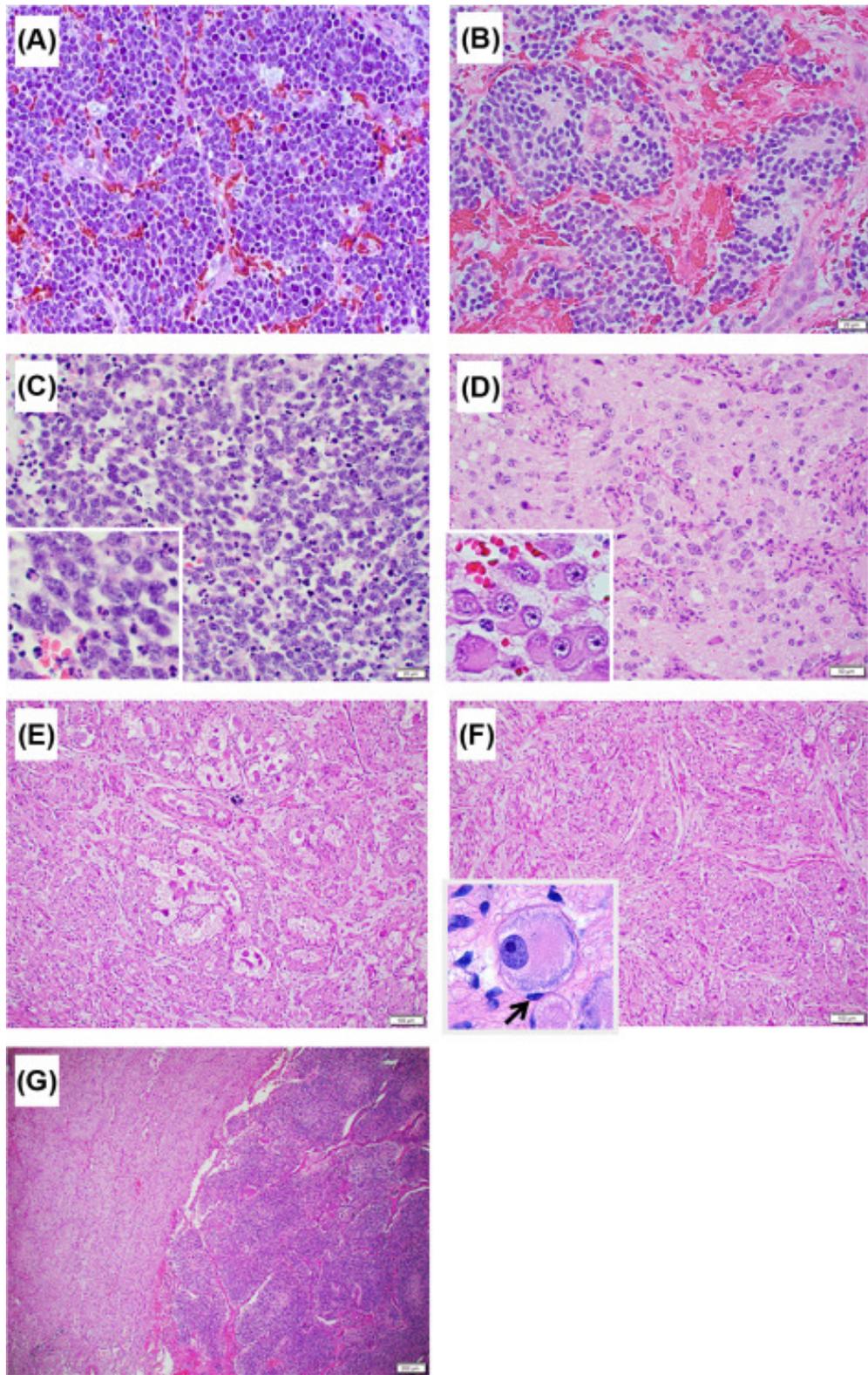


Figure 1.3: Examples of histology of neuroblastoma subtypes. Images A-C: A range of "true" undifferentiated neuroblastomas, dominated by immature nerve cells. Image F: Benign ganglioneuroma formed of mature ganglion cells with significant stromal development. Image D: Differentiating neuroblastoma subtype, demonstrating ongoing cell maturation. Image E: Intermixed ganglioneuroblastoma, showing an intermediate disease state. Image G: Nodular ganglioneuroblastoma, showing distinct cell lines at different stages of maturation. Re-use from Shimada and Ikegaki (2019) granted by Elsevier under Copyright Clearance Center RightsLink® License #5474751103309

Within the totality of cancer Neuroblastoma is considered a rare disease. However it remains the most common extra-cranial solid cancer of childhood (Matthay et al. 2016) and the most common solid cancer diagnosis within the first year of life (UKHSA 2021), accounting for approximately 100 new cases every year in the United Kingdom (UK). The vast majority of Neuroblastomas are believed to arise sporadically through random mutations, though some familial genetic pre-dispositions have been identified in around 1-2% of cases world wide (Matthay et al. 2016). In the majority of cases the detailed mechanisms of oncogenesis are not well understood, however the sub-group of progenitor neural crest cells and tumour microenvironment at the time of malignant mutation are thought to have a significant impact on the resulting cancer cells (Garner and Beierle 2016), which display a broad range of appearances under histology as shown in Figure 1.3. Neuroblastoma is therefore best understood not as a single disease but as a family of related cancers that range from "true" neuroblastomas, dominated by immature nerve cells (images A-C of Figure 1.3); to relatively benign ganglioneuromas, showing significant stromal development and comprised almost entirely of mature ganglion cells (image F); with less aggressive differentiating neuroblastomas (image D) and intermediate ganglioneuroblastomas (image E) somewhere between these extremes. In some instances, multiple cell lines may be involved within a single primary disease site. These are typically referred to as nodular ganglioneuroblastomas and show distinct nodular sections under histology (image G).

The prognosis of Neuroblastoma, commonly assessed through post treatment event free survival (EFS), has been strongly linked to a number of case specific factors. These include; the degree of histological subtype; mitosis-karyorrhexis index (MKI), a measure of cell turnover rate; and patient age at diagnosis (Shimada and Ikegaki 2019). In general, the more immature the disease cells appear on histology, the older the patient, and the more aggressively the disease replicates, the poorer the prognosis. However, the interaction of these three factors is fairly complex and the International Neuroblastoma Pathology Classification (INPC) system (Shimada et al. 2001) was proposed to formalise the stratification of histology, as shown in Figure 1.4. The difference in prognosis for disease with favourable and unfavourable histologies under the INPC is significant, as shown in the Kaplan Meier curve in Figure 1.5. Further developments in oncogenetics have since revealed the importance of certain tumour

genes, such as those that amplify production of the proteins N-Myc and anaplastic lymphoma kinase (ALK). These, alongside other genetic markers, have been shown to contribute to significantly poorer patient prognosis (Nakazawa 2021).

In addition to the INPC, two staging systems are used for the prognostic stratification of Neuroblastoma; the INSS (Brodeur et al. 1988) and the International Neuroblastoma Risk Group Staging System (INRGSS) (Cohn et al. 2009). The INSS is a post-surgical staging system developed by the Children's Oncology Group (COG) in the United States of America (USA); it originally combined age and the results of surgical resection in order to derive 6 stages; 1, 2A, 2B, 3, 4 and 4S. Stage 1 requires the visible disease to be localised to one area only, have been fully excised and to have shown no evidence of spread across the midline or to lymph nodes outside of the primary mass. Both stage 2 categories are identical to stage 1 but either were not fully resected (stage 2A) or show local lymphatic spread that does not cross the midline (stage 2B). Stage 3 is disease where either the primary mass or localised lymphatic disease has crossed the midline but there are no distant metastases. Stage 4S is a special category of metastatic disease which has a much higher probability to undergo spontaneous remission; the patient must be under one year of age and show no unfavourable histologic or oncogenetic features, any tumour and nodal disease must not cross the midline and further metastatic spread must be limited only to the liver and skin with very limited (<10% of cells) involvement observed in bone marrow biopsy. All other Neuroblastomas are considered stage 4 disease and show distant metastatic spread to lymph nodes, bones, bone marrow, skin and/or other organs. While these staging groups were found to provide good stratification of patient risk, INPC grading was shown to offer additional sub-stratification within each stage, shown in Figure 1.6. Later iterations of the INSS would move to incorporate the INPC and oncogenetic features such as n-Myc protein amplification and deoxyribonucleic acid (DNA) ploidy to better delineate the clinical risk groups.

Despite the incorporation of additional features, the INSS retained the significant downside of only providing post-surgical staging. It was therefore unsuitable for the stratification of these patients prior to commencing treatment; e.g. for triaging initial treatment and/or defining entry into front-line clinical trials. The INRGSS was developed as a consensus pre-treatment staging system, to combine the approach of

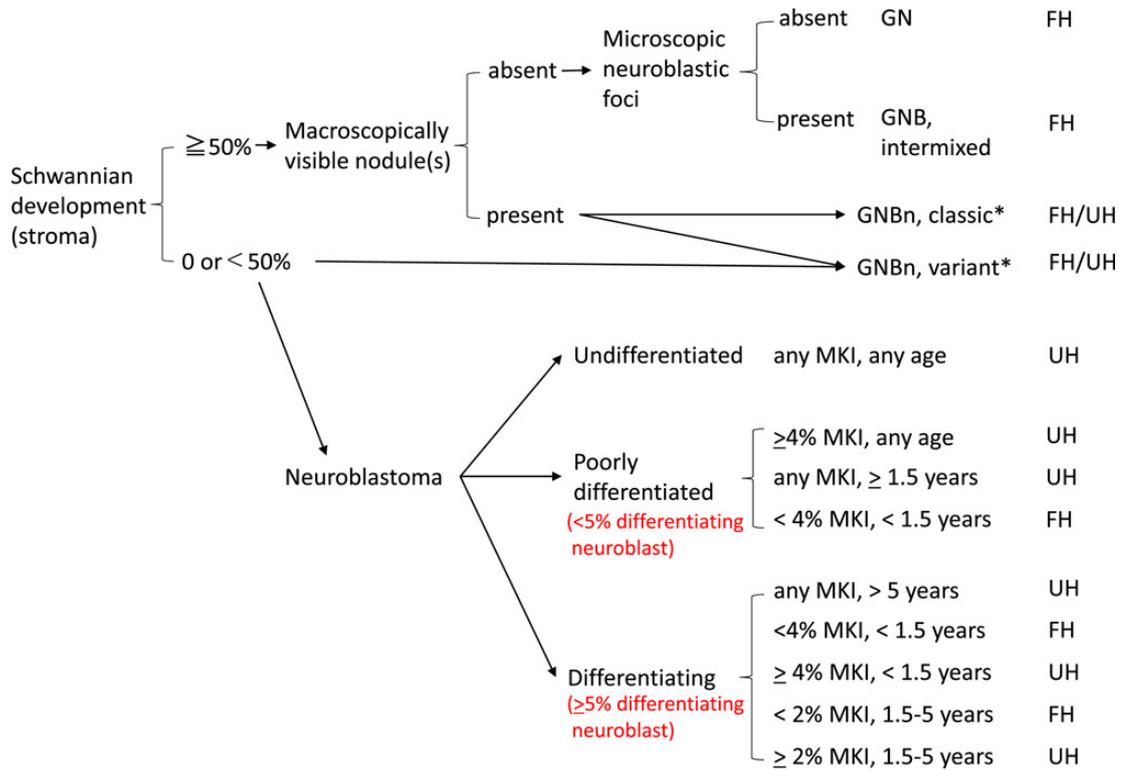


Figure 1.4: Stratification of Neuroblastoma histology under the International Neuroblastoma Pathology Classification (INPC). Classification into favourable histology (FH) or unfavourable histology (UH) proceeds according to the degree of cell development, disease subtype [Neuroblastoma, ganglioneuroma (GN), ganglioneuroblastoma (GNB) and nodular ganglioneuroblastoma (GNBn)], degree of differentiation and mitosis-karyorrhexis index (MKI). Re-use from Nakazawa (2021) granted by John Wiley and Sons under Copyright Clearance Center RightsLink® License #5471020259376

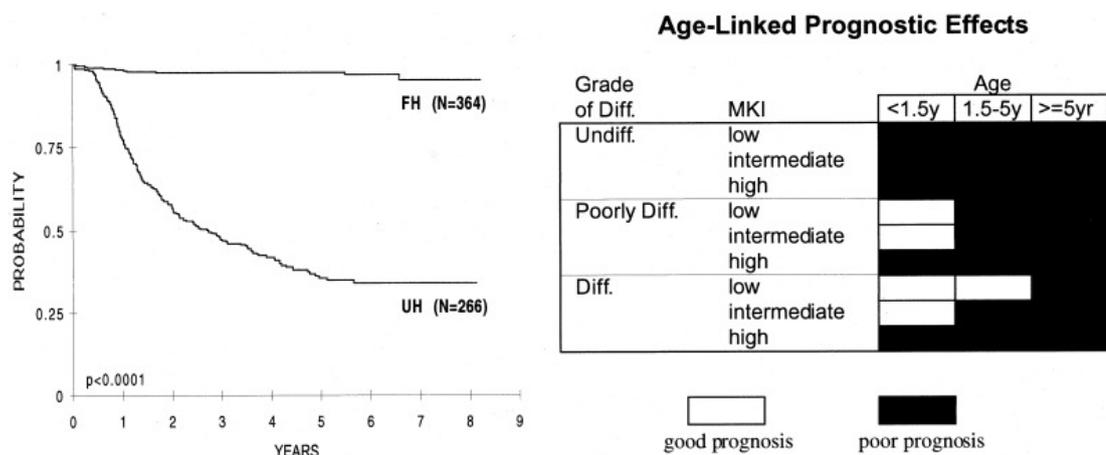


Figure 1.5: Prognostic effects of histology, differentiation, mitosis-karyorrhexis index (MKI) and patient age. Left: Kaplan Meier curve showing impact of favourable/unfavourable histology on event free survival (EFS). Right: Prognostic impact of grade of differentiation, mitosis-karyorrhexis index (MKI) and age at diagnosis. Re-use from Shimada et al. (2001) granted by John Wiley and Sons under Copyright Clearance Center RightsLink® License #5484940930986

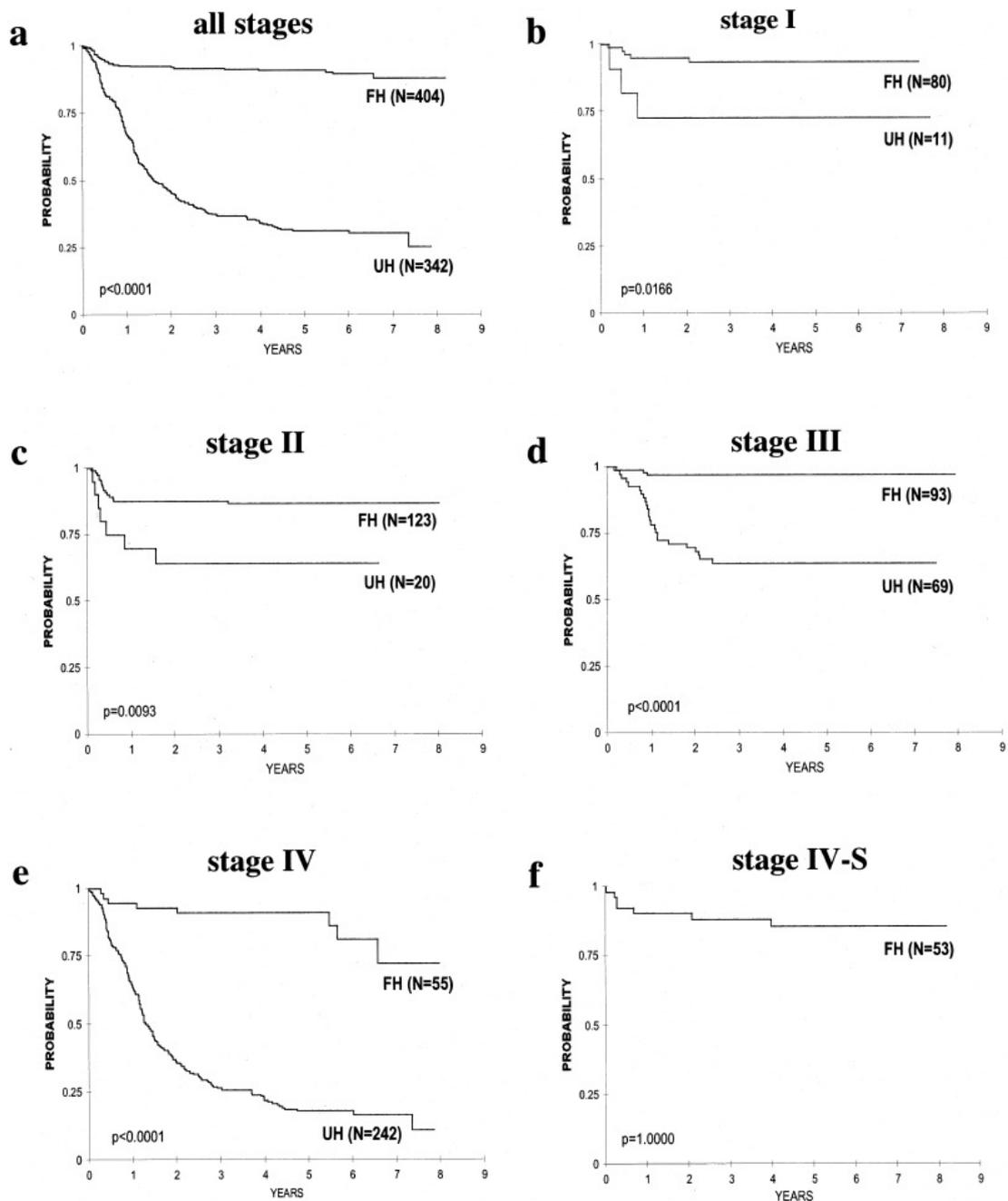


Figure 1.6: Kaplan Meier curves showing the effect of favourable and unfavourable histologies on event free survival (EFS) separated by International Neuroblastoma Staging System (INSS) stage. a) All INSS stages combined; b-e) INSS stages 1-4; f) INSS stage 4S. Re-use from Shimada et al. (2001) granted by John Wiley and Sons under Copyright Clearance Center RightsLink® License #5484940930986

INRG Stage	Age (months)	Histologic Category	Grade of Tumor Differentiation	MYCN	11q Aberration	Ploidy	Pretreatment Risk Group
L1/L2		GN maturing; GNB intermixed					A Very low
L1		Any, except GN maturing or GNB intermixed		NA			B Very low
				Amp			K High
L2	< 18	Any, except GN maturing or GNB intermixed		NA	No		D Low
					Yes		G Intermediate
	≥ 18	GNB nodular; neuroblastoma	Differentiating	NA	No		E Low
					Yes		H Intermediate
			Poorly differentiated or undifferentiated	NA			I Intermediate
				Amp			N High
M	< 18			NA		Hyperdiploid	F Low
	< 12			NA		Diploid	I Intermediate
	12 to < 18			NA		Diploid	J Intermediate
	< 18			Amp			O High
	≥ 18						P High
MS	< 18			NA	No		C Very low
					Yes		Q High
					Amp		R High

Figure 1.7: International Neuroblastoma Risk Group (INRG) consensus pre-treatment classification schema showing stratification from INRG stages to 16 distinct clinical subgroups within 4 pre-treatment risk categories. Re-use from Cohn et al. (2009) granted by Wolters Kluwer Health, Inc. under Copyright Clearance Center RightsLink® License #5478190865648

the INSS with that of the International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) group, who had developed a pre-surgical risk stratification approach based on image-defined surgical risk factors and n-Myc protein amplification status. Under the INRGSS, imaging and biopsy are used to first categorise each case into one of four INRG stages (L1, L2, M and MS) based on surgical risk factors and metastatic spread as observed on imaging. Stages L1 and L2 are both categories of localised disease with L2 containing one or more surgical risk factors, while stages M and MS are essentially equivalent to INSS stages 4 and 4S. As shown in Figure 1.7, these staging groups are then combined with patient age and a number of histologic and oncogenetic features for classification into 16 distinct risk subgroups, themselves placed in one of four pre-treatment risk categories based on 5-year EFS. These categories are Very Low Risk (EFS>85%), Low Risk (75%<EFS≤85%), Intermediate Risk (50%≤EFS≤75%) and High Risk (EFS<50%). Approximately 50% of patients diagnosed with Neuroblastoma are classified as high risk under the INRGSS (DuBois et al. 2022).

High risk neuroblastoma is amongst the most intensely treated of all cancers, with front line therapy typically split into induction, consolidation and maintenance phases

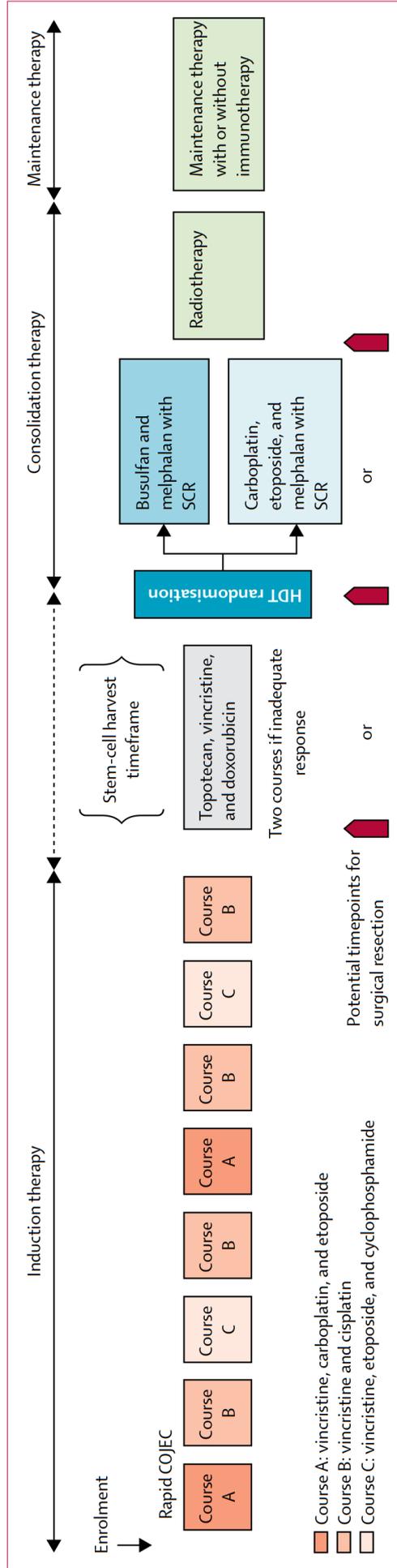


Figure 1.8: Treatment schedule for International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) High Risk Neuroblastoma 1 Clinical Trial. Re-use from Ladenstein et al. (2017) granted by Elsevier under Copyright Clearance Center RightsLink® License #5471031096496

as shown in Figure 1.8, which details the treatment schedule for SIOPEX High Risk Neuroblastoma 1 trial (Ladenstein et al. 2017). Induction incorporates multiple courses of overlapping systemic chemotherapy, each consisting of several cytotoxic agents used in combination. The goal of induction is to shrink primary disease in preparation for surgical resection and to eliminate as much metastatic disease as possible. Consolidation therapy consists of very high dose systemic chemotherapy to clear metastatic disease, supplemented with external beam radiotherapy where indicated to assist with control of stubborn primary or secondary tumours. Autologous stem cell harvest, performed at the end of induction treatment, is required to permit the safe recovery of immune function after consolidation therapy. Finally, the patient will be placed on maintenance therapy, often including the use of immunotherapy, to control any remaining disease. Typically, patients undergo bone marrow biopsies and imaging before and after induction chemotherapy, high dose chemotherapy and radiotherapy. Imaging assessments are also undertaken every three months while on maintenance therapy. If patients achieve remission, surveillance scans are typically undertaken every 3-6 months for several years after discharge. Scintigraphic imaging is key to this surveillance.

1.2 Scintigraphic Imaging

The gamma camera is a scintigraphic imaging technology first developed in the late 1950s by Karl Anger (Anger 1958) and used in medicine to image gamma-ray emitting radiopharmaceuticals within the human body. The technology has the capability to image a large range of gamma-ray energies (80-600keV) and as such the chemical tracer and radioisotope combination can be selected dependent on the physiological system of interest.

While the processing electronics of gamma cameras have significantly advanced since the first analogue models, the core imaging technology has remained unchanged since its inception. Figure 1.9 shows a schematic cross section of a traditional scintillation gamma camera, with a parallel hole collimator. Gamma-ray photons enter through the collimator, which has lead septa oriented perpendicular to the scintillation crystal plane. The collimator is designed to associate incoming gamma rays with position by absorbing those outside of a relatively narrow acceptance angle. Collimated gamma

rays may then proceed to the scintillation crystal, which has a relatively high stopping power across the energy ranges typically used for imaging. Interactions within the crystal generate pulses of photons that are proportional to the energy deposited in the crystal. These photons pass through the light guide (usually an optically transparent grease) and strike the photocathode of the photomultiplier tubes (PMTs) to generate accelerated cascades of electrons, resulting in a measurable electric signal. Following this amplification stage, the resulting electrical signals can be digitised and processed using modern electronics. Providing the set of PMTs have been appropriately calibrated, the summed signal resulting from a single photo cascade is proportional to the gamma ray energy, allowing for events of undesirable energy to be excluded. The distribution signals produced by the PMTs are approximately gaussian around the position of interaction within the crystal, and therefore the position of the gamma-ray event can be derived from a weighted sum of signals from the PMTs.

A modern general purpose gamma camera typically has two detector heads to permit simultaneous acquisition of two planar images, usually from anterior and posterior positions either side of the patient, and the capability to acquire in dynamic, gated or wholebody planar acquisition modes as required for the specific examination. However, each of these planar imaging techniques suffers from the several deleterious effects on image quality (Bailey and Willowson 2013); the super-position of gamma-ray events from all depths in the patient, a natural consequence of 2-dimensional (2D) imaging techniques; the distance-dependent spatial resolution of parallel hole collimation, which acts to enlarge and blur objects at greater distance from the detector; photon absorption (often called attenuation) within the patient, which acts to reduce the relative intensity of uptake of sites at greater tissue depths; and photon scatter within the patient, which both contributes to loss of photons and generates false events at incorrect spatial positions within the final image.

3-dimensional (3D) gamma-ray emission images which eliminate the superposition of gamma-ray events can be acquired using the technique of Single Photon Emission Computed Tomography (SPECT). To achieve this, the detector heads are rotated around the patient and relatively short planar images, known as projections, are acquired at a specified angular spacing (typically one projection per 3°). These projections are then reconstructed using a tomographic reconstruction algorithm, the

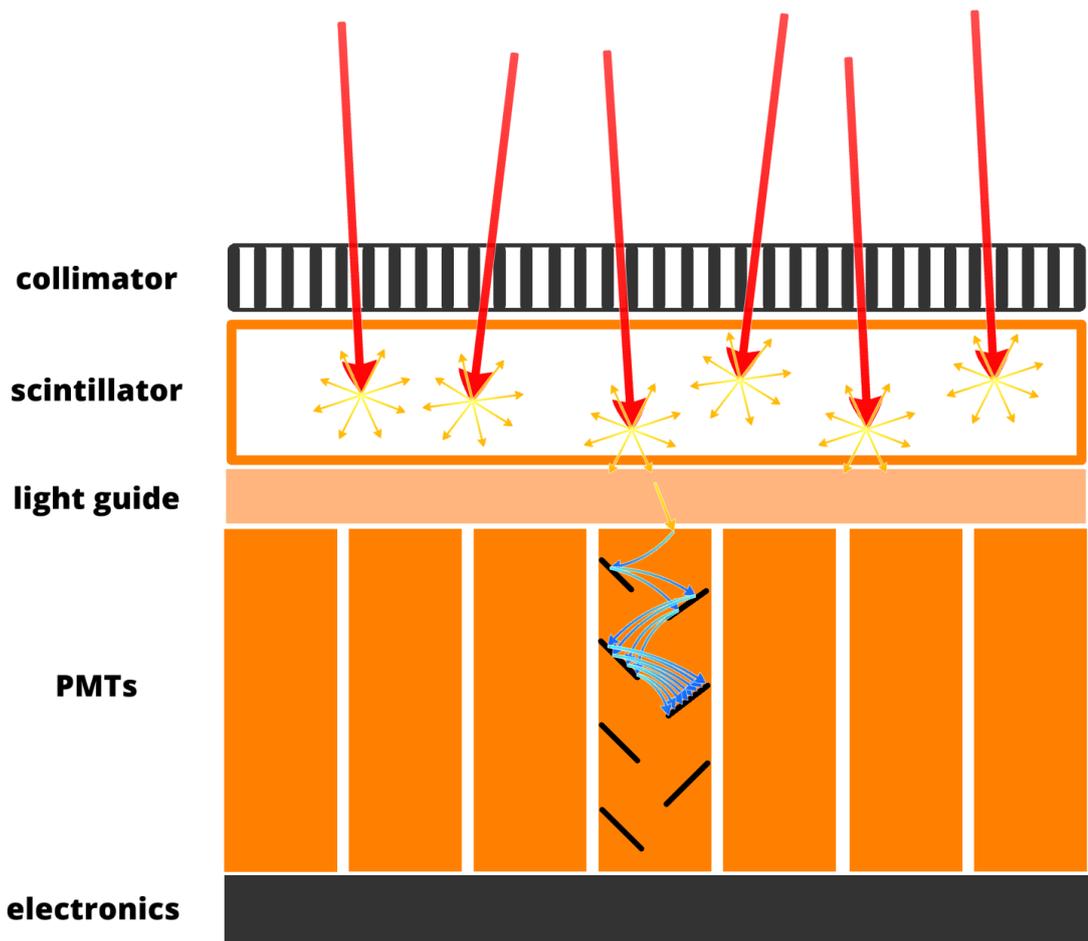


Figure 1.9: Schematic of a gamma camera detector. Gamma-ray photons (red arrows) enter through the collimator, which has lead septa oriented perpendicular to the scintillation crystal plane. Interactions within the scintillator crystal generate pulses of photons (yellow-orange gradient arrows) that are proportional to the energy deposited in the crystal. These photons pass through the light guide (usually an optically transparent grease, shown here much thicker than in reality) and strike the photocathode of the photomultiplier tubes (PMTs) to generate accelerated cascades of electrons (blue arrows), resulting in a measurable electric signal. Following this amplification stage, the resulting electrical signals are processed using analogue or digital electronics.

most common of which is an iterative algorithm known as Ordered Subsets Expectation Maximisation (OSEM) (Hudson and Larkin 1994). This multiplicative algorithm generates the statistically most likely model of the underlying activity distribution through iterative comparison with a subset of projections, facilitated by transformations for forward-projection (the generation of projections from a 3D volume) and back-projection (generation of a 3D volume from a set of projections), as shown in Figure 1.10.

Following an initial guess (usually a uniform distribution of gamma-ray events), which forms the first "current best" estimate of the 3D volume, a subset of acquired projec-

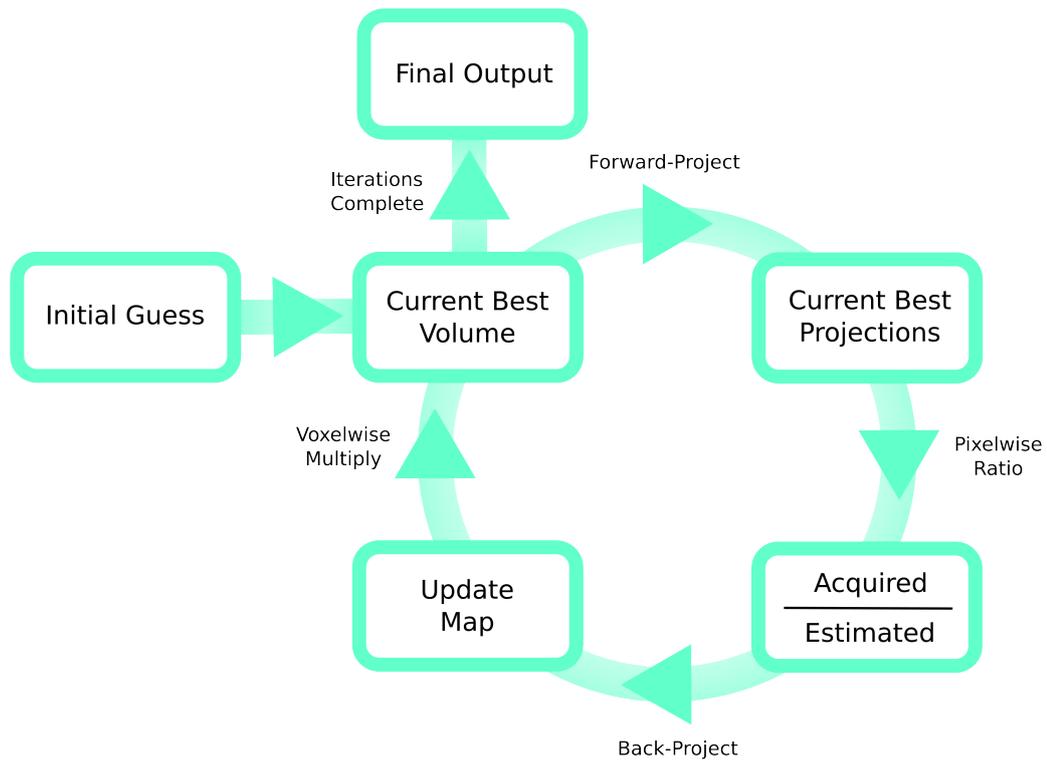


Figure 1.10: Operational flow of an Ordered Subsets Expectation Maximisation (OSEM) algorithm. Following an initial guess, which forms the initial best estimate of the 3D volume, a subset of acquired projections are selected and the estimated volume forward-projected at each of the matching angular positions. The pixelwise ratio between acquired and estimated projections are calculated and these ratios are then backprojected to create a multiplicative update map. Simple voxel-wise multiplication of the initial guess and the correction map generates a new updated volume which can then be used for comparison with the next subset of projections.

tions are selected and the estimated volume forward-projected at each of the matching angular positions. The pixelwise ratios between acquired and estimated projections are calculated and these ratios are then backprojected to create a multiplicative update map. Simple voxel-wise multiplication of the current best estimate and the update map generates a new updated volume which can then be used for comparison with the next subset of projections. Following the application of corrections from each subset (i.e. comparison to all projections once only) the algorithm is said to have completed one full iteration. At the time of writing, no true international standardisation of SPECT imaging exists, with most departments performing their own image optimisation processes according to the needs and preferences of their reporting clinicians. However in most instances 4-16 full iterations are completed, each typically of 8-16 subsets, and a Butterworth or Gaussian post-filter is then applied to reduce high spatial frequency noise, which is amplified by repeated application of the algorithm.

The successful integration of Positron Emission Tomography (PET) and Computed Tomography (CT) into the hybrid modality Positron Emission Tomography/Computed Tomography (PET/CT) (Beyer et al. 2000) inevitably spawned the development of hybrid Single Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) systems (Bocher et al. 2000). This integration benefits SPECT imaging in two main ways; it offers straightforward localisation of SPECT activity to specific anatomical structures within the patient, something often found to be challenging for reporting clinicians; and the ability to generate a patient specific attenuation correction (AC) map. This can be incorporated into the forward- and back-projection transformations of an iterative reconstruction algorithm to improve reconstruction accuracy (Ritt et al. 2011). The only real disadvantage of SPECT/CT compared to SPECT alone is an increased radiation exposure to patients, an aspect that has delayed its adoption amongst paediatric imaging centres.

1.3 Neuroblastoma Imaging

Neuroblastoma patients undergo an abundance of medical imaging throughout diagnosis, treatment and surveillance (Kushner 2004). This imaging has a direct impact on the patient management; identifying risk factors relevant for surgical resection (Brisse et al. 2011), indicating the need for changes to treatment regimen based on disease progression or response (Matthay et al. 2003) and providing the evidence for inclusion in clinical trials (Matthay et al. 2010).

The two most important imaging modalities for Neuroblastoma are Magnetic Resonance Imaging (MRI) and ^{123}I -labelled meta-iodobenzylguanidine (MIBG) scintigraphy (Mueller et al. 2013). These modalities are highly complementary; MRI offers high sensitivity for detection of bone marrow metastases (Pfluger et al. 2003) and provides exquisite soft tissue contrast for detailed visualisation of primary and secondary masses (Dumba et al. 2015). In contrast, functional imaging with ^{123}I -MIBG provides high specificity (Olivier et al. 2003) and permits distinction between sites of viable disease and post therapeutic changes (Sharp et al. 2011). ^{123}I -MIBG also operates as a diagnostic analogue of ^{131}I -MIBG, a molecular radiotherapy treatment used to deliver large radiation doses to treatment resistant MIBG-avid neuroblastomas (Kayano and Kinuya 2018).

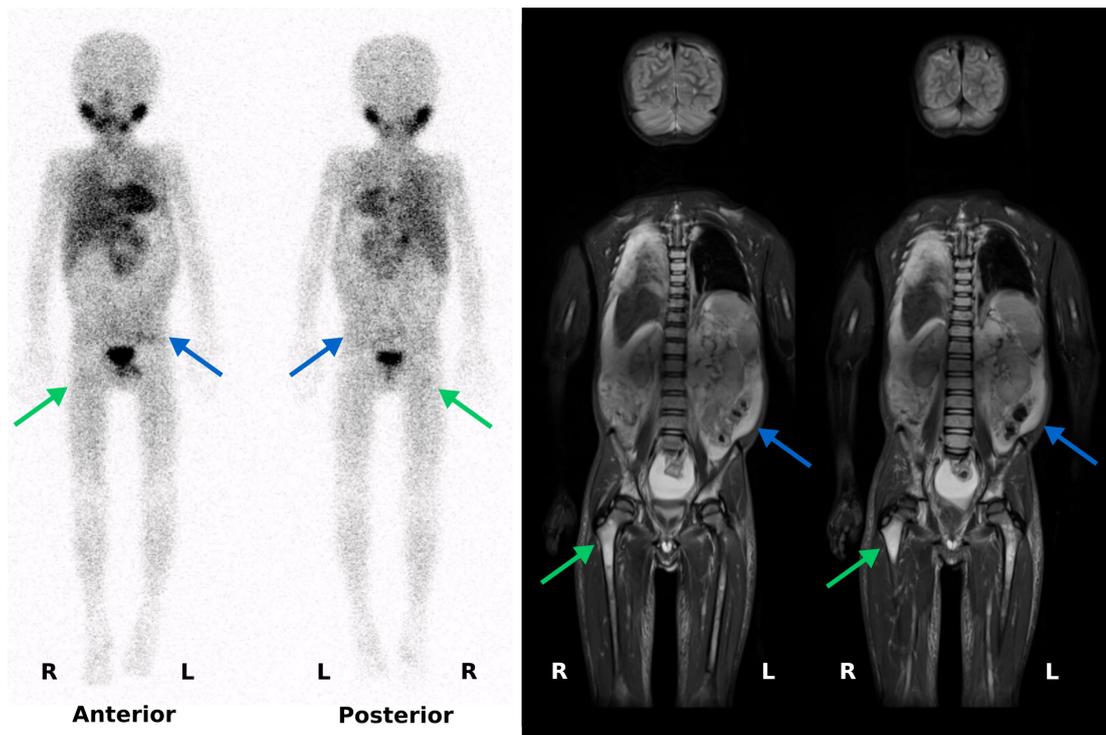


Figure 1.11: ^{123}I -MIBG planar wholebody scintigraphy compared to wholebody MRI. Left: planar wholebody scintigraphy showing patchy, ill-defined uptake in the region of the right femur (green arrows) and within the abdomen (blue arrows). Identification and characterisation of uptake sites can often be very challenging in planar scintigraphy due to poor spatial resolution, poor count statistics and superposition of activity at different depths in the patient. Right: Two slices of coronal plane T2-weighted wholebody MRI showing strong, clear enhancement in the right femoral head consistent with neuroblastic disease deposits (green arrows) and cystic uptake indicative of ascites within the abdomen (blue arrows).

At the outset of this thesis, most paediatric imaging departments relied almost entirely on the use of planar scintigraphy for imaging of ^{123}I -MIBG in Neuroblastoma. While this technique offers a relative speed advantage compared to wholebody MRI and is widely available, it presents a number of significant drawbacks in terms of image quality; as shown in Figure 1.11. Identification and characterisation of uptake sites can often be very challenging in planar scintigraphy due to poor spatial resolution, poor count statistics and superposition of activity sites at different depths in the patient. True quantification of uptake is also impossible in planar scintigraphy due to depth-dependent photon losses, a result of absorption and scattering within the patient. These limitations of planar scintigraphy strongly motivated the research described within this thesis.

1.4 Research Direction

The Royal Hospital for Children (RHC) in Glasgow is the largest paediatric hospital in Scotland and one of the four largest paediatric hospitals in the UK. RHC receives regional, national and UK-wide referrals for the treatment of neuroblastoma. At RHC, the Consultant Paediatric Radiologists were keen to improve the quality of neuroblastoma imaging through the fusion of ^{123}I -MIBG SPECT and MRI datasets. At the same time, there was a clinical aspiration from the Consultant Paediatric Oncologists for quantitative assessment of ^{123}I -MIBG uptake. This was in order to better guide patient management through improved assessment of prognosis and treatment response.

I therefore researched Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) hybrid imaging to improve image quantification and clinical reporting. I developed the hypothesis that the tissue information contained within MRI datasets would be sufficient to provide adequate AC of SPECT, as discussed in Chapter 2. I also developed two hypotheses regarding the clinical use of fused SPECT/MR; (1) that fused SPECT/MR would enable more neuroblastoma lesions to be identified and (2) that more neuroblastoma cases would be classified as unfavourable using fused SPECT/MR than planar imaging alone; both discussed in Chapter 3. When SPECT/CT became available for use in paediatric imaging, I chose to research the quantitative imaging of neuroblastoma using ^{123}I -MIBG SPECT/CT, as discussed in Chapter 4.

Chapter 2

A Simulation of MRI Attenuation Correction

2.1 Introduction

SPECT and PET provide unique functional imaging of the chemical pathways of the body. Hybrid imaging with integrated CT is now well established and has been shown to have greater clinical efficacy than the individual modalities used in isolation (Pfannenbergl et al. 2003; Even-Sapir et al. 2001; Horger and Bares 2006; Schulthess et al. 2006). More recent developments include the launch of simultaneous Positron Emission Tomography/Magnetic Resonance (PET/MR), however there is still significant debate on the clinical utility of this new hybrid modality (Yankeelov et al. 2012; Dickson et al. 2014).

If interest in PET/MR can be said to be discouraged by the lack of a "killer application", by comparison interest in SPECT/MR is almost non-existent; indeed there are only a handful of papers in the literature that mention SPECT/MR (Goetz et al. 2008; Hamamura et al. 2010; Ha et al. 2010; Marshall et al. 2011). Of these the entirety concern pre-clinical or animal imaging and only one attempts to assess the SPECT reconstruction error. This would seem to be the result of a number of contributing factors; the aforementioned lack of a clear role for PET/MR, the ongoing challenge of Magnetic Resonance Attenuation Correction (MRAC) in PET/MR (Keereman et al. 2013; Wagenknecht et al. 2013), the technical challenges associated with integrating current generation SPECT camera technology with MRI and the significant mismatch in cost base for these two modalities.

However, despite these significant barriers to entry there are a number of compelling arguments which suggest SPECT/MR could be a powerful hybrid modality. The ability of Magnetic Resonance (MR) to provide exquisite soft tissue detail for partial volume correction or anatomical priors would be of more immediate use to SPECT, as where the spatial resolution is typically poorer than in PET. SPECT also offers the capability of imaging multiple radio-tracers simultaneously; which could be leveraged to investigate multiple disease receptors while multi parametric MRI (e.g. T1-, T2- and diffusion-weighted imaging) fully characterises structural detail, all within a single examination.

An additional advantage of SPECT relative to PET, despite the typically lower photon energy, is a reduced overall sensitivity to photon attenuation; a natural consequence of single photon detection in SPECT as compared to dual photon detection in PET. This was demonstrated in the comparative work undertaken by Marshall et al. (2011); where MRAC of canine images was shown to be both more accurate and more precise in SPECT as compared to PET. Despite the animal setting, the techniques used for MRAC in this particular study mimic the segmentation-based approaches suggested by multiple authors investigating MRAC for PET (Martinez-Möller et al. 2009; Schulz et al. 2010; Steinberg et al. 2010), and those in clinical use in PET/MR today. The use of matching methodology would at first seem to suggest that the broad conclusions of the study could be extended to clinical SPECT/MR, but the relative size of canines compared to humans will reduce the impact of tissue attenuation for both SPECT and PET. Indeed, at the time of undertaking this work there was not a single peer reviewed study designed to assess the magnitude of clinical SPECT/MR reconstruction errors. The purpose of this study is therefore to initiate the development of this knowledge base.

The methodology adopted in this study to generate AC maps broadly mimics the work undertaken by Martinez-Möller et al. (2009); in that it involves the segmentation of CT component of a SPECT/CT into a simulation of Dixon-based MRAC, based on its Hounsfield Unit (HU) value of individual voxels. "Dixon-based" refers to the use of a Dixon MR sequence; a spectroscopic imaging technique that exploits the phase difference between hydrogen nuclei in predominantly adipose tissues versus those in predominantly non-lipid tissues and is commonly used to generate maps of "fat" and

"water" throughout the body (Dixon 1984; Coombs et al. 1997). The resultant maps of predominantly adipose tissues (e.g. white fat) and predominantly non-lipid tissues (e.g. muscle) can then be assigned bulk (tissue averaged) values for either photon attenuation or HUs to provide a first approximation of the attenuation or CT density of the soft tissues of the patient. Dixon sequences also have the advantage of being readily available on all clinical MRI scanners and being relatively fast to acquire. This combination of features makes these an obvious baseline sequence from which to develop a straightforward methodology for MRAC.

However, as with most MR sequences, both cortical bone and lung tissue produce very little signal leading to significant AC artefacts; particularly in the case of highly attenuating cortical bone. The solution first proposed by (Martinez-Möller et al. 2009) used a morphological closing process to fill the small holes in the segmented image with values equivalent to non-lipid tissue. Furthermore almost all PET/MR AC methods suggest segmentation of lung tissue as a necessary step (Marshall et al. 2011; Martinez-Möller et al. 2009; Schulz et al. 2010; Steinberg et al. 2010; Berker et al. 2012); however this requires more complex post processing, typically utilising some form of connected component analysis. Given the reduced sensitivity of SPECT to photon attenuation, it seems prudent to explore the necessity of such additional processing in the case of SPECT.

This study aims to evaluate the strengths and limitations of simple Dixon-based MRAC for clinical SPECT imaging, and to identify key areas for further development. Given the lack of true hybrid SPECT/MR imaging, Dixon equivalent pseudo-CT images were instead reverse engineered from clinically acquired ^{99m}Tc -methylene diphosphonate (MDP) bone scan SPECT/CT datasets. The derived pseudo-CT images were then used as AC maps for SPECT reconstruction and the resulting reconstructions were compared to the original reference reconstructions to assess the voxel-wise AC error. At the time of undertaking the study, ^{99m}Tc -MDP bone scans were the most commonly undertaken SPECT/CT examination at Glasgow Royal Infirmary (GRI). Given the ideal emission spectrum of ^{99m}Tc -MDP (and therefore minimal scatter and cross-talk) these examinations were deemed to be the most straightforward case for assessing Dixon-based MRAC. The broad range of body areas examined with bone scans was also identified as a key strength of this approach in assessing

potential applicability for use in SPECT/MR imaging of neuroblastoma.

The primary goal of this study is to evaluate the overall AC error for Dixon-based MRAC for SPECT. It is hypothesised that this will be less than equivalent published results for simulations of PET/MR. The secondary goals of the study are (1) to examine the contribution of each step of the Dixon process to the overall AC error and (2) to assess if it is possible to achieve a clinically useful AC without complex lung modelling. In lieu of such modelling, this study instead sets all voxels below the lower HU limit of adipose tissue to air equivalence (i.e. -1000 HU). Similarly, rather than attempt detection of cortical bone "holes", this study will instead set all voxels above the upper HU limit of non-lipid tissues to the bulk non-lipid tissue HU value, a process which simulates perfect segmentation of cortical bone "holes". The end result of these steps is a segmented CT equivalent to that achievable by basic Dixon MRI segmentation followed by post processing to detect and fill in cortical bone "holes".

2.2 Methods and Materials

Patient Selection

The Hermes Medical Solutions (Stockholm, Sweden) Nuclear Medicine database at GRI was queried for all patients undergoing SPECT/CT on the GE Healthcare (Chicago, Illinois) Optima 640 scanner as part of ^{99m}Tc -MDP bone scans over a one month period (August 2013). This returned a set of 11 SPECT/CT scans across a range of anatomical regions dependant on the site of uptake of interest; four scans were of the full thorax to assess the ribs and/or thoracic spine, four scans were of the lower thorax and abdomen to assess the thoraco-lumbar spine, two scans were of the abdomen and pelvis to assess lumbosacral spine and/or pelvis and one scan was of the knees. Prior to their inclusion in this study all patients were verified to have consented for the use of their imaging for departmental research and development.

Imaging Protocols

All SPECT/CT imaging was acquired in accordance with the GRI Nuclear Medicine Optima 640 bone SPECT/CT protocol. SPECT acquisitions were acquired with the gantry configured to 180° orientation and using step-and-shoot acquisition mode.

Each SPECT acquisition consisted of 120 total projections per full rotation of the gantry and 15 seconds per projection. Primary photopeak (126-154keV), lower scatter and upper scatter energy windows were acquired as per the GE recommended acquisition protocol. Hybrid low dose CT scans were acquired immediately following SPECT acquisition and covered the full SPECT acquisition field of view using a peak tube voltage of 120kV, reference tube current of 30mA, 1 second gantry rotation time, a 10mm beam width composed of 4 x 2.5mm slices and a helical pitch of 1.25. CT scans were reconstructed into 2.5mm transverse slices using the GE "Bone" kernel.

SPECT reconstructions were undertaken with the manufacturer recommended "Volumetric MI Evolution for Bone" reconstruction protocol; an iterative OSEM based algorithm using pre- and post-reconstruction smoothing filters, 2 subsets and 10 iterations. The reconstruction process incorporated a proprietary resolution recovery algorithm (GE Evolution), triple energy window based scatter correction and non-uniform attenuation correction derived from the hybrid CT data.

Data Analysis

SPECT acquisitions and CT volumes were anonymised and downloaded from the Hermes Medical Solutions database in Digital Imaging and Communications in Medicine (DICOM) format. All image processing was undertaken using the open source programming language Python with open source numeric, scientific and image processing libraries including Numpy, Scipy, Matplotlib and Pydicom.

The eleven patient CT volumes were combined to create a HU histogram. Following visual inspection of the histogram, the central soft tissue region (between -500HU and +500HU) was fitted with an overlapping bi-Gaussian function to empirically model the distribution of adipose and non-lipid soft tissues. Following this the fitted mean was selected as the best candidate for the bulk HU value for each tissue region. Multiple ranges were then derived for each tissue; these corresponded to ranges of $\mu \pm 2\sigma$, $\mu \pm 3\sigma$ and $\mu \pm 4\sigma$. In each case the internal intersection of the fitted Gaussians was used to define the internal boundary between the two soft tissue regions.

For each set of soft tissue ranges, a simple 3-region pseudo-CT was created by manipulation of the underlying HU values of each CT scan according to the following

steps:

1. To simulate Dixon segmentation, the soft tissue ranges were replaced with their corresponding bulk (mean) values.
2. To simulate morphological closing of small holes (i.e. left by cortical bone), any voxel values equal to or above the upper limit of the non-lipid region were set to the non-lipid bulk value.
3. All voxel values below the lower limit of the adipose region were assumed to be lung or background and replaced with an air equivalent (-1000 HU).

Once all three steps were applied, the designation *dixsim_Xsd* was appended to the series for identification, where *X* identifies the width of soft tissue range in standard deviations. Three additional pseudo-CT variants *segsoft*, *nobone* and *nolung* were also created. These variants consisted of each of the steps 1-3 above performed in isolation. Each of the 132 pseudo-CT volumes were then saved in DICOM format and loaded onto the GE Xeleris workstation for reconstruction. Figure 2.1 shows a flowchart of the pseudo CT generation process.

SPECT/CT reconstructions were repeated for each patient using each of the generated pseudo-CT volumes for attenuation correction. Other than the introduction of the manipulated CT dataset, all reconstruction settings and inputs remained unchanged from the original reference reconstruction. Each pseudo-CT corrected SPECT reconstruction was then compared voxel-by-voxel with the corresponding reference SPECT/CT reconstruction. Scatter plots and Bland-Altman plots of the voxel comparisons were produced in order to visualise the data, assess trends in the difference data and to establish 95% limits of agreement for each of the pseudo-CT reconstructions for each patient.

2.3 Results

The combined histogram of CT voxel values for all 11 patients is shown in Figure 2.2. The aggregated histogram was dominated by the central soft tissue peaks which were well approximated by a pair of overlapping Gaussian curves. In contrast, lung and

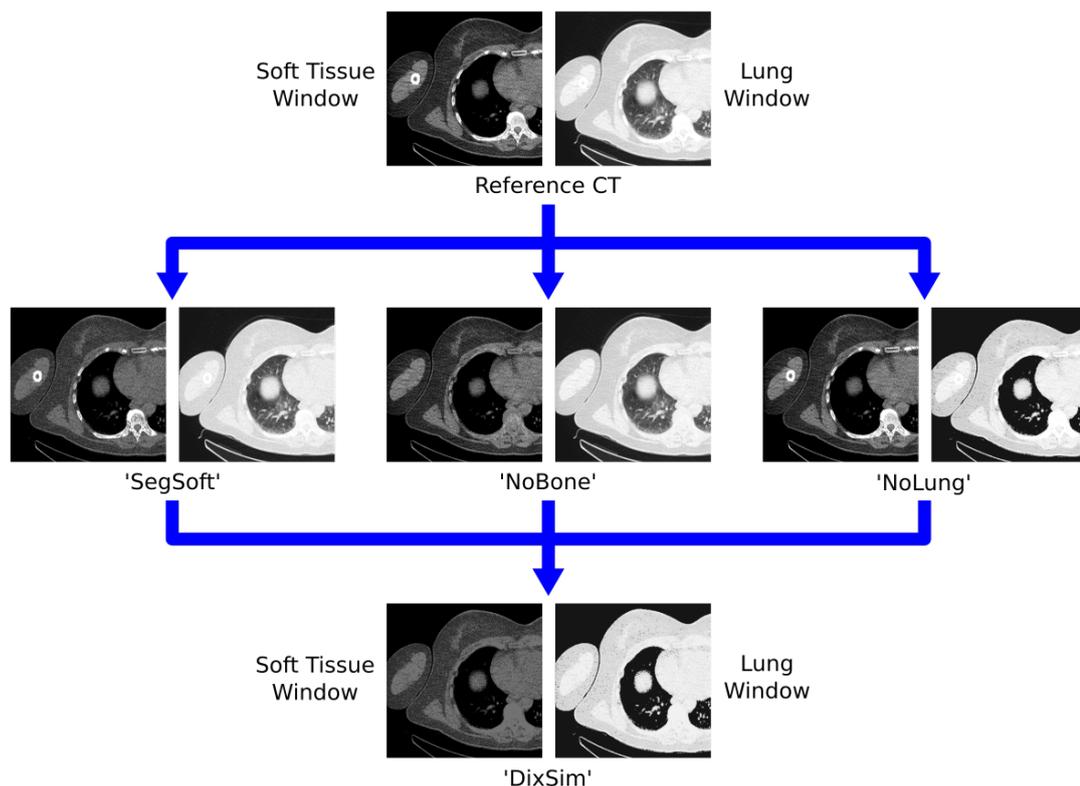


Figure 2.1: Example of pseudo-Computed Tomography (CT) generation workflow. Upper: Original reference CT shown with soft tissue and lung windows. Centre left: 'SegSoft' step showing soft tissue segmented into adipose and non-lipid tissues. Centre: 'NoBone' step showing bone replaced with non-lipid soft tissue. Centre right: 'NoLung' step showing lung replaced with air. Lower: Full Dixon simulation showing all three steps applied simultaneously.

bone tissues ranges showed limited prominence and were not considered suitable for direct modelling via curve fitting. The adipose (red) and non-lipid (blue) distributions were fitted with mean voxel values of -105HU and 25HU and standard deviations of 55HU and 65HU respectively. The point of intersection of the two gaussian curves was found to be -50HU. These fitted parameters were used to define adipose and non-lipid tissue ranges from the point of intersection out to 2, 3 and 4 standard deviations from the mean as shown in Table 2.1.

Each pair of soft tissue ranges shown in Table 2.1 were used to generate a full Dixon simulation pseudo-CT ('DixSim') and three variant pseudo-CT images ('SegSoft', 'NoBone' and 'NoLung') as described in Section 2.2. Examples of each step of the HU replacement algorithm applied to a single slice through the thorax are shown in Figure 2.3. Under visual inspection, the SPECT volumes reconstructed using the manipulated CT images showed only minor qualitative changes when compared to the original reference SPECT reconstruction. No notable reconstruction artefacts

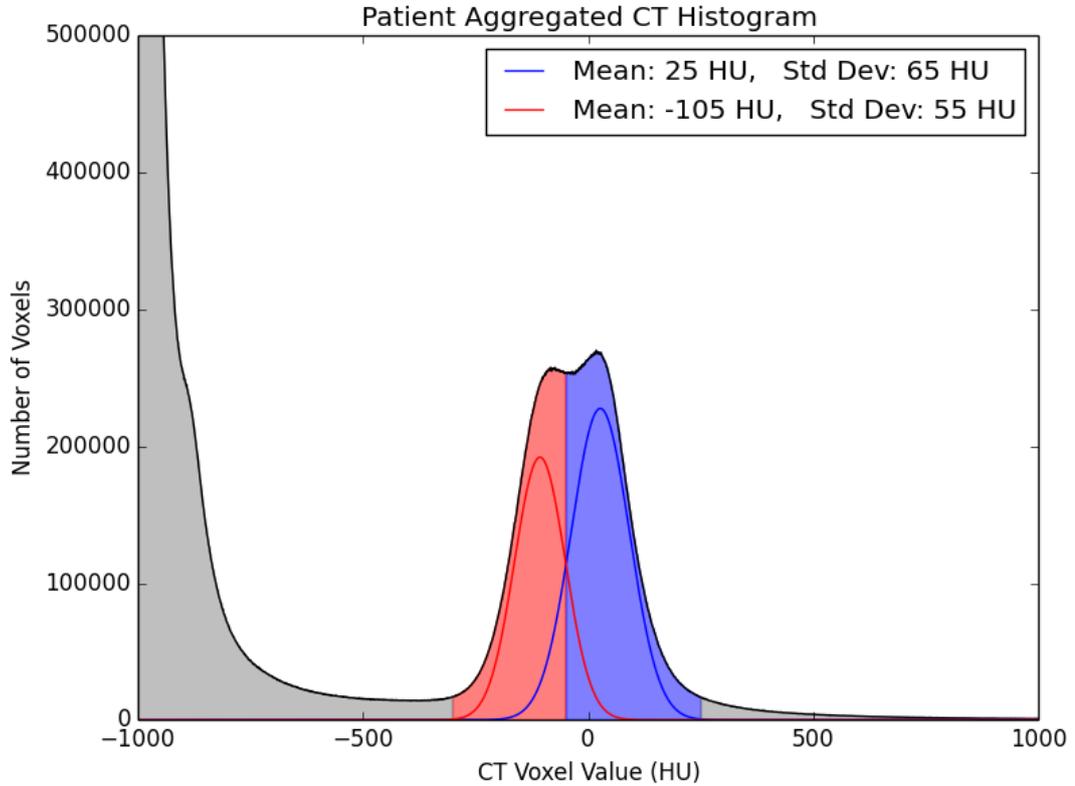


Figure 2.2: Hounsfield Unit (HU) histogram of all 11 patient low-dose Computed Tomography (CT) scan voxel values. Blue curve: Gaussian fit of non-lipid soft tissue region with mean of 25HU and standard deviation of 65HU. Red curve: Gaussian fit of adipose soft tissue region with mean of -105HU and standard deviation of 55HU. Red and blue regions: Adipose and non-lipid regions from -50HU to the edge of the 3σ boundary.

Tissue Region	Mean (HU)	Std. Dev. (HU)	Tissue Ranges (HU)		
			2σ	3σ	4σ
Adipose	-105	55	[-215, -50)	[-270, -50)	[-325, -50)
Non-Lipid	25	65	[-50, 155)	[-50, 220)	[-50, 285)

Table 2.1: Soft tissue region ranges for 2σ , 3σ and 4σ distances from the mean with a shared internal boundary at -50HU. Each region is defined inclusive of its lower bound and exclusive of its upper bound, as indicated by closed and open brackets respectively.

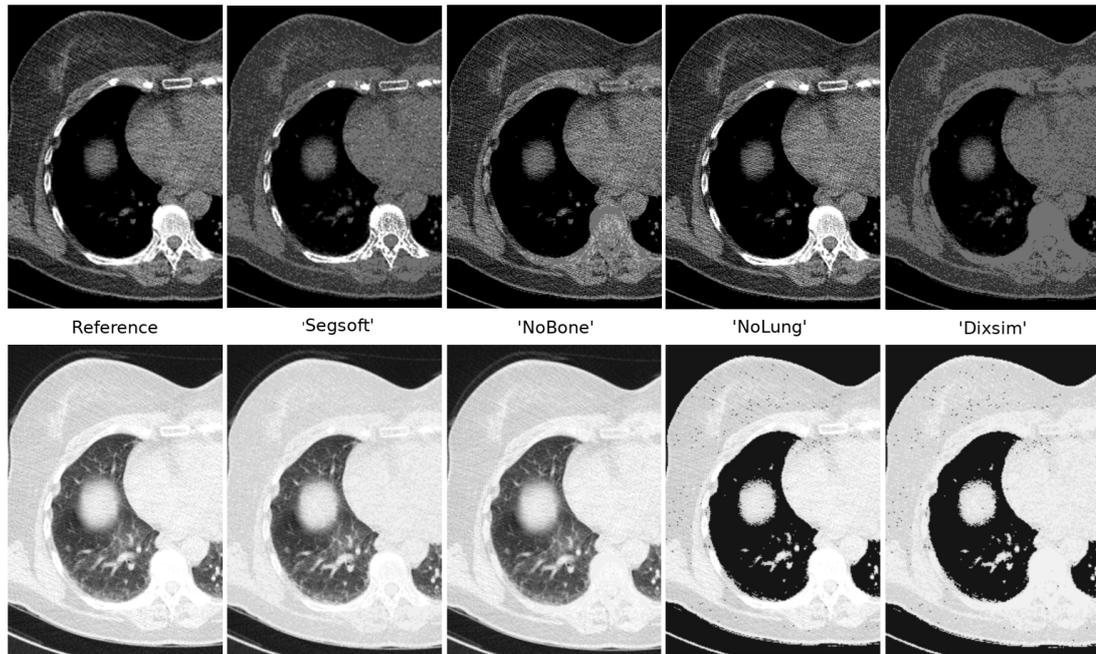


Figure 2.3: Comparison of pseudo-CTs generated with soft tissue regions defined out to 3σ from the mean for a single slice through the thorax. Upper row: Slices shown using standard CT soft tissue window. Lower Row: Same slice shown using standard CT lung window. From left to right: The original reference CT image, 'SegSoft' step showing soft tissue segmented into adipose and non-lipid tissues, 'NoBone' step showing bone replaced with non-lipid soft tissue, 'NoLung' step showing lung replaced with air, Full Dixon simulation with all three steps applied simultaneously.

were detected in any of the datasets for any variant pseudo-CT. Subjectively it was not possible to determine which variant reconstruction was under review without consulting the image header. Some example reconstructed slices are shown Figure 2.4.

Figures 2.5 and 2.6 respectively show voxel plots for example subjects with and without lung coverage in their original CT. Within these figures, each plot shows the voxel-by-voxel difference in SPECT uptake between variant reconstruction and reference reconstruction, plotted against the reference SPECT uptake. Each plot also includes a linear least-square fitting of the plotted data. While the majority of cases clearly demonstrated nonlinear components to the dispersion, the gradient of the least square fit was used as a general indication of the average difference in uptake relative to the unmodified reference SPECT. From visual inspection of plots a consistent pattern emerged across the dataset; the largest deviations in uptake occurred in 'DixSim' and 'NoLung' variants, with 'NoBone' variants showing a much smaller deviation, and 'SegSoft' variants showing high dispersion and gradients very close to zero.



Figure 2.4: Visual comparison of a single reconstructed Single Photon Emission Computed Tomography (SPECT) slice after reconstruction with each of the five attenuation correction maps. Upper left: SPECT reconstructed using original reference Computed Tomography (CT). Upper right: SPECT reconstructed using full Dixon simulation pseudo-CT. Bottom left: SPECT reconstructed using 'SegSoft' pseudo-CT, where soft tissue is segmented into adipose and non-lipid tissues. Lower centre: SPECT reconstructed using 'NoBone' pseudo-CT, where bone is replaced with non-lipid soft tissue. Lower right: SPECT reconstructed using 'NoLung' pseudo-CT, where lung is replaced with air.

Table 2.2 shows the gradient of least-squares fit ($\mu \pm \sigma$) averaged across six distinct groups per reconstruction; in addition to region width, each subject was also grouped on the basis of lung coverage. This produced two groups; those with at least some lung coverage included within the low-dose CT scan (8/11 patients) and those with no lung coverage (3/11 patients). Across all six groups the full Dixon simulation ('DixSim') resulted in mean differences in the range 8-15%, with lower average differences observed for those patients with no lung tissue within the CT coverage (8.2-11.0%) compared to those with larger contributions from lung tissue (12.5-15.1%). Within both lung and non-lung subject groups the lowest mean differences were observed for the 4σ -width regions while the highest mean differences were observed for the 2σ -width regions.

In agreement with visual inspection, these data show that the mean voxel differences in the full ('DixSim') simulation were dominated by contributions from the replacement of lung tissue with air ('NoLung'). In cases with significant lung contributions, 'NoLung' showed mean differences ranging from 9.5-11.7% compared to mean differences of 12.5-15.1% for 'DixSim'. In cases with no lung contribution, the 'NoLung'

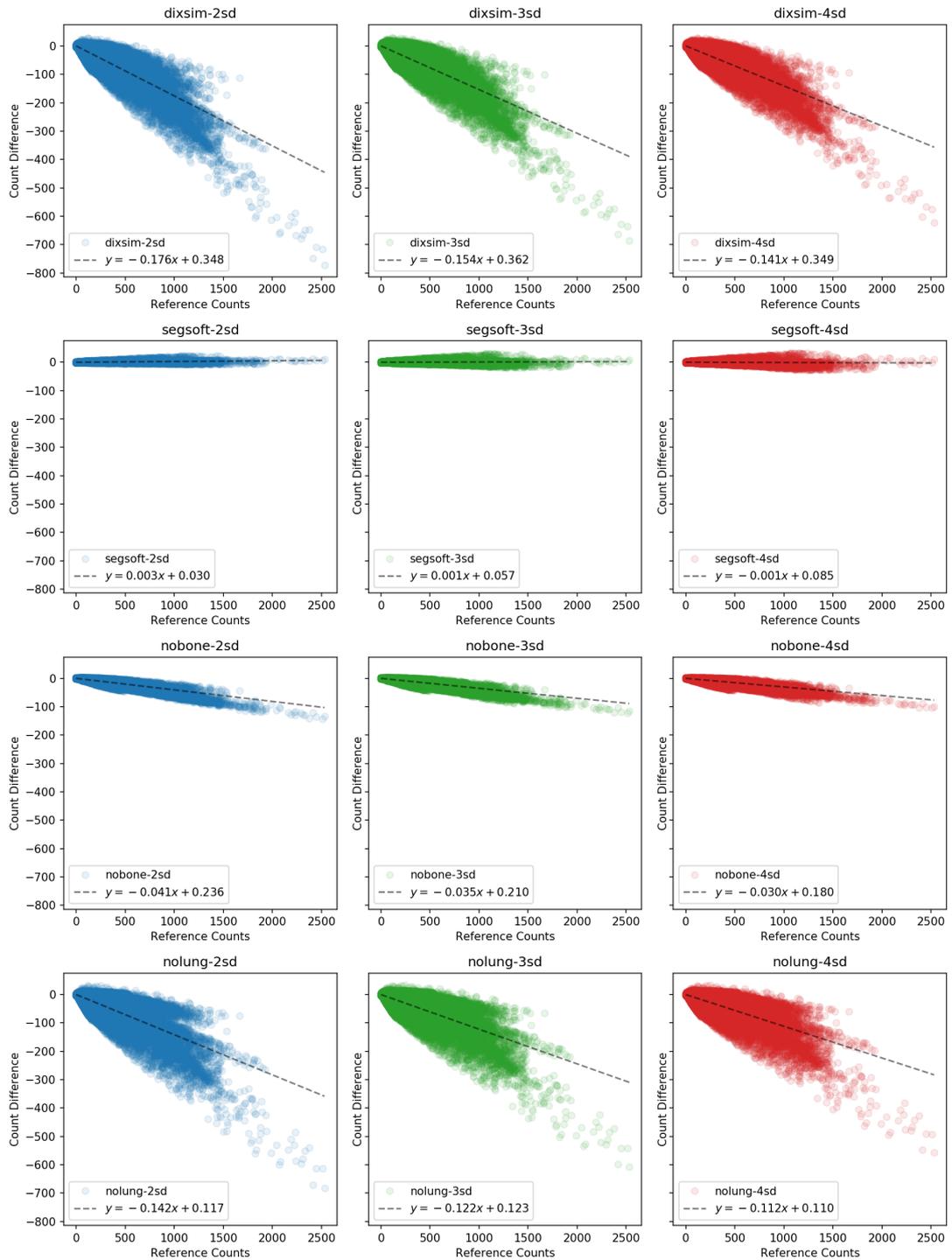


Figure 2.5: Bland-Altman style voxel plots for a single patient with significant lung volume within the Single Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) coverage (CT of full thorax). Each plot shows the voxel-wise difference in SPECT uptake between variant reconstruction and reference reconstruction (vertical axis) versus the voxel value of the reference SPECT reconstruction (horizontal axis).

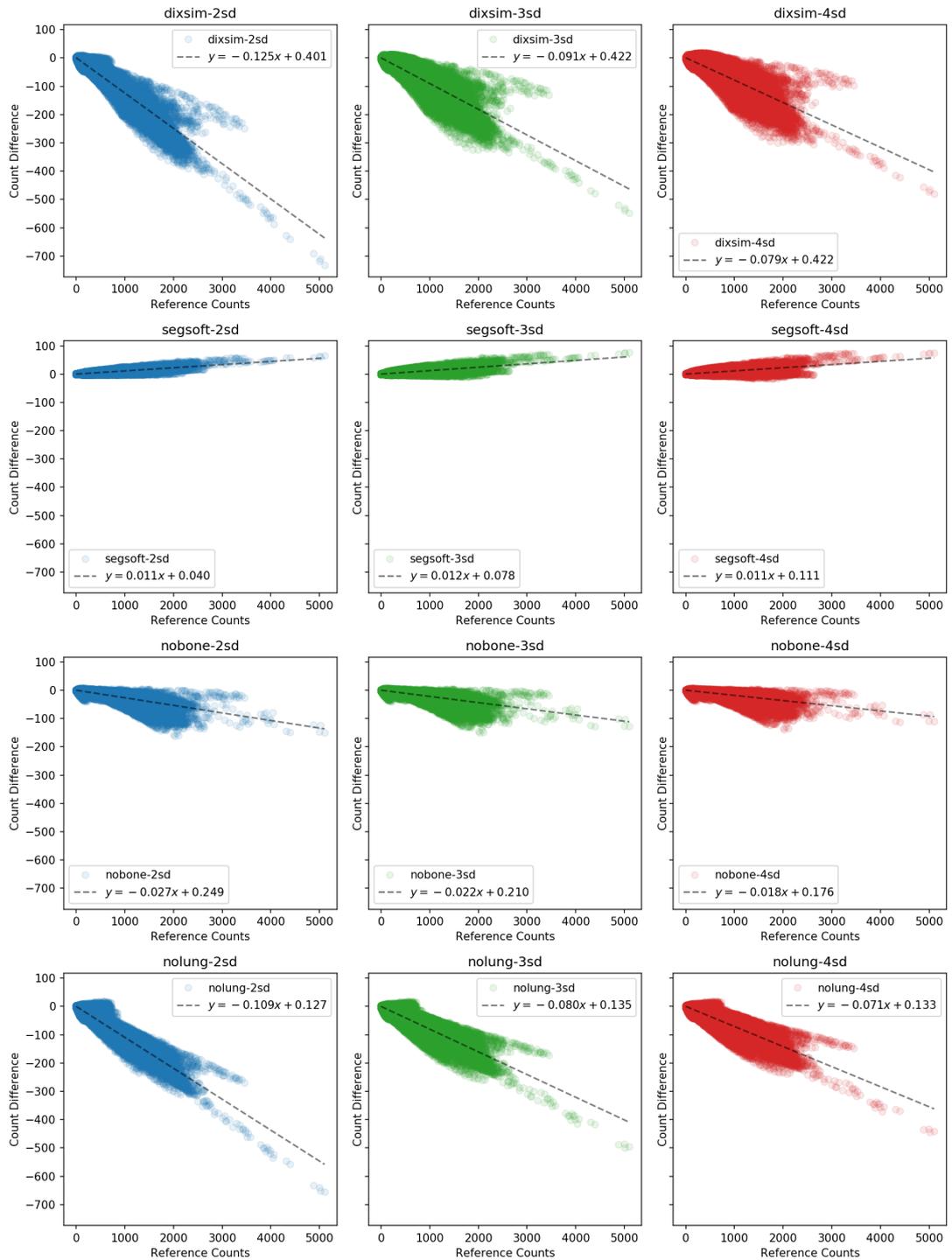


Figure 2.6: Bland-Altman style voxel plots for a single patient with no lung volume within the Single Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) coverage (CT of lower half of abdomen and pelvis). Each plot shows the voxel-wise difference in SPECT uptake between variant reconstruction and reference reconstruction (vertical axis) versus the voxel value of the reference SPECT reconstruction (horizontal axis).

Lung Coverage	Region Width	Least-Squares Fitted Gradient ($\mu \pm \sigma$)			
		DixSim	SegSoft	NoBone	NoLung
Some-Full	2σ	-0.151 ± 0.014	0.001 ± 0.004	-0.036 ± 0.010	-0.117 ± 0.015
	3σ	-0.135 ± 0.014	-0.001 ± 0.005	-0.032 ± 0.010	-0.103 ± 0.013
	4σ	-0.125 ± 0.014	0.001 ± 0.005	-0.027 ± 0.010	-0.095 ± 0.012
None	2σ	-0.110 ± 0.018	0.004 ± 0.007	-0.031 ± 0.007	-0.083 ± 0.032
	3σ	-0.090 ± 0.003	-0.004 ± 0.009	-0.027 ± 0.008	-0.066 ± 0.019
	4σ	-0.082 ± 0.003	0.002 ± 0.010	-0.024 ± 0.008	-0.060 ± 0.015

Table 2.2: Mean and standard deviations of least-squares fitted gradients for subjects grouped by lung coverage, pseudo-CT variant and region width.

mean differences reduced to 6.0-8.3% compared to 8.2-11.0% for 'DixSim'. In contrast, the soft tissue segmentation simulation ('SegSoft') showed mean gradients very close to zero for all combinations of lung coverage and region width. A two-factor Analysis of Variance (ANOVA) confirmed that there were no statistically significant differences between either the lung coverage groups ($p=0.058$) or region width groups ($p=0.37$) for this variant at the 95% confidence level. Finally, the replacement of bone with soft tissue ('NoBone') produced mean voxel differences of 2.4-3.6% across both lung coverage groups, a much smaller contribution than either the full simulation or replacement of lung tissues. A two-factor ANOVA confirmed that statistically significant differences were present between region width groups ($p=0.007$) but not between lung coverage groups ($p=0.06$) for this variant.

2.4 Discussion

That the visual evaluation of the various simulations found no distinguishable features is not particularly surprising given the poor spatial resolution of SPECT imaging. However, despite such simple implementation, the "Segsoft" and "NoBone" simulations showed encouragingly very low uptake errors of 0.1-0.4% and 2.4-3.6% respectively. In contrast the "NoLung" and "DixSim" uptake errors were markedly greater at 6.0-11.7% and 8.2-15.1% respectively. For comparison, Martinez-Möller et al. (2009) found that bone lesion uptake reduced by an average of 8% following an equivalent simulation of Dixon AC for PET and Berker et al. (2012) found the aggregate PET/MR Standardised Uptake Value (SUV) error to be 10.4% across 5 patients who underwent true Dixon-based PET/MR segmentation. Furthermore Marshall et al. (2011) found the mean SPECT/MR Dixon AC error to be 9% in canines; though the

equivalent measure in PET/MR Dixon of canines was reported at 17%, which was much higher than in the other reference studies.

The results of Martinez-Möller et al. (2009) and Berker et al. (2012) would seem to suggest that, at least in cases with no lung tissue coverage, this study was able to achieve essentially equivalent performance to both simulated and real-world Dixon PET/MR. However for cases inclusive of lung tissues, the mean AC error was approximately 50% greater than was demonstrated in these published studies. This large difference would seem to provide strong evidence that the proposed methodology contains one or more serious flaws.

The mean errors measured for each individual step of the Dixon MRAC simulation showed that the dominant source of error in "Dixsim" originated in the "NoLung" step, where lung tissues were replaced with air. However, even in cases without lung coverage this contribution seemed to remain dominant, with the mean errors extracted from the "NoLung" reconstructions being equivalent in value to upwards of 70% of those derived from the full "Dixsim" reconstruction. For this component to remain so dominant, even in cases with no lung tissue, a large proportion of the uptake error must arise from a source other than lung tissue. On further review of Figure 2.2, the most likely source of this error would seem to be the definition of the lower bound of the lipid soft tissue region, as there is clearly a wide valley between the lipid peak and the air dominated region at -1000HU. Although there is some indication of a lung dominated region between -500 and -900HU this is quite poorly defined in comparison to the soft tissue region. The definition of this lower bound, combined with the noisy nature of low-dose CT scans almost certainly caused the "speckling" apparent in the "NoLung" and "DixSim" pseudo-CTs (shown in Figure 2.3). On reflection, it seems likely that repeating the HU modelling and segmentation process with the addition of spatial pre-filters would improve the boundary definition.

As this "lower-bound" error must logically be the only source of error contributing to the "NoLung" values in Table 2.2 for cases with no lung coverage it seems reasonable to estimate the true mean error associated with the replacement of lung tissue with air by the difference in "NoLung" mean gradient between lung coverage groups. Based on the "NoLung" mean errors recorded in Table 2.2, the true lung replacement error

would should be expected to be of the order 3.4-3.7%, which compares favourably with the larger "lower-bound" error of 6.0-8.3%. Following this argument to its logical conclusion, subtraction of this estimated "lower-bound" error from "DixSim" results should estimate the size of the best case result for this simple two region Dixon AC method in SPECT/MR. For cases with no lung coverage the implied best case mean uptake errors are in the range 2.2-2.7%; while cases with lung coverage have implied best case mean uptake errors in the range 6.5-6.9%. Although these are crude estimations, both of these figure ranges would represent AC errors superior to those achieved for PET/MR in Martinez-Möller et al. (2009), Berker et al. (2012) and Marshall et al. (2011).

It is important to acknowledge that this study was not a simulation of real-world Dixon SPECT/MR but instead simulated Dixon SPECT/MR with perfect discrimination of adipose, non-lipid and bony tissues through the use of HU values. It is therefore important to consider the sources of error present in real-world Dixon-based MRAC and how these may increase the best-case prediction estimated above. In real world implementations there will remain the possibility of mis-segmentation of adipose or non-lipid tissue values due to unavoidable factors such as image noise or breathing artefacts. These types of errors are particularly relevant to the case of cortical bone replacement, where post processing of real-world Dixon MRI has a chance of erroneously filling narrow air or gas pockets (such as those within airways, sinuses or the gut) with soft tissue, which is much more attenuating than air. While there is no published data which assess these errors for SPECT/MR, the work of Berker et al. (2012) showed that the addition of ultra-short echo time (UTE) imaging enabled the more accurate segmentation of cortical bone and improved PET/MR mean uptake errors from 10.4% to 1.7% in head and neck patients, a difference in mean error of 8.7%. Assuming that SPECT/MR of the trunk would be subject to at worst 50% of this error (due to single vs dual photons), real world uptake errors would be estimated to be in the range 6.6-7.1% in cases with no lung tissues and 10.9-11.3% in cases with lung tissues. In cases with no lung tissues this range would represent AC errors superior to those achieved for PET/MR. For cases with lung tissues this range would represent inferior performance to PET/MR and an additional tissue segment for lung tissue would most likely be required to improve quantitative accuracy.

2.5 Conclusion

This study simulated a 2-tissue Dixon-based AC method for SPECT/MR using ^{99m}Tc -MDP SPECT/CT scans. Lung tissue replacement initially seemed to dominate the overall uptake error. In cases with no lung tissue this approach was initially able to achieve equivalent performance to published measurements 3-tissue Dixon MRAC of PET. Following more extensive analysis a systematic error was discovered that resulted from the poor definition of the lower bound of the lipid soft tissue region, rather than from true lung tissue replacement. If this source of systematic error were fully resolved, the best case mean uptake error for SPECT/MR was anticipated to be much lower than those achieved in published studies of 3-tissue Dixon MRAC for PET. Assuming magnitude of SPECT/MR uptake errors resulting from mis-segmentation of cortical bone were limited to half the magnitude of those observed in PET/MR, it is expected that a lung tissue segment, and suitable post-processing, would be required to achieve MRAC sufficient for quantitative accuracy.

2.6 Reflection

This chapter laid the ground work for further development of MRAC for SPECT and for further SPECT/MR integration. Initially some follow-up studies were anticipated, including an exploration of the use of machine learning models for the generation of more sophisticated pseudo-CTs from MRI; and the development of local specialist MR sequences for AC. These studies were expected to result in a further clinical pilot study to acquire co-registered SPECT/CT and MRI to demonstrate the feasibility of MRAC for SPECT. This planned work was confounded by the challenging technical issues associated with acquiring sufficiently well co-registered CT and MR imaging, as well as some logistical issues encountered in trying to obtain access to prior research data acquired for the development of MR-only radiotherapy planning. These challenges led to significant delays in this arm of the project, which permitted for a more significant investigation of fused SPECT/MR and semi-quantitative Neuroblastoma scores; discussed in Chapter 3. Ultimately, any further work in the domain of SPECT/MR integration was abandoned when SPECT/CT became available locally at RHC and the project focus shifted to SPECT uptake quantification; discussed in Chapter 4.

Chapter 3

Impact of Fused SPECT-MRI on Neuroblastoma Scores

3.1 Introduction

As discussed in Section 1.3, diagnostic imaging with MRI and ^{123}I -MIBG scintigraphy provides highly valuable and highly complementary information for the diagnosis and management of neuroblastoma (Mueller et al. 2013). Pioneering work by Pfluger et al. (2003) demonstrated that combined reporting of MRI and MIBG maximised sensitivity and specificity of lesion detection, as each individual modality's shortcomings were matched and corrected for by the strengths of their counterpart. However this work was unable to comment on the impact of combined reporting on patient prognosis and management due to its lesion-based study design (as opposed to a case-based study design). In the decade following this work, improvements to imaging and computing technology provided greater access to SPECT and image co-registration and fusion within hospitals. While the spread of these technologies should have led to a rapid uptake of combined reporting, a lack of in-hospital expertise in image fusion and the inherent technical challenges (predominantly patient compliance/motion, poor anatomical definition in SPECT, coping with non-rigid deformations) slowed its dissemination. Furthermore, MRI and Nuclear Medicine departments are typically managed separately; meaning that scans are often separated in time by days or weeks, rendering co-registration of scans markedly more challenging.

As a dedicated children's hospital with an integrated imaging department and dedicated on-site clinical physics support, the RHC in Glasgow (then operating as Royal Hospital for Sick Children (RHSC)) was in an ideal position to overcome these issues

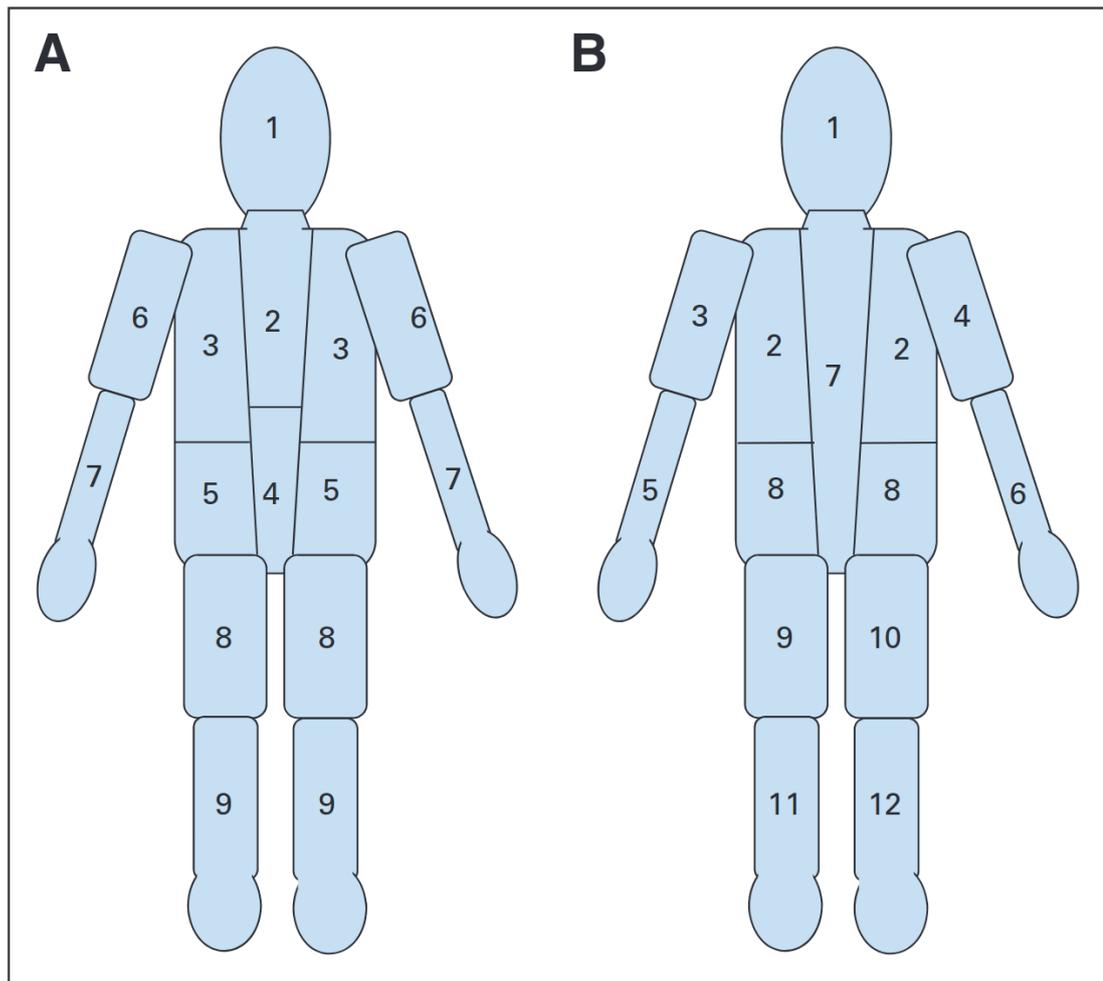


Figure 3.1: Anatomical regions used in Curie (A) and International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) (B) scoring systems. Replication from Decarolis et al. (2013) granted by Wolters Kluwer Health, Inc. under Copyright Clearance Center RightsLink® License #5473711183833

and began working towards a unified approach to neuroblastoma imaging in 2012 with the introduction of a routine MRI and MIBG fusion service. As this service gained recognition amongst referring oncologists, the service progressed to an integrated booking model with anaesthetic support provided for both MRI and MIBG. The case for this move was helped significantly by the foresight of the design of the departments; the Nuclear Medicine department had previously been fitted with anaesthetic gas supply to enable scanning of patients under General Anaesthetic (GA), and the co-location of both MRI and Nuclear Medicine departments allowed for straightforward movement of patients under GA between the departments.

Concurrent with this service transition, semi-quantitative scoring systems for MIBG scans became established within Neuroblastoma trials as an essential common framework for concise assessment of disease burden, response and progression (Messina

et al. 2006; Matthay et al. 2010). Two methods became widespread; the Curie score (Ady et al. 1995) and the SIOOPEN score (Lewington et al. 2009). Both scoring systems utilise similar score methodologies; divide the patient scan into a number of anatomical regions then assign each region a numerical score based on the disease burden identified within each region. Where each score differs is in the division of regions (see Figure 3.1) and the range of possible score values; the Curie score uses 9 skeletal regions and a single patient-wide soft tissue region, each of which is scored from 0-3 score units; while the SIOOPEN score uses 12 skeletal regions (i.e. no soft tissue contribution) and each region is scored from 0-6 score units. In both cases the scores for individual regions are then summed to form the overall patient score. The use of these scores in major clinical trials led to large studies that established their efficacy as prognostic indicators for high risk disease (Ladenstein et al. 2011; Decarolis et al. 2013; Yanik et al. 2013). The work of Decarolis et al. (2013) in particular demonstrated stratification into lower risk patient groups for Curie scores of ≤ 2 score units or SIOOPEN scores of ≤ 4 score units at first diagnosis and patients with no residual disease (i.e. scores of zero score units) following induction chemotherapy.

Despite their proven capability in prognostic assessment, a fair criticism of Curie and SIOOPEN scoring systems is their reliance on planar MIBG scintigraphy. Although not considered during score validation studies, SPECT or hybrid SPECT/CT had already been shown to be useful in resolving lesions of dubious significance (Rozovsky et al. 2008) and had become widely recommended for use in Neuroblastoma imaging (Brisse et al. 2011; Sharp et al. 2011). Furthermore anecdotal evidence from conference discussions has made it clear that despite the validation of SIOOPEN and Curie scores being based purely on planar imaging, in clinical practice Neuroblastoma scores are typically compiled with the benefit of the full interpretation of SPECT imaging if available.

Given this context, the fused SPECT/MR database at RHC offered the opportunity to explore two novel domains; the impact of anatomically fused SPECT imaging on planar Neuroblastoma scores and the prognostic impact of fused SPECT/MR on Neuroblastoma patients. The work of Pfluger et al. (2003), and local anecdotal observations from clinical practice, suggested that fused SPECT/MR would identify more Neuroblastoma lesions on average than planar imaging and therefore would

result in statistically larger Neuroblastoma scores compared to planar imaging alone (Hypothesis 1). By extension it was further hypothesised that these on average higher neuroblastoma scores would result in more cases being classified as unfavourable on the basis of Neuroblastoma score with fused SPECT/MR than with planar imaging alone (Hypothesis 2).

In order to explore these two hypotheses (hereafter H1 and H2 respectively) two case-based experiments were undertaken at RHC; the first of these was a blinded-clinical audit of Neuroblastoma scores with a local clinical observer. Following presentation of this work at a meeting of the British Nuclear Medicine Society (BNMS) (Wallace et al. 2016), a follow-up blinded observer study with an expert external observer was undertaken. At the time of embarking on these experiments there were no case-based studies investigating either hypothesis within the scientific literature.

3.2 Experiment 1: Blinded Clinical Audit

3.2.1 Introduction

At the outset of this work it was unclear whether a single centre such as RHC would have sufficient data to inform on either hypothesis H1 or H2. To begin exploring the problem domain, a pilot study was devised in the form of a blinded observer study with a local clinical observer. The study was staged over two sittings, one for scoring of planar imaging alone, and a second for scoring of planar imaging with the addition of fused SPECT/MR. The ordering of cases was randomised and scoring sessions were separated by more than 7 days in order to minimise the potential for short term memorisation of images and scores. In order to replicate the study conditions of Decarolis et al. (2013) and Yanik et al. (2013), and to maximise the efficacy of blinding, the observer was also blinded to clinical history and therefore analysis was based on consideration of imaging alone.

3.2.2 Methods and Materials

Patient Selection

The Hermes Medical Solutions paediatric nuclear medicine database at the RHC was queried for all cases of ^{123}I -MIBGscintigraphy undertaken between January 2013 and November 2014 inclusive; a time period where the nuclear medicine imaging protocol was held constant. This initial dataset was reduced to exclude patient time-points without SPECT imaging and to exclude time points with SPECT imaging but with no MRI imaging acquired within 14 days of the NM imaging, as these cases were deemed less likely to result in robust image fusion. From the available patient datasets, the earliest scan date for each patient was selected for use in the blinded clinical audit. Following this selection process 15 patient datasets were selected; each dataset contained ^{123}I -MIBGplanar scintigraphy, ^{123}I -MIBG SPECT images and a selection of MRI images.

Imaging Protocols

All Nuclear Medicine imaging was performed using a Siemens Symbia S dual headed gamma camera using Medium Energy Low Penetration (MELP) collimators and acquired approximately 24 hours following isotope administration. All nuclear medicine imaging was acquired using the Siemens default ^{123}I photopeak energy window (15% width centred on 159keV) with adjacent upper and lower scatter windows of equal width acquired for SPECT acquisitions. Planar imaging consisted of wholebody sweep images supplemented by anterior/posterior spot views of the abdomen and left/right lateral spot views of the skull. The wholebody sweep was acquired with a scan velocity of 10cm/min and all spot views were acquired to a total of 500k counts. SPECT imaging consisted of 1-2 SPECT acquisitions of 128 total projections per full rotation of the gantry and 20 seconds per projection. All SPECT acquisitions were acquired with the gantry configured to 180° orientation and using step-and-shoot acquisition mode. SPECT scans were reconstructed using Siemens proprietary FLASH3D reconstruction software, which is based on a standard OSEM iterative reconstruction algorithm and incorporates 3D resolution recovery correction for the collimator point-spread function. Each SPECT acquisition was reconstructed using 8 iterations and 12 subsets (i.e. 96 effective total iterations) with triple energy window scatter correction

and had an 8mm 3-dimensional Gaussian filter applied post reconstruction.

All MRI images were acquired on a Siemens Avanto 1.5T scanner as part of two imaging protocols undertaken for all patients; a metastatic survey protocol and a pre- and post- contrast primary mass imaging protocol. Sequences acquired as part of the metastatic survey included coronal-plane T1-weighted Turbo Spin Echo (TSE) and T2-weighted Short Tau Inversion Recovery (STIR) whole-body composited images. Sequences acquired as part of the primary mass imaging included pre- and post- contrast T1-weighted TSE images. After review of the available imaging for each dataset, the most suitable MRI images for fusion and display were selected and included in the study. T1-weighted imaging was typically used to achieve spatial co-registration and both T1-weighted and T2-weighted imaging were displayed with SPECT preferred for image review purposes. To achieve spatial co-registration, an expert operator co-registered the SPECT images onto the MRI images using a Siemens multimodality workplace and Siemens 3D fusion software. A combination of automatic rigid registration and manual registration using visual comparison were used to achieve spatial co-registration.

Observer Audit Design

Each patient dataset was assigned a pair of independently generated random numbers between 1 and 15. These numbers were used to generate pseudo-anonymous subject identifiers in the form "1XX" and "2YY" where XX and YY were replaced with two-digit representations of the pair of randomly assigned numbers. Using the Hermes Medical GOLD Browser, each dataset was anonymised to remove all identifying information and the patient name and patient ID were replaced with the assigned "2YY" identifier. A copy of the planar scintigraphic component of each dataset was similarly anonymised and the patient name and patient ID were replaced with the assigned "1XX" identifier.

Using Hermes Medical Solutions Hybrid Viewer a local clinical observer reviewed each set of planar images in sequential (i.e. randomised) order from identifier 101 through 115. At a separate later sitting the same observer reviewed each set of planar and fused SPECT/MR images in sequential (i.e. alternatively randomised) order from identifier 201 through 215. An example of the images presented to the observer is shown in Figure 3.2. Each uniquely identified dataset was scored using both the SIOPEN and

Subject	Timepoint	SIOPEN Scores			Curie Scores		
		Planar	+Fusion	Diff	Planar	+Fusion	Diff
1*	Diagnosis	0	0	0	1	1	0
2	Other	0	0	0	0	0	0
3	End Induction	0	0	0	0	1	0
4	Diagnosis	40	40	0	23	23	0
5	Other	0	0	0	1	0	-1
6	End Induction	56	56	0	26	26	0
7	Other	16	17	+1	11	12	+1
8*	Diagnosis	0	0	0	0	1	+1
9	End Induction	8	14	+6	7	10	+3
10	Other	13	17	+4	12	14	+2
11	Diagnosis	1	5	+4	2	7	+5
12	Other	1	1	0	1	1	0
13	Other	0	1	+1	0	1	+1
14	Diagnosis	44	44	0	21	23	+2
15	Other	0	0	0	2	2	0
Mean		11.9	13.0	+1.1	7.1	8.1	+1.0
Std. Dev.		18.9	18.8	1.9	9.3	9.4	1.5
Median		1.0	1.0	0.0	2.0	2.0	+1.0

Table 3.1: Table of International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) and Curie total scores for planar imaging alone, with the addition of fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) and score differences for local observer scoring of 15 patients at a range of clinical points. Subjects 1 and 8, highlighted with *, were found to be non-high risk Neuroblastoma cases and were subsequently excluded from analysis.

on patient prognosis if its addition caused the case to be reclassified into a lower or higher risk group based on SIOPEN or Curie score.

3.2.3 Results

The recorded score totals ranged between 0 and 56 score units for SIOPEN scores and between 0 and 26 score units for Curie scores as shown in Table 3.1. Total score distributions both with and without the addition of fused SPECT/MR showed a strong positive skew as shown in Figure 3.3. 60% (9/15) of cases scored on planar imaging alone and 53% (8/15) of cases score with the addition of fused SPECT/MR received a total score of zero score units respectively.

The addition of fused SPECT/MR images resulted in paired differences from 0 to +6 score units for the SIOPEN scores and -1 to +5 score units for Curie scores, also shown in Table 3.1. Both SIOPEN and Curie score paired differences showed some positive

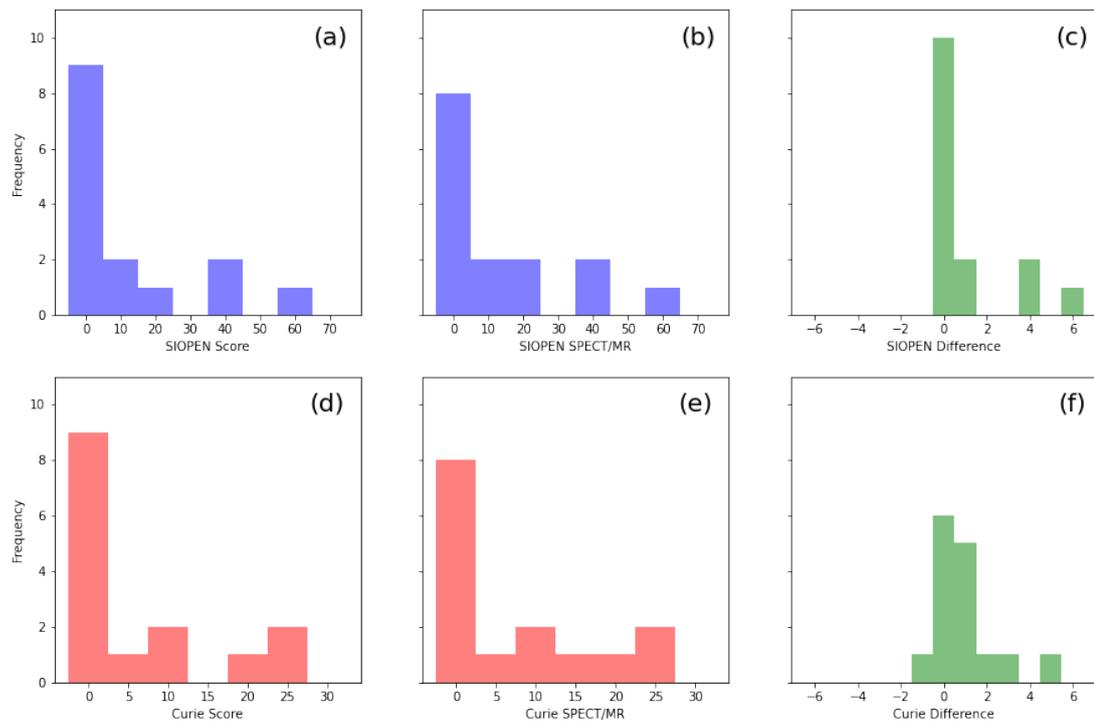


Figure 3.3: Histograms of local clinical observer scoring 15 cases at a range of clinical time points. Plots (a) through (c) show International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) scores and plots (d) through (f) show Curie scores. Plots (a) and (d) show scores derived from planar imaging alone; plots (b) and (e) show scores with the addition of Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) images; plots (c) and (f) show the paired differences in total scores.

skew in their distribution as shown in 3.3. However the SIOPEN paired differences showed a much stronger positive skew, with 67% (10/15) of cases showing zero score differences compared to 40% (6/15) for Curie scores. These visual impressions were reinforced by summary statistics, with SIOPEN scores showing a mean difference of +1.1 score units but a median difference of zero score units, while Curie score differences showed equal mean and median score differences of +1.0 score units. Bland-Altman plots, as shown in Figure 3.4, showed no clear evidence of trend in the paired differences as total scores increased. For both score systems the larger differences were found within the lower half of the score range but small differences were observed at both low and high score values.

The statistical significance of paired differences in both scores were assessed using Wilcoxon Signed-Rank tests. This non-parametric method was selected on the basis of the visual and numerical evidences of non-normality, the low number of cases and the ordinal nature of both score systems. Wilcoxon Signed-Rank tests of paired differences were found to be statistically significant at the 95% level for both SIOPEN scores

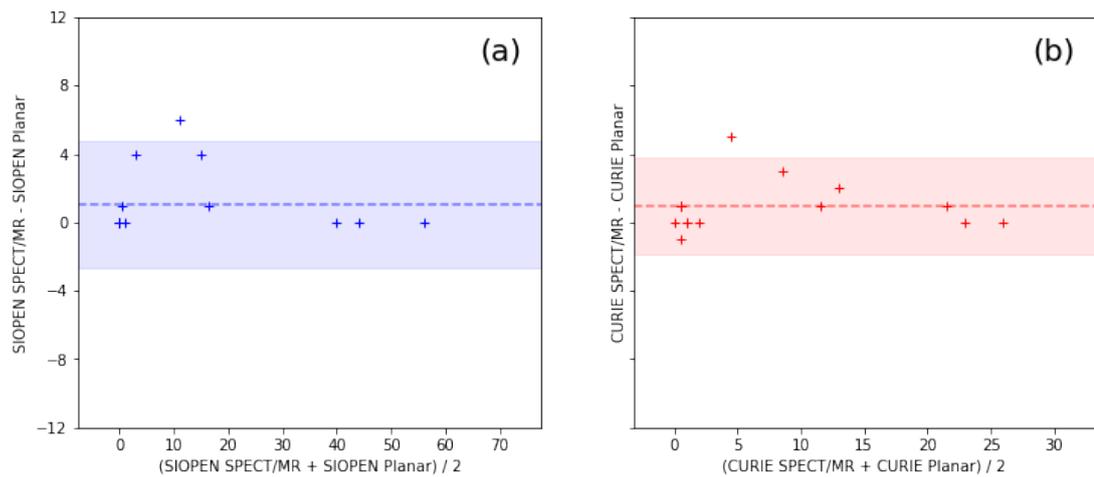


Figure 3.4: Bland-Altman plots of local clinical observer scoring 15 cases at a range of clinical time points. Graphs detail total International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) scores (a) and total Curie scores (b). Mean paired differences in total score are represented by dashed lines and 95% limits of agreement are represented by shaded areas

($p=0.039$) and Curie scores ($p=0.021$) across the full 15 case dataset. However, during analysis two cases were found to correspond to diagnoses other than High Risk Neuroblastoma; one was diagnosed with Low Risk Neuroblastoma and the other with Weaver Syndrome, a rare genetic overgrowth disorder which exhibits an increased risk of development of Neuroblastoma (Coulter et al. 2008). As SIOPEN and Curie scores were validated for use in high-risk Neuroblastoma only, it seemed prudent to repeat the statistical analyses with these cases excluded. Statistically similar results were obtained for the reduced dataset, with Wilcoxon signed-rank tests finding statistical significance for both SIOPEN scores ($p=0.039$) and Curie scores ($p=0.030$).

Five cases of high-risk Neuroblastoma included in the audit were scored at timepoints where SIOPEN and Curie scores have been validated; three at diagnosis and two at end of induction chemotherapy. The remaining cases were scored at a broad range of clinical timepoints and included a mix of disease states including refractory disease, relapsed disease and disease in remission. One case at diagnosis (Case 11 in Table 3.1) showed a shift from lower to higher prognostic risk group on both SIOPEN and Curie scores with the addition of fused SPECT/MR (SIOPEN score: $1 \rightarrow 5$, Curie score: $2 \rightarrow 7$). This change in score corresponded to the localisation of an area of ill-defined faint abdominal uptake; thought to be physiological colonic activity on planar imaging; to diffuse uptake within the right sacral ala as shown in Figure 3.5.

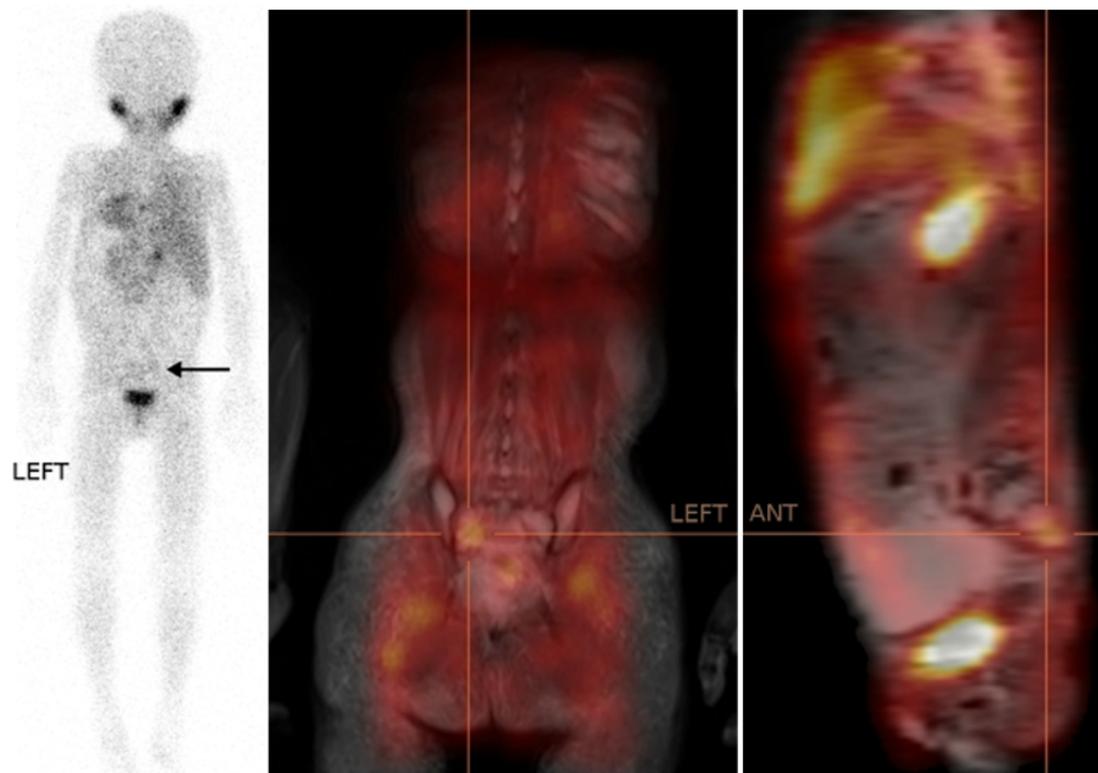


Figure 3.5: Key finding from case 11. Ill-defined abdominal uptake thought to be physiological activity within the colon on wholebody scintigraphy (Left - Posterior View) was revealed to be diffuse skeletal uptake within the right sacral ala on fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) (Centre - Coronal Slice, Right - Sagittal Slice). These findings increased the International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) and Curie scores from lower to higher prognostic risk categories.

One further case at the end of induction chemotherapy (Case 3 in Table 3.1) also showed a shift from lower to higher prognostic risk group on Curie score alone (SIOPEN Score: 0→0, Curie Score 0→1). This change in score corresponded to a region of unilateral uptake in the vicinity of the right adrenal gland in an otherwise negative planar scan. This uptake was scored as either normal liver uptake or benign adrenal hypertrophy; a normal variant of the healthy contralateral adrenal gland following adrenal mass resection. However fused SPECT/MR revealed this to be a residual adrenal mass which had shrunk to small size during induction chemotherapy, as shown in Figure 3.6. At the time of scoring this case, the observer noted that the lack of clinical history limited their ability to accurately interpret the planar imaging of this uptake. On closer review, the lack of history was very likely to have confounded this case and also similarly confounded the scoring of case 5 and case 8. As this was considered a potential source of bias towards discovery of score differences between planar and SPECT imaging, all statistical analyses was repeated with these

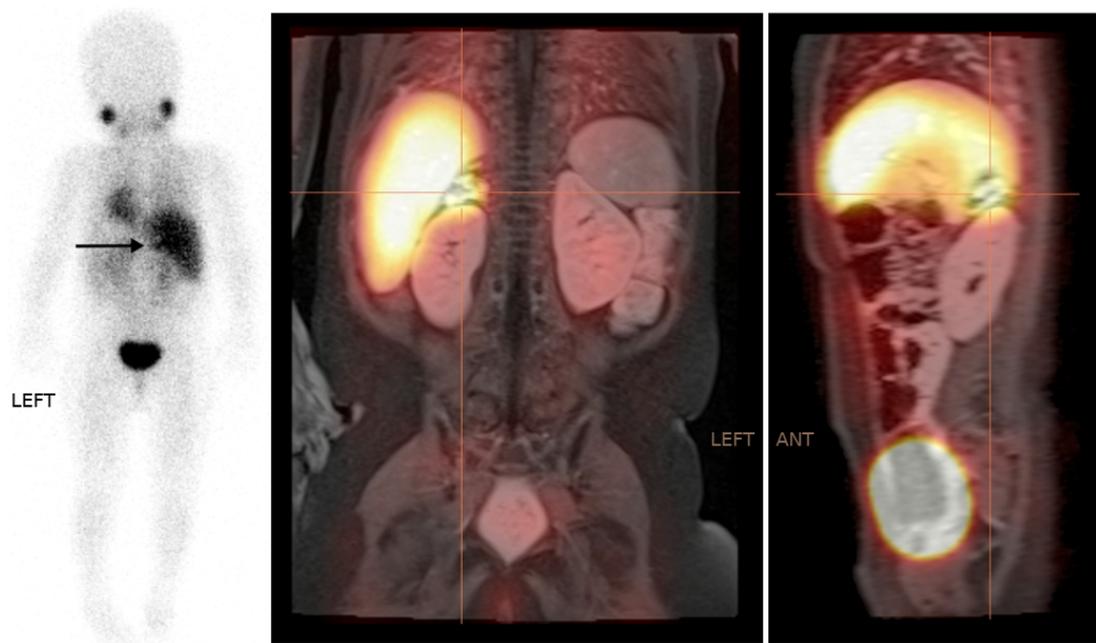


Figure 3.6: Key finding from case 3. Unilateral uptake in the vicinity of the right adrenal gland thought to be post surgical benign adrenal hypertrophy on wholebody scintigraphy (Left - Posterior View) was revealed to be residual shrunken adrenal mass on Fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) (Centre - Coronal Slice, Right - Sagittal Slice). These findings increased the Curie score from lower to higher prognostic risk category.

three cases excluded (note that case 8 had previously been excluded on the basis of diagnosis). In this repeat analysis both SIOPEL ($p=0.039$) and Curie ($p=0.028$) paired score differences remained statistically significant at the 95% level. However this potential confounding influence necessitated that case 3 be discounted as a true reflection of the prognostic impact of fused SPECT/MR.

There were two additional cases showing changes in SIOPEL and Curie scores that were not categorised as prognostically significant on the basis of score changes that nonetheless demonstrated clinically significant findings. The first of these (Case 13 in Table 3.1) is shown in Figure 3.7. This case demonstrated clear identification of a lesion within the left orbit on fused SPECT/MR that was not readily apparent on planar imaging alone. This finding increased both SIOPEL and Curie scores (0→1 score units in both cases) but occurred during post-remission surveillance so was not included in prognostic risk analysis. However the confirmation of disease relapse in a case otherwise considered to be in remission constitutes a highly clinically significant finding that would provoke a marked change in patient management.

The second such case (Case 14 in Table 3.1) is shown in Figure 3.8. This case

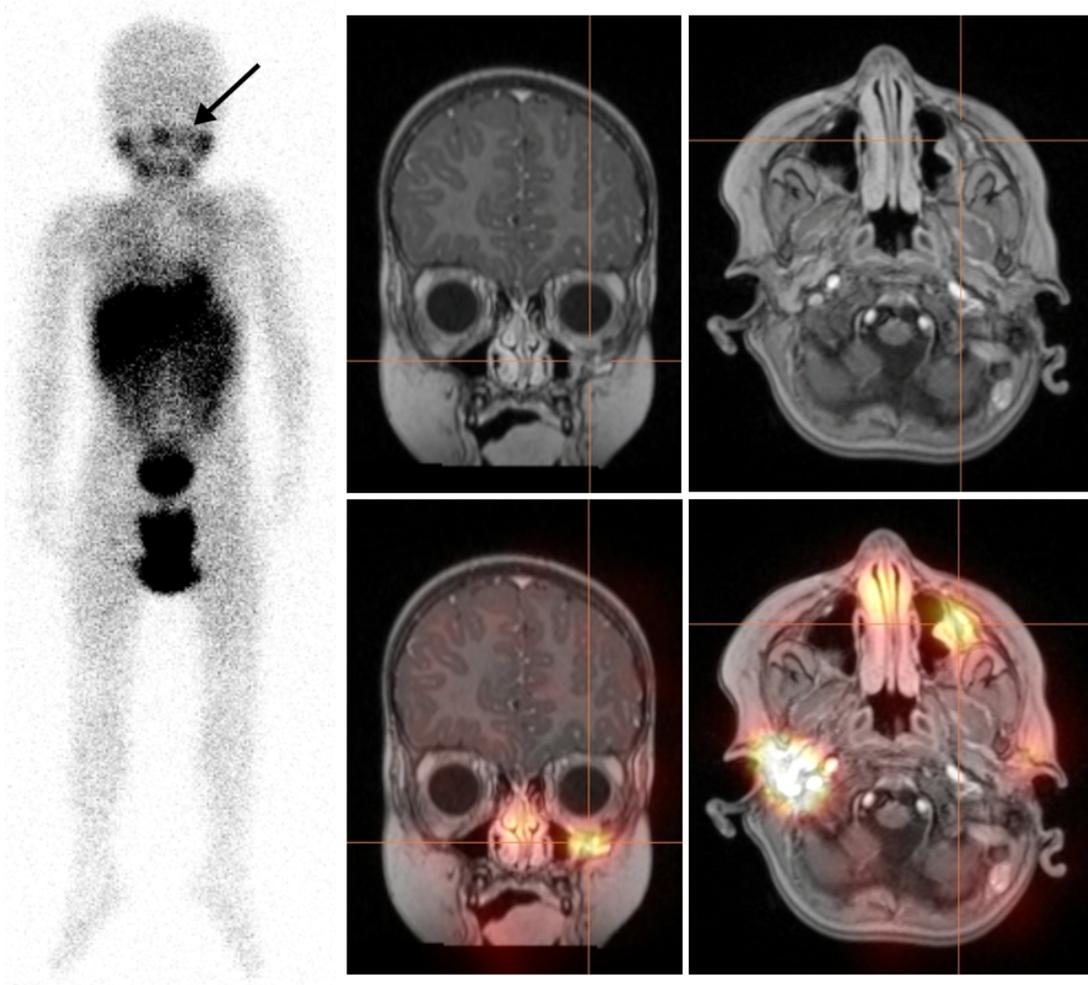


Figure 3.7: Key finding from case 13. Slightly elevated uptake in left perinasal region on wholebody anterior view showed significantly increased prominence and localisation to left orbital bone on fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) imaging.

showed areas of diffuse uptake, assumed to be skull vault metastases on whole-body imaging, that were subsequently localised to extensive dural effusions on fused SPECT/MR. While this finding occurred during diagnosis and also increased the Curie score (21→23 score units) the underlying disease burden in the patient was otherwise high and therefore no change in prognostic risk group was indicated. However this finding was noted as highly clinically significant, since it indicated an increased risk of brain injury due to elevated intracranial pressure and an increased risk of metastatic spread into the brain from the dural space.

3.2.4 Discussion

At the time of undertaking this study there existed a common perception among clinicians that fused SPECT/MR and other hybrid techniques would contribute only

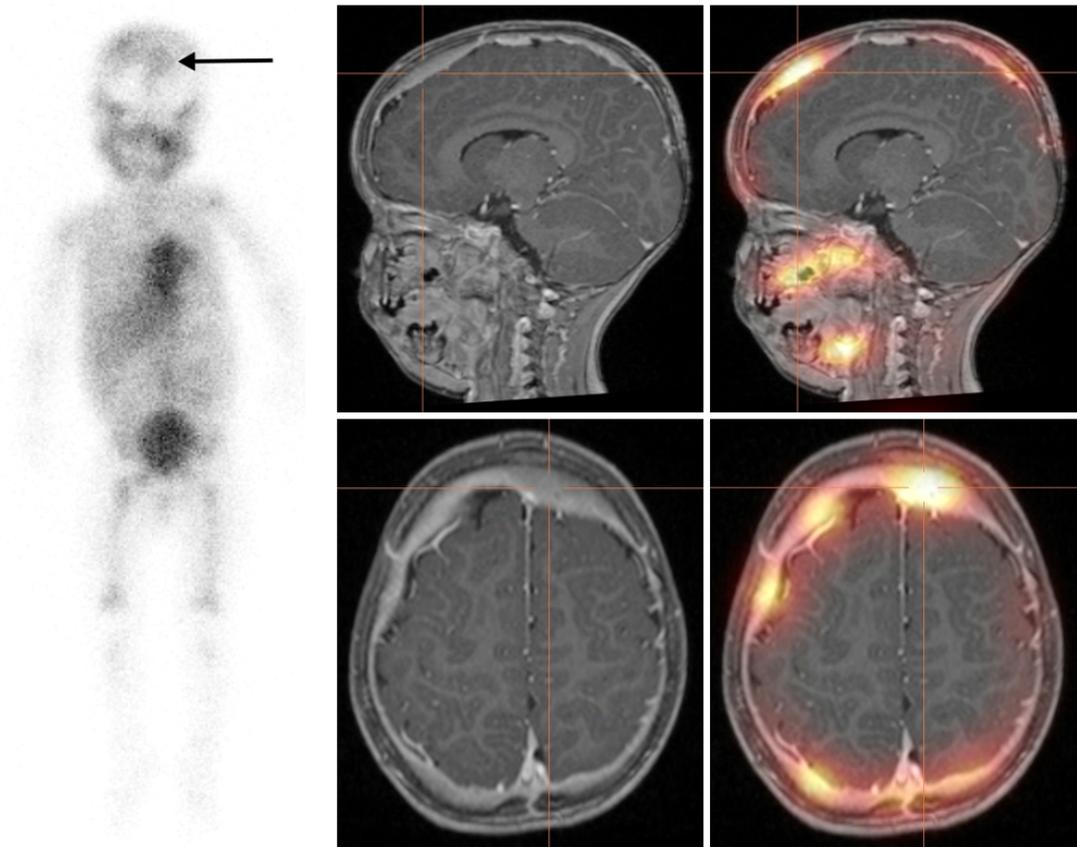


Figure 3.8: Key finding from case 14. Diffuse uptake in skull vault assumed to be entirely skeletal in nature on wholebody imaging was localised to widespread dural effusions on fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) imaging.

small differences to Neuroblastoma scores. This was based on an assumption of limited impact on skeletal scores, the main component of each scoring system. However this small single-observer study was nonetheless able to find statistically significant score differences with the addition of SPECT/MR to both SIOOPEN and Curie scoring methods at the 95% level. The observed changes showed an even split between soft tissue scores and skeletal scores; four cases showed changes in soft tissue alone, four cases showed changes in skeletal disease alone and a single case showed changes in both skeletal and soft tissue scores. While a number of the observed skeletal changes were mere refinements to diffuse uptake scores, the study was also able to find evidence of prognostically significant score changes in a single case (Case 11, highlighted in Figure 3.5). Taken together, these findings demonstrate clear contradiction of conventional wisdom and show evidence supporting both hypotheses H1 and H2, as postulated in Section 3.1. With both scoring systems showing statistically significant differences with the addition of fused SPECT/MR it was also clear that

neither SIOPEX nor Curie scores offered any inherent advantages towards mimicking the observed differences.

In addition to showing statistically significant score differences across a range of time-points, this study also showed patterns of uptake that were recognised as clinically significant for patient management in two cases (Cases 13 and 14, highlighted in Figures 3.7 and 3.8 respectively). While not significant in terms of validated score changes, these observations nonetheless provide strong evidence that localised tomographic imaging would have a clinically significant impact on patient management if adopted as the standard of care; in close agreement with the work of (Pfluger et al. 2003). These cases also provide strong evidence that patient management would be improved by application of more advanced imaging to the head; a finding not described in prior publications. Given the obvious need to image the patient trunk to depict adrenal and para-spinal disease, this finding would seem to encourage the implementation of multiple field of view (FOV) SPECT imaging as the standard of care. These findings directly led to a revision of imaging protocol at RHC to include multi-FOV SPECT for all patients under GA and for all suitably compliant non-GA patients.

The findings of this study are particularly interesting in light of the observation that clinicians do not typically issue scores based on planar imaging alone; indeed should any fused or hybrid imaging be available, all such supplementary imaging is reviewed prior to scoring. Based on the results of this study, it would be expected that small systematic differences will exist between centres that make extensive use of fused imaging and centres that do not. On this basis, caution should be observed where locally reported score thresholds are used in multicentre clinical trials either as gatekeepers for recruitment or for disease response assessment. A consistent approach to imaging, interpretation and scoring would seem essential to avoid biasing such measurements towards increased scores in centres with more extensive use of advanced imaging techniques. Independent confirmation of all Neuroblastoma scores would seem a prudent approach to ensure internal consistency should trials use them in this fashion.

It is important to recognise the limitations of the design of this single centre study.

Firstly the sample size is small, a fact which may have contributed to the distinctly non-normal distribution of the SIOOPEN and (to a lesser extent) Curie score differences; necessitating the use of typically less powerful non-parametric analyses. The study design also included a sequential case pool, which contributed to the broad range of clinical timepoints and two cases of non high-risk disease being included; factors which both limited the number of cases eligible for consideration of prognostically significant score changes and reduced the overall disease burden across the dataset. The omission of clinical history also had an unfortunate confounding interaction with the range of timepoints included in the study, and this led to the necessary exclusion of three cases with potentially suspicious adrenal uptake. This confounding effect occurred due to the importance of surgical status in interpreting adrenal gland uptake; a lone healthy adrenal gland is much more likely to show signs of benign adrenal hypertrophy and confirmation of the presence of a contralateral adrenal was only visible with the addition of fused SPECT/MR. In these cases the MRI imaging was effectively providing the clinical history, at least as regards adrenal surgical status, in addition to soft tissue characterisation. Despite these limitations, and the necessary exclusion of four cases from the final statistical analyses, this small study nonetheless was able to identify significant findings relevant to both hypotheses under consideration.

The remaining limitations of the study relate to the clinical observer and the registration of the images. While the observer had extensive experience interpreting clinical imaging of Neuroblastoma they lacked experience in Neuroblastoma scoring, as neither SIOOPEN nor Curie scores were in routine use at RHC at the time of undertaking this study. In the weeks prior to taking part in this work, training cases distributed via presentation at the European Symposium of Paediatric Nuclear Medicine (ESOPNM) were completed by the observer to familiarise them with both scoring methods. Additionally, with only a single observer there was no opportunity to comment on inter-observer agreement in scores. This had been noted as a key strength of the planar scoring methodology at validation in Decarolis et al. (2013), with the study demonstrating Pearson's correlation coefficients of 0.96 and higher for all combinations of scoring method and clinical timepoint.

The registration of images was also achieved through the visual assessment of an expert operator with significant experience fusing SPECT and MRI images rather

than by a true spatial registration based on markers or serial scanning. This process introduced additional sources of error such as subjective visual assessment (i.e. user bias) and the possibility of non-rigid deformations to the patient between scans (e.g. different head positions, limb positions, phases of the breath cycle or difference in scanner bed/patient support geometries). These errors could potentially contribute to misinterpretation of the fused images. However, the use of an expert human operator also allowed for the generation of multiple fused images (e.g. one for the head, one for the trunk) to attempt to correct for any such non-rigid deformations. The use of an expert human operator for fusion was considered unavoidable due to the non-existence of clinical SPECT/MR scanners and the extremely limited options available for MR compatible fiducial marker/patient positioning systems which allow for radioactive markers.

3.2.5 Conclusion

This small single observer study demonstrated clear statistical evidence in support of hypothesis H1; that Neuroblastoma scores are higher with the addition of fused SPECT/MR to planar imaging. It also found a single case in support of hypothesis H2; that increased scores due to fused SPECT/MR would increase prognostic risk when compared to validated score thresholds. Further examples of the clinical impact of fused SPECT/MR on patient management were also noted. Overall this study shows that fused SPECT/MR has a statistically and clinically significant impact on Neuroblastoma scores, and that SPECT/MR has a larger clinical impact outside of the scoring process. Expansion of the role of fused SPECT/MR in Neuroblastoma would seem to be indicated. However given the small case pool and single local observer, it would seem prudent to repeat the study with a larger case pool focused on a single clinical timepoint and an expert external observer in order to confirm these findings.

3.3 Experiment 2: Blinded Observer Study

3.3.1 Introduction

In order to address the limitations of experiment 1, a follow-up study was proposed with changes to the study protocol such that; (1) all subjects were scored at diagnosis

and (2) a single external expert clinical observer was used for scoring. The first of these changes was expected to address the primary weaknesses of the prior study design; namely low case numbers suitable for prognostic assessment, a majority of cases having zero scores and the ambiguity of adrenal uptake in the absence of clinical history. The use of an expert observer was essential to provide an authoritative confirmation of the prior study. As a secondary benefit, the external observer was also expected to remove any potential biases due to local familiarity with cases, though this was not perceived to be a particular weakness of the prior study. Ethical approval was sought for the proposed follow-up research study and this was granted via proportionate review by a Research and Ethics committee.

3.3.2 Methods and Materials

Patient Selection

In order to maximise patient inclusion, the selection process of the follow up experiment considered the full historical dataset of the SPECT/MR fusion service. At the time of undertaking scoring, this record spanned November 2012 to March 2017. The inclusion process followed a similar procedure to that described in Section 3.2.2; the minimum inclusion criteria were an acquired ^{123}I -MIBG SPECT scan and MRI dated within 14 days of the NM imaging. As with the previous study, the earliest scan meeting these criteria was selected for each patient within the dataset.

The inclusion query returned a series of 32 subjects which was then reduced by exclusion criteria rejecting (1) any cases acquired at clinical time-points other than initial diagnosis and (2) any cases where SPECT and MRI images were not suitable for co-registration. Following the application of these exclusion criteria, 17 subjects remained for entry into the study, including 5 cases previously scored by the local observer. Two of these cases were the cases of low risk Neuroblastoma and Weaver syndrome included in the prior study and excluded in Section 3.2.4. While these cases should not be considered for statistical analysis of validated score differences, they were included to increase the number of cases scored by both observers for assessment of inter-observer score differences and to measure expert evaluation of adrenal uptake. Each subject dataset entered into the study contained ^{123}I -MIBG planar scintigraphy (wholebody acquisition with optional supplementary spot views), one or more ^{123}I -

MIBG SPECT fields of view and MR imaging suitable for fusion with reconstructed SPECT images.

Imaging Protocols

Five cases were included within the dataset of experiment 1. For these cases the nuclear medicine acquisition parameters were identical to those described in Section 3.2.2. For the remaining 14 cases acquired after November 2014 the nuclear medicine acquisition parameters differed only in two areas; the wholebody acquisition speed was reduced to 5cm/min in order to comply with updated imaging guidelines published by the European Association of Nuclear Medicine (EANM); and two SPECT fields of view were acquired where patient compliance permitted. The increased coverage was used to acquire SPECT of the entire trunk (equivalent to chest/abdo/pelvis) and/or entire head depending on patient size and disease burden evident on wholebody imaging.

Following the move to a new hospital campus in June 2015 and an upgrade of the Hermes Medical Solutions system, MIBG SPECT reconstruction had switched to being performed using Hermes Medical Solutions Hybrid Recon 1.1.2, which offered significant advantages for multiple field of view reconstruction. To ensure a consistent reconstructed SPECT appearance throughout the experiment, all SPECT acquisitions were reconstructed using this software. SPECT reconstructions were performed using 8 iterations and 16 subsets (128 effective iterations) with resolution recovery and a 3D Gaussian post reconstruction filter with a Full Width at Half Maximum (FWHM) of 0.9mm was applied post reconstruction. Notably, the Hermes Medical Solutions Hybrid Recon software only offers scatter correction via a Monte Carlo algorithm using a patient attenuation map (e.g. CT scan) so in the absence of such a map, all reconstructions were undertaken with resolution recovery applied but no scatter correction.

All MRI images were consistent with the sequences described in 3.2.2. SPECT and MRI images were spatially co-registered by an expert operator using the Hermes Medical Solutions Hybrid 3D software application. A combination of automatic rigid registration and manual co-registration based on visual assessment were used to align the images. After co-registration, suitable T1-weighted and T2-weighted images for fusion were selected for display in the observer audit.

Observer Study Design

The design of the observer study was identical to that described in 3.2.2; each set of planar imaging was copied and pseudo-anonymised using two sets of independently generated random numbers (in this case 3XX and 4YY where XX and YY were replaced with the randomly generated numbers from 01 to 15). The SPECT and MRI images were then pseudo-anonymised using the second of these random number lists. As with the previous study, each uniquely identified dataset was then scored by the observer. The planar-only datasets (3XX) were scored in a single sitting and the planar+SPECT datasets (4YY) were scored in a second later sitting.

Statistical Analysis

Statistical analyses were performed following similar methodology to that described in 3.2.2; visual and numerical inspection were undertaken using histograms, Bland-Altman plots and summary statistics. Wilcoxon signed rank tests were used to assess the statistical significance of the pairwise differences in total score. However unlike in the previous experiment, all but two cases were able to be considered in the analysis of prognostic risk group reclassification due to the improved exclusion criteria and expanded inclusion period. While this experiment only involved a single observer, it was noted that there was a small overlap in cases previously scored by the local observer in experiment 1 (Section 3.2) and those to be scored by the expert observer in this experiment. Therefore a small scale inter-observer comparison was undertaken for these cases using mean score differences and Pearson Correlation Coefficient (PCC) analysis.

3.3.3 Results

The recorded score totals ranged between 0 and 57 score units for SIOOPEN scores and between 0 and 27 score units for Curie scores as shown in Table 3.2. Distributions of total scores with and without the addition of fused SPECT/MR showed a marked reduction in positive skew compared to the prior study, as shown in Figure 3.9 A much greater proportion of cases received scores in the upper half of the score range compared to experiment 1 and furthermore only a small minority of cases (3/15 for SIOOPEN scores and 1/15 for Curie scores respectively) received a total planar score of zero score units compared to the 9/15 cases with zero planar scores in the prior

Subject	#FOVs	WB Speed (cm/min)	SIOPEN Scores			Curie Scores		
			Planar	+Fusion	Diff	Planar	+Fusion	Diff
1	2	5	46	48	+2	21	22	+1
2	1	5	0	0	0	0	0	0
3	1	10	39	40	+1	22	23	+1
4*	1	10	0	0	0	0	1	+1
5	1	5	0	0	0	1	1	0
6	2	10	37	36	-1	15	16	+1
7	2	5	33	32	-1	14	13	-1
8	1	5	57	57	0	27	26	-1
9	1	5	0	4	+4	1	2	+1
10	1	10	6	12	+6	5	8	+3
11	1	10	39	38	-1	23	21	-2
12	2	5	46	49	+3	20	23	+3
13	2	5	41	46	+5	23	24	+1
14	1	5	5	6	+1	5	6	+1
15	2	10	54	48	-6	26	23	-3
16*	1	10	0	0	0	1	1	0
17	2	5	32	28	-4	20	18	-2
		Mean	25.6	26.1	+0.6	13.2	13.4	+0.2
		Std. Dev.	20.8	20.0	3.2	9.8	9.3	1.7
		Median	33	32	0.0	15	16	+1.0

Table 3.2: International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) and Curie total scores for planar imaging alone, with the addition of fused SPECT/MR and score differences for expert observer scoring of 17 patients at diagnosis. Subjects 4 and 16, highlighted with *, were non-high risk Neuroblastoma cases included for the purposes of inter-observer comparison and were excluded from statistical analysis of score differences.

experiment.

The addition of fused SPECT/MR images resulted in paired differences from -6 to +6 score units for SIOPEN scores and -3 to +3 score units for Curie scores, also shown in Table 3.2. Both SIOPEN and Curie score paired differences showed an approximately normal distribution, with the majority of differences clustered close to zero and little asymmetry towards either extreme. Both scoring systems showed a larger range of differences compared to the previous local observer study, with negative score differences being larger in magnitude and more common for both SIOPEN and Curie paired differences. These visual impressions agreed with the summary statistics which showed SIOPEN and Curie mean score differences (+0.6 and +0.2 score units respectively) and median score differences (0 and +1 score units respectively) were close to zero. Bland-Altman plots, as shown in Figure 3.10, showed some evidence of negative differences increasing as average scores increased but positive differences

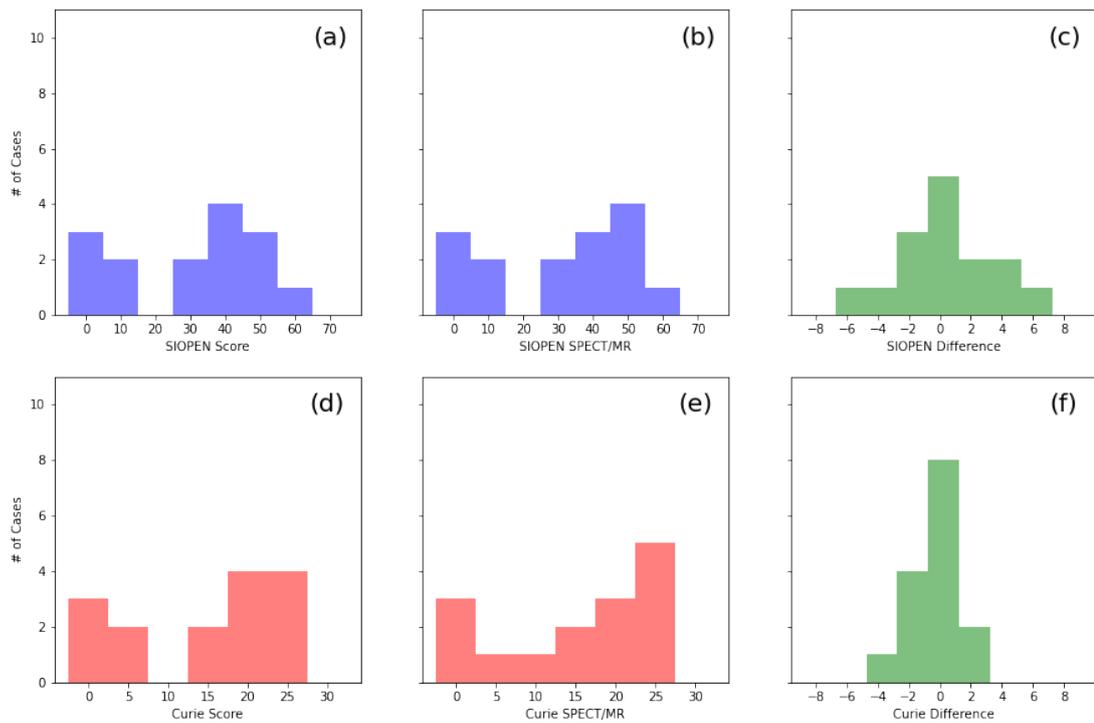


Figure 3.9: Histograms of expert observer scoring 15 cases of high-risk neuroblastoma at diagnosis. Graphs detail total International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) scores (a) and Curie scores (d) derived from planar imaging alone; total SIOPEN scores (b) and Curie scores (e) with the addition of Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) images and paired differences in total SIOPEN scores (c) and Curie scores (f).

demonstrated similar magnitudes across the full score range.

As with the prior experiment, Wilcoxon Signed-Rank tests were used to assess the statistical significance of score differences; with the addition of fused SPECT/MR; for the 15 cases of high-risk neuroblastoma. However these tests were unable to demonstrate statistically significant differences for either score system at the 95% level ($p=0.39$ and $p=0.32$ for SIOPEN and Curie scores respectively). Following this primary analysis of score differences, the score difference data were grouped to examine the effect of both the number of SPECT FOVs and the wholebody acquisition speed on the differences in scores. As shown in Table 3.3, both SIOPEN and Curie scores exhibited lower average score differences for cases where wholebody imaging was acquired at 5cm/min compared to when it was acquired at 10cm/min. The standard deviations of differences were also larger for cases where two SPECT fields of view were acquired or where the wholebody speed was increased. A two-factor ANOVA was performed for each scoring system but neither analysis was able to demonstrate a statistically significant difference for either grouping parameter (SIOPEN: $p=0.48$,

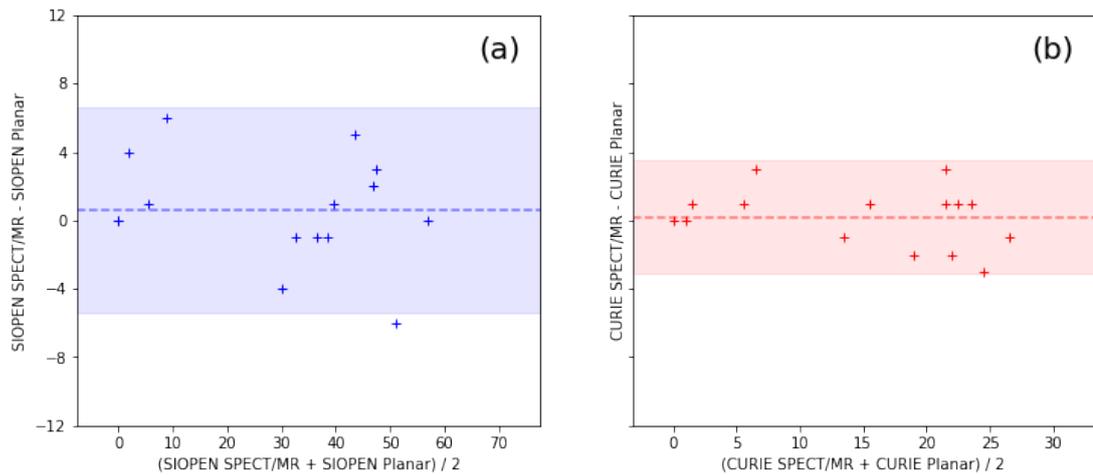


Figure 3.10: Bland-Altman plots of expert observer scoring 15 cases of high-risk neuroblastoma at diagnosis. Graphs detail total International Society of Paediatric Oncology Neuroblastoma Europe (SIOPEN) scores (a) and total Curie scores (b). Mean paired differences in total score are represented by dashed lines and 95% limits of agreement are represented by shaded areas.

$p=0.31$; Curie: $p=0.86$, $p=0.81$; for wholebody speed, number of SPECT fields of view respectively).

No cases in this study demonstrated a prognostically significant change in Neuroblastoma scores with the addition of fused SPECT/MR. However, four cases were flagged as containing clinically significant findings despite not showing prognostically significant changes in neuroblastoma scores. The first of these (Case 15 in Table 3.2) was also included in the previous study and in both studies the case was flagged as containing clinically significant findings due to the potential dural uptake shown on SPECT/MR. Images of this case are shown in Figure 3.8 and described in Section 3.2.3.

Cases 7 and 9 were both flagged with potentially clinically significant findings within or adjacent to the lungs. In case 7, shown in Figure 3.11, the marked uptake was initially scored as skeletal rib disease on planar imaging alone but was localised to the left upper lobe on fused SPECT/MR. Based on the SPECT/MR appearances, this region was judged to most likely represent non-specific uptake within a left upper lobe atelectasis and was therefore not scored as soft tissue uptake on fused SPECT/MR. However, the observer noted that the relatively poor quality of MRI characterisation within the lungs could not fully exclude metastatic uptake and in a clinical situation they would recommend further imaging with CT. Following study completion, it was confirmed

SIOPEN Score Differences					
WB Speed	1FOV	2FOV	WB Speed	1FOV	2FOV
5cm/min	0	+2	10cm/min	+1	-1
	0	-1		+6	-6
	0	+3		-1	
	+4	+5			
	+1	-4			
Mean	+1.0	+1.0	Mean	+2.0	-3.5
Std. Dev.	1.7	3.5	Std. Dev.	3.6	3.5

Curie Score Differences					
WB Speed	1FOV	2FOV	WB Speed	1FOV	2FOV
5cm/min	0	+1	10cm/min	+1	+1
	0	-1		+3	-3
	-1	+3		-2	
	+1	+1			
	+1	-2			
Mean	+0.2	+0.4	Mean	+0.2	+0.4
Std. Dev.	0.8	1.9	Std. Dev.	0.8	1.9

Table 3.3: Grouped SIOPEN (upper table) and Curie (lower table) score differences with summary statistics for 15 cases of high-risk Neuroblastoma. Data are grouped by wholebody acquisition speed (rows) and number of SPECT fields of view (columns).

that diagnostic CT was available and this imaging showed localised atelectasis with no suspicious features. The atelectasis was also noted to have fully resolved on follow-up CT imaging three months later. Fusion of SPECT uptake onto diagnostic CT confirmed the localisation of the uptake to the left upper lobe atelectasis.

The clinically significant features in case 9 were broadly similar to those of case 7. On planar imaging alone, there was some suggestion of increased general uptake in the posterior right thoracic region but nothing specific enough to merit a SIOPEN or Curie score, as shown in Figure 3.12. On fused SPECT/MR appearances were mixed; the transverse view showed diffuse right rib uptake but the coronal views showed more of a suggestion of uptake within the lung body; and while the sagittal view showed more focal uptake within the rib, it also showed appearances of diffuse uptake extending into the lung body. Given the thick coronal slices, on balance the observer judged the uptake as positive for skeletal MIBG. However, as with case 7, it was noted that further imaging with CT would have been requested in order to conclusively rule as to the nature of the uptake. Following study completion review of diagnostic CT

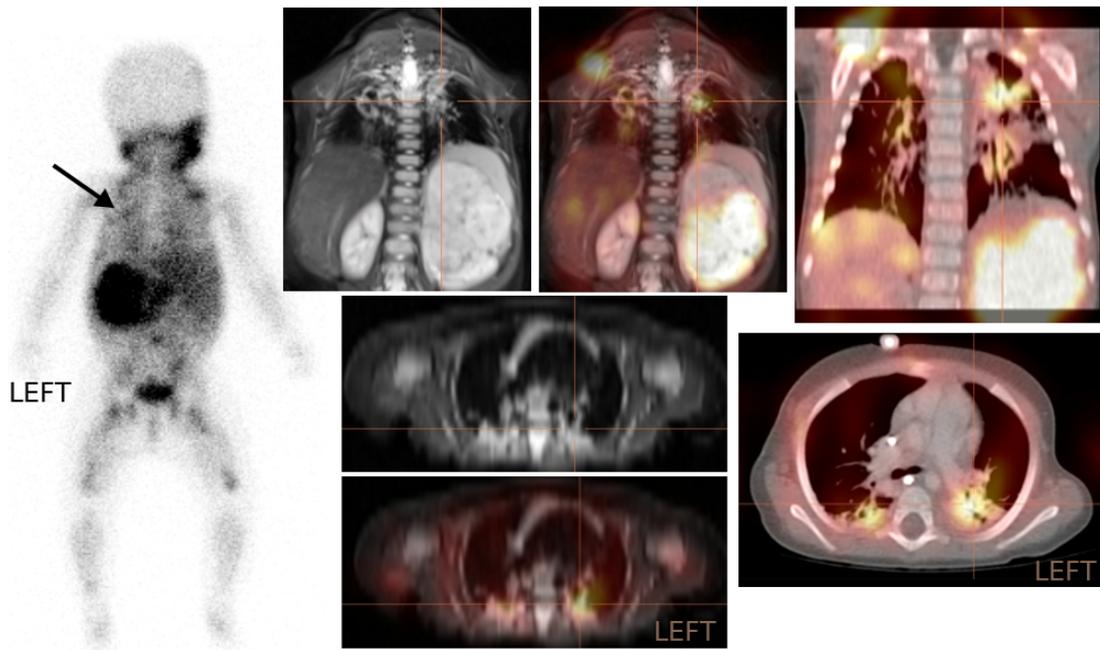


Figure 3.11: Clinically significant finding in case 7. Left - Posterior wholebody scintigraphy showing activity initially scored as rib uptake. Centre - Coronal and transverse plane fused SPECT/MR showing uptake corresponding to an area of soft tissue within the left upper lobe; fused SPECT/MR was deemed insufficient to fully characterise the uptake. Right - Coronal and transverse plane fused SPECT/CT showing localisation of uptake to region of atelectasis which was shown to fully resolve on follow-up imaging three months later.

imaging confirmed the presence of atelectasis in the vicinity of this uptake and fusion with SPECT imaging confirmed the localisation of the uptake to this site.

The final case with clinically significant features was case 10, which featured the largest positive increase in SIOPEN score in the study and the joint largest increase in Curie score. As shown in Figure 3.13, this case featured massive ascites which contributed to the planar imaging being significantly reduced in quality. Furthermore, the ascites appeared to show generalised low grade MIBG uptake which was only correctly characterised after review of fused SPECT/MR. A region of more intense ascital uptake was mistaken for diffuse skeletal uptake within the pelvis on planar imaging but this was clearly cystic on fused SPECT/MR. The intensity of this uptake was sufficient that the observer was concerned for the possibility of peritoneal seeding of tumour rather than mere drainage and this activity was therefore scored as soft tissue uptake during Curie scoring. As a further result of the reduced image quality, the observer found rib and femoral uptake on review of SPECT/MR that they felt were not clearly apparent on planar imaging. These additional findings are shown in Figure 3.14.

Subject	SIOPEN Scores					
	Expert Observer			Local Observer		
	Planar	+Fusion	Diff	Planar	+Fusion	Diff
4	0	0	0	0	0	0
10	6	12	+6	1	5	+4
11	39	38	-1	40	40	0
15	54	48	-6	44	44	0
16	0	0	0	0	0	0
Mean	19.8	19.6	-0.2	17.0	17.8	0.8

Subject	Curie Scores					
	Expert Observer			Local Observer		
	Planar	+Fusion	Diff	Planar	+Fusion	Diff
4	0	1	+1	0	1	+1
10	5	8	+3	2	7	+5
11	23	21	-2	23	23	0
15	26	23	-3	21	23	+2
16	1	1	0	1	1	0
Mean	11.0	10.8	-0.2	9.4	11.0	1.6

Comparison		PCC	
Expert Observer	Local Observer	SIOPEN	Curie
Planar	Planar	0.985	0.985
Planar	+SPECT/MR	0.989	0.993
+SPECT/MR	Planar	0.974	0.968
+SPECT/MR	+SPECT/MR	0.987	0.996

Table 3.4: Inter-observer comparison of 5 cases scored in common between experiments 1 and 2. Upper table: Comparison of SIOPEN scores and differences for expert and local observers. Middle table: Comparison of Curie scores and differences for expert and local observers. Lower table: Pearson Correlation Coefficient (PCC) for each combination.

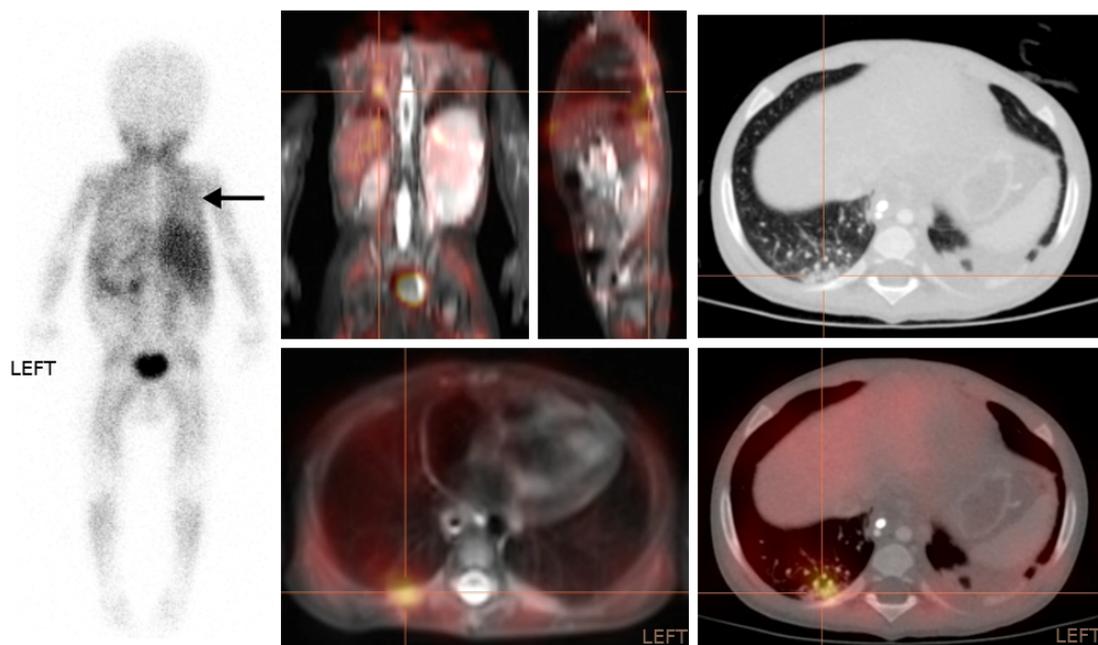


Figure 3.12: Clinically significant finding in case 9. Left - Posterior wholebody scintigraphy showing no obvious specific uptake within the posterior right thoracic region. Centre - Coronal, sagittal and transverse plane fused Single Photon Emission Computed Tomography/Magnetic Resonance (SPECT/MR) showing diffuse uptake localised to rib with lower intensity spread into the lung body. Right upper - Transverse slice of diagnostic CT with lung windows showing localised consolidation. Right Lower - Transverse slice of fused SPECT/CT showing localisation of uptake to region of consolidation.

Inter-observer comparison of five cases scored in common between Experiments 1 and 2, shown in Table 3.4, suggested on average the expert observer assigned slightly higher scores to planar imaging than the local observer (SIOPEN: 19.8 vs 17.0, Curie: 11.0 vs 9.4) and that better agreement between observers was achieved with the addition of fused SPECT/MR for the local observer. While larger score differences remained for the SIOPEN scoring of cases 10 and 15 with the addition of fused SPECT/MR (7 and 4 score units respectively), Curie scoring showed only very small differences in total scores when both observers incorporated fused SPECT/MR findings. PCC analysis also agreed with these numerical observations. The PCC was 0.985 regardless of score system with both observers using planar imaging alone, while improvement was noted for both scores with the addition of fused SPECT/MR for the local observer (SIOPEN: 0.989, Curie: 0.993) and for both observers (SIOPEN: 0.987, Curie: 0.996). The worst PCC for both scoring systems was found for the local observer scoring planar imaging alone and the expert observer scoring fused images (SIOPEN: 0.974, Curie: 0.968).

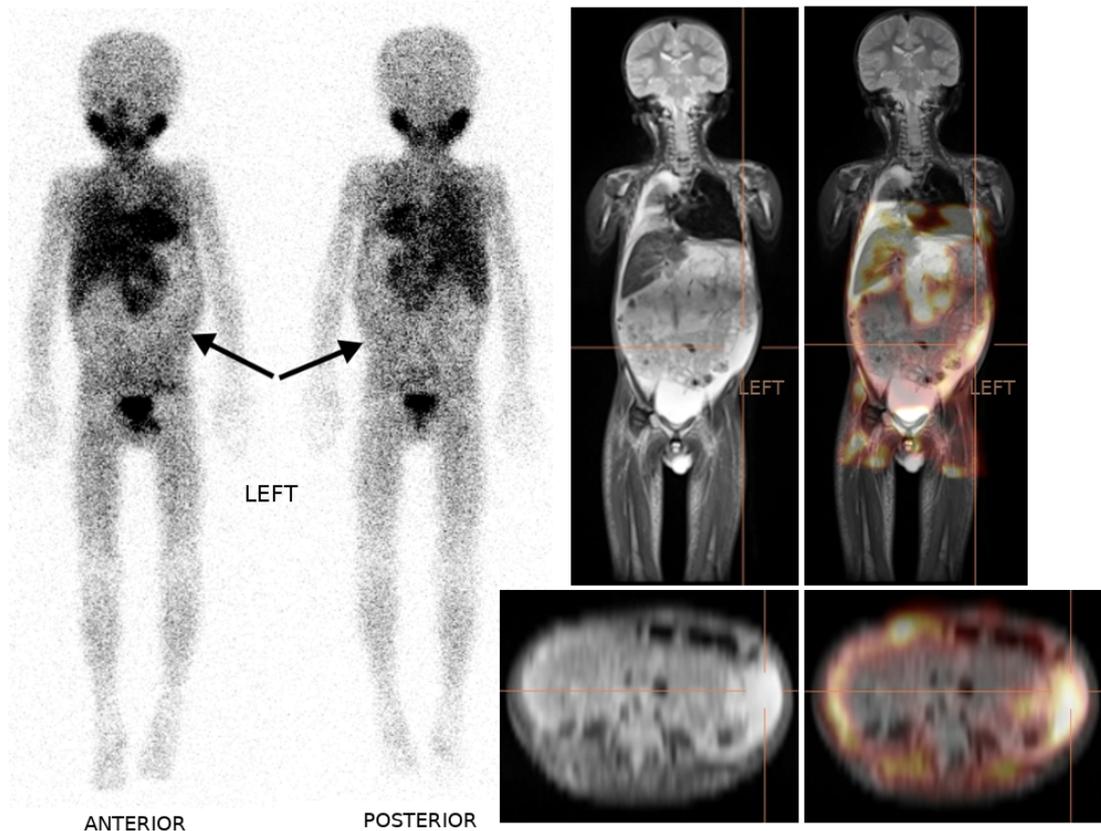


Figure 3.13: Clinically significant findings in case 10. Left - Anterior and Posterior whole-body scintigraphy showing "grainy" low count appearances And low grade atypical activity visible in inferior medial and left lateral abdomen, initially scored as skeletal uptake within the pelvis. Centre - T2-weighted STIR coronal and transverse MRI images showing cystic (white) enhancement within abdominal space characteristic of marked ascites. Right - Fused SPECT/MR images showing localisation of low grade MIBG uptake to ascites in left lateral aspect of abdomen.

3.3.4 Discussion

The improved design of this second observer study was able to address many of the flaws noted in Section 3.2.4. This was evidenced by the higher average total scores, more normally distributed score difference data, elimination of the adrenal gland interpretation confounder, and a threefold increase in cases eligible for consideration in the assessment of prognostically significant change (Hypothesis 2). However despite these improvements in design, this study was unable to show any statistically significant differences in either SIOPEN or Curie scores with the addition of fused SPECT/MR (Hypothesis 1) or show any cases with a prognostically significant change in either Neuroblastoma score (Hypothesis 2).

There are three plausible sources for the disparity of outcome between these two

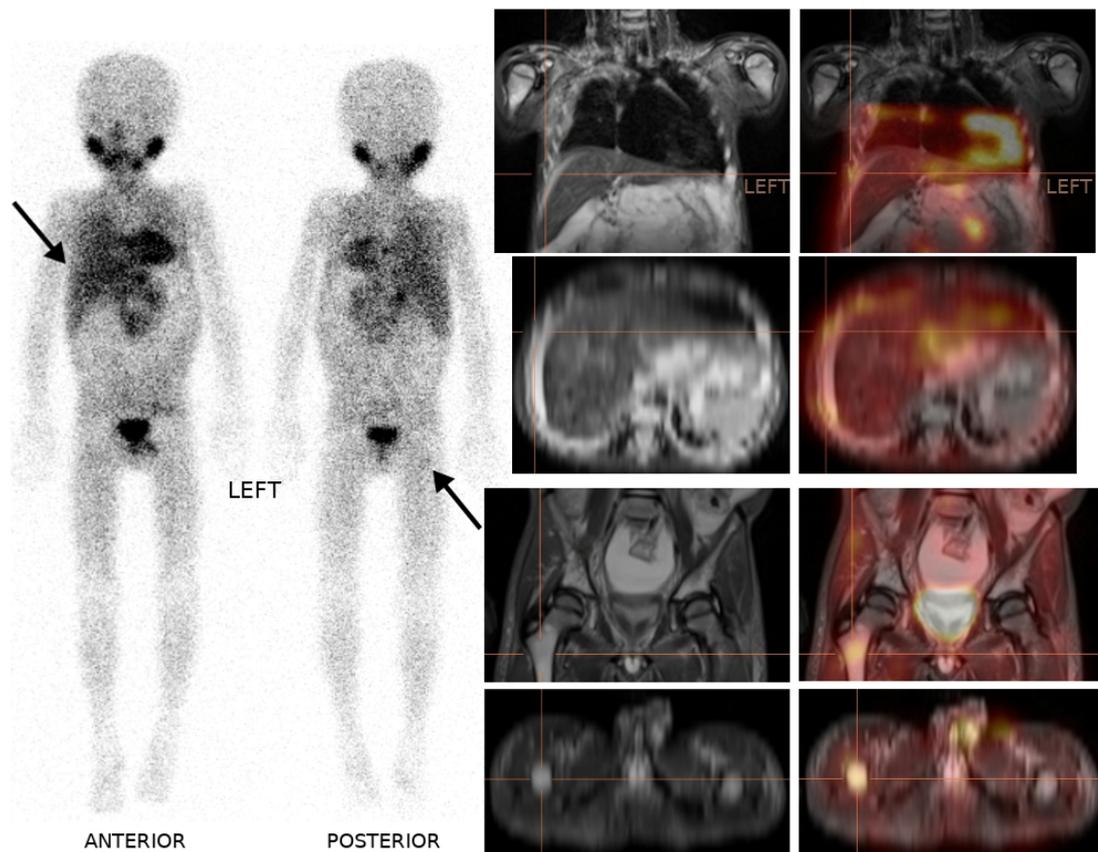


Figure 3.14: Clinically significant findings in case 10 (continued). Left - Anterior and Posterior wholebody scintigraphy showing low grade, noisy uptake in right femur initially thought to be insignificant and unclear appearances in thoracic cage. Centre - T2-weighted STIR coronal and transverse MRI images localised to right lateral ribs (upper pair) and right femoral head (lower pair). Right - Fused SPECT/MR images showing increased uptake localised to right thoracic cage (upper pair) and right femoral head (lower pair).

experiments; differences in the imaging parameters, differences in observer scoring, or genuine differences in underlying score populations. The only imaging parameter differences were a shift to a slower wholebody sweep speed (10/15 cases) and a shift to multiple SPECT fields of view (7/15 cases). While the grouped data shown in Table 3.3 appears to show some differences in mean and standard deviation across both factors, the accompanying post-hoc two-factor ANOVA was unable to demonstrate any statistically significant differences between these groups. Furthermore the group which provides the most similar summary statistics to experiment 1 is not consistent between the two score systems; SIOPEN score differences in experiment 1 are most similar to the (5cm/min, 1FOV) group while Curie score differences in experiment 1 are most similar to the (10cm/min, 1FOV) or (5cm/min, 2FOV) groups depending on whether the group mean or standard deviation is given priority. Even if the failure of the ANOVA to find significance was due to the small sample size, the fact that the closest group to the prior experiment for both scores was not the group with matched

imaging parameters (10cm/min, 1FOV) would seem to suggest that a difference in imaging parameters can be reasonably excluded (or at least down-weighted) as a source of explanation for the inability of this second study to reach significance.

Examination of the data shown in Table 3.4 would seem to suggest that there were some genuine differences in observer scoring between the two experiments. While the correlation of scores between the experiments was shown to improve with the provision of fused SPECT/MR to both observers, similar improvement was noted when only the local observer had access to fused images. Conversely, when only the expert observer had access to fused images the correlation of the datasets worsened, suggesting a systematically different approach to scoring of planar and SPECT/MR data. The pattern of score totals in the inter-observer tables, although a very small sample size, gives a general impression that the expert observer was more likely to assign larger scores to planar imaging alone when compared to the local observer. This may have been the result of an additional "hedging" process undertaken by the expert observer when scoring planar imaging.

After the initial region-by-region scoring pass was complete, the expert observer considered any interpretation difficulties and on multiple occasions opted to "hedge" several SIOPEN scores to account for interpretative uncertainty. For example, in one case with extensive disease including a single lesion in each shoulder joint, both lesions were considered borderline between a score of 1 (single focal lesion) and a score of 4 (diffuse lesion of less than 50% of the skeletal region) due to very slight elongations/asymmetries of the lesions. This meant that, despite being very small lesions, these skeletal sections could range from a relatively small combined score of 2 through to a high combined score of 8 score units. As there was some doubt about how to interpret both lesions, primarily due to the poor quality of the planar imaging, the expert observer opted to assign a score of 5 for both regions combined, effectively averaging between the possible interpretations of the images.

On further discussion regarding this technique and its justification, the expert observer described this step as a necessary correction for the "main weakness" of the SIOPEN score. In contrast to the Curie method, the SIOPEN method can produce large score differences for relatively small changes in the interpretation of small skeletal lesions.

Due to the poor spatial resolution of nuclear medicine imaging, it can often be difficult to robustly assess if a lesion is truly focal (a score of 1) or shows a small amount of elongation, indicative of a small area of diffuse disease (a minimum score of 4). Since the Curie method does not differentiate between diffuse and focal lesions until disease spreads over 50% of a score region, the Curie technique shows greater stability and requires less "hedging".

Importantly, this weakness is not detailed within any publications describing or validating these methods (Ady et al. 1995; Lewington et al. 2009; Ladenstein et al. 2011; Decarolis et al. 2013; Yanik et al. 2013). Furthermore, anecdotally the expert observer appeared to rely on "hedging" less often with the addition of fused SPECT/MR images. This was likely due to the resolution recovery algorithm incorporated into the SPECT reconstruction; which results in improved spatial resolution. This reduction of "hedging" by the expert observer would be consistent with Pearson Correlation Coefficients shown in Table 3.4 that indicated more similar scoring approaches from both observers when fused SPECT/MR was available. Despite some advantages to the Curie score, the need for any users (particularly experts) to resort to "hedging" to overcome the shortcomings of the available imaging should be considered a major weakness of both semi-quantitative scoring methods.

A further source of observer difference would be the differing institutional approaches; at RHC fused images are used routinely for diagnostic assessment of Neuroblastoma cases, while in their own clinical practice the expert observer makes use of fused imaging only sparingly. This means that the local observer almost certainly had considerably less experience at interpreting planar MIBG scans without the reference of SPECT/MR images. While some cases were common to both experiments, the overlapping dataset is a very small sample and is therefore subject to significant potential bias. This renders robust inter-observer comparison impossible and, in combination with the differences in planar scoring approaches followed by the observers, clearly underlines the limitations of the single observer approach used in these experiments. Multiple observers from different centres (ideally ≥ 3) are recommended for any further investigations of Neuroblastoma scores.

As these experiments lack a gold standard reference for neuroblastoma scores it is

not possible to truly differentiate whether these differences arise from underscoring by the local observer or systematic over-scoring by the expert observer. However, given the expert observer's extensive experience in assessing Neuroblastoma scores for international trials and the consistency of the published score validation in Decarolis et al. (2013), it seems plausible to assume that the expert observer's planar scores should be considered a defacto gold standard in this case (despite the use of score "hedging"). This would suggest that the most likely interpretation of the results is that the local observer generally underscored planar imaging, which artificially increased the score differences due to the addition of fused SPECT/MR in experiment 1 and explains the lack of significance on repeat analysis. Nonetheless, it remains encouraging that the local and expert observers showed improved correlation of total scores when fused SPECT/MR images were available. This suggests that such imaging can potentially act to equalise the insights of observers with different levels of experience. In light of these findings, and with the absence of more objective methods of disease burden assessment, it would seem reasonable to recommend the use of fused or hybrid imaging to complement planar imaging when assessing neuroblastoma scores. Although this study did not reach statistical significance, scores were changed with the addition of fused SPECT/MR, and these scores were more consistent between observers than planar imaging alone.

While there appears to be sufficient evidence to suggest the change in observer played a key role in the negative result of experiment 2, there is also the possibility that the pre-treatment patient population does in fact show a smaller average difference in neuroblastoma scores than the mixed population assessed in experiment 1. As highlighted in Figure 3.13, SPECT imaging seems to be more sensitive than planar imaging when uptake is low intensity. After the onset of treatment any responsive disease will be reduced in uptake, potentially to low intensity, and any new emergent site of disease will also be low intensity. This will mean that there will generally be a greater abundance of low intensity disease sites in patients following the onset of treatment, and fused imaging may produce larger score differences at these time points. Unfortunately this study cannot provide further evidence to support this line of argument and further work would be required to better separate the effect of observer and clinical time-point on score differences. Based on the reduced size of score differences found by experiment 2, a much larger dataset of the order 250

patients would seem to be required to guarantee a future study reached significance. As the current rate of new Neuroblastoma diagnoses in the entire UK is only around 10 cases per year, any follow-up work at RHC would require multi-centre participation with multiple years of prospective recruitment.

Interestingly, a larger study examining the impact of SPECT/CT on Neuroblastoma scores was published during the undertaking of experiment 2 (Cerny et al. 2016). In this paper the authors concluded that MIBG SPECT/CT was able to identify more lesions than planar imaging alone in high risk Neuroblastoma, and that SIOPEX and Curie scores were significantly different when assessed with SPECT/CT compared to planar imaging alone. The three main differences between this published study and experiment 2 were the larger dataset (45 patients, 35 of which were high risk); the use of multiple scans per patient (213 scans total); and the use of hybrid SPECT/CT instead of fused SPECT/MR. While the larger dataset alone would act to increase the power of the statistical analysis, the use of an average of 4.7 scans per patient would place the preponderance of data at later clinical time-points than that of experiment 2. The conclusions of this study, while not directly comparable to experiment 2 due to the difference in image modality, provide some additional weight to the prediction that given a large enough dataset at least Hypothesis 1 would be proven true for fused SPECT/MR. However it should be noted that the dataset reported was formed of multiple scans per patient and is therefore not a true independent dataset. It seems likely that this study is subject to some statistical biases that would be best controlled for in an even larger multi-centre dataset.

Finally, the use of SPECT/CT as presented in Cerny et al. (2016) likely offers some advantages over the fused SPECT/MR used in experiments 1 and 2. Particularly of note are the superior lung characterisation of CT and the superior spatial co-registration achievable with a truly hybrid scan. In the cases of dubious lung uptake, highlighted in Figures 3.11 and 3.12, it is clear that hybrid SPECT/CT would provide superior imaging to fused SPECT/MR. Furthermore, throughout both experiments 1 and 2 there were only a small number of cases where MRI tissue characterisation of soft tissue was more important to the scoring process than robust visualisation of the extent and localisation of skeletal disease. This is yet another area where it seems likely that SPECT/CT will provide a more robust complement to planar imaging than

fused SPECT/MR. SPECT/CT also offers significant benefits for SPECT image quality through the use of attenuation correction, patient-based scatter correction and the potential for absolute quantification of SPECT images.

3.3.5 Conclusion

While this study markedly improved upon the design of experiment 1, it failed to demonstrate any evidence for either Hypotheses 1 or 2. Although there were imaging protocol differences between both experiments, these changes did not appear to be sufficiently impactful to explain the difference in outcomes. While there may be genuine patient population differences responsible for the observed smaller average differences in Neuroblastoma scores, the small overlapping dataset between both experiments would seem to suggest observer experience and technique were the most likely sources of discrepancy between the studies. Despite this inconsistency the observers showed the strongest correlation of scores when fused SPECT/MR was available, indicating that increased use of fused imaging would improve score and reporting consistency. Even in the absence of hybrid SPECT and MRI scans, retrospectively fused SPECT/MR changed Neuroblastoma scores and reduced the need of the expert observer to compensate the scoring process for uncertain interpretation of unclear and low grade lesions. Where available, use of hybrid SPECT/CT is recommended due to its superior lung and skeletal characterisation in addition to superior spatial co-registration.

3.4 Reflection

The experiments described in this chapter demonstrate the clinical impact of the improved uptake characterisation available with fused SPECT/MR compared to that of planar scintigraphy for Neuroblastoma. However, this chapter also demonstrated the drawbacks of fusing non-hybrid scans and the limitations of semi-quantitative Neuroblastoma scores more generally. This work arose naturally from the integrated same-day, single-GA scanning protocol developed at RHSC; which was influential both in the development of new Neuroblastoma scanning guidelines and in the design of the new RHC imaging and theatre facilities. This work also clearly demonstrated the need for true, quantitative, hybrid SPECT/CT in Neuroblastoma and presents

a bridge between the earlier SPECT/MR focused work of Chapter 2 and the later quantification focused work of Chapter 4.

Chapter 4

Wholebody Quantitative SPECT-CT for Neuroblastoma

4.1 Introduction

In contrast to haematological cancers, solid tumours respond stubbornly to systemic treatment (O et al. 2016). If disseminated throughout the body (as is typically the case with high risk Neuroblastoma) cases often remain incurable by surgical intervention even after a multiplicity of toxic therapies. In order to improve this rather bleak clinical picture, it is essential to expedite the development of novel treatments and to direct changes to clinical care based on individual patient responses. The gold standard measures of treatment response in oncology are outcome data, such as EFS or Overall Survival (OS), which are most commonly used in Phase III clinical trials. These measures are highly time consuming to acquire and expensive to collate at scale; limiting their utility for the timely and cost effective development of novel therapeutics. Furthermore, while Phase III trials are often used to optimise treatment pathways, they are necessarily narrow in focus so cannot assist the decision making of clinicians treating patients who fall outside of the trial inclusion criteria. In order to meet the demands of treatment development and clinical practice; reliable, outcome-linked, noninvasive, *in vivo* response metrics are required.

The World Health Organisation (WHO) developed the Response Evaluation Criteria in Solid Tumours (RECIST) (Therasse et al. 2000), as a standardised framework for assessing disease burden and treatment response precisely in order to address this need. RECIST uses the metric of lesion size, which has the advantage of being disease agnostic, and recommends CT as the modality of choice for performing measurements.

This recommendation is based on the superior reproducibility of CT compared to MRI, operator independence compared to ultrasound and 3D size measurement compared to plain film chest x-ray or clinical (caliper) measurement.

The revised RECIST methodology (RECIST 1.1) requires a baseline assessment of all disease sites, required to be performed within 4 weeks of treatment commencement and recommended to be performed via contrast enhanced CT survey (Eisenhauer et al. 2009). Under RECIST 1.1 all non-nodal lesions with maximal diameter of at least 10mm are considered "measurable", as are all pathologically enlarged lymph nodes with short axis diameter of at least 15mm (Eisenhauer et al. 2009). Up to five measurable lesions and nodes are then selected as "target lesions", following the work of Bogaerts et al. (2009), who identified five lesions as optimal for disease burden assessment. The target lesion selection process is based on a combination of lesion size (largest is preferred), location (a maximum of two per organ) and expected reproducibility for repeated measurement (lesions unsuitable for repeated quantification are discounted) (Eisenhauer et al. 2009). The diameters of the selected lesions (long axis for non-node lesions, short axis for nodes) are then summed to generate a figure representative of disease burden. Reassessment following one or more treatments permits response assessment to be objectively defined based on the change to the summed diameters of target lesions, as described below:

- Complete Response: Disappearance of all target lesions, all pathological lymph nodes must have reduction in short axis to less than 10mm.
- Partial Response: At least a 30% reduction in summed target lesion diameters as compared to baseline.
- Progressive Disease: The emergence of new lesions, or at least a 20% relative increase *and* at least a 5mm absolute increase in summed target lesion diameter, both as compared to the smallest summed target lesion diameters recorded throughout treatment.
- Stable Disease: Neither sufficient shrinkage for Partial Response, nor sufficient increase to qualify as Progressive Disease.

As the initial RECIST framework was being developed, hybrid PET/CT was postulated (Kinahan et al. 1998) and then demonstrated in a clinical environment (Beyer et al. 2000). The combination of modalities permitted attenuation and scatter correction of PET imaging as well as localisation of PET uptake to anatomical structures on CT. These features combined with the use of ^{18}F -labelled fluorodeoxyglucose (FDG); allowed PET/CT to deliver high sensitivity, high resolution, truly quantitative maps of glycolytic metabolism (Kapoor et al. 2004). The ability to directly examine the metabolic rate of cancer cells, a disease agnostic measure linked to growth rate and aggressiveness, was nothing short of a paradigm shift for oncology (Hofman and Hicks 2016).

Such was the clinical utility of ^{18}F -FDG PET/CT, that it quickly became the primary tool for detection and evaluation of a broad range of cancers and a PET based response assessment framework, known as Positron Emission tomography Response Criteria in Solid Tumours (PERCIST) was rapidly developed and published nearly in parallel with the first major revision of RECIST (Wahl et al. 2009). Practically, PERCIST is straightforward; requiring only the measurement of physiological background FDG activity and the measurement of the "peak" uptake in the single most active tumour site within the patient. Though, in keeping with RECIST, the measurement of FDG activity in up to the five "hottest" tumour sites at baseline with a maximum of 2 lesions per organ has also been suggested (Wahl et al. 2009; O et al. 2016; O and Wahl 2018).

Under PERCIST, physiological background uptake is defined by both the mean and standard deviation of FDG activity within a 3cm diameter spherical volume of interest (VOI) positioned in the medial right lobe of liver, or in cases with extensive liver metastases, by the mean and standard deviation of FDG uptake measured within a 1cm diameter and 2cm long cylindrical VOI placed in the centre of the blood-pool within the descending thoracic aorta (Wahl et al. 2009; O et al. 2016; O and Wahl 2018). "Peak" tumour uptake measurements are defined as the 1ml volume (defined as a 1.2cm diameter spherical VOI) with the highest mean uptake i.e. the mean uptake of the most metabolically active 1ml of tumour (Wahl et al. 2009; O et al. 2016; O and Wahl 2018).

PERCIST requires both background and tumour FDG uptake to be measured in units of SUV scaled to lean body mass (LBM), interchangeably expressed as SUV_{lbm} , SUV/LBM or Standardised Uptake Value scaled to lean body mass (SUL). SUV units are used extensively in quantitative PET studies as they are designed to be representative of the radiopharmaceutical uptake normalised to both administered activity and patient size. SUVs are calculated for the voxels of a PET image as shown in equation 4.1 below:

$$SUV/W^* = c_i \times \frac{W^*}{A_i} \quad (4.1)$$

where c_i is the voxel-wise activity concentration provided by the PET scanner (usually measured in kBq/ml); A_i is the administered activity decay corrected to the time of imaging; and W^* represents some measure of patient size such as body weight (BW), body surface area (BSA) or LBM (Kinahan and Fletcher 2010). PERCIST prefers the use of LBM in SUV scaling as FDG does not accumulate within white fat and therefore hepatic SUL shows superior inter-patient reproducibility compared to SUV/BW or SUV/BSA (Sugawara et al. 1999).

In order to ensure reproducible PET scans, PERCIST specifies a number of "assessibility criteria" which includes a 50-70 minute uptake time between administration and scan; use of the same PET/CT scanners and reconstruction methods for all scans; careful control of patient fasting regimen with serum glucose measurements pre-scan; and at most a 20% difference in injected activity. In addition to these considerations, PERCIST defines image-based reproducibility criteria based on the background VOI. A series of scans cannot be assessed with PERCIST if the relative difference in hepatic SUL is greater than 20% (relative to the largest liver measurement) or if the absolute difference is greater than 0.3 SUL units (Wahl et al. 2009; O et al. 2016).

As with the RECIST criteria, PERCIST establishes a minimum threshold for measurability of target lesion at baseline (Wahl et al. 2009; O et al. 2016; O and Wahl 2018). However under PERCIST this is defined as a minimum SUL relative to the patient specific background measurement; with slightly different definitions used in cases with liver and aortic backgrounds respectively, as shown in equations 4.2 and

4.3 below:

$$SUL_{min,liver} = 1.5 \times \mu_{liver} + 2.0 \times \sigma_{liver} \quad (4.2)$$

$$SUL_{min,aorta} = 2.0 \times \mu_{aorta} + 2.0 \times \sigma_{aorta} \quad (4.3)$$

where μ_{liver} , σ_{liver} and μ_{aorta} , σ_{aorta} are the mean and standard deviation pair for hepatic and aortic background VOIs respectively. The use of such a threshold is intended to ensure that meaningful changes to tumour SUL remain measurable within the useful dynamic range of the PET/CT scanner (i.e. above background) (O et al. 2016).

PERCIST recommends recording the percentage change in tumour peak SUL relative to baseline for the target lesion; this data combined with number of weeks since treatment permits waterfall plots and further disease specific analysis. Additionally, PERCIST provides broad response categories designed to mirror the response categorisation of RECIST, as described below (Wahl et al. 2009; O et al. 2016; O and Wahl 2018):

- Complete Metabolic Response: No new disease sites detected, reduction of FDG uptake to less than that of the mean liver SUL and all disease sites indistinguishable from the surrounding background tissues.
- Partial Metabolic Response: At least a 30% reduction in the most intense target lesion *and* at least a 0.8 SUL unit reduction in uptake, both as compared to baseline. Furthermore no new FDG avid lesions should be visible and no non-target lesion should show an increase in FDG uptake greater than 30% compared to baseline.
- Progressive Metabolic Disease: The emergence of new FDG avid lesions, or the unequivocal progression of non-target lesions or at least a 30% relative increase *and* at least a 0.8 SUL unit increase in any target lesion.
- Stable Metabolic Disease: Neither sufficient reduction for a Partial Metabolic Response, nor sufficient increase to qualify as Progressive Metabolic Disease.

PERCIST also defines a set of exploratory measurements which are recommended to be recorded in PERCIST compliant studies so as to progress the scientific literature. These include the tumour maximal SUV, PET derived metabolic volume (a number of potential SUV thresholds are suggested) and total lesion glycolysis (mean tumour uptake \times total tumour volume) (Wahl et al. 2009; O et al. 2016; O and Wahl 2018). The core and/or exploratory metrics of PERCIST have been shown to be associated with, or predictive of, outcomes in numerous cancers including; non-Hodgkin's lymphoma (Manohar et al. 2012); oesophageal cancer (Yanagawa et al. 2012); colorectal cancer (Skougaard et al. 2013); small-cell lung cancer (Ziai et al. 2013); breast cancer (Riedl et al. 2017); and melanoma (Lee et al. 2022).

One criticism of the PERCIST methodology is that the definition of peak tumour uptake is defined across too large a volume to be useful in cancers which exhibit markedly heterogeneous disease sites with small active foci, such as in Hodgkin's Lymphoma (Hasenclever et al. 2014; Barrington and Kluge 2017). To address this issue, an alternative framework known as "qPET" was proposed by Hasenclever et al. (2014) as a more objective replacement for the Deauville scale; a qualitative scale based on visual comparison of FDG uptake in tumour compared to that of normal liver and bloodpool background (Gallamini et al. 2009; Meignan et al. 2009).

In contrast to PERCIST, qPET uses alternative definitions for peak tumour uptake and hepatic background; peak tumour uptake is defined as the mean of the maximal tumour voxel and the three hottest adjacent voxels (extracted via an automated analysis tool created in collaboration with Hermes Medical Solutions); while hepatic background is defined as the mean uptake within a 30ml cuboidal volume with length:width:height ratio of 2:2:1 placed within the medial right lobe of liver. After extraction of tumour and background measurements, the qPET value is then calculated as the ratio of tumour peak uptake to liver mean uptake. Similar to the use of the single maximal voxel (SUV_{max}), "peak" uptake defined in this manner is considered less subject to partial volume effects but more influenced by image noise, voxel size and reconstruction algorithm (Wahl et al. 2009). The alternative background VOI definition also offers the advantage of straightforward VOI placement even in young children, while maintaining excellent reproducibility; quoted as a coefficient of variation of 3% between upper right lobe, lower right lobe and left lobe across a

subset of 150 patients (Hasenclever et al. 2014).

A semi-quantitative approach somewhat reminiscent of qPET was used in SPECT imaging of neuroblastoma as part of the pioneering work of Fendler et al. (2013) who investigated the relationship between MIBG uptake and Neuroblastoma histopathology. In this study, a number of histopathological and oncogenetic risk factors identified by the INPC; Schwannian stromal development, neuroblastic differentiation, proliferation rate, n-Myc amplification and p1 deletion (Shimada et al. 1999); were recorded for resected tumours and compared to Tumour-to-Liver Count-Rate Ratio (TLCRR) assessed on pre-surgical MIBG SPECT scans. Within this study, TLCRR was calculated as the ratio of the maximal voxel value within the tumour site and the maximal voxel value within a 100ml volume cubic VOI positioned in the right lobe of liver. Using this measure of tumour uptake, it was found that stroma-poor tumours (i.e. "true" Neuroblastomas) had a significantly higher TLCRR than either stroma-dominant (Ganglioneuromas) or stroma-rich tumours (Ganglioneuroblastomas). TLCRR was also found to be significantly but weakly correlated with neuroblastic differentiation but was not found to be significantly correlated with the presence of either oncogenic marker. Perhaps the most clinically significant finding of the study was a statistically significant link between higher TLCRR on follow-up scans and higher MKI, a measure of tumour proliferation rate known to be strongly linked to patient outcomes (Joshi et al. 1992; Taggart et al. 2011; George et al. 2012).

Despite these potentially exciting results, there are several key criticisms that can be brought against the methodology of Fendler et al. (2013). Firstly, the measurement of TLCRRs were performed using SPECT data acquired on a SPECT-only scanner and therefore without the benefits of AC which are available through the use of SPECT/CT. This is of course an understandable flaw given the year of publication and the lack of access to SPECT/CT in paediatric centres at this time. However the lack of attenuation correction necessarily introduces multiple confounders to the accurate measurement of TLCRRs, including dependence on patient size and VOI/tumour depth. The authors also fail to describe any corrections applied during SPECT reconstruction, such as resolution recovery or scatter correction, or whether any pre- or post-reconstruction filters were used. Without these details it is difficult to assess how large an impact image noise or spatial resolution would have on tumour and hepatic

uptake. Finally, the measurement of hepatic uptake by extracting the maximal voxel value within a large VOI, as opposed to calculating the mean uptake within the organ, is a seemingly unconventional choice which is not explicitly justified or explored within the publication. However, it may be the case that this was the most reasonable choice of metric for the study, given the lack of AC and the commonality of image noise contributions to both tumour and liver maximal voxel values.

With quantitative SPECT reconstruction packages commercially available from Hermes, Siemens and GE, it was only a matter of time until quantitative neuroblastoma imaging was published within the scientific literature. The first such paper, published by Brady and Shulkin (2019), presented an analysis of phantom and patient SUV values for scans reconstructed using the Hermes Medical Solutions commercial package SUV-SPECT. The phantom study consisted of contrast recovery coefficients (CRC) measured using a National Electrical Manufacturers Association (NEMA) International Electrotechnical Commission (IEC) body phantom imaged on a Siemens Symbia Intevo scanner using Low Energy High Resolution (LEHR) Collimators. The patient study consisted of a cross-scanner comparison of SUV values measured in certain physiological tissues (salivary glands, heart, liver, adrenal glands and bladder) and neoplastic deposits in 111 scans across 43 unique patients aged between 9 months and 17 years at the time of scan. The first 42 scans were acquired on a GE Infinia Hawkeye using Medium Energy collimators, the remaining 69 scans were acquired on a Siemens Symbia Intevo using LEHR collimators and 28 patients were scanned on both scanners throughout the study; albeit at different timepoints. While coverage varied from scan to scan, each physiological tissue was represented in 30-102 individual scans across both scanners and no statistically significant differences in SUV values between scanners were found for any examined physiological tissue. Across the 111 scans, 75 sites of neoplastic tissue were also identified and grouped by location (Upper body bony, thorax, spine, abdomen and pelvis) rather than separated by scanner. Neoplastic SUV values showed significant variance; with min and max values ranging across an order of magnitude; and there was no obvious pattern of uptake based on location. Within the paper, neoplastic uptake was also presented longitudinally for a select number of cases, including that shown in Figure 4.1, a case with 10 scans between diagnosis and week 73 of treatment. In this case the primary tumour maximum SUV was plotted alongside tumour length and liver mean SUV.

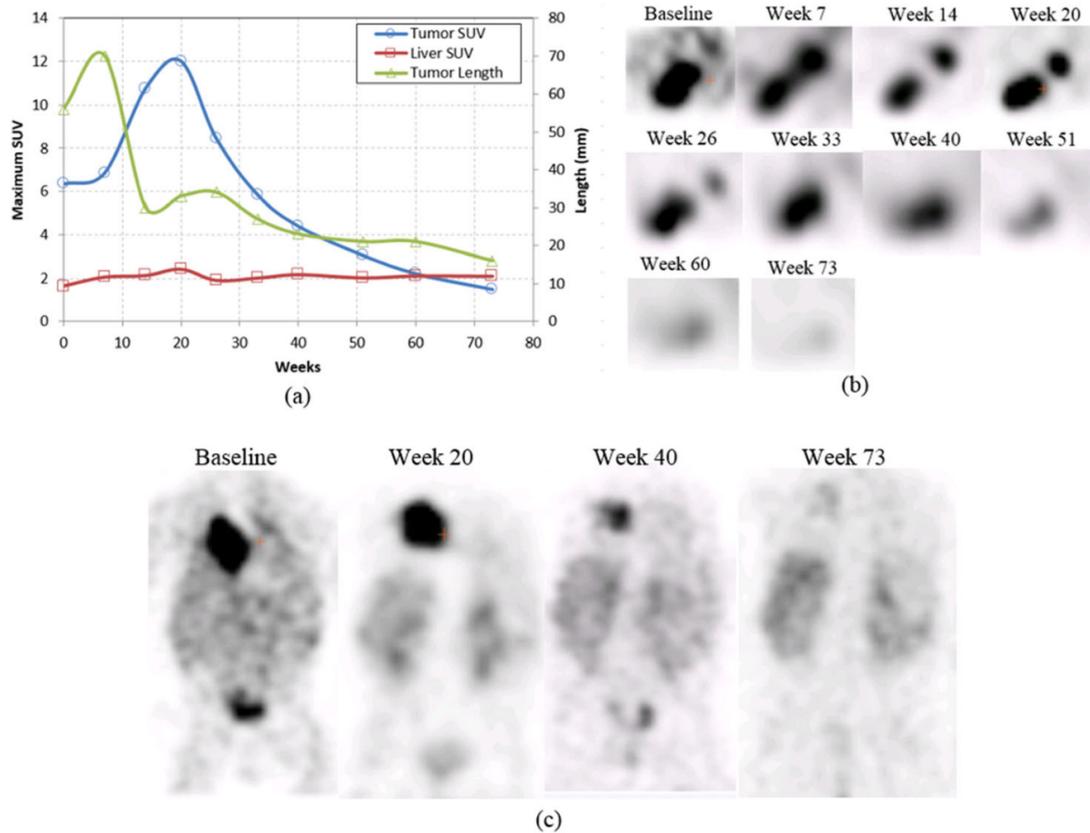


Figure 4.1: Longitudinal tumour and liver uptake in a single case from diagnosis to week 73 of treatment. Replicated in full from Brady and Shulkin (2019) under Creative Commons License (Attribution 4.0 International) as described at <https://creativecommons.org/licenses/by/4.0/legalcode>.

This case showed an initial surge in maximum SUV in response to treatment while tumour length simultaneously reduced. Notably the liver mean SUV remained very consistent across the 10 scans with a mean of 2.1 SUV/BW.

While the work of Brady and Shulkin (2019) provided the first published insight into quantitative SPECT/CT of Neuroblastoma, some of the technical details necessary to replicate or improve upon the work were absent from the publication. Justifications for significant changes to the acquisition protocol, such as change of collimator, are not described and the selection of reconstruction parameters remains unexplored. Ultimately, if changes to quantitative uptake of ^{123}I -MIBG are indicative of response to treatment, then it would be desirable to create a robust quantitative approach comparable to PERCIST or qPET to standardise its reporting. A key component of PERCIST is its strongly assessability criteria; a combination of both technical specifications that must be followed to generate suitable imaging and physiological reference ranges that confirm that image quality standards have been achieved. The latter of

these requires a strong focus on the methodology of measurement, including definition of appropriate measurement units, to ensure a common approach is undertaken. This is in contrast to many areas of Nuclear Medicine where local standard of practice can vary dramatically from centre to centre.

Following from the recommendations of Chapter 3, and concurrent to the publication of Brady and Shulkin (2019), local protocols for wholebody quantitative SPECT/CT of Neuroblastoma were under development at RHC. These were advanced through a set of three studies focussed on acquisition, reconstruction and clinical measurement in quantitative SPECT/CT respectively. These studies aimed to address some of the gaps in the scientific literature and enable the centre to progress imaging towards wholebody quantitative SPECT/CT of Neuroblastoma and contribute to standardisation of these techniques.

4.2 Collimator Selection

4.2.1 Introduction

There has long been debate surrounding the optimal collimator selection for MIBG imaging of neuroblastoma and imaging of Iodine-123 in general. A product shift by gamma camera manufacturers towards higher sensitivity, lead-foil based, low energy collimators in the early 2000s inevitably increased the levels of septal penetration observed in the imaging of emissions slightly higher than Technetium-99m (140keV); such as those generated by Krypton-81m (190keV) and Iodine-123 (159keV). Following this product shift, Nuclear Medicine departments moved away from imaging of Krypton with low energy collimators and publications advocating for the use of medium energy collimators for ^{123}I -MIBG emerged from nuclear cardiology (Inoue et al. 2003) and later from oncology (Snay et al. 2011). However, the use of low energy collimators in Iodine-123 imaging remained widespread; in part due to the ongoing clinical utility of high resolution SPECT (and therefore low energy high resolution collimators) in applications requiring the visualisation of small physiological structures. These include; imaging of the dopamine transporter system of the brain using ^{123}I -loflupane ("DaTscan") (King et al. 2016); and pre-surgical detection of parathyroid adenomas through dual-isotope subtraction of ^{99m}Tc -MIBI (thyroid and parathyroid)

and ^{123}I -Iodide (thyroid only) (Woods et al. 2017).

As access to SPECT/CT became more widespread, and inevitably reached paediatric centres, the discussion of collimator selection re-emerged in the context of ^{123}I -MIBG for Neuroblastoma. The most rigorous publication on this topic is an examination of lesion detectability published by Gregory et al. (2017); where the authors used a channelised Hotelling observer (a form of automated observer) to perform receiver-operator characteristic (ROC) analysis on planar images of an anthropomorphic phantom. They concluded that despite the loss of spatial resolution, medium energy collimators remained the superior collimator for planar imaging of ^{123}I -MIBG across all manufacturers. A further key point of discussion was that the use of medium energy collimators would provide ^{123}I -MIBG imaging more comparable to post-treatment ^{131}I -MIBG therapy imaging; which require heavy high energy collimators with relatively low spatial resolution; and therefore would be more suitable in the context of pre-treatment planning and dosimetry.

Following on from this work, and after relocation to new premises at the RHC, a quantitative evaluation of the available parallel-hole collimators for the Siemens Symbia Intevo SPECT/CT scanner was undertaken. The goals of this undertaking were to (1) quantify the magnitude of Iodine-123 septal penetration occurring for each collimator set, (2) assess the distance dependence of the sensitivity of each collimator, and (3) examine the impact of septal penetration on the acquired energy spectra. In order to achieve the latter of these observations, a reproducible methodology for count-rate scaling of Siemens energy spectra was developed.

4.2.2 Methods and Materials

Phantom Specification and Fabrication

A poly-methyl methacrylate (PMMA) gamma camera sensitivity phantom was specified and fabricated in collaboration with the Department of Clinical Physics and Bioengineering (DCPB) Medical Devices Unit (MDU) Mechanical Workshop team. The phantom design was an optimised revision of an existing phantom fabricated in the 1990s for use in Department of Health Gamma Camera Assessment Team (GCAT) tests performed throughout Scotland. The purpose of both the original and revised

phantoms, are to provide a thin fillable chamber with suitably low self attenuation to closely replicate NEMA sensitivity measurements in air, as specified in (NEMA 2012), with a low risk of contamination of the collimator face.

The revised phantom, schematic shown in Figure 4.2, consists of a 5mm deep 100x100mm rounded-corner square hollow; machined out of a 9mm thick 120x120mm section of precision skimmed PMMA. The hollowed out section was then enclosed with a 4mm thick 120x120mm precision skimmed section of PMMA sheeting and sealed using a combination of glue and countersunk PMMA screws to create a 5mm thick internal air chamber. Small threaded holes were drilled in the centre of one edge and supplied with PMMA screws and rubber seals to permit filling and emptying. Finally, the phantom was supplied with two sets of four precision elevation legs (50mm and 100mm length respectively) that can be affixed via threaded recesses machined into the edge region of the main body. This revised design therefore replicates the main features of the previous phantom but with reduced chamber thickness (5mm compared to 6mm); equal and reduced wall thicknesses (4mm compared to 5 and 6mm); lower tolerance for surface deviation ($\leq 0.1\text{mm}$ after skimming); entirely PMMA design (compared to the original phantom fitted with metal filling screws); and a choice of two precision leg lengths (compared to only 100mm legs).

Phantom Filling and Imaging

After a period of leak testing using non-radioactive water, the sensitivity phantom was filled with 40.7MBq of ^{123}I -Iodide, sufficiently low as to remain within the linear count-rate response range of the gamma camera, and was allowed to reach an equilibrium distribution of activity over a period of 1 hour before scanning. Syringe activity and calibration time were recorded before and after filling of the phantom. Activity was measured to 0.1MBq precision using a Capintec CRC-25R isotope calibrator with a 2mm copper filter insert and instrument calibration traceable to the National Physical Laboratory (NPL).

All images were acquired with a Siemens (Munich, Germany) Symbia Intevo dual headed SPECT/CT gamma camera using either MELP, LEHR or Low Energy All Purpose (LEAP) collimator pairs. The detector heads were positioned in 0° and 180° orientations and the filled phantom was placed on the lower collimator face in the

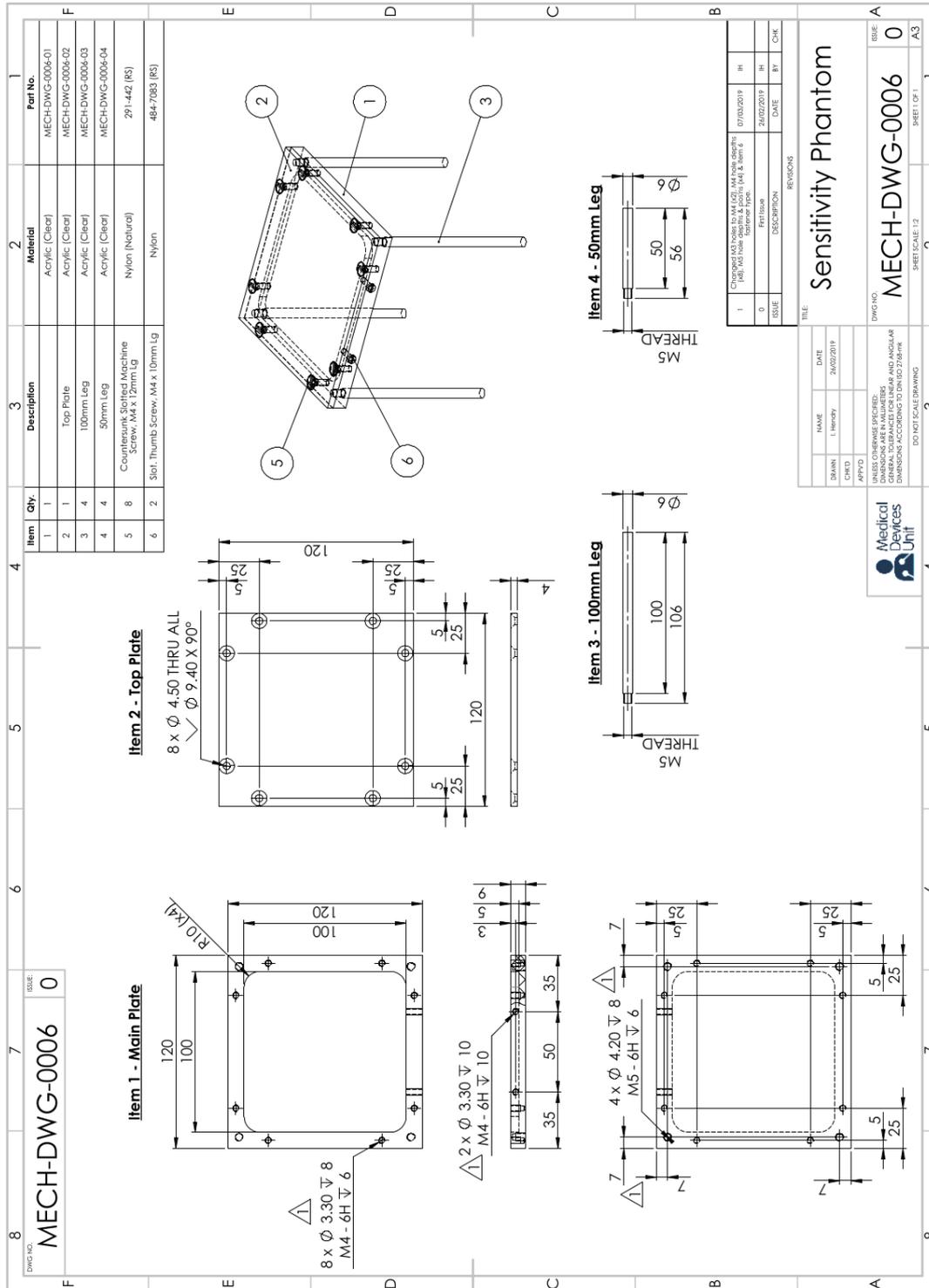


Figure 4.2: Schematic of poly-methyl methacrylate (PMMA) precision sensitivity phantom, designed by the Department of Clinical Physics and Bioengineering (DCPB) Medical Devices Unit (MDU) Mechanical Workshop Team according to author specification. Schematics show full assembled unit and cross sectional views of individual components: Item 1 - 9mm thick main plate with fillable hollow, filling channels, top plate screw holes and leg attachment points; Item 2 - 4mm thick top plate with countersunk screw holes; Item 3 - 100mm leg with M5 threaded mount; Item 4 - 50mm leg with M5 threaded mount.

approximate centre of the detector field of view. Using a combination of the interchangeable precision legs, detector radial positioning and a precision skimmed 50mm thick block of PMMA, the phantom was precisely positioned with 50mm separation between its upper surface and the upper collimator face. A 120 second planar image was acquired using the Symbia Intevo default ^{123}I photopeak energy window (15% width centred on 159keV) and a 1024×1024 matrix size. Immediately following image acquisition, an energy spectrum was acquired and saved using the gamma camera pulse height analyser (PHA).

The phantom was then repositioned at increasing increments of 50mm from the collimator face and image and spectrum acquisitions were repeated at each distance. This procedure was continued until the maximal separation of the collimators was reached; 600mm for LEHR and LEAP collimators and 550mm for the thicker MELP collimator set. Finally, a 120 second background image was acquired with the phantom stored outside of the shielded scan room. The entire experimental process was then repeated for each of the two remaining collimator sets.

Data Analysis

Planar images were reviewed using Hermes Medical Solutions (Stockholm, Sweden) Hybrid Viewer. For each phantom image, the collimator, collimator-phantom distance, acquisition time and total image counts were recorded. For each background image the collimator and total image counts were recorded. These image data, and the pre- and post-fill phantom activity measurements, were loaded into a Python (Rossum 1995) equipped Jupyter Notebook (Kluyver et al. 2016) for further analysis and visualisation. Analysis was completed using various opensource data handling and visualisation libraries including matplotlib (Hunter 2007), numpy (Harris et al. 2020), pandas (McKinney 2010) and scipy (Virtanen et al. 2020).

The apparent system sensitivity was first calculated for each phantom image according to Equation 4.4:

$$S_i = \frac{C_i - C_{bkg}}{\Delta t} \times \frac{1}{A_i} \quad (4.4)$$

where S_i is the image-derived sensitivity measurement, measured in cps/MBq; C_i and C_{bkg} are the total phantom and background image counts respectively; Δt is the

image acquisition duration; and A_i is the decay corrected phantom activity at the time of the phantom image acquisition.

For each image, the collimator-phantom separation was corrected for the detector-collimator offset (47.6mm for MELP, 32.1mm each for LEHR and LEAP) and the sensitivity was plotted against the detector-phantom separation for visual inspection. Following a similar methodology to NEMA septal penetration analysis (NEMA 2012), the relationship between sensitivity and detector-phantom separation was fitted using a Levenberg-Marquardt non-linear least-squares algorithm combined with an exponential fitting function as shown in Equation 4.5:

$$S_i(d) = S_0 + S_1 \times e^{-\ln(2) \times d / D_{1/2}} \quad (4.5)$$

where $S_i(d)$ is the image-derived sensitivity at detector-phantom separation d , S_0 and S_1 are the fixed (photopeak) and distance dependent (septal penetration) components of sensitivity respectively and $D_{1/2}$ is the half-value-distance, i.e. the distance over which the septal penetration component of sensitivity decreases by a factor of 2.

Following curve fitting of sensitivity, and in keeping with NEMA methodology, the penetration fraction (PF) was then modelled using the formula shown in Equation 4.6:

$$PF(d) = 100 \times \left(1 - \frac{S_0}{S_i(d)} \right) \quad (4.6)$$

where $PF(d)$ is the septal penetration (expressed as a percentage) at distance d and the remaining symbols share the same definitions as in Equation 4.5. Using this model, septal penetration was plotted against detector-phantom separation for visual inspection. The septal penetration percentage at 100mm from the detector (PF_{100}) was also calculated in order to provide a representative measure of septal penetration and scatter for each collimator at a clinically relevant distance.

Finally, the energy spectra acquired for each collimator and distance combination were converted into sensitivity measurements, so as to allow direct comparison of spectra between the three collimators. The Symbia Intevo Pulse Height Analyser

(PHA) interface records until a single energy channel (typically within the photopeak) reaches 35,000 counts, and only energy and total count values are provided in exported datafiles. Therefore the total count values for each channel were converted into sensitivity measurements using the formula shown in Equation 4.7:

$$S_e = S_i \times \frac{c_e}{\sum_{e=E_{min}}^{E_{max}} c_e} \quad (4.7)$$

where S_e and c_e are the sensitivity and total counts respectively in the channel corresponding to energy e ; E_{max} and E_{min} are the upper and lower bounds of the energy window used to acquire the sensitivity phantom images and S_i is the image-derived sensitivity calculated from the immediately preceding phantom image.

After conversion into channel sensitivity measurements, the collimator-phantom separation was corrected for the detector-collimator offset and the relationship between channel sensitivity and detector-phantom separation was then fitted using a Levenberg-Marquardt non-linear least-squares algorithm combined with the exponential fitting function shown in Equation 4.8:

$$S_e(d) = S_{0,e} + S_{1,e} \times e^{-\ln(2) \times d / D_{1/2}} \quad (4.8)$$

Where $S_e(d)$ is the sensitivity in the channel at energy e at detector-phantom separation d ; $S_{0,e}$ and $S_{1,e}$ are the fixed (photopeak) and exponential (septal penetration) components of channel sensitivity at energy e ; and $D_{1/2}$ is the common half-value-distance for the collimator of interest.

4.2.3 Results

For each of the three collimators tested, the highest sensitivity was measured at smallest separation of the phantom and collimator face, as shown in Table 4.1. At 50mm separation, the calculated sensitivity for LEAP and LEHR collimators (213.9 and 170.9 cps/MBq respectively) were much greater than for the MELP collimator (111.1 cps/MBq). However as the phantom was moved increasingly distant from the detector, the sensitivity of the MELP collimator was only slightly reduced while

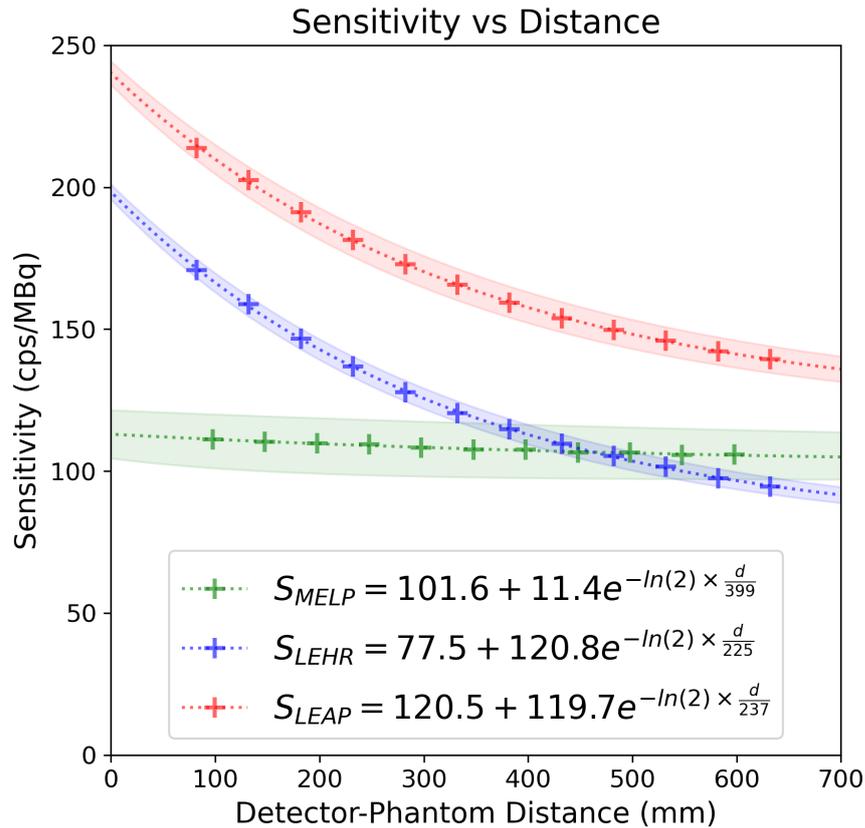


Figure 4.3: System sensitivity versus detector-phantom distance for each of three collimators; Medium Energy Low Penetration (MELP) (green), Low Energy High Resolution (LEHR) (blue) and Low Energy All Purpose (LEAP) (red). Phantom measurements, shown as markers, were modelled as photopeak emission (constant component) plus component due to septal penetration (exponential component). Model best fit shown as dashed line with 95%-confidence limits represented by the shaded region.

both LEAP and LEHR sensitivities substantially declined. For collimator-phantom separations of 450mm or greater the LEHR collimator sensitivity was lower than the sensitivity of the MELP collimator, though the LEAP collimator sensitivity remained greater than either LEHR and MELP throughout the full range of measurement. At 550mm, the largest collimator-phantom distance common to all three collimators, sensitivity values were calculated as 142.3, 97.5 and 105.9 cps/MBq for LEAP, LEHR and MELP collimators respectively.

Figure 4.3 shows phantom measurements and exponential sensitivity models for LEAP, LEHR and MELP collimator sets plotted against detector-phantom separation. The fitted parameters for both LEAP and LEHR collimators (also listed in Table 4.1) showed near identical similar exponential components (S_1) and similar half-value-distances ($D_{1/2}$), but the LEHR collimator had a significantly smaller fixed component

Phantom Distance (mm)	Collimator Sensitivity (cps/MBq)		
	LEAP ($\sigma = 0.6$)	LEHR ($\sigma = 0.4$)	MELP ($\sigma = 0.4$)
50	213.9	170.9	111.1
100	202.5	158.8	110.4
150	191.2	146.7	109.8
200	181.4	136.9	109.4
250	172.8	127.9	108.2
300	165.7	120.5	107.6
350	159.4	114.8	107.5
400	153.9	109.6	106.6
450	149.7	105.4	106.5
500	146.0	101.5	105.7
550	142.3	97.5	105.9
600	139.5	94.6	-

Parameter	Value for each Collimator		
	LEAP	LEHR	MELP
S_0 (cps/MBq)	120.0 ± 1.3	77.5 ± 0.8	101.6 ± 3.4
S_1 (cps/MBq)	119.9 ± 0.8	120.8 ± 0.5	11.4 ± 2.8
$D_{1/2}$ (mm)	239 ± 7	226 ± 4	399 ± 105
PF ₁₀₀ (%)	42.6 ± 0.1	53.4 ± 0.0	8.6 ± 1.7

Table 4.1: Sensitivity measurements and model parameters for each of Low Energy All Purpose (LEAP), Low Energy High Resolution (LEHR) and Medium Energy Low Penetration (MELP) collimators. Upper table: Sensitivity measurements made in 50mm increments of increasing collimator-phantom separation, with uncertainties derived from poisson standard deviations and measurement precision using error propagation. Lower table: Sensitivity model parameters ($\mu \pm \sigma$) fitted against detector-phantom separation, with uncertainties derived from fit covariance and input σ values; and septal penetration fractions (expressed as %) at 100mm detector-phantom separation, with means and standard deviations derived from sensitivity model parameters; for each of LEAP, LEHR and MELP collimators.

(S_0) compared to the LEAP collimator. The fitted model for the MELP collimator showed a substantially smaller exponential component than that of either low energy collimator (11.4 cps/MBq compared to 119.9 and 120.8 cps/MBq for LEAP and LEHR respectively), and a half-value-distance approximately twice as large (399mm compared to 239 and 226mm respectively). The constant component for the MELP collimator (101.6 cps/MBq) was 15.6% lower than that of the LEAP collimator (120.5 cps/MBq) and was 31.1% higher than that of the LEHR collimator (77.5 cps/MBq).

The dispersion of both LEAP and LEHR curves was very low, with standard errors of fixed and exponential sensitivity components in the range 0.5-1.3 cps/MBq and corresponding coefficients of variation (CoVs) in the range 0.4-1.1%. The standard errors of half-value-distance for LEAP and LEHR were slightly larger; at 7mm and 4mm respectively; resulting in CoVs of 2.9% and 1.8%. In contrast, the standard errors for S_0 and S_1 for the MELP collimator were 3.4 cps/MBq and 2.8 cps/MBq respectively; more than double the largest observed amongst the low energy collimators; and the standard error of $D_{1/2}$ for the MELP collimator was 105mm, an order of magnitude larger than the equivalent dispersion for LEAP and LEHR collimators. The corresponding CoVs for all three parameters for the MELP collimator were correspondingly larger, at 3.3%, 25% and 26% respectively.

Figure 4.4 shows the septal penetration fraction, derived from the sensitivity models and plotted against detector-phantom separation for each of the three collimators. The LEAP and LEHR collimators showed very large septal penetration fractions close to the detector, with maximal (theoretical) penetration fractions of 50.0% and 60.9% respectively. In contrast, the MELP collimator showed significantly lower septal penetration contribution even close to the detector with a maximal theoretical penetration fraction of 10.1%. Representative real-world septal penetration performance was measured for each collimator as the penetration fraction at 100mm from the detector, as shown in both Figure 4.4 and Table 4.1, with resulting values of PF_{100} found to be $42.6 \pm 0.1\%$, $53.4 \pm 0.0\%$ and $8.6 \pm 1.7\%$ for LEAP, LEHR and MELP respectively. As with sensitivity modelling, the MELP standard deviation was much greater at 1.7% compared to 0.1% and $<0.1\%$ for LEAP and LEHR respectively.

Figure 4.5 shows sensitivity spectra acquired at 50mm, 150mm and 450mm collimator-

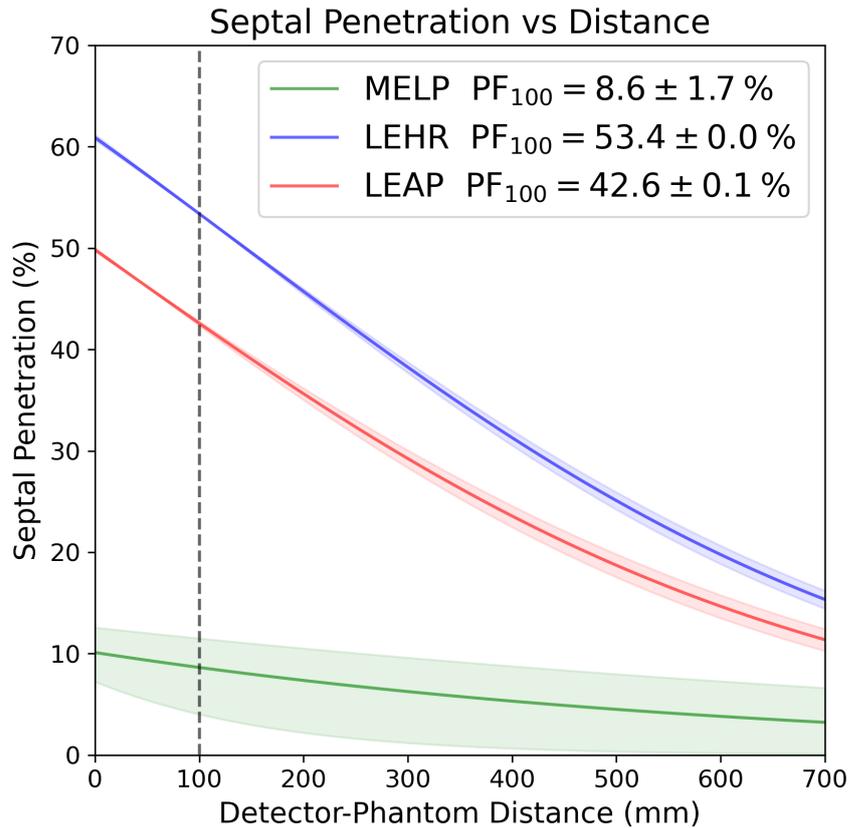


Figure 4.4: Septal penetration derived from sensitivity model versus detector-phantom distance for each of three collimators; Medium Energy Low Penetration (MELP) (green), Low Energy High Resolution (LEHR) (blue) and Low Energy All Purpose (LEAP) (red). Model best fit shown as solid line with 95%-confidence limits represented by the shaded region.

phantom separation for each of the three collimators of interest. The LEAP collimator showed the highest sensitivity within the photopeak region at all distances with both LEHR and MELP collimators showing photopeaks consistently below the LEAP spectrum. As separation increased, and in keeping with the image sensitivity measurements, the MELP collimator showed only a slight reduction in photopeak height; while the photopeak height for both LEAP and LEHR significantly declined as separation increased. The LEHR photopeak dropped below the MELP photopeak at collimator-phantom distances of 450mm or greater. Both LEAP and LEHR collimators also showed markedly higher response rates outside of the photopeak region than were observed for the MELP collimator. There were four major common features identified within both low energy collimator spectra in addition to the expected photopeak and Compton region. These were; a large sensitivity spike in the 75-85keV range (consistent with characteristic x-ray emission from Lead); some higher energy peaks consistent with known high energy emissions at 440, 529 and 539keV; a large Compton continuum "background" region, extending from approximately 40keV through

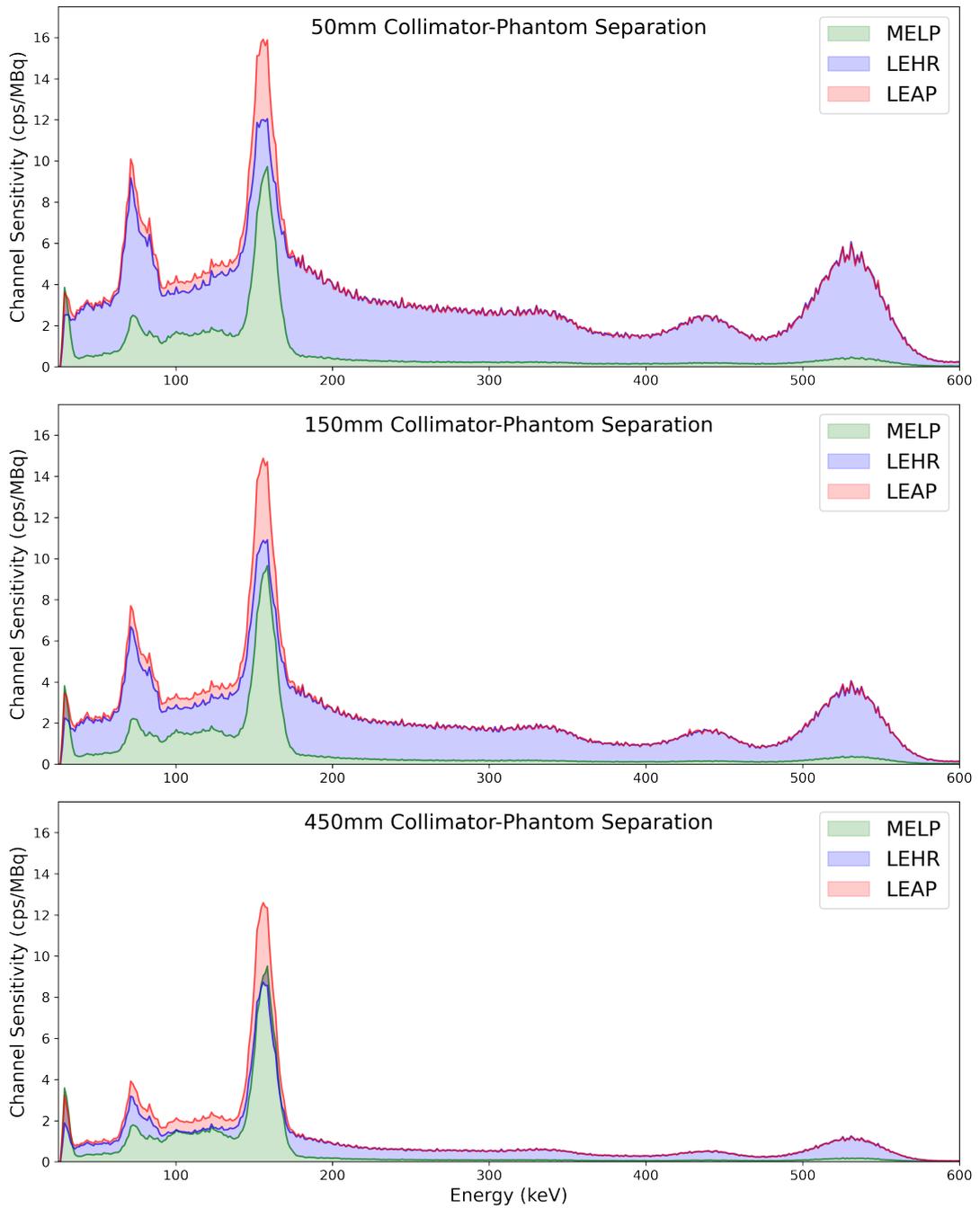


Figure 4.5: Sensitivity spectra acquired with phantom positioned at 50, 150 and 450mm distant from for each of three collimators; Medium Energy Low Penetration (MELP) (green), Low Energy High Resolution (LEHR) (blue) and Low Energy All Purpose (LEAP) (red). Each spectra is shown as a solid line with shading added for ease of visualisation.

to 350keV; and a steeper "background prominence" region extending approximately 60keV either side of the main photopeak. All of these features substantially decreased in intensity as the distance between phantom and collimator increased.

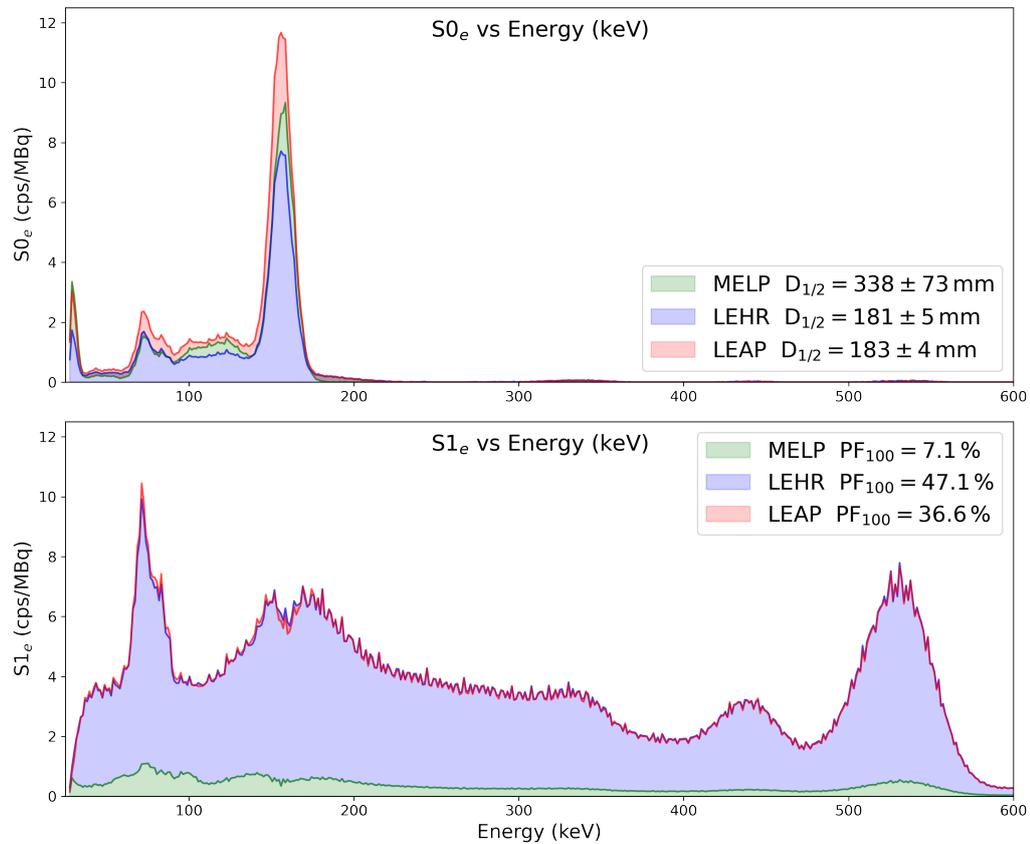


Figure 4.6: Fitted sensitivity spectra model parameters for each of Low Energy All Purpose (LEAP), Low Energy High Resolution (LEHR) and Medium Energy Low Penetration (MELP) collimators using a common half-value distance parameter for each collimator of interest. Upper: Photopeak (constant) component of each sensitivity spectrum plotted, legend shows shared half-value-distances for each collimator. Lower: Septal penetration (exponential) component of each sensitivity spectrum, legend shows septal penetration fractions at 100mm for each collimator of interest.

Figure 4.6 shows the decomposition of the acquired energy spectra into photopeak (constant) and septal penetration (exponential) components; achieved through non-linear least-squares fitting of the relationship between channel sensitivity and detector-phantom separation. Each of the three collimators showed very similar distributions for the photopeak component ($S_{0,e}$). The main observable features were; a strongly dominant photopeak at 159keV with a small preceding x-ray escape peak region around 130keV; a K- α x-ray emission peak at 27keV; a small constant component of Lead characteristic X-ray emissions at 75-85keV; and extremely small, essentially negligible contributions from the low intensity, higher energy emissions at 440, 529 and 539keV. In contrast, the septal penetration component spectra ($S_{1,e}$) showed

similar distributions for the LEAP and LEHR collimators but a markedly reduced distribution for the MELP collimator. LEAP and LEHR collimators showed all of the non-photopeak features previously described for Figure 4.5. Though, in a departure from previously described spectra, the "background prominence" region appeared as two overlapping peaks each positioned adjacent to the photopeak, with a separation of approximately 30keV.

The spectral decomposition method derived some model parameters with somewhat different values to those obtained directly from image sensitivity modelling. In particular, half-value-distances and septal penetration fractions results were both found to be lower (15-23% and 12-17% respectively) than those shown in Table 4.1. The standard errors on the half-value-distances ($\pm 73\text{mm}$, $\pm 5\text{mm}$ and $\pm 4\text{mm}$ for MELP, LEHR and LEAP respectively) were also found to be lower than those previously derived, and the corresponding CoVs were very similar for LEHR and LEAP collimators (2.7% and 2.1% respectively) but slightly reduced for the MELP collimator (22%).

4.2.4 Discussion

In comparison to the relatively qualitative assessment presented in Snay et al. (2011), this work thoroughly evaluates and models the Iodine-123 sensitivity and septal penetration for the three most common parallel hole collimators available on the Siemens gamma camera platform. It is clear from the data presented in Table 4.2.3 and Figures 4.3 and 4.4 that, at clinically relevant distances, the LEAP and LEHR collimators show significantly higher septal penetration than the MELP collimator. Indeed, while naive measurements of Iodine-123 system sensitivity at 100mm from the collimator face would erroneously indicate that LEAP and LEHR were approximately 2x and 1.6x more sensitive than the MELP, their "true" photopeak sensitivities (represented by their respective S_0 values) are actually only 1.18x and 0.76x compared to MELP. This places the MELP collimator firmly in between the LEAP and LEHR in terms of useful sensitivity.

Perhaps the most troubling effect of septal penetration on the low energy collimators is the significant distance-dependence it imposes upon their respective sensitivities. With penetration fractions in the range of 40-60% and half-value-distances around 220-230mm, both LEAP and LEHR collimators have considerable septal penetration

components at close range, with a majority of LEHR counts deriving from penetration events at typical imaging distances. Both low energy collimators will also show highly variable count rates from objects across the clinically relevant range of distances from the detector, even without the presence of any intervening attenuating material. This is a significant departure of sensitivity from the ideal case (constant response at all distances) and contributes to poorer quality acquisitions of both planar and SPECT data. In contrast, the MELP collimator showed only a small minority of events resulting from septal penetration, even at the smallest phantom-collimator separations, which contributed to an almost uniform response with distance in air, much closer to ideal imaging conditions throughout the clinically relevant range of distances.

An examination of energy spectra shown in Figure 4.5 revealed the MELP collimator to produce energy spectra close to ideal under all conditions, while LEAP and LEHR collimators showed significant departures from ideal spectra at anything other than the largest distances assessed. The sensitivity decomposition spectra shown in Figure 4.6 provided additional evidence for the superiority of MELP over LEHR in terms of photopeak sensitivity; represented by the higher photopeak visible in the MELP $S_{0,e}$ distribution as compared to LEHR; and also revealed the complexity of the septal penetration distributions for both low energy collimators.

Both LEAP and LEHR $S_{1,e}$ distributions were near identical, and featured large high energy peaks corresponding to known low intensity emissions at 440, 529 and 539keV; as well as a corresponding Compton scatter region extending across the photopeak range and ending with a Compton edge in the expected range for these emissions (356keV for a 529keV photon scattering through 180°). While these high energy photons have a combined intensity of only 2% of all Iodine-123 emissions (0.4%, 1.3% and 0.3% respectively), these photons are essentially uncollimated by the septa of the foil collimators which are only 0.16mm thick. A value of only 5% of one half-value-layer of lead at 500keV (3.2mm). The uncollimated nature of these emissions dramatically increases their prominence compared to collimated photopeak events, which are themselves only a small percentage of incident photopeak emissions.

The very large characteristic x-ray emission peaks at 75-85keV are an unsurprising ap-

pearance in the septal penetration spectra for LEAP and LEHR, given the appearances of Figure 4.5 in this energy range. The high intensity of this spike reflects the large number of photons able to penetrate the thin septa and interact with the lead shielding of the detector housing, creating the characteristic x-ray emission spike at these energies. Finally, the "background prominence" region has the slightly asymmetric appearance of a lead back-scatter peak, caused by high energy photons Compton scattering within the detector shielding and subsequently re-entering the detector crystal. The small valley between the apparent "peaks" of this region is too coincidentally aligned with the 159keV photopeak to be anything other than an artefact arising from the fitting process in the region dominated by photopeak events.

Overall, these complex features reinforce the difficulty in providing adequate corrections for septal penetration when using LEAP or LEHR collimators for Iodine-123. For planar imaging, such corrections are undoubtedly impossible to achieve due to the distance dependence of these septal penetration effects. However, for SPECT/CT, Monte-Carlo simulation of both collimator and detector shielding interactions could be incorporated into the reconstruction algorithm in order to compensate for these undesired events within the photopeak energy window. Indeed, at the time of undertaking this work Hermes Medical Solutions was in development of advanced Monte-Carlo based collimator models for specific collimator/isotope pairing including Siemens LEHR and Iodine-123. This pairing was clearly aimed at improving Iodine-123 imaging for the most common applications in adult nuclear medicine.

Unfortunately this approach has a number of drawbacks for neuroblastoma imaging. Firstly, there remains an ongoing requirement for high quality planar to accurately calculate image-based prognostic scores, such as the SIOPEN score discussed in Chapter 3, and thus the optimisation of planar imaging remains a significant objective of any Neuroblastoma service. As the MELP collimator is clearly indicated for optimised planar imaging, the use of these advanced LEHR models would require the collimator to be changed mid exam. While this may be feasible in older patients without general anaesthetic, it is entirely impractical to switch collimators during general anaesthetic sessions; and these sessions account for the vast majority of patient scans through diagnosis and first line treatment of neuroblastoma. This would lead to differential imaging for patients based on age and anaesthetic use, which is extremely undesirable,

as it would lead to inconsistent image reporting across serial patient scans and would increase the risk of user error when selecting collimators.

These correction methods are also unproven in clinical practice and poorly understood compared to standard reconstruction methods. In the interest of establishing a robust baseline for quantitative SPECT/CT of neuroblastoma, it is desirable to use well-established and widely available algorithms to increase the breadth of applicability across different centres. Furthermore, the recent thesis of Brown (2018); which evaluated the Hermes Medical Solutions collimator corrections; concluded that these corrections for LEHR collimators introduced a position-dependency to lesion uptake in ^{123}I -MIBG SPECT/CT images. Clearly, further investigations and perhaps refinements are required before they can be recommended for quantitative oncology.

Given these practical and quantitative difficulties, and the superiority of the MELP collimator in terms of true photopeak sensitivity (compared to LEHR) and distance-dependence (compared to both LEAP and LEHR), it seems clear that the MELP collimator offers the optimal neuroblastoma imaging solution for Siemens gamma camera platforms.

4.2.5 Conclusion

This study of Siemens collimator performance showed that the septal penetration of Iodine-123 emissions was significantly greater for both LEAP and LEHR collimators than for the MELP collimator. Both low energy collimators demonstrated penetration fractions of 40-60% at clinically relevant distances, which dramatically reduced their "true" photopeak sensitivity. In contrast the MELP showed less than 10% of counts arising from penetration events at similar distances. Decomposition of sensitivity clearly demonstrated the superior detection of photopeak counts by the MELP collimator compared to the LEHR, though the LEAP still retained an advantage of 20% of total photopeak counts collected.

Both LEAP and LEHR sensitivity also showed a strong distance dependence that was considered incompatible with high quality planar imaging required for SIOPEN scoring, consistent with the work of Gregory et al. (2017). This strong distance dependence was also considered unsuitable for use in quantitative SPECT/CT without

advanced corrections for penetration effects that were clinically unavailable at the time of writing. Based on these factors, good quality quantitative SPECT/CT of Neuroblastoma on a Siemens system must require the use of the MELP collimators.

4.3 Phantom Evaluation

4.3.1 Introduction

Following the prior evaluation of collimator sensitivity and septal penetration, a thorough investigation of SPECT/CT reconstruction parameters was undertaken. The two closely linked goals of this experiment were (1) to determine the optimal parameters (or range of parameters) for quantitative SPECT/CT reconstruction and (2) to confirm which voxel-based measurements seem best suited for use in quantitative SPECT/CT of Neuroblastoma.

4.3.2 Methods and Materials

Phantom Selection and Filling

A Data Spectrum Flangeless Deluxe Jaszczak Phantom, as shown in Figure 4.7, was used with all inserts removed and was filled with 89.6MBq of ^{123}I -Iodide. A NEMA IEC PET Body Phantom, also shown in Figure 4.7, was used without the lung insert and was filled with a total of 170.5MBq ^{123}I -Iodide so as to achieve a sphere-to-background activity concentration ratio between 4:1 and 5:1. The true sphere-to-background ratio was later confirmed as 4.45:1 through counting of 1ml samples using a PerkinElmer (Waltham, Massachusetts, United States) Wizard 3 1480 Gamma Counter. Syringe activity and calibration time were recorded before and after filling both phantoms. All activities were measured to 0.1MBq precision using a Capintec CRC-25R isotope calibrator with a 2mm copper filter insert and instrument calibration traceable to the National Physical Laboratory (NPL). Phantoms were weighed before and after filling, using Class III scales with a 20g measurement error, to determine total contained water mass and therefore the total enclosed volume. At the time of scanning, both phantoms contained activities that were sufficiently low as to remain within the linear count-rate response range of the gamma camera and both were allowed to reach equilibrium activity distributions over a period of at least 1 hour between filling and



Figure 4.7: Photographs of Single Photon Emission Computed Tomography (SPECT) phantoms. Right: National Electrical Manufacturers Association (NEMA) International Electrotechnical Commission (IEC) Positron Emission Tomography (PET) Body Phantom with bone/lung insert removed; used for optimisation of Single Photon Emission Computed Tomography (SPECT) reconstructions. Left: Data Spectrum Flangeless Deluxe Jaszczak Phantom with all inserts removed; used for calibration of Single Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) reconstructions

scanning.

Image Acquisition and Reconstruction

SPECT/CT acquisitions of both phantoms were acquired with a Siemens (Munich Germany) Symbia Intevo dual headed SPECT/CT gamma camera using the manufacturer default ^{123}I primary photopeak energy window (15% width centred on 159keV) with a 128x128 acquisition matrix size and medium energy low penetration (MELP) collimators. All SPECT acquisitions contained 128 projections over 360° in step-and-shoot mode with the detectors starting in $0^\circ/180^\circ$ orientation and automatic contouring. Both phantom acquisitions were also acquired to 25 kcounts in each image of the first position. This was designed to achieve parity with a typical clinical ^{123}I -MIBG 24 hour SPECT acquisition of the trunk (approximately 3 million total counts on average) and also to provide sufficient count data to achieve a good quality SUV SPECT calibration as specified by Hermes Medical Solutions (at least 2 million counts). Hybrid CT exposures were performed with Siemens Caredose4D automatic exposure control, a quality reference tube current of 8mAs, 130kV peak tube voltage, 0.6 second gantry rotation time, a 9.6mm (16x0.6mm) beam size and a helical pitch

of 1.5.

Both phantom SPECT/CT acquisitions were reconstructed using Hermes Medical Solutions Hybrid Recon Oncology 3.2.1 using Hermes default CT attenuation correction, resolution recovery and Monte-Carlo scatter correction parameters. Each phantom was reconstructed with a combination of range of iteration schemes and Gaussian post-filters. The total effective iterations ranged from 16 to 128 in steps of 16 (i.e. 1x16 to 8x16), while post-filter FWHMs ranged from 6mm to 21mm in steps of 3mm. For each combination of reconstruction parameters, an SUV-SPECT calibration factor (measured in cps/MBq) was generated from the Jaszczak phantom acquisition using Hermes Medical Solutions (Stockholm, Sweden) calibration factor calculation software and a 70% cylindrical VOI. This calibration factor was then applied to the matched NEMA IEC phantom reconstruction to produce 48 distinct quantitative reconstructions each using the same acquisition data and different reconstruction parameters.

Data Analysis

The reconstructed NEMA IEC phantom images were analysed using the open source image computing platform 3D Slicer (Fedorov et al. 2012). In an approach adapted from NEMA PET image quality analysis (NEMA 2001) spherical VOIs matched to sphere size (10, 13, 17, 22, 28 and 37mm diameters respectively) were positioned at the centre of each corresponding sphere using the CT image to direct accurate placement, as shown in Figure 4.8. 26 sets of six concentric spherical VOIs were also placed within the background region of the phantom; 13 were placed within the plane of the spheres following the 12 designated background positions described by NEMA PET image quality methodology with one additional set of VOIs positioned centrally due to the absence of the lung insert. These 13 VOI sets were then duplicated and moved 50mm along the cross-sectional axis of the phantom into a plane containing no hot spheres. In total, 162 VOIs(6 spheres and 156 background) were created. 3D slicer was then used to extract absolute activity concentration statistics (number of voxels, mean, standard deviation, 1ml peak, minimum, maximum, median, first quartile and third quartile) from all 162 VOIs applied to each of the 48 SPECT reconstructions. These VOI measurements were loaded into a Python (Rossum 1995) equipped Jupyter Notebook (Kluyver et al. 2016) for further analysis

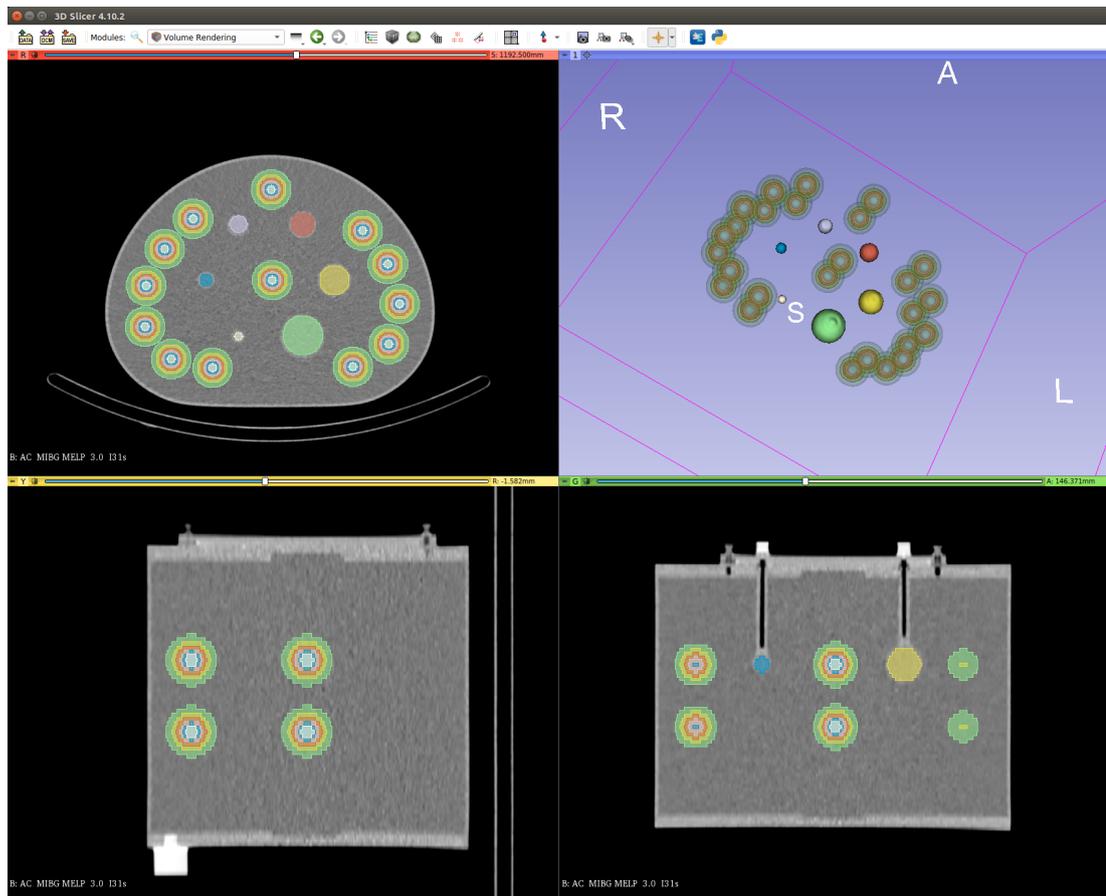


Figure 4.8: Visualisation of volume of interest (VOI) analysis of National Electrical Manufacturers Association (NEMA) phantom performed using 3D Slicer (Fedorov et al. 2012). Spherical VOIs matched to sphere size (10, 13, 17, 22, 28 and 37mm diameters respectively) were positioned at the centre of each corresponding sphere using the CT image to direct accurate placement. 26 sets of six concentric spherical VOIs were also placed in the background region of the phantom; 13 within the plane of the spheres matching the placement of background regions of interest as described in NEMA (2001); and 13 duplicates shifted 50mm along the cross-sectional axis of the phantom to a plane not containing the spheres.

and visualisation using various opensource data handling and visualisation libraries including matplotlib (Hunter 2007), numpy (Harris et al. 2020), pandas (McKinney 2010) and scipy (Virtanen et al. 2020).

Contrast recovery and background variation were calculated for each SPECT reconstruction according to Equations 4.9 and 4.10 respectively, both adapted from NEMA PET image quality methodology (NEMA 2001):

$$Q_s = 100 \times \frac{\mu_s / \mu_{B,s} - 1}{C_S / C_B - 1} \quad (4.9)$$

$$V_s = 100 \times \frac{\sigma_{B,s}}{\mu_{B,s} - 1} \quad (4.10)$$

where Q_s , V_s and μ_s are the contrast recovery, background variability and VOI mean respectively for the sphere of size s ; $\mu_{B,s}$ is the mean across all background VOIs of size s (i.e. mean of means); $\sigma_{B,s}$ is the standard deviation of mean uptake across all background VOIs of size s (i.e. standard deviation of means); and C_S and C_B are the known activity concentrations of the spheres and background volume respectively.

Finally, the extracted absolute activity measurements were collated for each size of sphere and background VOI. Activity boxplots were constructed for each combination of reconstruction parameters, combined with 1ml peak measurements and were compared visually and numerically to the known reference activity concentrations for hot spheres and background at the time of scanning.

4.3.3 Results

Figure 4.9 shows contrast recovery and background variability measurements plotted against total effective iterations for each combination of sphere diameter and filter FWHM. Contrast recovery showed a strong dependence on sphere diameter, with contrast increasing with sphere size for any fixed combination of iterations and filter FWHM, although this narrowed to less than an a 2% difference between the 28mm and 37mm spheres when using the 6mm filter and 128 total iterations. For the 10mm sphere, contrast recovery was less than 0% (i.e. not detectable above background) and closely overlying for all combinations of reconstruction parameters. The 10mm

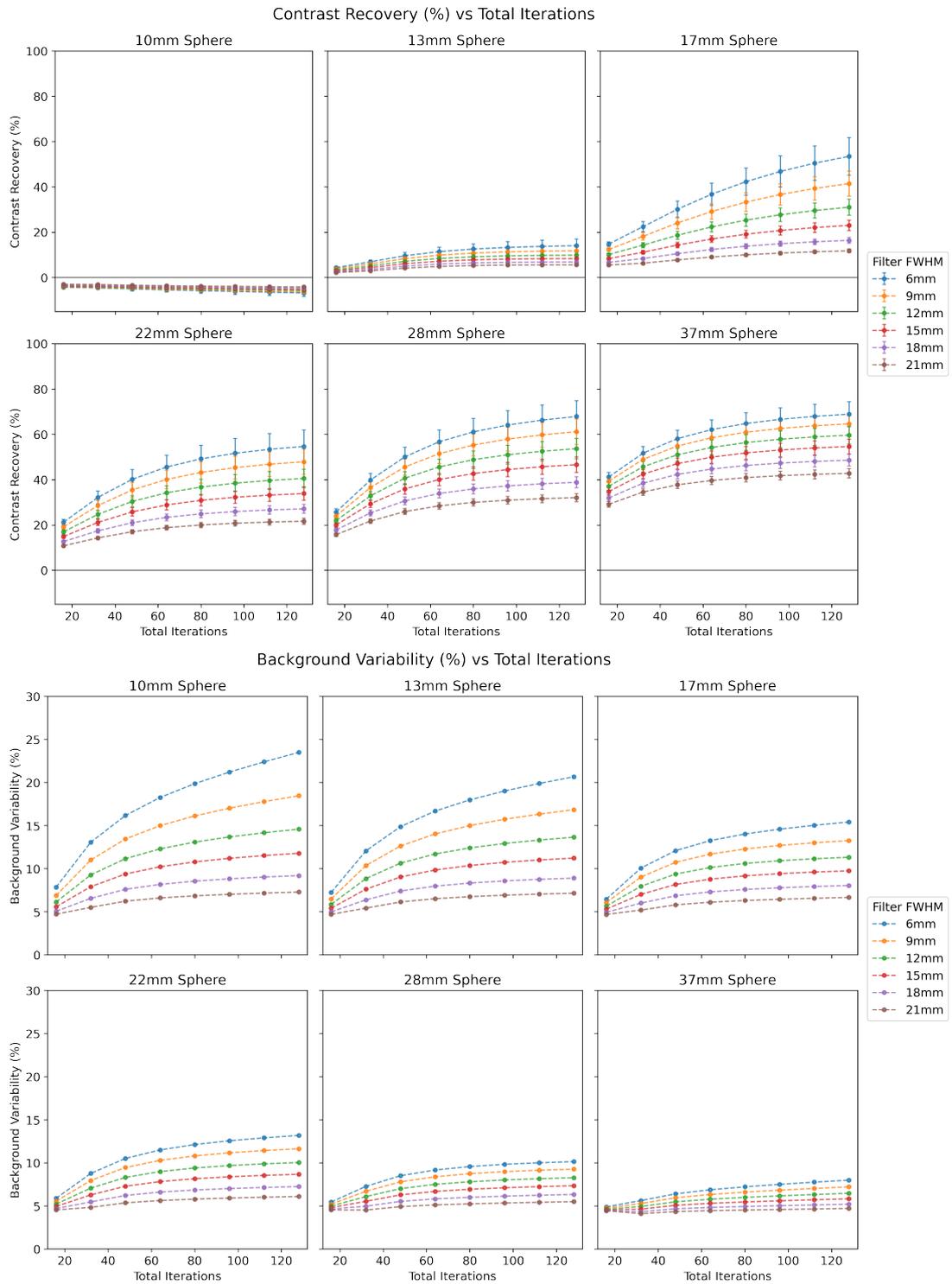


Figure 4.9: National Electrical Manufacturers Association (NEMA) contrast recovery (upper) and background variability (lower) plotted against total effective iterations for each combination of volume of interest (VOI) diameter and filter Full Width at Half Maximum (FWHM).

sphere also uniquely showed a slight downwards trend in contrast recovery as the number of iterations increased. For all other spheres, contrast recovery increased as total iterations increased and as filter FWHM decreased.

The spread of contrast recovery measurements also increased as total iterations increased. None of the plotted lines crossed throughout the range of measurement, which gave the appearance of "contours" of contrast recovery for each selected filter. For the 28mm spheres, each 3mm step in filter FWHM produced an approximately constant corresponding step in contrast recovery that steadily increased with iterations; 3-4% between adjacent FWHM "contours" at 32 iterations to 5-6% at 64 iterations to 6-7% at 128 iterations. This was near identical for the 22mm sphere, with the exception of the step from 18-21mm FWHM which trailed these step changes by 0.5-1%. The 13mm and 17mm spheres showed a strong trend of increasing separation between adjacent FWHM "contours" as filter FWHM decreased; i.e. with the use of sharper filters. Though the overall spread of measurements for the 13mm sphere was very low in contrast to the 17mm sphere. The 17mm sphere showed both the largest spread of contrasts and the single largest step change in contrast with FWHM (11% when moving from 6mm to 9mm FWHM at 128 iterations). Uniquely among the detectable spheres, the 37mm sphere showed a narrowing of these "contours" as filter FWHM decreased.

Background variability plots, also shown in Figure 4.9, similarly showed a "contour-like" appearance; with variability for all sphere sizes increasing monotonically with total iterations and decreasing filter FWHM. The effect of VOI size was particularly dominant in these data, with the range of variability results at 128 iterations reaching 8-24% for 10mm VOIs compared to only 5-8% for 37mm VOIs. For the 10-17mm VOIs, each 3mm reduction in filter FWHM produced noticeably larger increases to variability; compared to approximately constant step changes in variability for the 22-37mm VOIs. This gave the impression of a multiplicative effect of VOI size and filter FWHM on variability.

Figures 4.10, 4.11 and 4.12 show 1ml peak values (where applicable) and boxplots of absolute activity concentrations for each hot sphere. Each individual subplot shows these measurements at each of the eight iteration totals for a single sphere diameter

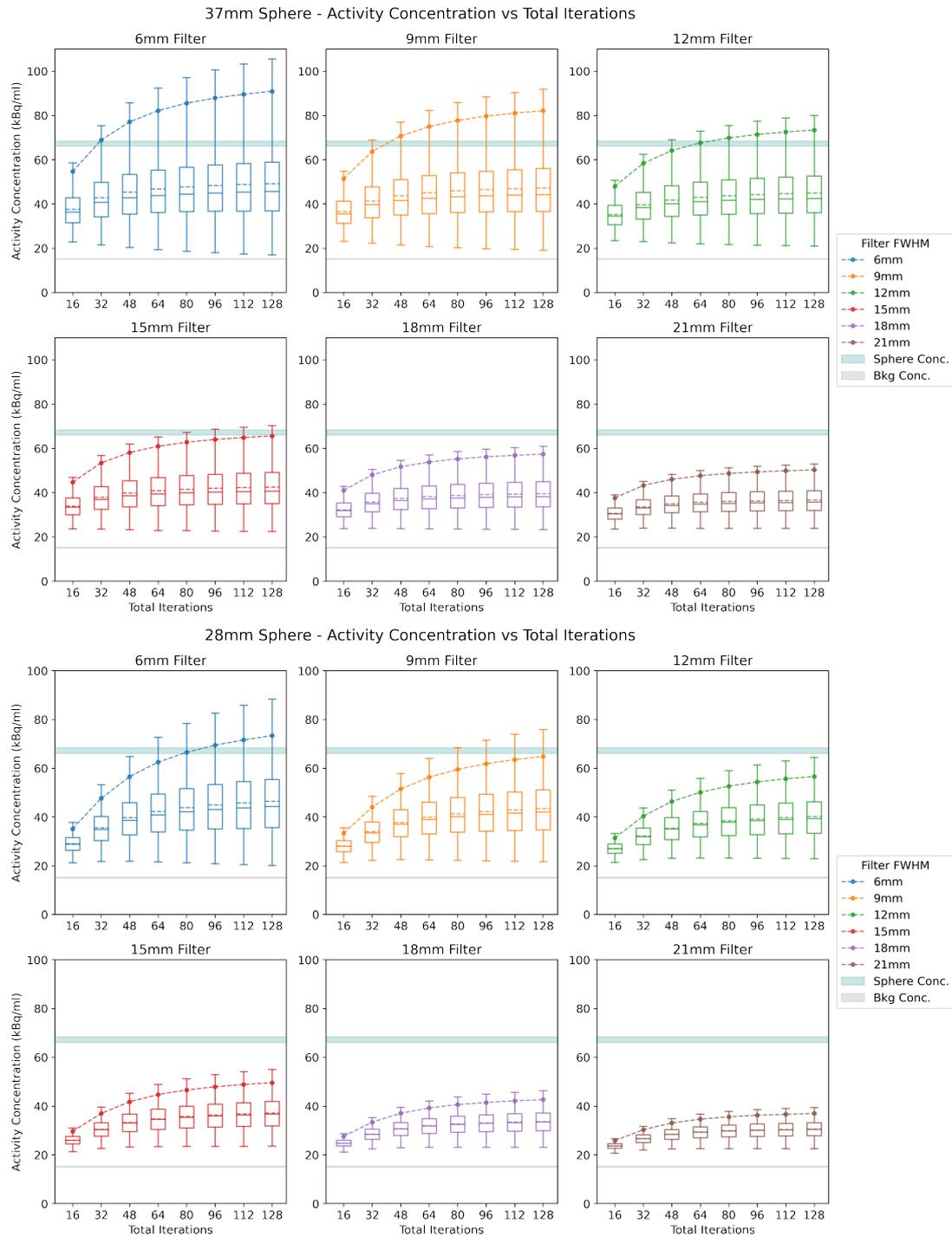


Figure 4.10: 1ml peak values and boxplots of absolute activity concentrations for 37mm diameter sphere (upper plot) and 28mm diameter sphere (lower plot) for each filter Full Width at Half Maximum (FWHM) plotted against total effective iterations. 1ml peak values are shown as points connected by dashed lines and each boxplot shows both the median (solid line) and mean (dashed line). The teal and grey reference bands correspond to the 95% confidence interval of the true underlying sphere and background activities respectively.

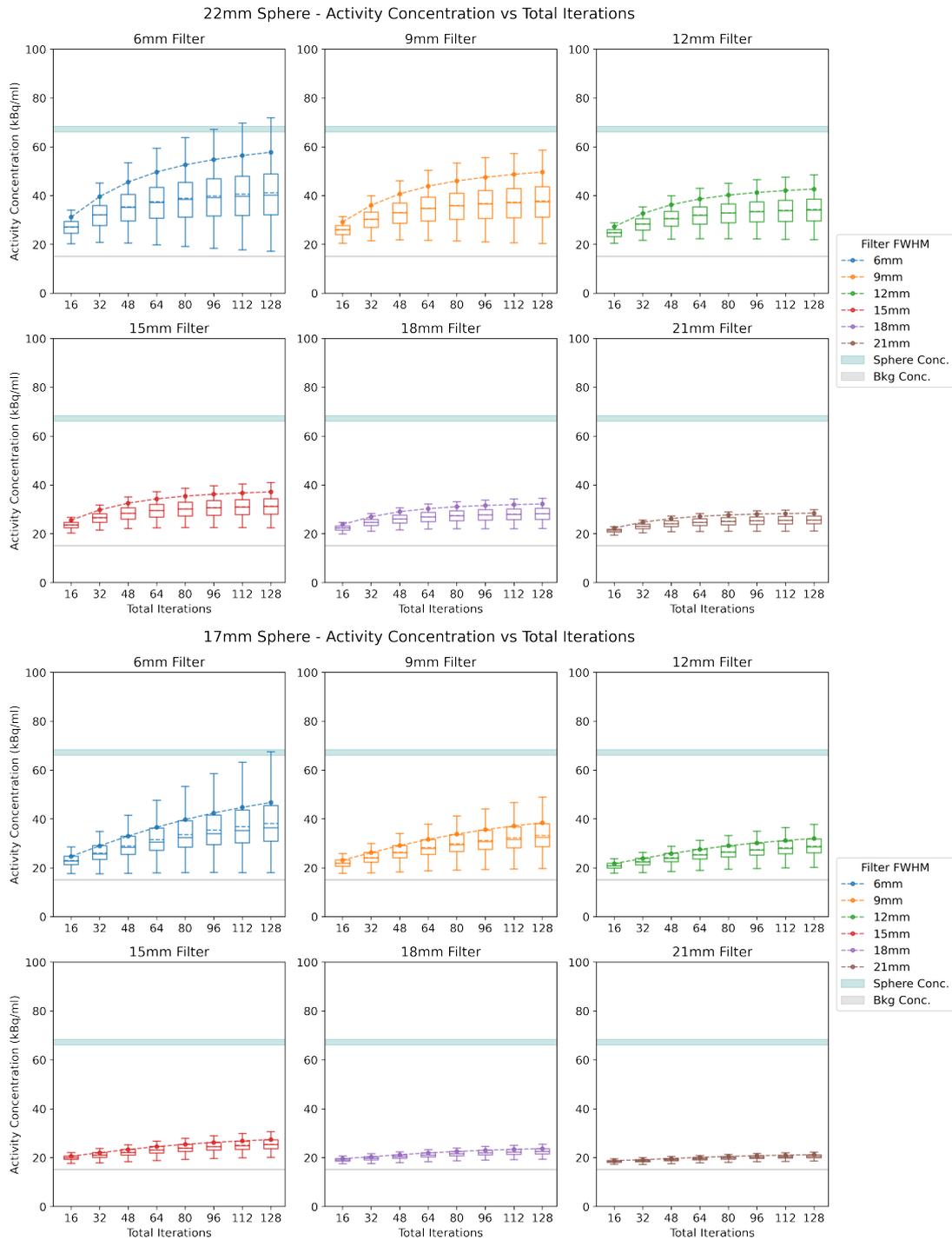


Figure 4.11: 1ml peak values and boxplots of absolute activity concentrations for 22mm diameter sphere (upper plot) and 17mm diameter sphere (lower plot) for each filter Full Width at Half Maximum (FWHM) plotted against total effective iterations. 1ml peak values are shown as points connected by dashed lines and each boxplot shows both the median (solid line) and mean (dashed line). The teal and grey reference bands correspond to the 95% confidence interval of the true underlying sphere and background activities respectively.

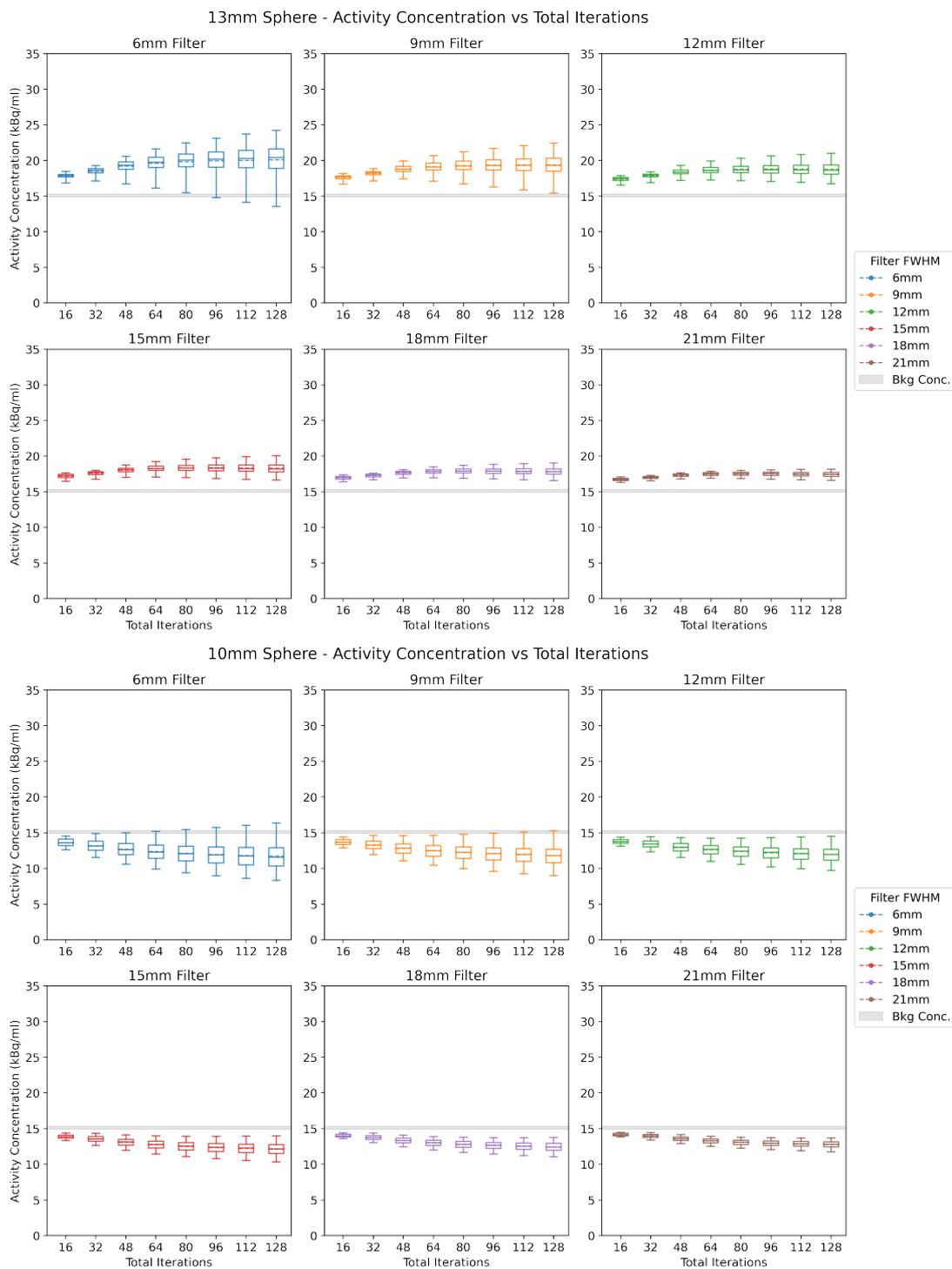


Figure 4.12: Boxplots of absolute activity concentrations for 13mm diameter sphere (upper plot) and 10mm diameter sphere (lower plot) for each filter Full Width at Half Maximum (FWHM) plotted against total effective iterations. Each boxplot shows both the median (solid line) and mean (dashed line). The grey reference band corresponds to the 95% confidence interval of the true underlying background activity concentration. The teal reference band for the true underlying sphere activity is not shown due to axis rescaling.

and single filter FWHM. The filter FWHM colour scheme from Figure 4.9 has been retained for clarity. The teal and grey reference bands correspond to the 95% confidence intervals of the true underlying sphere and background activity concentrations respectively.

Despite a strong dependence on sphere diameter, mean and median sphere concentrations (dashed and solid lines inside each box region respectively) remained substantially lower than the true underlying sphere activity concentration for all sphere sizes and all combinations of reconstruction parameters. Consistent with the contrast recovery plots shown in Figure 4.9, the closest mean activity concentration (49.2 ± 1.1 kBq/ml) was achieved for the largest sphere diameter (37mm), using the highest number of iterations (128) and smallest filter FWHM (6mm); though this was still only 73% of the true underlying sphere activity at the time of the scan (67.2 ± 0.6 kBq/ml). With the exception of the 10mm sphere, which showed a reduction in mean concentration with increased iterations, each sphere showed the largest (and therefore most accurate) mean and median concentrations at 128 total iterations.

The sphere minimum concentration generally showed a much weaker dependence on reconstruction parameters than the mean or median. For 13-37mm spheres and 12-21mm filters the minimum concentration showed effectively no change after 48 iterations and remained substantially above the background 95% confidence band. The 6mm and 9mm filters showed decreasing minimum concentration at ≥ 48 and ≥ 64 iterations respectively. For some spheres, this caused the minimum concentration to approach the underlying background concentration (15.1 ± 0.1 kBq.ml) at high iterations. E.g. for the 37mm sphere with the 6mm filter the minimum concentration reached 17.0 kBq/ml at 128 iterations, only 12% above background. The 13mm sphere uniquely displayed a minimum concentration that dropped below that of the background confidence band, albeit only for the 6mm filter and 112 or more total iterations.

In contrast to the mean, median and minimum concentrations the 1ml peak and maximum concentrations showed very strong dependence on sphere size, total iterations and filter FWHM; and certain parameter combinations resulted in peak and/or maximum concentrations reaching or surpassing the true underlying sphere concentration.

Filter FWHM (mm)	Iteration Limit	
	1ml Peak Concentration	Maximum Concentration
6	<32	<32
9	<48	32
12	64	48
15	128	80
18	≥ 128	≥ 128
21	≥ 128	≥ 128

Table 4.2: Iteration limits for 1ml peak concentration and maximum concentration; defined as the largest number of iterations for which the measured quantity does not exceed the true underlying sphere activity concentration.

E.g. the 37mm sphere with 6mm filter FWHM and 128 iterations recorded a distribution maximum of 105.5 kBq/ml and a 1ml peak concentration of 90.9 kBq/ml (157% and 135% of true sphere uptake respectively). This over-correction of the upper half of the sphere distribution also corresponded to the boxplots showing the largest upwards biasing of the mean relative to the median, a clear indication of skew. While over-correction was observed in the 22-37mm sphere, in the 22mm and 28mm spheres this always occurred at a greater number of iterations than for the 37mm sphere given same filter size. Table 4.2 shows the iteration limits for 1ml peak concentration and maximum concentration; defined as the largest number of iterations for which the measured quantity does not exceed the true underlying sphere activity concentration for the 37mm sphere.

Figures 4.13, 4.14 and 4.15 show 1ml peak values (where applicable) and boxplots of absolute activity concentrations for each diameter of background VOI. As with Figures 4.10, 4.11 and 4.12; each individual subplot shows the results for a single sphere diameter and single filter FWHM across the full range of total iterations and the grey reference band corresponds to the 95% confidence intervals of the true underlying background activity concentration.

The mean and median background concentrations showed a similar pattern across all filter sizes; below 64 total iterations the means and medians were larger than the true background concentration, at 64-80 iterations (80 for 37mm VOIs and 64 for all other VOI sizes) the means and medians entered the 95% confidence band and ≥ 80 the mean remained within the 95% confidence band while the median shifted downwards.

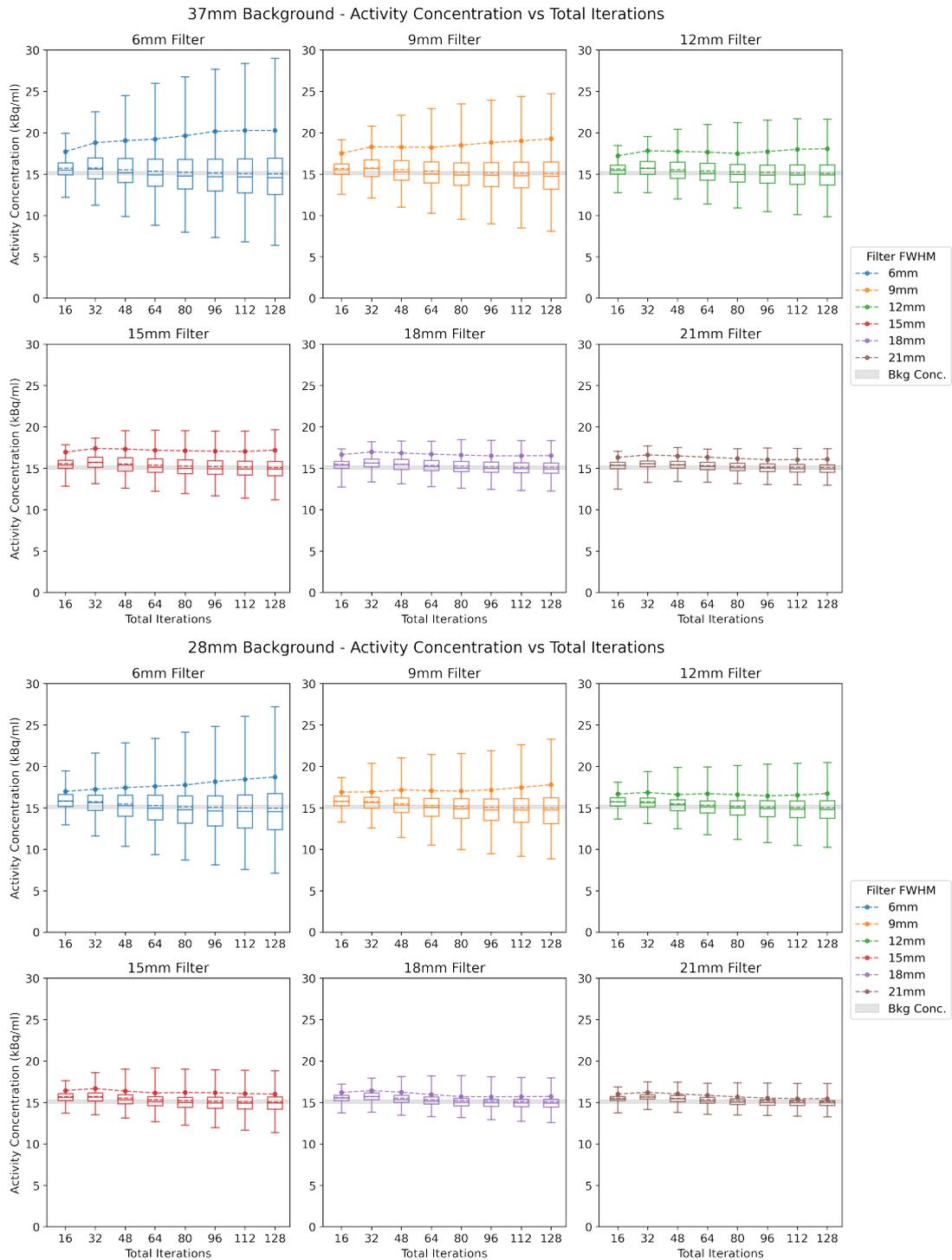


Figure 4.13: 1ml peak values and boxplots of absolute activity concentrations for 37mm diameter background volume of interest (VOIs) (upper plot) and 28mm diameter background VOIs (lower plot) for each filter Full Width at Half Maximum (FWHM) plotted against total effective iterations. 1ml peak values are shown as points connected by dashed lines and each boxplot shows both the median (solid line) and mean (dashed line). The grey reference band corresponds to the 95% confidence interval of the true underlying background activity concentration.

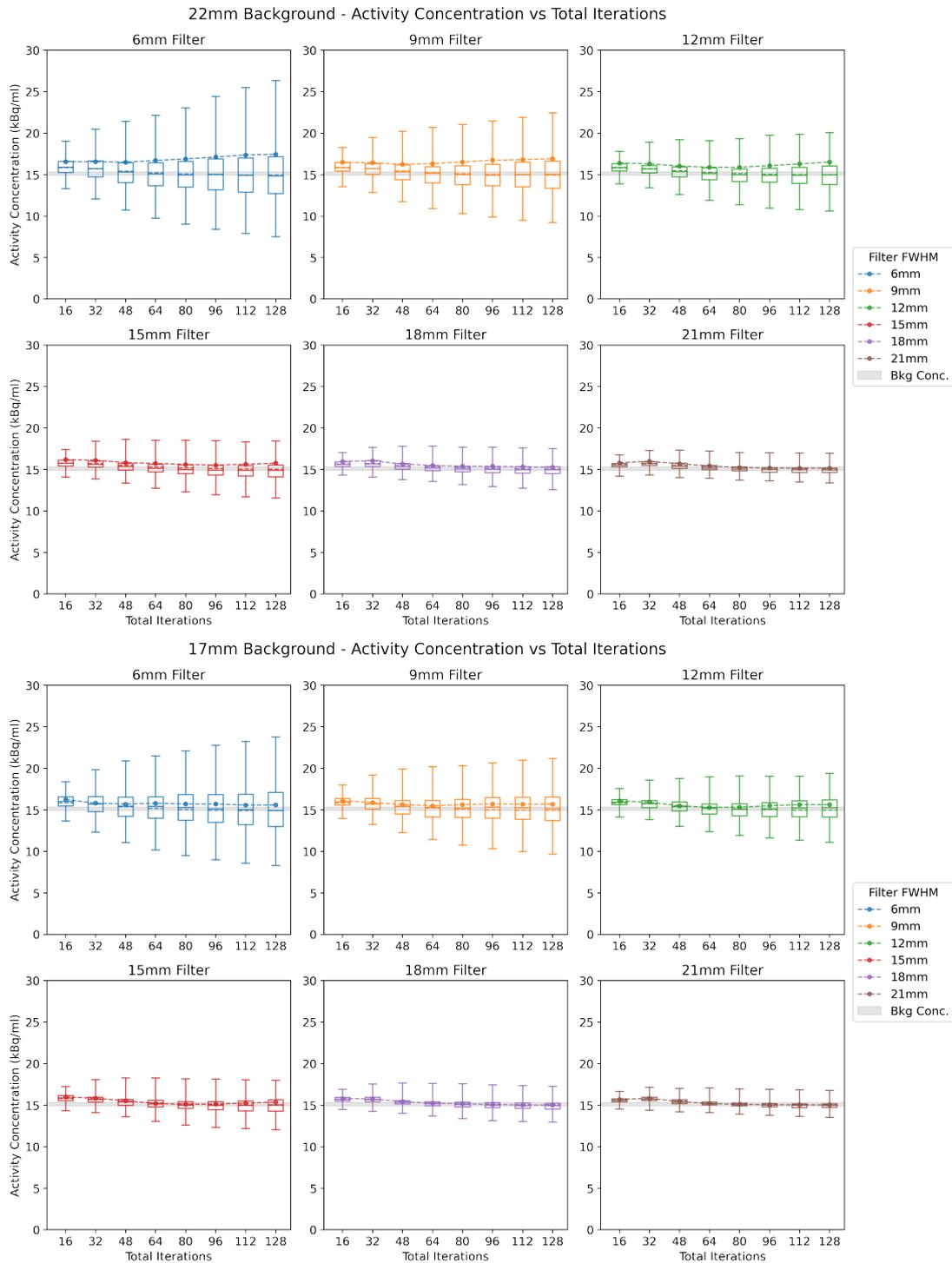


Figure 4.14: 1ml peak values and boxplots of absolute activity concentrations for 22mm diameter background volume of interest (VOIs) (upper plot) and 17mm diameter background VOIs (lower plot) for each filter Full Width at Half Maximum (FWHM) plotted against total effective iterations. 1ml peak values are shown as points connected by dashed lines and each boxplot shows both the median (solid line) and mean (dashed line). The grey reference band corresponds to the 95% confidence interval of the true underlying background activity concentration.

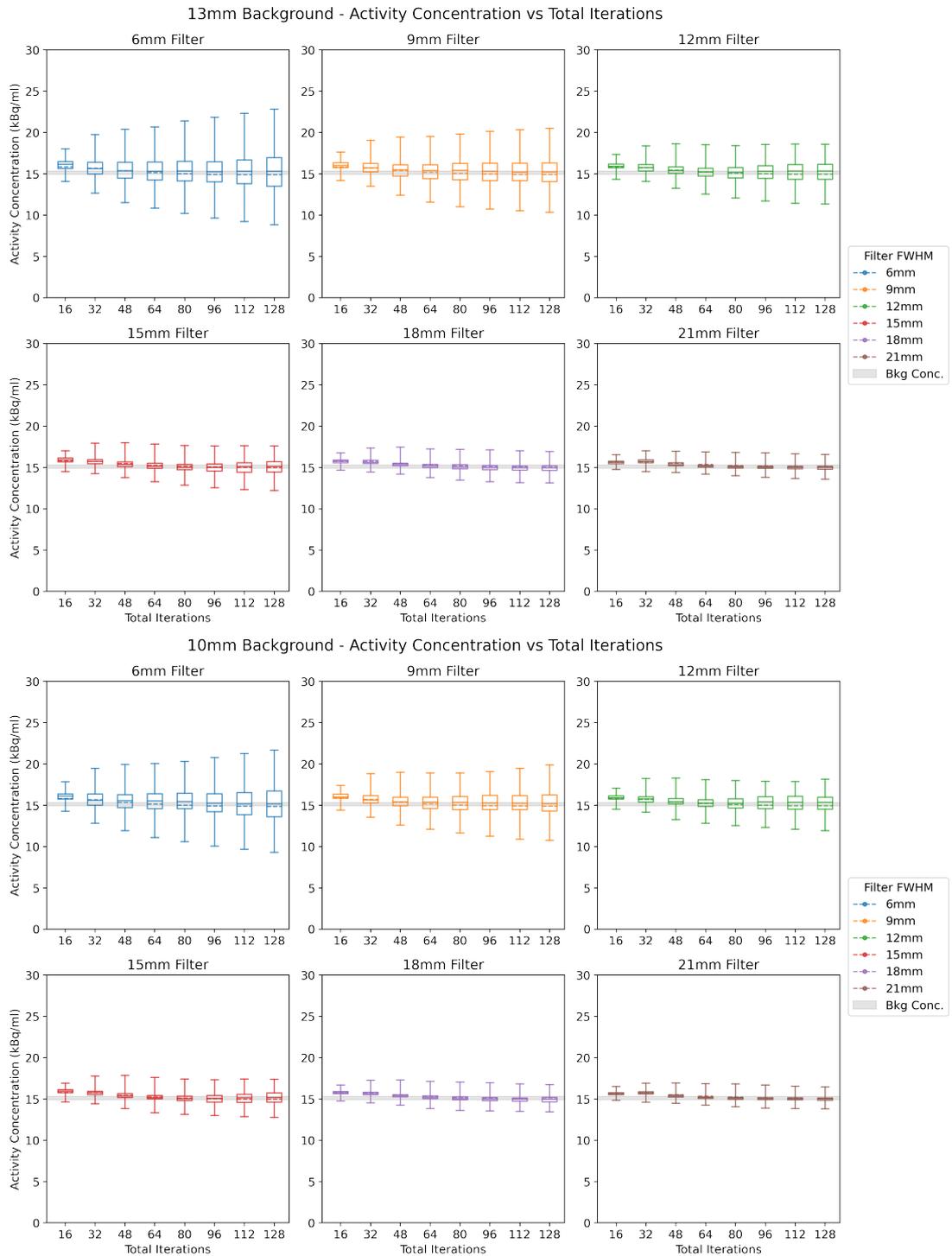


Figure 4.15: Boxplots of absolute activity concentrations for 13mm diameter background VOIs (upper plot) and 10mm diameter background volume of interest (VOI)s (lower plot) for each filter FWHM plotted against total effective iterations. Each boxplot shows both the median (solid line) and mean (dashed line). The grey reference band corresponds to the 95% confidence interval of the true underlying background activity concentration.

For 12-21mm filters the median also remained within 95% confidence band for all VOI sizes but for 6-9mm FWHMs both the 37mm and 28mm VOIs medians dropped outside of the true background concentration band. These instances corresponded to a larger upwards biasing of the mean relative to the median; similar as was observed for the hot spheres under conditions of over-correction.

The spread of each distribution, indicated by the range and interquartile range (IQR), also showed similar behaviour across all VOI sizes. Both range and IQR increased from 16 to 64 iterations where it remained stable for 12-21mm filters but continued increasing for 6-9mm FWHMs. When using the same filter size, the range and IQR were very similar for the three largest VOI sizes, and also for the three smallest VOI sizes; though the 37mm VOIs had a slightly larger range at 6mm than the 28mm or 22mm VOIs. The 1ml peak concentration also showed a similar response to increasing iterations as the range and IQR; increasing with iterations for 6mm and 9mm filters but halting or at ≥ 64 iterations for 12-21mm FWHMs. For the 17mm, the 1ml peak essentially followed the mean and median concentrations while for the 22mm VOIs the peak corresponded closely to the third quartile concentration.

4.3.4 Discussion

It is not surprising that the NEMA analyses in Figure 4.9 showed that both contrast recovery and background variability increased for higher total iterations and smaller filter FWHMs. These observations match the widely held intuition that more applications of the resolution recovery algorithm (i.e. more iterations) will act to amplify the high spatial frequency components of the underlying image data (signal) but also amplify the high frequency noise in the reconstruction. The smaller the filter then applied post reconstruction, the greater proportion of those high frequencies (both signal and noise) that are retained within the final image. While these plots underline the need to strike a balance between maximising signal and minimising noise, they do not provide any clear evidence to aid in identifying which parameters are optimal for doing so. Instead, the contrast recovery plots indicate that it is desirable to use the smallest filter and largest number of iterations possible in order to maximise the recovered signal in small objects. Background variability plots show that, providing the VOI size is sufficiently large, the variability can remain relatively low even at small

filter sizes and high iteration counts. Taken in isolation, these NEMA measurements would seem to suggest that quantitation should be optimised by the use of a 6mm filter FWHM and 128 total iterations, providing a 37mm diameter background reference VOI is used. However, the resulting images using these parameters are extremely noisy and unsuitable for clinical use so it is clear that this conclusion must be erroneous.

This false impression derived from the standard NEMA analyses results from the underlying unsuitability of the VOI mean at the sphere sizes of interest; something abundantly demonstrated by the absolute activity concentration distributions shown in Figures 4.10, 4.11 and 4.12. In these plots, the mean concentration does not approach the underlying sphere activity, even for the largest sphere, regardless of the reconstruction parameters used. Conversely, both the maximum voxel and the 1ml peak concentrations reach and/or surpass the 95% confidence band for the 37mm and 28mm diameter spheres under certain filter and iteration combinations. These alternative measures therefore seem much more promising for quantitative measurement of uptake of objects in the 10-37mm size range. This result makes some intuitive sense given the limited spatial resolution of the MELP collimators; known to be 10.9-11.3mm at 10cm in air from prior acceptance testing. In effect, the limited resolution increases the minimum distance required to transition between the colder background and hotter spheres, leaving the peripheral voxels of the sphere much closer to the background activity concentration and down-weighting the mean when measured over the whole sphere. However, 1ml peak and maximum voxel concentrations can instead be extracted from the interior voxels of the larger spheres; therefore avoiding this transition region and these limitations.

Resolution effects are also the most likely cause of the over-correction of 1ml peak and maximum voxel concentrations observed at high iterations. The upwards biasing of the mean relative to the median in Figure 4.10 provides evidence that the hottest voxels within the spheres are being disproportionately amplified, rather than the resolution recovery correction uplifting the voxel values in the periphery of the sphere. Due to these effects, use of 1ml peak or and/or maximum voxel values requires that the number of iterations must be limited as detailed in Table 4.2 so as to avoid over-correction. On reflection, it is difficult to see how a purely SPECT-based resolution recovery correction, such as the geometric point spread function used here,

could overcome this algorithmic limitation. However, more advanced reconstruction techniques with modelling of patient specific anatomy based on the acquired CT data (similar to those featured in Siemens xSPECT Bone) could perhaps be leveraged to improve upon this.

Unlike the poor accuracy achieved by the use of mean concentration in the hot spheres, Figures 4.13, 4.14 and 4.15 showed that the VOI mean was well suited as a measure of background activity concentration providing a minimum requisite number of iterations were undertaken. Depending on the size of the background VOI this was either 80 iterations (37mm VOI) or 64 iterations (all other VOI diameters). This of course presupposes that the object of interest is much larger than the VOI used and that this VOI is placed in a sufficiently central position that the transition region at the organ boundary can be safely ignored. The background VOIs used in this experiment were placed following NEMA specifications requiring that no background VOI is placed within 15mm of the phantom walls. It would seem reasonable and practicable to require that this same separation is required for background VOIs placed in large reference organs such as the liver. However, this likely makes aortic reference VOIs impractical to measure in paediatric patients.

With both upper and lower iteration limits defined, it becomes feasible to formally exclude certain combinations of reconstruction parameters as quantitatively prohibitive. For instance, neither the 6mm nor 9mm filters are simultaneously compatible with accuracy in both hot spheres and background, as the iteration requirements for accurate 1ml peak or maximum voxel values in Table 4.2 are below the minimum 64 iterations required for accurate background measurement with a ≤ 28 mm background VOI. Similarly, the 12mm filter FWHM is only suitable for use at precisely 64 iterations, but maximum voxel values in large lesions (≥ 37 mm) should not be considered quantitatively accurate and the background VOI must be no larger than a 28mm diameter sphere. The 15mm filter can be used for accurate quantification at 80-128 total iterations with no limitation of background VOI size, however 80 iterations is the only quantitatively viable combination for accurate measurement of maximum voxel values. If the primary quantitative measurement were instead 1ml peak, then 128 iterations would be recommended to maximise quantitative accuracy. While the 18mm and 21mm filters could be used for 1ml peak or maximum voxel values at

Sphere Diameter (mm)	Absolute Activity Concentration (kBq/ml)		
	12mm Filter ×64 Iterations	15mm Filter ×128 Iterations	15mm Filter ×80 Iterations
37	67.6	65.6	62.8 (67.2)
28	50.1	49.6	46.6 (51.2)
22	38.6	37.2	35.4 (38.7)
17	27.6	27.4	25.4 (28.0)
13	(19.9)	(20.1)	(19.5)
10	(14.2)	(14.0)	(13.9)
BKG 28	15.3	15.1	15.2

Sphere Diameter (mm)	Sphere Relative Recovery Coefficients (%)		
	12mm Filter ×64 Iterations	15mm Filter ×128 Iterations	15mm Filter ×80 Iterations
37	99.1	96.9	90.8 (99.2)
28	65.9	66.2	59.9 (68.6)
22	44.1	42.4	38.5 (44.8)
17	23.3	23.6	19.4 (24.4)
13	(8.7)	(9.6)	(8.2)
10	(-2.1)	(-2.1)	(-2.5)

Table 4.3: Absolute activity concentration measured in all spheres and the 28mm background volume of interest (VOI)s (upper table) and the resulting sphere relative recovery coefficients (lower table) for each of the quantitatively viable reconstruction parameter combinations. Results shown inside brackets indicate maximum voxel values or recovery coefficients derived using maximum voxel values. Results shown without brackets indicate 1ml peak values (upper table, spheres) or mean concentrations (upper table, background) or recovery coefficients derived using these measures (lower table).

128 total iterations, the recovered sphere concentrations are strictly inferior to those available using the 12mm or 15mm filters in the combinations listed above so neither can be recommended based on these data. Notably neither 18mm or 21mm filters produced any over-corrected measurements in the 37mm sphere, and may be able to recover slightly higher concentrations by using increased iterations beyond the maximum of 128 used in this experiment. However in both instances the trend suggests diminishing returns compared to simply switching to a suitable parameter set using 12mm or 15mm filter FWHMs.

The upper table in Table 4.3 shows the absolute activity concentrations measured in all spheres and the 28mm background VOIs using each of the quantitatively viable reconstruction parameter combinations recommended above. The lower table shows the relative recovery of each sphere derived using the measurements taken from the

upper table and Equation 4.9. In both tables results shown inside brackets indicate maximum voxel values or recovery coefficients derived from maximum voxel values, while results without brackets indicate either 1ml peak values or recovery coefficients derived using 1ml peak values. The only exceptions to this are the values listed in the "BKG 28" row, which are VOI mean values extracted from the set of 28mm background VOIs. This was the largest background VOI which was compatible with all of the quantitatively viable reconstructions. Overall, these results show that there was very little difference between the recovered concentrations achieved with each parameter set, providing the maximum voxel value was used for the larger spheres for the 15mm filter at 80 iterations. Given that the background variability in the 28mm sphere was found to be 8.0-8.5% across the three parameter sets, differences of a few percent in relative recovery should be considered statistically negligible.

Of the parameter sets investigated, the 15mm FWHM at 80 total iterations offers the most flexibility for development of a quantitative framework for MIBG SPECT/CT, as it permits accurate measurement of both 1ml peak measurements (as used in PERCIST) and maximum voxel values (more similar to qPET methodology); with the latter remaining quantitatively stable up to an object size of 37mm. Above this object size there may be an over-correction of maximum voxel values, but for such objects the 1ml peak could be used instead. This would provide a cushion of around 9% relative recovery over and above the 37mm sphere. Although no larger spheres were tested in this experiment, this level of overhead would seem fairly appropriate based on the observation of peak/mean ratios in 28mm and 37mm background VOIs which were found to be 6.3% and 12.0% respectively. Both of the 15mm parameter sets also permit greater flexibility in terms of background VOI size. The 3cm diameter liver background VOI used in PERCIST is well approximated by the 28mm diameter VOI used in this phantom study, while the cuboidal liver VOI defined in qPET is equivalent in volume to a 38.6mm diameter sphere and so can be considered statistically similar to the 37ml diameter background VOI. At 80 or 128 iterations, both of these background VOI configurations are quantitatively viable.

It was notable that none of the candidate parameter sets were able to detect the 10mm sphere, as indicated by the negative relative recovery values, and absolute concentration values slightly below the the 28mm background VOIs. This is not entirely

surprising, given that this sphere diameter is smaller than the inherent spatial resolution of the MELP collimator, measured at 10.9-11.3mm during acceptance testing. It seems reasonable to expect that signal will only be detected for objects below the collimator spatial resolution when they have significantly higher sphere:background contrast ratios e.g. 20:1 or 50:1; orders of magnitude very unlikely to be encountered in clinical practice. This means that a SPECT/CT quantitative framework for MIBG would be limited to assessing lesions at most 11-12mm across, slightly larger objects than can be assessed under RECIST or qPET. While a switch to a higher resolution collimator (e.g. LEHR with 7mm spatial resolution at 10cm) would undoubtedly offer some improvement to 10mm sphere recovery, this would still remain strongly constrained by partial volume effects; due to constituting only 7 voxels at 128x128 matrix size. Under these conditions not only does the spatial resolution of the imaging system limit the recovery of signal from the sphere but all sphere voxels straddle some combination of the sphere centre, sphere walls and the immediately surrounding background volume; leading to extremely pronounced partial volume effects. Quantitative accuracy would also be compromised by the substantial septal penetration effects detailed in Section 4.2. These effects are of a large enough magnitude to render the LEHR inferior to the MELP for planar imaging (Gregory et al. 2017); making the LEHR undesirable for general use in Neuroblastoma. In SPECT/CT, use of the LEHR would require collimator specific septal penetration modelling to be incorporated into the reconstruction algorithm, a feature that remains unavailable at the time of writing.

Finally, it should be noted that these measurements were derived from a single set of SPECT data acquired at a single time-point. This methodology would benefit from multiple repeat acquisitions to measure repeatability as well as multiple full repetitions on a number of distinct Siemens scanners to confirm the general reproducibility of the results.

4.3.5 Conclusion

This work has demonstrated that quantitatively meaningful measurements of absolute activity concentration can be made in MIBG SPECT/CT using Hermes Medical Solutions (Stockholm, Sweden) SUV-SPECT. Under conditions that closely mimic clinical

MIBG scanning, objects in the size range 13-37mm were resolvable, but the mean measured across the full extent of such objects poorly represented the true underlying activity due to resolution effects. Alternative quantitative measurements such as 1ml peak and maximum voxel value performed much better, but could be overwhelmed by image noise if reconstruction parameters were not carefully controlled. Optimal reconstruction settings were determined to be 80 total iterations using an OSEM algorithm with geometric resolution recovery, CT attenuation correction and Monte Carlo scatter correction followed by a 15mm FWHM 3D Gaussian post filter. Quantitative methodologies that mirror PERCIST and qPET for SPECT/CT would both seem to be feasible under these imaging conditions and using these reconstruction parameters. A clinical evaluation should now be undertaken to confirm the feasibility of such measurements in real world patient scans.

4.4 Clinical Evaluation

4.4.1 Introduction

Following collimator selection and SPECT/CT phantom studies, a clinical evaluation of quantitative ^{123}I -MIBG SPECT/CT was undertaken using the Neuroblastoma SPECT/CT archive at the RHC. The goals of the study were; (1) to quantify the normal range of SPECT/CT uptake measured in the right lobe of liver, left lobe of liver and in the brain for the local population and compare these results to that of Brady and Shulkin (2019); (2) to determine which measurement units provide the most consistent baseline for the liver and brain; (3) to make practical recommendations for the use of these uptake measurements in the liver and brain. Tumour values were not considered at this stage due to the expected relatively small patient sample.

4.4.2 Methods and Materials

Patient Selection

The CRIS Radiology Information System at the RHC was retrospectively queried for all instances of ^{123}I -MIBG imaging undertaken from the initiation of the MIBG SPECT/CT service in November 2018 until December 2021 inclusive. The initial query returned a dataset of 77 scans of 25 patients across this time period. The body

of the radiology report text was used to exclude scans where the patient was not under GA (24 scans) and where the CT component of the scan was abandoned (3 scans). The remaining dataset consisted of reports for 50 examinations of 13 patients scanned under GA during the period of interest. For each identified examination, the Hermes Medical Solutions (Stockholm, Sweden) nuclear medicine database was queried for a planar wholebody acquisition, multi-field of view SPECT acquisition and corresponding low-dose CT scan acquired for the full extent of the SPECT coverage. All images were pseudo-anonymised and moved to a dedicated storage archive for review, reconstruction and analysis. A pseudo-anonymised record was created for each examination which included the disease stage at diagnosis, clinical time point, presence of liver metastases on same-day MRI and any surgical interventions affecting the anatomy of the liver. Patient age, height, weight, uptake time (time between administration and scan) and decay corrected activity at the time of scanning were also added to the pseudo-anonymised record for each scan.

Image Acquisition and Reconstruction

All imaging was acquired on a Intevo 16 SPECT/CT scanner approximately 24 hours following isotope administration. Patients were positioned with arms down in order to maximise similarity of patient positioning to same-day MRI examinations. All nuclear medicine imaging was acquired using the Siemens default ^{123}I primary photo-peak energy window (15% width centred on 159keV) and MELP collimators. Planar wholebody scans were acquired with a scan velocity of 5cm/min, automatic contouring, a 256x1024 matrix size and included the full extent of the patient. All SPECT acquisitions were acquired using a "wholebody tomo" workflow set to acquire 2-3 contiguous SPECT acquisitions acquired with the gantry configured to 180° orientation. The system automatically coordinated bed and detector movement between successive SPECT acquisitions. Each SPECT acquisition contained 128 projections over 360° with a 128x128 matrix size and was acquired in step-and-shoot mode, with automatic contouring and 20-25 seconds of accumulation per projection. SPECT coverage included the entire patient head and continued to either mid-thigh, mid-calf or the full extent of the patient. The extent of SPECT coverage was dependent on patient height and visible disease extent on wholebody planar imaging. Hybrid CT scans were acquired immediately following SPECT acquisition and matched the full

SPECT coverage. All CT exposures were performed with Siemens Caredose4D automatic exposure control, a quality reference tube current of 8mAs, 0.6 second gantry rotation time, a 9.6mm (16x0.6mm) beam size and a helical pitch of 1.5. Peak tube voltage was set according to the local optimised exposure protocol; 80kV was used for patients aged 0-4 years, 110kV was used for patients aged 5-9 years and 130kV was used for patients aged 10+ years.

Quantitative SPECT reconstructions were performed using Hermes Medical Solutions (Stockholm, Sweden) Hybrid Recon 3.2.1, incorporating the quantitative reconstruction package SUV-SPECT. All reconstructions were performed using 5 iterations and 16 subsets (80 effective iterations) with resolution recovery, CT attenuation correction and Monte-Carlo-based scatter correction. A 3D Gaussian post filter with a FWHM of 15mm was applied post reconstruction. Decay correction and calibration factor scaling were also applied during reconstruction to convert SPECT voxel values into absolute ^{123}I activity concentration. Patient height, weight and administered activity were supplied to the reconstruction application to allow activity concentrations to be converted into standard uptake values scaled to body weight (SUV/BW) and standard uptake values scaled to lean body mass (SUV/LBM) as required.

Study Analysis

Each reconstructed SPECT/CT examination was loaded into Hermes Medical Solutions Hybrid Viewer 5.0.1 and the images reviewed along side the anonymised radiology report text to provide clinical context. The presence and location of any congenital, surgical or pathological abnormalities that impacted the liver (e.g. liver metastases, liver adjacent mass, liver resection) were recorded for each case. The scans of the youngest patients (i.e. those with the smallest livers) were used to determine the largest spherical VOI size which was able to be easily positioned within both the right and left lobes of liver, while preserving a 15mm surrounding margin of liver tissue. A 22mm diameter spherical VOI was found to be optimal for these cases and was used for all further VOI analysis.

To maintain consistency between phantom and clinical analyses, VOI placement and numerical analysis was undertaken using the open source image computing platform 3D Slicer (Fedorov et al. 2012). VOI placement was performed using CT images only

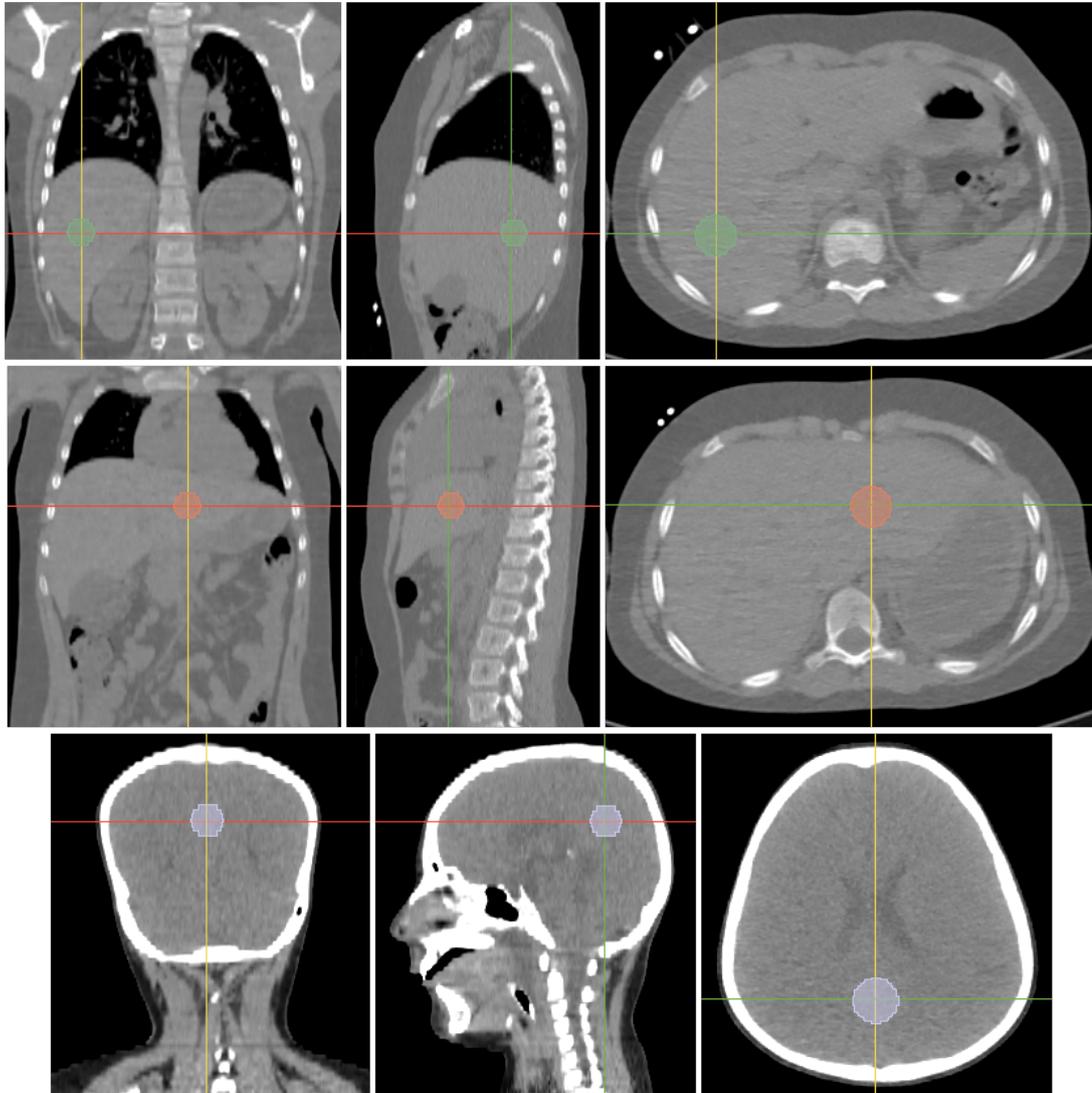


Figure 4.16: Placement of 22mm volume of interest (VOI)s in right liver (upper images, green VOI), left liver (middle images, orange VOI) and brain (lower images, blue-grey VOI) in a patient with grossly abnormal anatomy. In such cases the right liver VOI was positioned in the medial-posterior aspect of the right lobe of liver, approximately half way between the dome and the inferior margin; the left liver VOI was placed within the medial left lobe of liver, approximately midway between the superior and inferior margins at the midline of the patient; and the brain VOI was placed in the medial parietal lobe, approximately midway between the lateral ventricles and the posterior skull vault. Placement was adapted consistently on a per-patient basis to account for pathological or anatomical abnormalities such that VOI placement was matched across an entire set of patient scans.

and was determined on a per-patient basis so that positioning remained consistent across all serial scans of the same patient. Unless otherwise indicated due to abnormal anatomy, the right liver VOI was positioned in the medial-posterior aspect of the right lobe of liver, approximately half way between the dome and the inferior margin; the left liver VOI was placed within the medial left lobe of liver, approximately midway between the superior and inferior margins at the midline of the patient; and the brain VOI was placed in the medial parietal lobe, approximately midway between the lateral ventricles and the posterior skull vault. Following determination of optimal VOI positioning, the three 22mm diameter spherical VOIs were placed in each patient scan as shown in Figure 4.16. The anatomically positioned VOIs were then automatically copied to the quantitative SPECT reconstructions and absolute activity concentration statistics (number of voxels, mean, standard deviation, 1ml peak, minimum, maximum, median, first quartile and third quartile) were exported to text file for further analysis.

The exported VOI measurements were loaded into a Python (Rossum 1995) equipped Jupyter Notebook (Kluyver et al. 2016) for further analysis and visualisation using various opensource data handling and visualisation libraries including matplotlib (Hunter 2007), numpy (Harris et al. 2020), pandas (McKinney 2010) and scipy (Virtanen et al. 2020).

Height and weight were combined using Equation 4.11, derived from the work of Peters et al. (2011), to generate an estimation of lean body mass for each patient scan:

$$LBM = 3.8 \times 0.0215 \times W^{0.6469} \times H^{0.7236} \quad (4.11)$$

where LBM , W and H are the estimated paediatric lean body mass measured in kg, patient weight measured in kg and patient height measured in cm respectively. VOI statistics and were then rescaled using Equation 4.1 to generate measurements in units of SUV/BW and SUV/LBM.

VOI means and standard deviations were then combined into patient-level and cohort-level means and standard deviations using Equations 4.12 and 4.13, adapted from Altman et al. (2013):

$$\bar{\mu} = \frac{1}{N} \sum_i \mu_i n_i, \quad (4.12)$$

$$\bar{\sigma} = \sqrt{\frac{1}{N-1} \left(\sum_i \sigma_i^2 (n_i - 1) + \mu_i^2 n_i - \bar{\mu}^2 N \right)} \quad (4.13)$$

where n_i , μ_i and σ_i are the sample size, mean and standard deviation for the i -th subgroup; and N , $\bar{\mu}$ and $\bar{\sigma}$ are the sample size, mean and standard deviation aggregated over the combined group.

Boxplots, means and standard deviations of right and left lobes of liver and brain uptake were then compared visually and numerically for each of the three measurement units (kBq/ml, SUV/BW, SUV/LBM) at the levels of individual scans, same-patient groups and across the entire cohort. Mean uptake within each VOI was also plotted against uptake time and Bland-Altman analysis was undertaken to assess the mean difference between left and right lobes of liver.

4.4.3 Results

Table 4.4 records the number of scans, presence of liver metastases, age at time of scanning, uptake time between administration and scan, weight, height and administered activity at time of scanning grouped by patient. Included patients were aged between 4 months and 11.5 years of age and showed a correspondingly wide range of height (60-138cm), weight (6.1-38.9kg) and decay corrected activities at the time of scanning (19.5-78.0MBq). With the exception of three scans (patient 8 scan 1, 21.8 hours; patient 11 scan 6, 23.3 hours; patient 13 scan 3, 23.4 hours) all acquisitions were undertaken with at least 25 hours of uptake time between administration and scan and the longest uptake time recorded was 28.3 hours (patient 1, scan 6).

Only two patients had known liver metastases identified in any included scans while one patient had a history of liver metastases in previous scans. Patient 4 was stage 4S at diagnosis and only had a single scan within the dataset. Same-day MRI observed multifocal liver metastases in both right and left lobes but these metastases were not visible on the hybrid CT images due to the low-dose nature of the CT scanning protocol. It was therefore not feasible to guarantee the exclusion of liver metastases from left and right liver VOIs when placed on CT. Patient 8, also an instance of stage 4S Neuroblastoma, had a clinical history of liver metastases present at diagnosis but none were detected in the post-treatment scans included in this dataset.

Patient	Scans	Liver Metastases	Age at Scan (y)	Uptake Time (h)	Height (cm)	Weight (kg)	Activity at Scan (MBq)
1	6		4.4 – 5.8	25.7 – 28.3	110 – 123	19.7 – 23.3	43.4 – 49.1
2	2		5.9 – 6.2	25.7 – 26.6	115 – 118	18.3 – 19.6	40.4 – 46.8
3	6		4.6 – 7.0	25.9 – 27.1	100 – 111	19.1 – 23.5	39.8 – 47.5
4	1	Multifocal	0.4	26.3	60	6.1	19.5
5	2		3.5 – 4.7	25.6 – 26.0	91 – 99	14.3 – 16.5	35.0 – 40.9
6	7		0.9 – 2.7	25.1 – 27.2	77 – 91	7.8 – 11.4	22.3 – 31.3
7	9		4.1 – 6.2	25.5 – 27.0	102 – 115	16.7 – 22.0	40.0 – 49.6
8	1	Previous	0.4	21.8	66	7.8	28.7
9	1		3.5	26.4	85	12.4	28.8
10	2		8.1 – 8.8	26.8 – 27.5	135 – 138	37.5 – 37.7	63.3 – 66.2
11	6		10.3 – 11.4	23.3 – 27.5	133 – 138	31.4 – 38.9	58.1 – 78.0
12	4	Right Lobe	1.3 – 1.9	25.2 – 27.2	80 – 88	10.7 – 12.5	27.8 – 30.3
13	3		2.6 – 3.0	23.4 – 27.4	90 – 94	11.3 – 12.0	27.9 – 34.2
Full Cohort	50	-	0.4 – 11.4	21.8 – 28.3	60 – 138	6.1 – 38.9	19.5 – 78.0

Table 4.4: Number of scans, presence of liver metastases, age at time of scanning, uptake time between administration and scan, weight, height and administered activity at time of scanning grouped by patient.

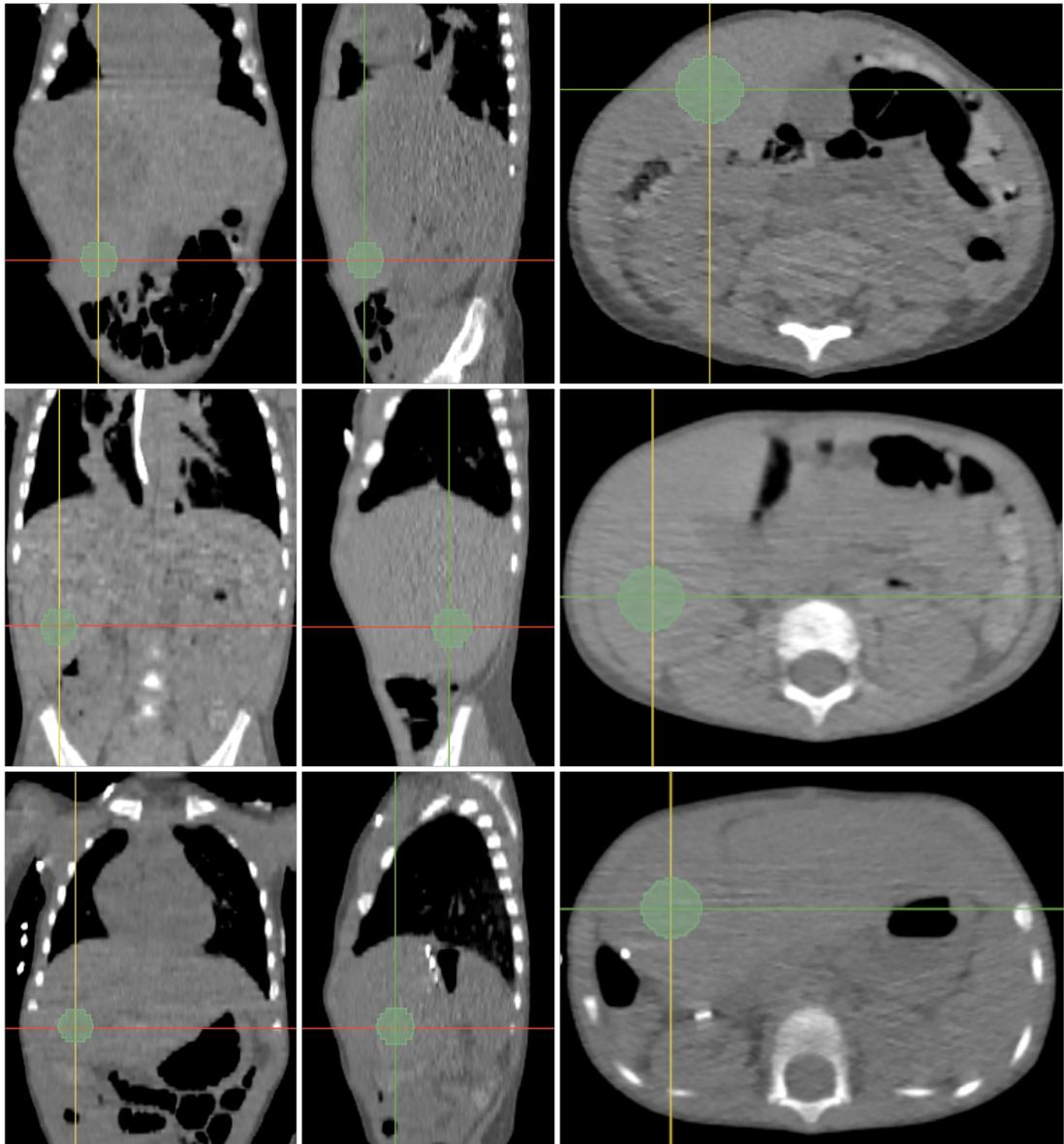


Figure 4.17: Patient 12 - the lone instance of liver surgery within the dataset. At diagnosis the medial segments of the right lobe of liver were heavily infiltrated by a large mass which also compressed all remaining liver tissue and displaced it outwards. Therefore the only section of right lobe tissue suitable for volume of interest (VOI) placement was the non-infiltrated portion of segment VI, which had been pushed further inferior and anterior than is anatomically typical (upper images). In the second scan, the mass had significantly reduced in size due to systemic treatment and segment VI had returned to a more normal position within the patient (middle images). Following the second scan the patient underwent a near complete resection of the right lobe of liver so as to remove the residual mass and all affected liver tissues (lower images). In the third and fourth scans the right lobe of liver had partially regenerated and the most similar segment of liver tissue was used for VOI placement.

Patient 12 was the lone instance of liver surgery within the dataset and this intervention rendered consistent placement of the right liver VOI impossible, as is shown in Figure 4.17. At diagnosis the medial segments of the right lobe of liver were heavily infiltrated by a large mass which also compressed all remaining liver tissue and displaced it outwards. As a consequence the only section of right lobe tissue suitable for VOI placement was the non-infiltrated portion of segment VI, which had been pushed further inferior and anterior than is anatomically typical. In the second scan the mass had significantly reduced in size due to systemic treatment and segment VI had returned to a more normal position within the patient. Following the second scan the patient then underwent a near complete resection of the right lobe of liver so as to remove the residual mass and all affected liver tissues. In the third and fourth scans the right lobe of liver had partially regenerated and the most similar segment of liver tissue was used for VOI placement.

Figures 4.18 - 4.20 show boxplots of the distribution of ^{123}I -MIBG uptake measured in VOIs placed in the right lobe of liver, left lobe of liver and in the brain for each of the 50 individual scans. Figure 4.18 shows the absolute activity concentration measured in kBq/ml; Figure 4.19 shows SUV/BW and Figure 4.20 show SUV/LBM. Boxplots are grouped by patient and shown in chronological order within each patient group. The right-most distribution in each plot shows the aggregate boxplot of uptake for all 50 patient-scans.

Absolute activity concentrations in the right and left lobes of liver, shown in Figure 4.18, ranged over an order of magnitude across the dataset with aggregate ranges of 1.2-9.2 kBq/ml and 1.4-10.8 kBq/ml respectively. Visually, uptake within the right lobe appeared generally lower than in the left lobe, an observation reflected in the slightly reduced aggregate median of 2.9 kBq/ml for the right lobe compared to 4.0 kBq/ml for the left. Both lobes of liver also tended to show concordance in the changes in uptake from scan to scan, with both left and right lobes showing high or low uptake for the same scan. Right liver uptake appeared more consistent than in the left, with only two scans showing median uptake above 6kBq/ml (the single scans of patients 4 and 8) both of which were the scans of the youngest patients in the dataset and 2/3 cases of INSS stage 4S Neuroblastoma included in the dataset. In contrast, the left lobe showed 10 scans from 6 patients with median uptake above

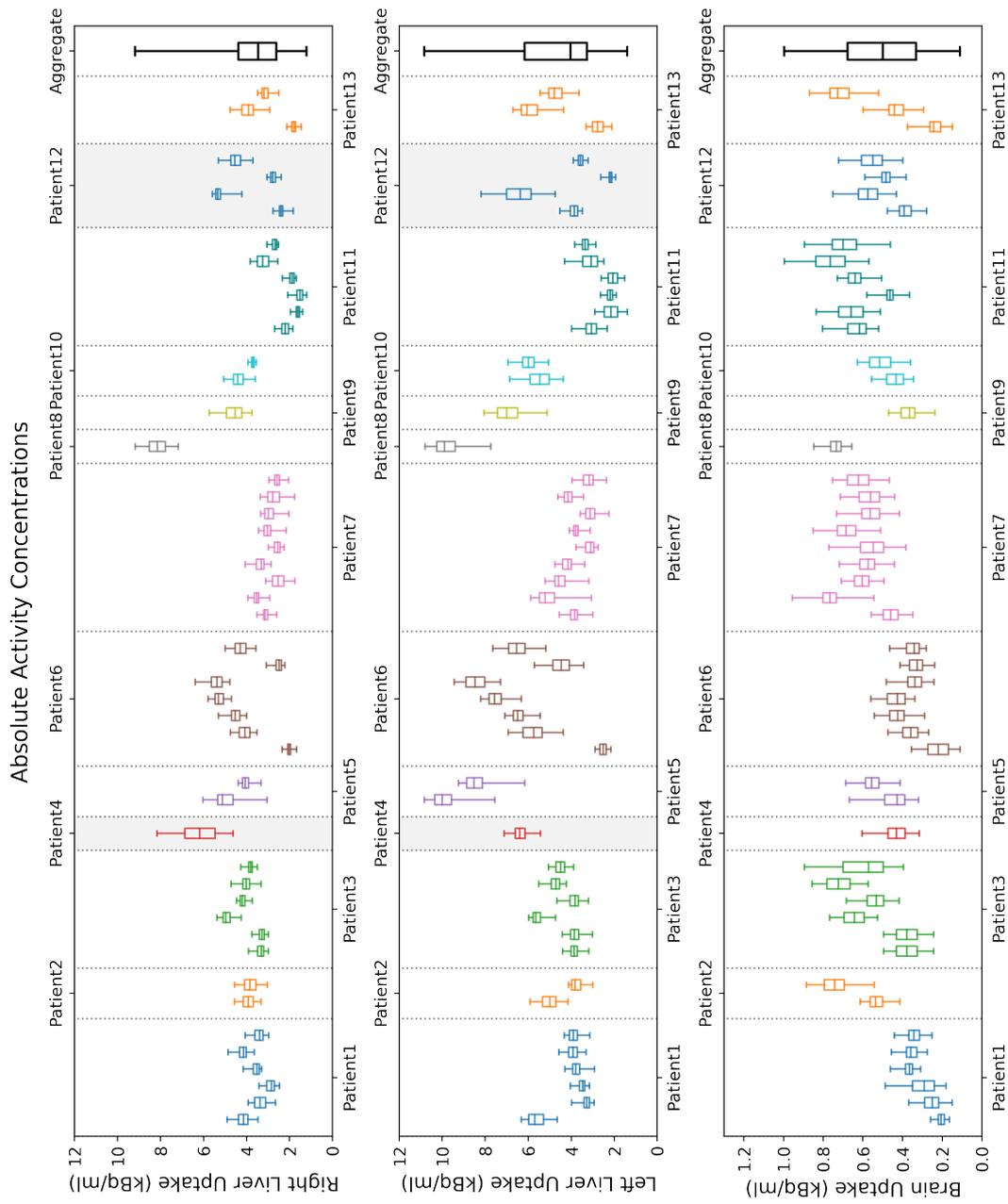


Figure 4.18: Distribution of ^{123}I -meta-Iodobenzylguanidine (MIBG) absolute activity concentration (kBq/ml) in 22mm VOIs positioned in right lobe of liver (upper plot), left lobe of liver (middle plot) and brain (lower plot) shown for each of the individual 50 scans, grouped by patient and shown in chronological order within each patient group. The final (right-most) boxplot shows the aggregate distribution of all scans combined. Grey shading indicates VOI measurements which may be compromised due to surgical resection or local presence of metastases.

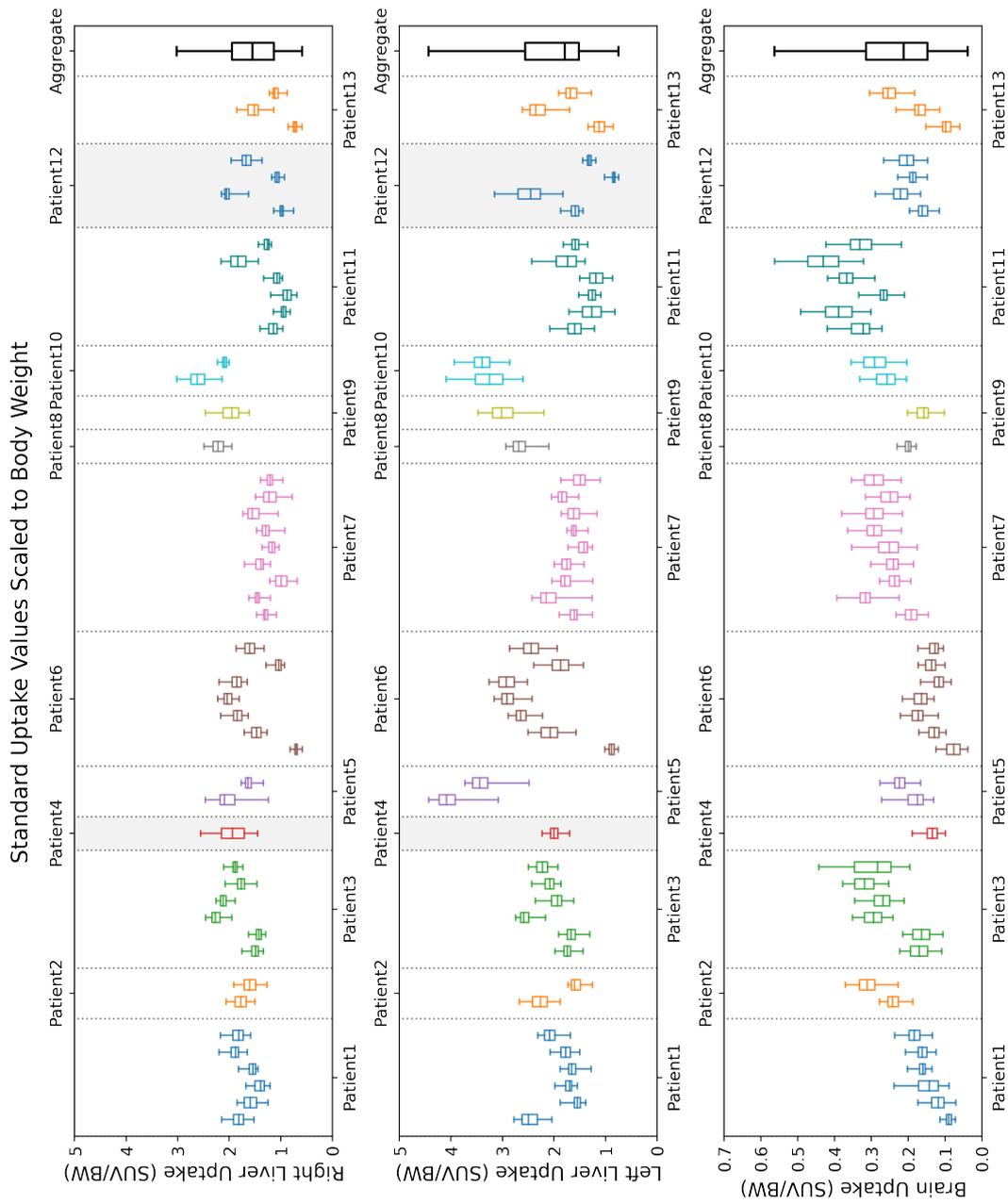


Figure 4.19: Distributions ^{123}I -meta-iodobenzylguanidine (MIBG) standard uptake values scaled to body weight in 22mm VOIs positioned in liver (upper plot) and brain (lower plot). Each boxplot shows the distribution within a single scan, grouped by patient and shown in chronological order within each patient group. The final (right-most) boxplot shows the aggregate distribution of all scans combined. Grey shading indicates volume of interest (VOI) measurements which may be compromised due to surgical resection or local presence of metastases.

Standard Uptake Values Scaled to Lean Body Mass

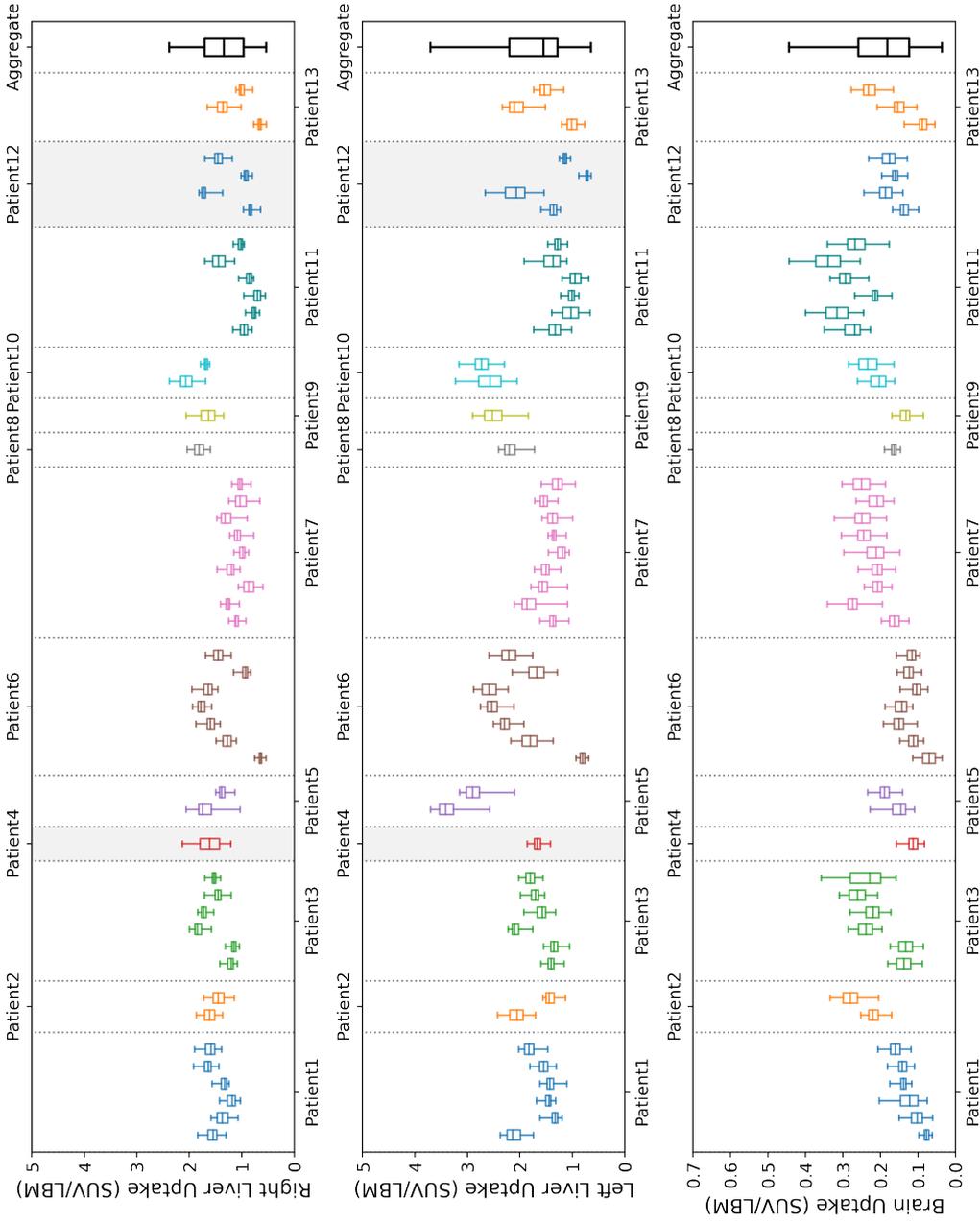


Figure 4.20: Distributions ^{123}I -meta-iodobenzylguanidine (MIBG) standard uptake values scaled to lean body mass in 22mm VOIs positioned in liver (upper plot) and brain (lower plot). Each boxplot shows the distribution within a single scan, grouped by patient and shown in chronological order within each patient group. The final (right-most) boxplot shows the aggregate distribution of all scans combined. Grey shading indicates VOI measurements which may be compromised due to surgical resection or local presence of metastases.

6kBq/ml, including those of patients 4 and 8, all scans of patient 5 and 9 (three total) and a subset of five total scans from patients 6 and 12. These differences could also be observed in the aggregate distributions for each lobe, as the left lobe showed a broader interquartile range (IQR) than the right lobe in both absolute (2.9 vs 1.8 kBq/ml) and relative (72% vs 51% of median) measures; combined with strong visual evidence of skew in both the upwards biased IQR and long upper tail. The right lobe showed no signs of skew other than a somewhat extended upper tail on account of the single outlying scan of patient 8.

For most patients, both right and left lobes of liver exhibited a fairly consistent level of uptake across all scans, although left lobe uptake was typically greater than the right. However patients 6 and 12 showed a much larger degree of variability in activity concentration from scan to scan. Of these, patient 6 underwent their first scan at 10 months of age and their final scan was performed at just under 3 years old, so the increase in uptake through scans 1-5 accompanied a similar rise in administered activity as the patient grew and increased in body weight. The reduced uptake shown in scan 6 corresponded to a slightly reduced administration activity (113 vs 128 MBq for scan 5) as a result of the patient weight falling precisely between two weight based scaling factors under the ARSAC Notes for Guidance rendering either factor a reasonable compromise. In the administration for scan 6, the staff member involved used the lower of the two factors, while the higher factor was used by different staff for scans 5 and 7. As discussed above, and shown in Figure 4.17, the increased variability in uptake in patient 12 was most likely the result of the liver undergoing substantial changes from scan to scan; from compression and infiltration in scan 1, to intense systemic treatment and decompression in scan 2, and then resection and regrowth in scans 3 and 4.

After rescaling to SUV/BW, as shown in Figure 4.19, both right and left lobes of liver showed a reduction in the total variation of uptake across the dataset. Visually, the appearances were of increased uptake for previously low activity concentrations and reduced uptake for the previously high activity concentrations; though for the left lobe patients 5 and 9 remained elevated and patient 10 also emerged as a high uptake case under SUV/BW scaling. These visual changes were reflected in the reduction of aggregate medians to 1.5 and 1.8 SUV/BW for right and left lobes respectively,

and the reduction of aggregate ranges to 0.6-3.0 and 0.8-4.4 SUV/BW respectively. Combining these measurements showed a contraction of the range:median ratio, down from 2.3:1 for both lobes under kBq/ml scaling to 1.6:1 for the right lobe and 2.1:1 for the left when rescaled to SUV/BW. This smaller relative change for the left lobe corroborates the visual impression that right lobe uptake was normalised much more effectively than the left lobe under SUV/BW scaling. The aggregate distribution for the left lobe retained clear signs of positive skew in both the IQR and the long upper tail, while the right lobe reduced in skew, remaining unskewed across its IQR and showing a reduced upper tail compared to absolute scaling.

Rescaling to SUV/LBM, shown in Figure 4.19, further narrowed the variation of uptake measurements in both right and left lobes across the dataset compared to SUV/BW scaling. Visually, the impressions were of a continuation of the VOI uptake changes observed in the switch from kBq/ml to SUV/BW scaling. For the right lobe, all cases appeared to be evenly distributed across a relatively narrow range, while the left lobe retained some some upwardly biased results in patient 5, 9 and 10 but also scans 3-5 of patient 6. These continued changes were reflected in the aggregate distributions for right and left lobes, with overall medians reducing to 1.3 and 1.6 SUV/LBM respectively; aggregate ranges also reducing to 0.5-2.4 SUV/LBM for the right lobe and 0.6-3.7 SUV/LBM for the left; and the left lobe distribution retaining visual evidence of positive skew in IQR and upper tail, while the right lobe distribution showed no visual signs of skew. Range:median ratios were also further reduced to 1.4:1 for the right lobe and 2.0:1 for the left; confirming the visual impression that the left lobe remained relatively more variable than the right under SUV/LBM scaling.

Activity concentrations measured in the brain VOIs were approximately an order of magnitude lower than those measured in the liver, with an aggregate range of 0.1-1.0 kBq/ml and a median of 0.5 kBq/ml. In contrast to both lobes of liver, no outlier cases were observed and the aggregate distribution showed no evidence of skew in either IQR or tails. Under SUV/BW scaling the distribution of brain uptake became visibly more variable across the cohort, and the aggregate distribution gained a degree of positive skew, despite the aggregate range and median reducing to 0.04-0.6 SUV/BW and 0.2 SUV/BW respectively. However the range:median ratio corroborated this observation, increasing from 1.8:1 for kBq/ml to 2.5:1 for SUV/BW. SUV/LBM

scaling slightly improved on these results, with the variability visibly and aggregate skew visibly reducing. The upper end of the range also decreased to 0.4 SUV/LBM, though the minimum and median remained static. This reduced the range:median ratio of 2.3:1, numerically confirming the visual impression of a slight improvement to inter-patient and intra-patient variability compared to SUV/BW.

Figure 4.21 shows scatter plots of mean uptake in right liver, left liver and brain VOIs plotted vs uptake time between administration and scan for each of the three units of interest (kBq/ml, SUV/BW and SUV/LBM). The single scan with an uptake time less than 22 hours was found to be amongst the largest absolute activity concentrations measured in all three VOIs but on SUV/BW or SUV/LBM scaling this scan fell within the middle to upper-middle of the uptake distribution, resulting in a fairly random distributed of means. No convincing trends were identified for any combination of VOI or measurement scale. Figure 4.22 shows the same uptake data plotted vs administered activity decay corrected to the time of the scan for each unit of interest. Some evidence of uptake trends were observed for left and right liver uptake when measured in kBq/ml but these were eliminated when the uptake data were rescaled to either SUV/BW or SUV/LBM. The brain VOI showed the opposite behaviour; absolute activity measured in kBq/ml showed no strong evidence of trend with administered activity while both SUV scales introduced an obvious increasing trend to the uptake data.

Figure 4.23 shows histograms of VOI standard deviations in right liver, left liver and brain for each measurement scale. All distributions showed a strong positive skew and were log-transformed to generate normal distributions prior to the calculation of the displayed means and 95% confidence limits. While all three VOIs showed clear changes in distribution in the change from kBq/ml to SUV/BW, distributions remained very similar under SUV/BW and SUV/LBM scales. Compared to the right lobe of liver, the left lobe showed a broader distribution and therefore a higher mean and wider confidence interval on all three measurement scales.

Figure 4.24 shows scatter plots of mean uptake in the left lobe of liver (upper row) and left-right paired mean differences (lower row) plotted against the mean uptake in the right lobe of liver for each of the 50 individual scans and each of kBq/ml, SUV/BW

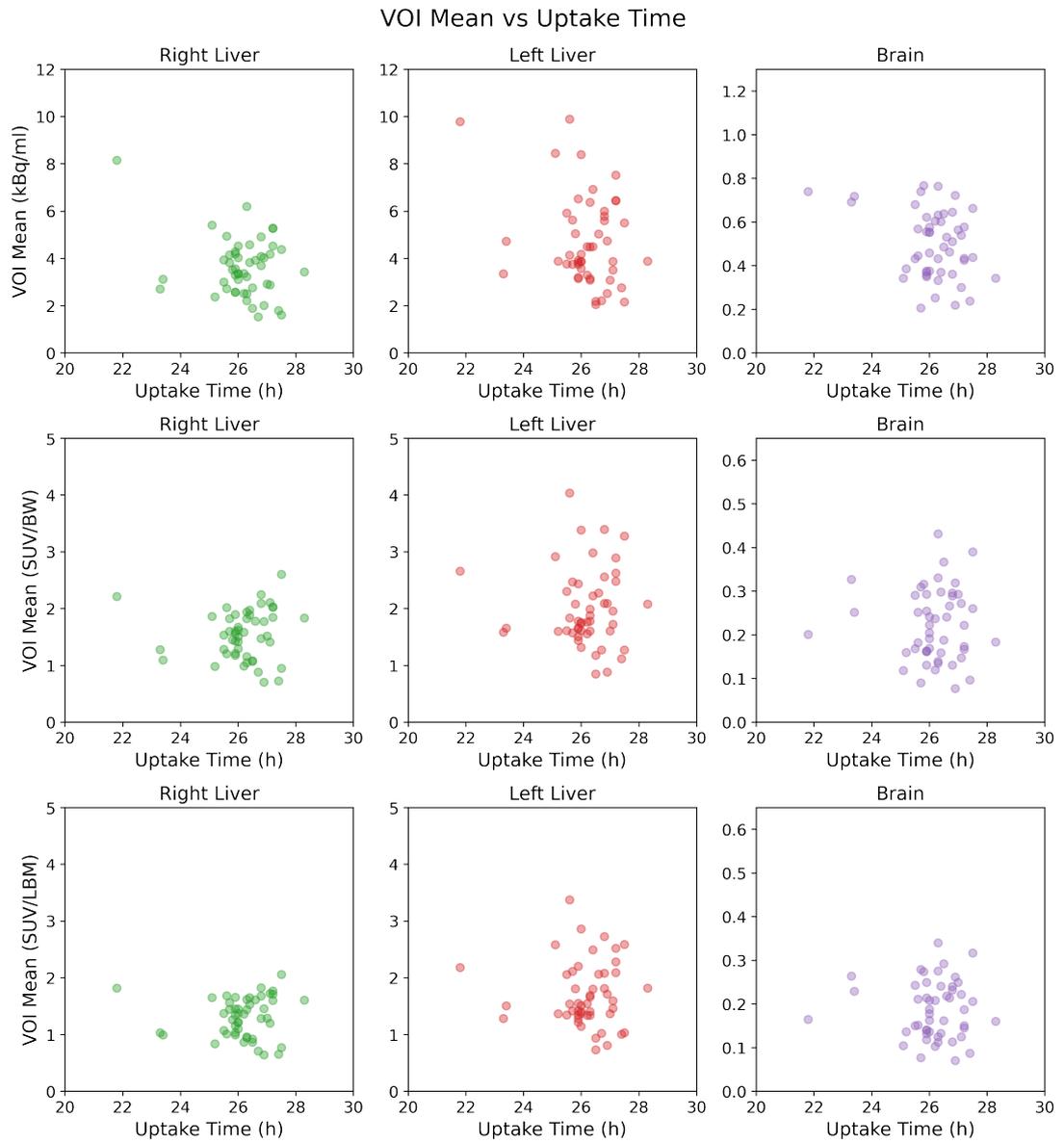


Figure 4.21: Scatter plots of mean uptake in VOIs placed in right liver (left column), left liver (middle column) and brain (right column) plotted vs uptake time between administration and scan. The upper row shows mean uptake scaled to kBq/ml, the middle row shows mean uptake scaled to SUV/BW and the lower row shows mean uptake scaled to SUV/LBM.

VOI Mean vs Decay Corrected Activity

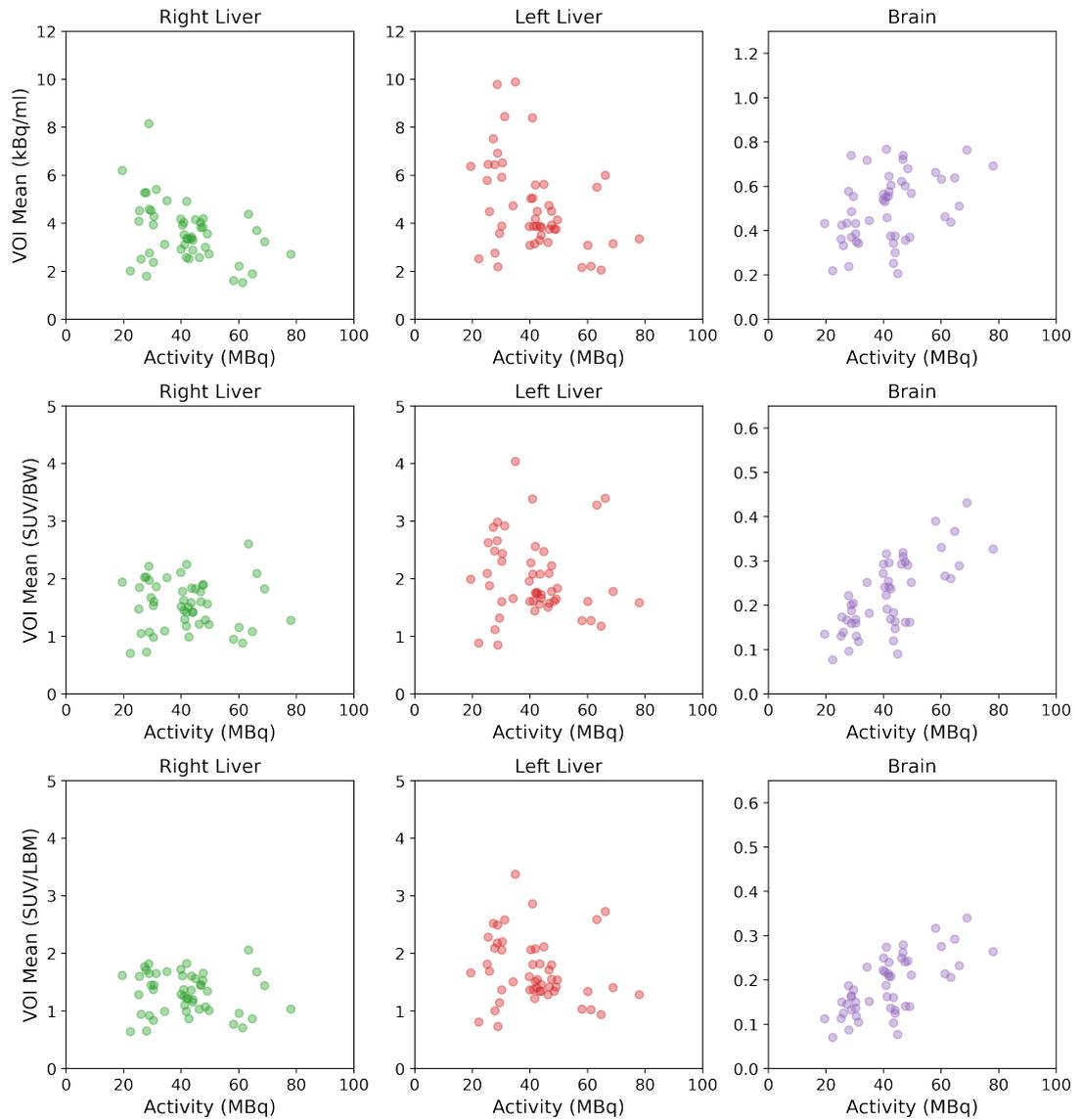


Figure 4.22: Scatter plots of mean uptake in VOIs placed in right liver (left column), left liver (middle column) and brain (right column) plotted vs administered activity decay corrected to the time of scan. The upper row shows mean uptake scaled to kBq/ml, the middle row shows mean uptake scaled to SUV/BW and the lower row shows mean uptake scaled to SUV/LBM.

VOI Standard Deviations

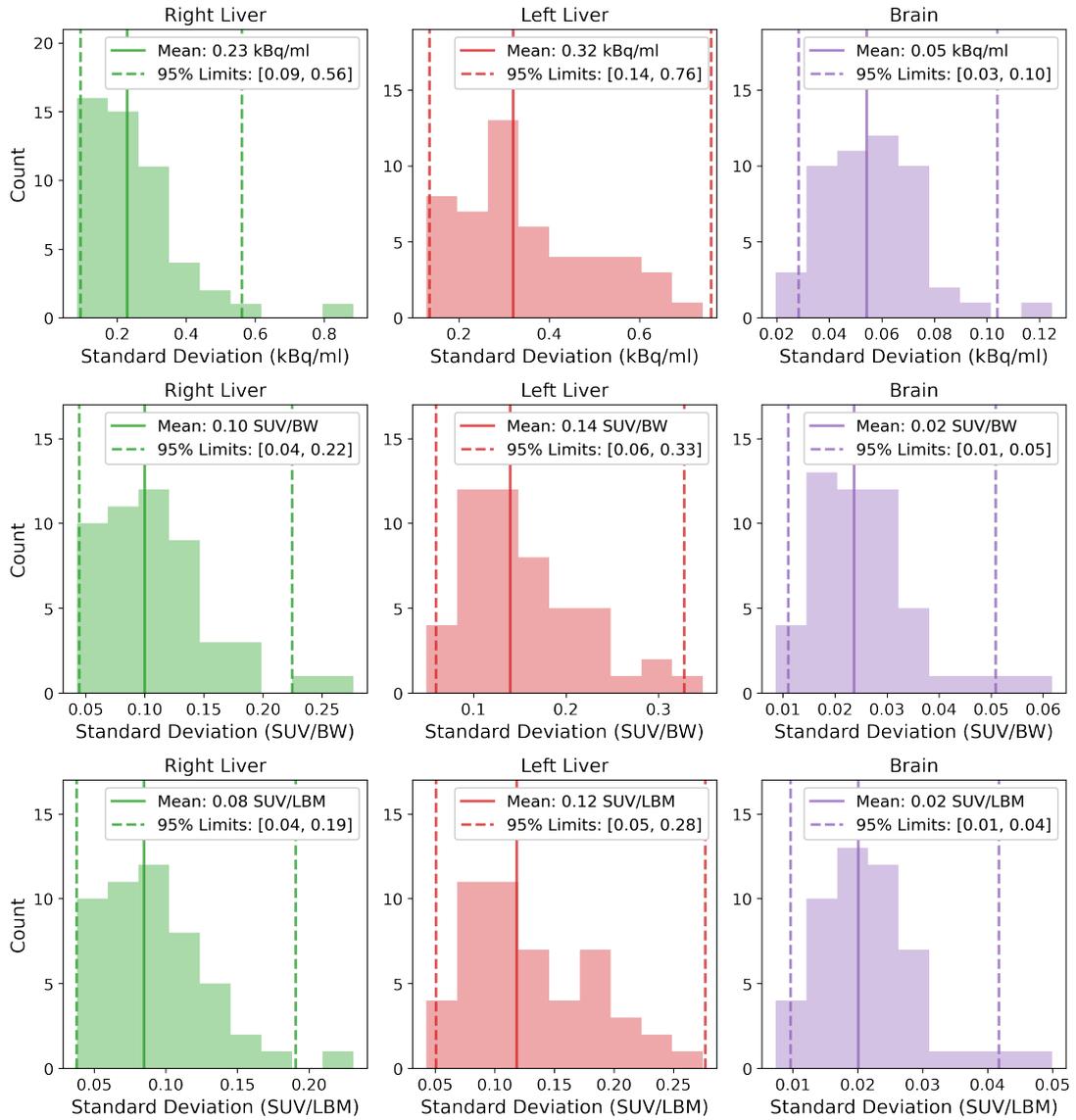


Figure 4.23: Histograms of the standard deviations of uptake measured in VOIs placed in right liver (left column), left liver (middle column) and brain (right column) across all 50 scans. Upper row shows VOI standard deviations scaled to kBq/ml, the middle row shows mean uptake scaled to SUV/BW and the lower row shows mean uptake scaled to SUV/LBM. Each plot shows the mean and 95% limits of measurement derived from the same data log-transformed into normal distributions.

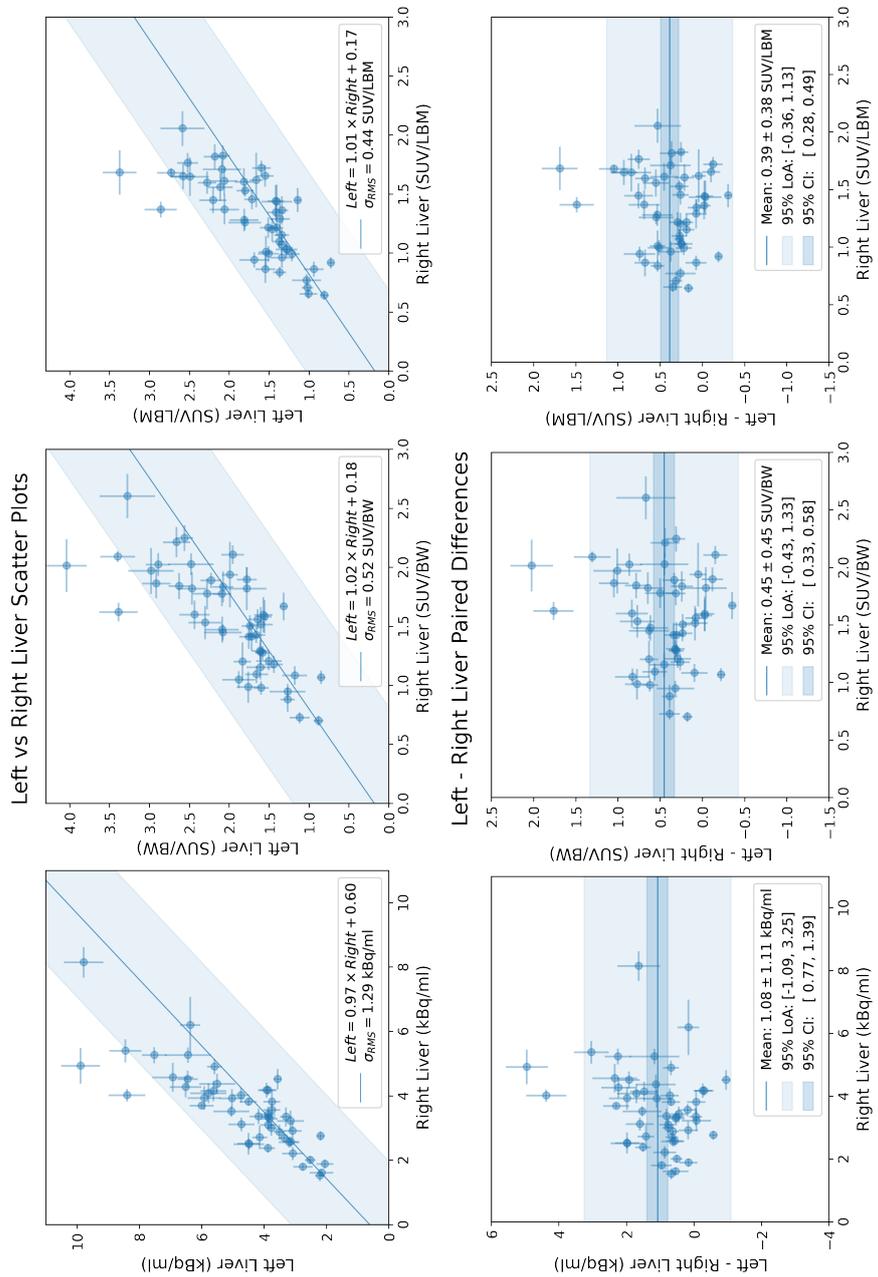


Figure 4.24: Scatter plots of mean uptake in the left lobe of liver (upper row) and left - right paired mean differences (lower row) plotted against mean uptake in the right lobe of liver for each of 50 scans. Plots show the data measured in either kBq/ml (left column), SUV/BW (middle column) and SUV/LBM (right column). Error bars represent either VOI standard deviations or error-propagated VOI standard deviations as appropriate. The solid lines indicate lines of best fit (upper plots) or lines of mean difference (lower plots) while shaded regions represent the 95% confidence band of the straight line fit (upper plots) or the 95% limits of agreement (lower plots).

and SUV/LBM measurement scales. For all measurement scales, left lobe uptake was strongly correlated with right lobe uptake, with the majority of datapoints distributed close to the line of equality, and an offset of 0.6 kBq/ml or 0.2 score units when measured using either SUV scale. Other than 2-3 outliers, all datapoints were located within the 95% confidence range of the line of best fit, generated using the root mean squared deviation from the fitted straight line (σ_{RMS}). Paired difference plots showed no clear evidence of correlation between paired differences and uptake in the right lobe of liver. However, both SUV scales showed a more even dispersion pattern than kBq/ml and were therefore more convincingly uncorrelated. Mean paired differences between left and right lobes were calculated as 1.08 kBq/ml, 0.45 SUV/BW and 0.39 SUV/LBM. While the 95% limits of agreement included zero difference, t-tests of paired differences showed statistical significance at the 99.9% level ($p < 0.001$) for all three units of measurement.

Table 4.5 shows VOI measurements in the liver and brain grouped and aggregated by patient using the weightings provided by Equations 4.12 and 4.13. Consistent with visual observations at the individual scan level, weighted means measured in left and right lobes of liver showed a wider range of values when scaled in kBq/ml but considerably narrowed following rescaling to SUV/BW and then additionally narrowed when rescaled to SUV/LBM. The weighted means also showed greater uptake in the left lobe of liver than in the right for all patients. Intra-patient paired mean differences between the left and right lobes were found to be 1.3 kBq/ml, 0.6 SUV/BW and 0.5 SUV/LBM. Paired t-tests found these differences to be statistically significant at the 99% level ($p < 0.01$) for all three measurement scales. The final two rows of the table show the patient groups combined into a full cohort aggregate mean, and a full cohort aggregate mean of only those cases without liver metastases and/or liver surgery.

4.4.4 Discussion

It was extremely encouraging to find that only one case (patient 12) posed any real difficulty for the consistent placement of VOIs within the liver, that being the only case undergoing significant liver resection concurrent with imaging. A 22mm diameter VOI was found to be practical for use in patients down to 4 months of age and would be suitable for use in even smaller livers, albeit with less surrounding tissue. This size

Patient	kBq/ml $[\mu \pm \sigma]$			SUV/BW $[\mu \pm \sigma]$			SUV/LBM $[\mu \pm \sigma]$		
	Left Liver	Right Liver	Brain	Left Liver	Right Liver	Brain	Left Liver	Right Liver	Brain
1	4.0 ± 0.8	3.6 ± 0.5	0.31 ± 0.07	1.9 ± 0.3	1.7 ± 0.2	0.14 ± 0.04	1.6 ± 0.3	1.5 ± 0.2	0.12 ± 0.03
2	4.4 ± 0.7	3.9 ± 0.3	0.64 ± 0.12	1.9 ± 0.4	1.7 ± 0.2	0.28 ± 0.04	1.7 ± 0.4	1.5 ± 0.2	0.25 ± 0.04
3	4.4 ± 0.7	3.9 ± 0.6	0.55 ± 0.15	2.1 ± 0.3	1.8 ± 0.3	0.26 ± 0.07	1.7 ± 0.3	1.5 ± 0.3	0.21 ± 0.06
4*	6.4 ± 0.3	6.2 ± 0.9	0.43 ± 0.06	2.0 ± 0.1	1.9 ± 0.3	0.14 ± 0.02	1.7 ± 0.1	1.6 ± 0.2	0.11 ± 0.02
5	9.2 ± 1.0	4.5 ± 0.6	0.50 ± 0.08	3.7 ± 0.4	1.8 ± 0.3	0.20 ± 0.03	3.1 ± 0.3	1.5 ± 0.2	0.17 ± 0.03
6	5.9 ± 1.9	4.0 ± 1.2	0.35 ± 0.08	2.2 ± 0.7	1.5 ± 0.5	0.13 ± 0.04	2.0 ± 0.6	1.3 ± 0.4	0.12 ± 0.03
7	3.9 ± 0.7	2.9 ± 0.4	0.60 ± 0.10	1.7 ± 0.2	1.3 ± 0.2	0.26 ± 0.05	1.4 ± 0.2	1.1 ± 0.2	0.22 ± 0.04
8	9.8 ± 0.6	8.2 ± 0.5	0.74 ± 0.04	2.7 ± 0.2	2.2 ± 0.1	0.20 ± 0.01	2.2 ± 0.1	1.8 ± 0.1	0.16 ± 0.01
9	6.9 ± 0.7	4.6 ± 0.5	0.37 ± 0.05	3.0 ± 0.3	2.0 ± 0.2	0.16 ± 0.02	2.5 ± 0.2	1.7 ± 0.2	0.13 ± 0.02
10	5.8 ± 0.6	4.0 ± 0.4	0.47 ± 0.07	3.3 ± 0.3	2.3 ± 0.3	0.27 ± 0.04	2.7 ± 0.2	1.9 ± 0.2	0.22 ± 0.03
11	2.7 ± 0.6	2.2 ± 0.6	0.64 ± 0.11	1.5 ± 0.3	1.2 ± 0.3	0.35 ± 0.06	1.2 ± 0.2	1.0 ± 0.3	0.28 ± 0.05
12*	4.0 ± 1.6	3.7 ± 1.2	0.50 ± 0.09	1.6 ± 0.6	1.4 ± 0.4	0.19 ± 0.03	1.3 ± 0.5	1.2 ± 0.4	0.17 ± 0.03
13	4.5 ± 1.4	2.9 ± 0.9	0.45 ± 0.20	1.7 ± 0.5	1.1 ± 0.3	0.17 ± 0.07	1.5 ± 0.5	1.0 ± 0.3	0.15 ± 0.06
Full Cohort	4.7 ± 1.9	3.6 ± 1.2	0.50 ± 0.16	2.0 ± 0.7	1.5 ± 0.4	0.22 ± 0.09	1.7 ± 0.6	1.3 ± 0.4	0.19 ± 0.07
* Excluded	4.7 ± 1.9	3.5 ± 1.2		2.0 ± 0.7	1.6 ± 0.4		1.7 ± 0.6	1.3 ± 0.4	

Table 4.5: volume of interest (VOI) measurements grouped by patient and combined using the weightings of Equations 4.12 and 4.13. Each row records the mean \pm standard deviation measured in the 22mm VOIs placed in left and right lobe of liver and in the brain using each of the three units of interest (kBq/ml, SUV/BW and SUV/LBM). The final two rows of the table show the patient groups combined into a full cohort measurement, and a full cohort measurement less those cases with known liver metastases and/or liver surgery (exclusion highlighted with *)

of VOI was also straightforward to place in cases with large abdominal masses where the liver was markedly compressed. In such instances a larger VOI, such as those described in qPET (Hasenclever et al. 2014) and PERCIST (O et al. 2016), would not be able to be positioned fully within the available liver tissue. This is particularly more of a concern for VOI placement in the left lobe which, being smaller, offers less volume even in anatomically normal cases. As Neuroblastoma can present at birth and commonly presents with very large abdominal masses, the flexibility to accommodate a wide range of patient sizes and clinical conditions is essential to the development of a robust quantitative methodology for imaging this disease.

The uptake times recorded for all but three scans were outwith the 20-24 hours recommended in current EANM guidelines for ^{123}I -MIBG scintigraphy of Neuroblastoma. This increased uptake time typically resulted from a combination of factors outwith departmental control. Primarily the long duration of wholebody scintigraphy (typically 20-35 minutes at 5cm/min) and multi-FOV SPECT/CT (25-30 mins per FOV) increased the reference time of the reconstructed SPECT scans by 1.5 to 3 hours after the actual arrival time of the anaesthetised patient, since the reference time of the reconstructed multi-FOV SPECT was set to that of the final SPECT acquisition. Delays to arrival time were also very common as MIBG imaging was the third step in the clinical process (Anaesthetic induction, MRI and then MIBG) and therefore any delay encountered during anaesthetic induction or during same-day MRI scanning (e.g. additional scans required) would also delay patient arrival time at Nuclear Medicine. Occasionally additional delays were introduced after the arrival of the patient in the nuclear medicine department, most commonly to allow for bladder catheterisation and emptying following visualisation of a very large high uptake urinary bladder on wholebody scintigraphy. Despite the increased uptake times Figure 4.21 showed no strong relationship between uptake time and liver or brain uptake within this dataset. These observations suggest that it is sufficient to simply acquire MIBG scans in a consistent 4-5 hour uptake time window in order to obtain good agreement of physiological uptake between serial scans. With only three cases in the 20-24 hour uptake time window, these data are insufficient to make a statistical comparison between EANM recommendations and the majority of local practice. However for new patients, alternative administration scheduling will be explored to improve conformity with EANM recommendations.

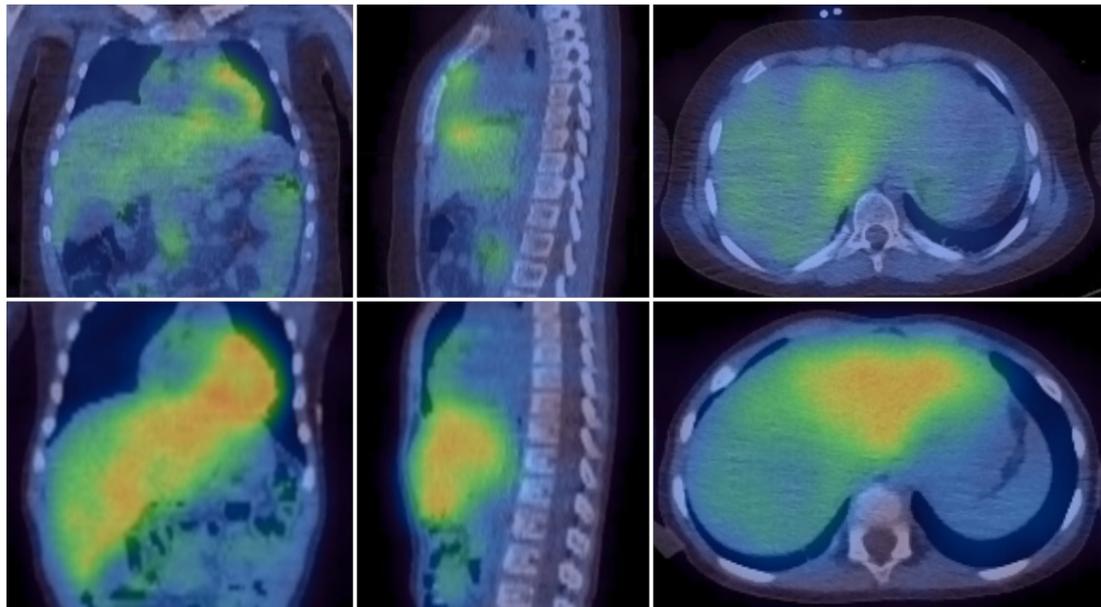


Figure 4.25: Upper Row: Balanced background liver uptake in both lobes of Patient 11. Lower Row: Intense diffuse uptake observed in the left lobe of liver of Patient 5.

Given the large range of patient ages, and therefore patient sizes in this study, it was anticipated that rescaling from units of absolute activity concentration (kBq/ml) to SUVs using either body weight or lean body mass would improve consistency across the dataset. As shown in Figure 4.22, some residual dependency on injected activity (itself dependent on patient size) remained in the absolute concentration measurements for right and left lobes of liver but this was resolved by the use of SUV units. It was initially quite surprising to observe increased variability for the brain VOI under SUV scaling. However, since the brain is a "cold" organ with no active or passive uptake mechanisms (MIBG cannot pass through an intact blood-brain barrier) these VOI measurements were essentially sampling the ambient "reconstruction noise" inside the patient. This apparent "uptake" was therefore already uncorrelated with the administered activity or patient size when measured in absolute units, and rescaling to SUV only acted to introduce these dependencies to the data. This effect can clearly be seen in the scatter plots of Figure 4.22 and is also visible in the trends of same-patient measurements in Figures 4.19 and 4.20; where later scans when the patients were larger and therefore received more activity were relatively increased compared to Figure 4.18. Overall these effects would seem to reduce the quantitative value of SUVs measured in the brain but absolute concentration measurements of this "cold background" could be used as a quality control check for SPECT/CT that includes the head.

SUV/LBM offered superior intra- and inter-patient variability for both lobes of liver compared to SUV/BW, indicated by tighter same-patient scan distributions and narrower aggregate distributions in Figure 4.20 compared to Figures 4.18 and 4.19. However neither SUV scale was able to eliminate the difference in uptake between the two lobes, which was found to be statistically significant both for the cohort of individual scans and the cohort of combined patient group means. While there was an obvious correlation between the uptake in the two lobes, indicative of the shared uptake mechanisms across the liver; the lack of correlation between the paired differences in lobar uptake and that of the right lobe showed that the difference in left lobe uptake arose through an additional, independent effect. This would be consistent with the persistent congenital differences in functional uptake described by Jacobsson et al. (2007), who first identified small but statistically significant differences in lobar uptake on ^{123}I -MIBG SPECT.

In the current study, the lobar differences remained small in the majority of cases, with mean differences of +0.45 SUV/BW and +0.39 SUV/LBM respectively. However there were two clear outlier scans with much larger lobar differences shown in Figure 4.24. Both of these outlier scans were of patient 5 and both demonstrated lobar differences of 1.8-2.0 SUV/BW and 1.5-1.7 SUV/LBM, which corresponded to left/right liver ratios of 2.0-2.1. These results were not consistent with the relative uptake ratios quoted in Jacobsson et al. (2007) (1.26 ± 0.12) and demanded an alternative explanation. On review, both scans showed intense diffuse uptake within the left lobe of liver. Images of this uptake pattern are shown in Figure 4.25 in comparison to images of patient 11 which shows balanced background liver uptake across both lobes. It was found that this pattern of uptake had previously been noted in non-quantitative SPECT/CT and presented in a published abstract (Bar-Sever et al. 2011). In that work the authors found that diffuse or focal uptake in the left lobe of liver was a normal variant in their dataset, as 64% (34/53) of presented scans demonstrated one of these patterns of uptake, but in only one case were liver metastases later proven.

Following this, the study data were re-assessed using similar criteria to that of Bar-Sever et al. (2011). Diffuse or focal uptake within the left lobe of liver was identified in 40% (20/50) of scans, while 54% (27/50) of scans showed balanced uptake across

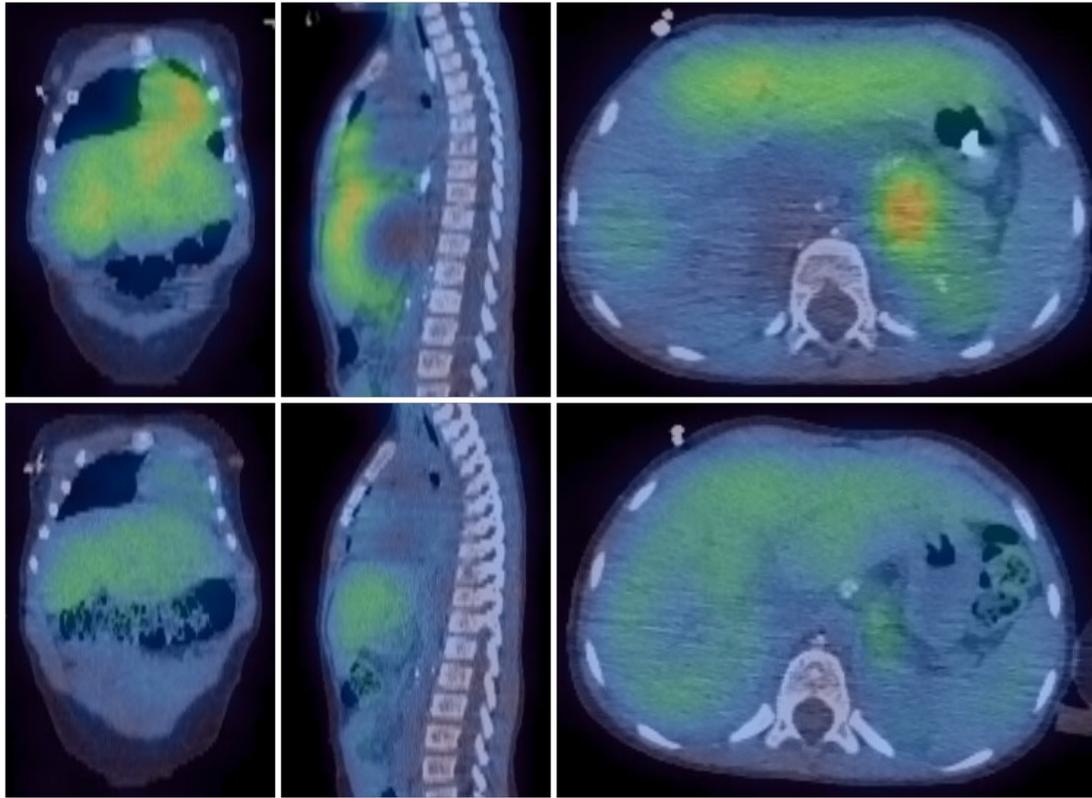


Figure 4.26: Upper Row: Increased diffuse uptake in the left lobe of liver in scan 3 of patient 7, under conditions of liver compression by a non-meta-iodobenzylguanidine (MIBG) avid central mass. Lower Row: Return to balanced uptake across both lobes in scan 4 of patient 7, following significant shrinkage of the mass with systemic therapy.

both lobes. The remaining 3 scans of patients 4 and 12 showed multiple foci of uptake consistent with the known liver metastases. Outside of these three cases with metastases, the uptake pattern in the liver appeared to be entirely consistent across all same-patient scans except in a small number of instances where significant anatomical or pathological change affected the liver. For example, compression of the left lobe of liver by a central mass was observed in scans 2 and 3 of patient 7 and this resulted in an increased visual intensity of uptake in the left lobe as shown in Figure 4.26. Figures 4.18-4.20 showed these scans to have increased median uptake in the left lobe compared to the right, consistent with visual appearances.

These observations confirm the normal variant status of the left lobe diffuse uptake pattern within the local population of RHC but also would seem to suggest another factor contributing to the much larger differences in lobar uptake observed in patient 5. The clinical history of this case showed an instances of stage 4S disease and a history of multifocal liver metastases at diagnosis. While no liver metastases were observed at the timepoints of the included scans, residual subclinical disease cells

within the liver cannot be excluded. There may also have been an impact on the development of liver vasculature, due to the disease or treatment at such an early stage of life (stage 4S disease must be identified at under 1 year of age) but with only a single example, such speculation is beyond the scope of this study. Future investigation of patients with high differential lobar uptake would seem prudent to establish potential causal factors.

While the left lobe showed a range of uptake patterns, the right lobe of liver showed consistent visual appearances across the cohort. When measured in SUV/BW, the cohort means for the right lobe of liver ($1.5-1.6 \pm 0.4$ SUV/BW) showed excellent agreement with the mean SUV/BW values for liver uptake published by Brady and Shulkin (2019) for both the GE Infinia with medium energy collimator (1.5 ± 0.6), and Siemens Intevo with LEHR collimator (1.6 ± 0.5). Assuming that the right lobe of liver was used consistently in Brady and Shulkin (2019), the current study would represent a very encouraging replication of results across departments on different continents. With this consistency of measurement, despite a multitude of potential confounders, it would seem very reasonable to propose the medial-posterior right lobe of liver as a reliable physiological reference location for Neuroblastoma, at least in cases with no history of liver metastases or liver surgery. While the left lobe of liver showed greater inter-patient variability in this study, it remained quite consistent in uptake across same-patient scans and could conceivably serve a role as an alternative reference VOI in cases where the right lobe was otherwise compromised.

The use of SUV/LBM would also be recommended for Neuroblastoma based on the results of this study, as this unit scale provided the most consistent and most normally distributed right liver uptake across the cohort. The method used in this study for estimation of lean body mass, taken from Peters et al. (2011), is only appropriate for pre-pubescent children and identification of a suitable cut-point to the sex-specific adult equations is required. The paper authors suggest the use of either an age-based transition at 13-14 years or a weight-based cut-point at approximately 50kg but do not make a definitive recommendation. Assuming a suitable cut-point is identified, this LBM scaling would be expected to offer superior consistency of SUVs through adolescence and adulthood, when MIBG non-avid white fat deposits typically increase, as MIBG like FDG does not accumulate in white fat.

The 95% confidence interval derived from cohort means and standard deviations shown in Table 4.5 places the expected range of right liver uptake at 0.5-2.1 SUV/LBM. Similarly, the 95% limits shown in Figure 4.23 suggest that the standard deviation within the right lobe of liver should not be expected to exceed 0.2 SUV/LBM for a 22mm diameter spherical VOI. Single scan right liver means or standard deviations outside these limits may indicate compromised scan quality, or the presence of pathology within the liver, and would certainly be reasonable grounds to investigate for potential confounding factors.

Intra-patient coefficients of variation, derived from the patient-wise group means and standard deviations in Table 4.5, ranged from 10-30% for patients with at least two scans in the dataset. At the lower end, this variability was an excellent match for the 10.8% CoV found by Paquet et al. (2004) for test-retest comparison of right liver uptake in ^{18}F -FDG PET, while at the upper end this dataset showed approximately 3x the level of variability. While some additional statistical variation should be expected in SPECT compared to PET, due to an inherently lower sensitivity, it was particularly notable that the upper limit of this range was essentially equivalent to the overall cohort CoV of 30.7%. This indicated that the variation of the most variable patients was equal to or greater than the variation between patients.

The two highest variability cases were patients 6 and 13, with CoVs of 29% and 30% respectively. Both of these patients shared an initial scan at the point of diagnosis which in both cases showed significantly lower uptake in both lobes of liver than in subsequent same-patient scans. While both scans also showed slightly reduced absolute uptake in brain VOIs in Figure 4.18, no obvious confounders for uptake (e.g. tissue injection sites) were observed in either case. Outside of these two scans, the only other scans acquired at diagnosis were the single scans of patients 4 and 8, which showed mean uptakes in the upper half of the uptake distribution. However, these cases were both cases of stage 4S disease and patient 4 had known liver metastases, so direct comparison would seem potentially spurious. On balance, the findings in patients 6 and 13 would seem to be in conflict with the consistent liver uptake presented in Brady and Shulkin (2019) and plausibly may indicate some staging-dependent impact of high disease burden on MIBG uptake on liver uptake. Based on these observations, robust intra-patient liver variation limits equivalent to

those of PERCIST cannot be recommended at this time and further investigation of liver uptake at diagnosis is indicated.

4.4.5 Conclusion

This clinical evaluation of quantitative ^{123}I -MIBG SPECT/CT revealed that the right lobe of liver was a superior physiological reference location compared to the left lobe of liver, on both a practical and quantitative basis. The cold background within the brain was also found to be very consistent across the patient sample when measured as an absolute activity concentration, with a normal range of 0.2-0.8 kBq/ml. While not of clinical interest, it may offer a potentially useful measurement of SPECT reconstruction quality.

Left liver uptake was shown to be more variable than the right lobe across the dataset regardless of unit scale, and was subject to a range of normal-variant uptake patterns. A small but statistically significant difference in mean uptake was detected between left and right lobes of liver, which was independent of the baseline level of uptake in the right lobe. Although of insufficient prevalence for statistical analysis, significantly reduced uptake in both lobes of liver was observed in 2/2 cases of stage 4 disease imaged at diagnosis but not in either case of stage 4S disease. This was suggestive of a potential staging-dependent effect of the initial disease burden on liver uptake which should be investigated further. In light of these observations, definition of intra-patient variability limits in the style of PERCIST was not considered prudent.

Despite these observations, the use of a 22mm diameter spherical VOI permitted straightforward placement in the medial right lobe of liver down to 3 months of age, even in the presence of large abdominal masses or following liver surgery. Only extensive liver involvement of Neuroblastoma prevented consistent placement of VOIs. Quantitatively, the right lobe was more reliable than the left lobe, both on an intra- and inter-patient basis. Right liver uptake measured in SUV/BW showed excellent agreement with previously published data, though SUV/LBM was shown to be a superior unit scale for consistent liver uptake measurement within same-patient scans and across the sample of 50 scans. Using the full dataset, a normal range of 0.5-2.1 SUV/LBM was defined for mean uptake measured in the medial right lobe of liver. Similarly an upper limit of 0.2 SUV/LBM was determined for the standard deviation

of right lobe uptake.

4.5 Reflection

This chapter builds a technical and clinical foundation with which reliable, truly quantitative ^{123}I -MIBG SPECT/CT of Neuroblastoma can be acquired. This work also aspires to contribute to the development of quantitative SPECT imaging throughout Europe and beyond. Expansions of this work to multiple camera manufacturers, collimators and reconstruction packages are intended with the ultimate goal of publication. It is hoped that such a publication will lead to standardisation of Neuroblastoma SPECT/CT and better enable multi-centre clinical trials of novel therapeutics for this disease.

Chapter 5

Conclusions & Future Work

5.1 Conclusions

This project has straddled a significant period of development of Nuclear Medicine imaging. At the outset of the project SPECT/CT was almost entirely absent from paediatric nuclear medicine departments and was still becoming established in the adult sector. At the time of writing hybrid SPECT/CT imaging systems have become ubiquitous throughout Nuclear Medicine and out-of-the-box absolute quantitative SPECT/CT is available from each of the major gamma camera manufacturers; albeit limited to specific isotopes. These changes to the practice of Nuclear Medicine have followed directly from the massive success and growth of PET/CT in the management of oncology patients, a rise built upon the diagnostic power of quantitative functional imaging and the utility of standardised frameworks for assessing disease burden and response. The overarching goal of this project was to improve upon the existing imaging techniques used in the management of Neuroblastoma, and ultimately to shift the direction of travel towards that of quantification and standardisation.

At the inception of this work, neuroblastoma cases were typically assessed separately on anatomical imaging and planar scintigraphy, while semi-quantitative scoring systems were only just entering use in clinical trials. Furthermore, on-site hybrid imaging was not locally available at the paediatric hospital so the research focus remained on the integration of SPECT and MRI through the simulation of MRAC (Chapter 2) and the assessment of the impact of SPECT/MR fusion on semi-quantitative Neuroblastoma scores (Chapter 3). Later, following the relocation to new hospital premises and the introduction of SPECT/CT to clinical practice, the research focus shifted to the development of absolute quantification of Neuroblastoma SPECT/CT (Chapter

4) with a view to laying the first foundations of a standardised framework.

The simulation study in Chapter 2 replicated some of the early pioneering work of the PET/MR literature for SPECT by replicating the tissue specific effects of Dixon spectroscopic MRAC from hybrid SPECT/CT images. Through this process, the study showed that even very basic SPECT/MR attenuation correction methodologies should be expected to reach the equivalent performance of PET/MR attenuation correction, in cases without significant lung coverage, providing the lower bound of lipid soft tissue was properly modelled. At the time of presenting at an international conference (Wallace et al. 2014), this simulation provided an assessment of the expected strengths and weaknesses of pseudo-CT generation from MRI as applied to SPECT. Given the technical limitations at the time, in particular the lack of comparison reference SPECT/CT at the children's hospital, the direct applications for clinical practice were deemed to be limited and the project focus was shifted to other topics. Developments in PET/MR and radiotherapy MRI planning have since delivered significant improvements to the creation of pseudo-CTs from MRI, such as machine learning-based voxel-wise pseudo-CT prediction recently proposed by Siemens (Hoesl et al. 2022). Based on the results of this early work, these modern techniques should be expected to offer superior attenuation correction performance for SPECT compared to PET.

The observer studies in Chapter 3 evaluated the novel same-day SPECT/MR fusion dataset acquired at the Royal Hospital for Sick Children in the context of semi-quantitative scoring of Neuroblastoma. Following national presentation of the clinically significant differences in interpretation and statistically significant differences scores between the use of planar and SPECT/MR fusion (Wallace et al. 2016), an expert observer was recruited to strengthen and formalise the work for publication. While statistical significance was not achieved in the follow-up expert observer study, the clinically significant observations were ultimately strengthened, with improved evaluation of additional disease sites in the face and skull being a particular point of note. Gratifyingly, following multiple presentations of this work at the ESOPNM, the most recent published European guidance on the imaging of Neuroblastoma now recommends combined interpretation of MRI and scintigraphic findings and describes SPECT or SPECT/CT as "an integral part of the ^{123}I -MIBG acquisition" (Bar-Sever

et al. 2018). For the local service, the observer dependence of score differences was taken as an example of the limitation of semi-quantitative scores, though the finding that SPECT/MR reduced score differences between observers was obviously welcome. The introduction of SPECT/CT and the development of true absolute quantification of uptake was clearly indicated by the findings of this study. As such, when SPECT/CT became available in the children's hospital, SPECT/CT quantification inevitably became the new focus of this research project.

Chapter 4 describes the development of quantitative SPECT/CT for Neuroblastoma through three distinct stages; assessment of collimator performance for Iodine-123, optimisation of SPECT/CT reconstruction for robust quantitative analysis, and evaluation of local normal ranges of ^{123}I -MIBG physiological uptake to identify criteria suitable for assessment of scan quality. In the first of these subprojects, a methodology to quantify photopeak and septal penetration components of sensitivity was developed and used to disambiguate the differences in "naive sensitivity" between Siemens low and medium energy collimators from their resulting image quality. The Siemens MELP collimator, despite having a lower "naive sensitivity" at 100mm collimator-phantom separation, was shown to have substantially reduced contribution from septal penetration and consequently a greater true sensitivity to photopeak emission than the typically recommended LEHR collimator. At the time of undertaking this work, the continued dependence on high quality planar imaging for Neuroblastoma scoring and the available corrections for SPECT reconstruction dictated the continued use of the MELP collimator for optimal clinical and quantitative imaging. Following presentation of this work at the European Symposium on Paediatric Nuclear Medicine, these findings were echoed in the latest EANM guidelines for imaging Neuroblastoma, which suggest medium energy collimators as a potential image quality improvement for individual departments to investigate (Bar-Sever et al. 2018).

In the follow-up phantom-based optimisation of quantitative SPECT performance, standard NEMA quality measurements of contrast recovery and background variability were found to be inadequate. A novel optimisation approach utilising absolute activity concentrations was developed to limit image noise to levels which permitted accurate quantification in both large background volumes and in measurements of interest in hot objects ranging from 13-37mm in size. This approach was able to

identify combinations of reconstruction parameters which provided accurate quantification for the measurement methodologies of PERCIST and/or qPET, deemed to provide the most promising starting points for any quantitative SPECT/CT system. A single combination of reconstruction parameters offered quantitative accuracy for both PET methodologies and was identified as the optimal parameter set for further clinical testing. This generalist parameter set is recommended for use in any department undertaking Neuroblastoma SPECT/CT using the Siemens MELP collimator and Hermes Medical Solutions Hybrid Recon for quantitative reconstructions, while the methodology is recommended as a reliable framework for task-specific optimisation of SPECT/CT for all manufacturers, collimators and reconstruction platforms. Prior to this investigation, the only justification for parameter selection was based on manufacturer recommendations e.g. as stated in Brady and Shulkin (2019). This work provides the means for the development of a scientific evidence base for standardised quantitative SPECT/CT across a range of applications, manufacturers and acquisition protocols.

The final clinical evaluation subproject was focussed on providing recommendations for effective standardisation of quantitative analysis of ^{123}I -MIBG SPECT/CT. At the outset of this subproject no published guidance on appropriate VOI sizes, physiological reference sites or measurement units for ^{123}I -MIBG SPECT/CT were available. The first clear recommendation from this work was the use of a 22mm diameter spherical VOI for measurement of liver uptake. Although slightly smaller than the VOI size recommended in PERCIST this was found to be well suited for use in either lobe of liver, in patients down to 3 months of age, and also in cases of significant liver compression. Given that Neuroblastoma is typically diagnosed in the first year of life, and most commonly presents with a large adrenal or central mass, this approach will allow for standardised analysis of physiological uptake across essentially the entire patient population. Lean body mass scaling of SUVs was also recommended over body weight scaled SUVs or absolute activity concentrations, as SUV/LBM provided the most consistent mean liver uptake of the three unit scales. This recommendation is consistent with the methodology of PERCIST and reflects the known lack of MIBG avidity in white fat cells. Based on the sample demographics in this investigation, the paediatric eLBM formulae of Peters et al. (2011) is recommended for use for SUV scaling in patients below 12 years of age and below 40kg. Further analysis of SUVs

in older and heavier paediatric patients is required to make further recommendations on the optimal handling of the transition from this sex-agnostic paediatric equation to adult sex-specific eLBM equations.

Finally, the right lobe of liver was recommended over the left lobe as a physiological reference site, due to a more consistent and more normally distributed uptake across the cohort. Right lobe uptake also showed excellent agreement with the results of Brady and Shulkin (2019) despite differences in administration and acquisition protocols between the centres. This provided sufficient confidence to define the first standardised quantitative ^{123}I -MIBG SPECT/CT quality control measures; a normal range for the mean uptake and an upper limit on the uptake standard deviation, both measured in the right lobe of liver. Importantly, these were both defined as single-scan reference measurements, rather than PERCIST-style limits for intra-patient variability in sequential scans. This mark of caution was the result of remaining uncertainty regarding large unexplained differences in liver uptake measured at diagnosis of two stage M patients compared to subsequent scans. These findings, consistent with a dependence of liver uptake on disease burden, were presented at the SIOPEN annual general meeting in November 2018 and at the ESOPNM in May 2019; both ahead of the publication of Brady and Shulkin (2019). Further investigation of these observations is clearly warranted but fell beyond the scope of this project.

Despite some unresolved questions in each problem domain, this research project has clearly delivered its stated goal of advancing quantitative imaging of Neuroblastoma. At each stage of the project novel contributions were made to the body of research, primarily through conference presentations, and the work undertaken at RHC was able to contribute at the forefront of the evolution of paediatric nuclear medicine. Under an initial focus on SPECT and MRI integration; pioneering work from the PET/MR literature was replicated in SPECT and the first known work on the impact of fused SPECT imaging on semi-quantitative Neuroblastoma scores was undertaken. Following a shift to a focus on quantitative SPECT/CT, collimator performance and septal penetration were analysed to a gamma-spectrum level, a novel methodology for task-specific quantitative optimisation was developed and optimised reconstruction parameters were identified, and recommendations of standardised methods and normal ranges for measurement of physiological reference uptake in the liver were assessed in

a serial patient cohort. These tools are expected to be of great assistance to centres attempting to achieve reliable quantification of MIBG scans for Neuroblastoma, and have more general applications in the wider domain of quantitative SPECT/CT.

5.2 Future Work

Due to time constraints there were several avenues of research that have been left for future investigation. The most direct of these was an expansion of the clinical evaluation in Section 4.4 to include more physiological tissues as well as primary and metastatic disease sites. For physiological tissues, this would provide additional normal ranges and recommendations for VOI sizes and measurements. For pathological sites, it would seem most fruitful to focus on the typical changes in 1ml peak uptake and/or maximum voxel uptake between diagnosis and the end of induction, consolidation and maintenance phases of treatment, as well as disease changes before and after ¹³¹I-MIBG molecular radiotherapy where undertaken. As a single centre, this would not likely provide sufficient caseload to perform direct comparison with outcome data, but such measurements would provide the first step towards adapting PERCIST and qPET style response metrics to Neuroblastoma, as well as permitting correlation of changes in MIBG uptake to lesion size changes and/or changes in semi-quantitative scores.

It would also be very useful to expand the dataset to include older non-GA patients, though this would require resolution of the transition points between the paediatric lean body mass formula of Peters et al. (2011); suitable for use in prepubescent children; and sex-specific adult formulae; such as those of Boer (1984). Assuming this can be straightforwardly resolved, further investigation of the relationship between liver uptake and MIBG-avid disease burden would be of particular interest as this remains an unresolved question from the work of Section 4.4.

Further advancement of SPECT/CT image quality should also be considered following the quantitative baseline established by this project. A major obstacle to achieving this goal is the continued requirement for high quality planar imaging to permit semi-quantitative scoring, as this necessitates very long acquisition times. A necessary first step to reducing reliance on planar imaging would be the acquisition of a proof

of concept dataset inclusive of both high quality planar imaging and full coverage SPECT/CT. Then planar-only scoring could be compared to SPECT/CT-only scoring and/or scoring of reprojected planar images. Assuming that such scoring was at least no worse than true planar imaging, a shift to a SPECT/CT-only imaging paradigm would become a reality. One of the key arguments for continued use of the Siemens MELP collimator in Section 4.2 was the need for high quality planar imaging. However, in a SPECT/CT-only paradigm, other collimators and reconstruction techniques that would permit high resolution images could be investigated in earnest.

Given the advent of clinical Cadmium-Zinc-Telluride (CZT) solid-state ring detectors (Desmonts et al. 2020), it seems very likely that a shift to a SPECT/CT-only imaging paradigm is on the horizon. These systems offer much improved sensitivity, spatial resolution and energy resolution with their main drawback being an inability to image higher energy gamma-ray emissions such as those from Iodine-131. The optimisation techniques proposed in Section 4.3 and standardisation approaches of Section 4.4 are obviously also applicable to ring detector imaging and will only become more relevant to maximising efficacy of SPECT/CT. Although the arrival of SPECT/CT to paediatric hospitals was delayed due to patient exposure concerns, the improved acquisition times and exquisite resolution of ring detectors should see some uptake in larger paediatric imaging centres with significant oncology work and multi-camera departments. RHC is positioned well to gain access to such technology relatively early and this would offer many research opportunities.

A return to a SPECT/MR integration focus of Chapters 2 and 3 may also be a useful direction of travel, now that quantitative ^{123}I -MIBG SPECT/CT has been established. Local research into this area would require the development of accurate co-registration of CT and MRI, itself a major area of research. Assuming that spatial co-registration can be achieved, a true evaluation of MRAC in Neuroblastoma imaging could be undertaken, as well as an investigation into the use of MRI soft tissue information in SPECT reconstruction to improve partial volume correction of small structures. Ultimately this may enable a case to be made for the physical integration of SPECT and MRI, perhaps with ring-detector technology. While SPECT still lags behind PET in terms of sensitivity and spatial resolution, it retains the advantage of relative insensitivity to photon attenuation which makes it ideally suited to integration with

MRI.

Finally, it seems likely that there is a future expanded role for the use of PET in the imaging of Neuroblastoma. An alternative PET formulation of MIBG using ^{18}F , meta-fluorobenzylguanidine, has been developed and small scale clinical trials are underway in Europe (Samim et al. 2022). Initial results look extremely promising, offering the tumour affinity of ^{123}I -MIBG combined with the advantages of modern PET/CT systems (well established quantification, fast full-body acquisition times, superior sensitivity and spatial resolution) with access to PET/CT being the only barrier to entry. Furthermore, the use of ^{177}Lu -dotatate molecular radiotherapy is set to expand in the management of Neuroblastoma (Sundquist et al. 2022) and this will increase demand for "theranostic" imaging with ^{68}Ga -dotatate. These clinical applications may also provide new justification for the expanded use of PET/MR, given the importance of MRI in the work-up of Neuroblastoma and the dose reduction available through the omission of CT exposure. Certainly it seems prudent that paediatric centres such as RHC ensure good access to PET in some form.

These exciting future developments underpin a broader scientific truth; regardless of the modality of choice, it is clear that quantitative imaging of neuroblastoma will continue to advance for many years to come.

Bibliography

- Ady, N. et al. (1995). “A new ^{123}I -MIBG whole body scan scoring method—Application to the prediction of the response of metastases to induction chemotherapy in stage IV neuroblastoma”. In: *European Journal of Cancer* 31.2, pp. 256–261. ISSN: 0959-8049. DOI: 10.1016/0959-8049(94)00509-4.
- Altman, D. et al. (2013). *Statistics with confidence: confidence intervals and statistical guidelines*. Second edition. Wiley. ISBN: 978-1-118-70250-5.
- Anger, H. O. (Jan. 1958). “Scintillation Camera”. In: *Review of Scientific Instruments* 29.1, pp. 27–33. ISSN: 0034-6748. DOI: 10.1063/1.1715998. URL: <https://aip.scitation.org/doi/10.1063/1.1715998> (visited on 01/31/2023).
- Bailey, D. L. and Willowson, K. P. (Sept. 14, 2013). “Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality”. In: *European Journal of Nuclear Medicine and Molecular Imaging*. ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-013-2542-4. URL: <http://rd.springer.com/article/10.1007/s00259-013-2542-4>.
- Bar-Sever, Z. et al. (2011). “Significance of MIBG liver uptake in children with neuroblastoma”. In: *Journal of Nuclear Medicine* 52 (supplement 1), pp. 1400–1400. ISSN: 0161-5505, 2159-662X. URL: https://jnm.snmjournals.org/content/52/supplement_1/1400.
- Bar-Sever, Z. et al. (2018). “Guidelines on nuclear medicine imaging in neuroblastoma”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 45.11, pp. 2009–2024. ISSN: 1619-7089. DOI: 10.1007/s00259-018-4070-8. URL: <https://doi.org/10.1007/s00259-018-4070-8>.
- Barrington, S. F. and Kluge, R. (2017). “FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 44 (Suppl 1), pp. 97–110. ISSN: 1619-7070. DOI: 10.1007/s00259-017-3690-8. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5541086/>.

- Berker, Y. et al. (2012). “MRI-Based Attenuation Correction for Hybrid PET/MRI Systems: A 4-Class Tissue Segmentation Technique Using a Combined Ultrashort-Echo-Time/Dixon MRI Sequence”. In: *Journal of Nuclear Medicine* 53.5, pp. 796–804. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.111.092577.
- Beyer, T. et al. (2000). “A Combined PET/CT Scanner for Clinical Oncology”. In: *Journal of Nuclear Medicine* 41.8, pp. 1369–1379. ISSN: 0161-5505, 2159-662X.
- Bocher, M. et al. (May 1, 2000). “Gamma camera-mounted anatomical X-ray tomography: technology, system characteristics and first images”. In: *European Journal of Nuclear Medicine* 27.6, pp. 619–627. ISSN: 1619-7089. DOI: 10.1007/s002590050555. URL: <https://doi.org/10.1007/s002590050555>.
- Boer, P. (Oct. 1984). “Estimated lean body mass as an index for normalization of body fluid volumes in humans”. In: *The American Journal of Physiology* 247.4, F632–636. ISSN: 0002-9513. DOI: 10.1152/ajprenal.1984.247.4.F632.
- Bogaerts, J. et al. (2009). “Individual patient data analysis to assess modifications to the RECIST criteria”. In: *European Journal of Cancer (Oxford, England: 1990)* 45.2, pp. 248–260. ISSN: 1879-0852. DOI: 10.1016/j.ejca.2008.10.027.
- Brady, S. L. and Shulkin, B. L. (2019). “Analysis of quantitative [I-123] mIBG SPECT/CT in a phantom and in patients with neuroblastoma”. In: *EJNMMI Physics* 6.1, p. 31. ISSN: 2197-7364. DOI: 10.1186/s40658-019-0267-6. URL: <https://ejnmmiphys.springeropen.com/articles/10.1186/s40658-019-0267-6> (visited on 06/10/2020).
- Brisse, H. J. et al. (2011). “Guidelines for Imaging and Staging of Neuroblastic Tumors: Consensus Report from the International Neuroblastoma Risk Group Project”. In: *Radiology* 261.1, pp. 243–257. ISSN: 0033-8419. DOI: 10.1148/radiol.11101352.
- Brodeur, G. M. et al. (Dec. 1988). “International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma”. In: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 6.12, pp. 1874–1881. ISSN: 0732-183X. DOI: 10.1200/JCO.1988.6.12.1874.
- Brown, C. M. (2018). “Application of novel corrections for quantification of 123I SPECT”. PhD thesis. University of Glasgow. URL: <https://eleanor.lib.gla.ac.uk/record=b3304905>.
- Cerny, I. et al. (2016). “Superiority of SPECT/CT over planar 123I-mIBG images in neuroblastoma patients with impact on Curie and SIOPEN score values”. In:

- Nuklearmedizin* 55.4, pp. 155–157. ISSN: 0029-5566, 2567-6407. DOI: 10.3413/Nukmed-0743-15-05.
- Cohn, S. L. et al. (Jan. 10, 2009). “The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report”. In: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27.2, pp. 289–297. ISSN: 1527-7755. DOI: 10.1200/JCO.2008.16.6785.
- Coombs, B. D. et al. (1997). “Two-point Dixon technique for water-fat signal decomposition with B0 inhomogeneity correction”. In: *Magnetic Resonance in Medicine* 38.6, pp. 884–889. ISSN: 1522-2594. DOI: 10.1002/mrm.1910380606.
- Coulter, D. et al. (2008). “Weaver syndrome and neuroblastoma”. In: *Journal of Pediatric Hematology/Oncology* 30.10, pp. 758–760. ISSN: 1536-3678. DOI: 10.1097/MPH.0b013e3181758974.
- Decarolis, B. et al. (2013). “Iodine-123 Metaiodobenzylguanidine Scintigraphy Scoring Allows Prediction of Outcome in Patients With Stage 4 Neuroblastoma: Results of the Cologne Interscore Comparison Study”. In: *Journal of Clinical Oncology* 31.7. ISSN: 0732-183X, 1527-7755. DOI: 10.1200/JCO.2012.45.8794.
- Desmots, C. et al. (Mar. 17, 2020). “Evaluation of a new multipurpose whole-body CzT-based camera: comparison with a dual-head Anger camera and first clinical images”. In: *EJNMMI Physics* 7, p. 18. ISSN: 2197-7364. DOI: 10.1186/s40658-020-0284-5. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7078403/>.
- Dickson, J. C. et al. (2014). “A comparison of CT- and MR-based attenuation correction in neurological PET”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 41.6, pp. 1176–1189. ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-013-2652-z.
- Dixon, W. T. (1984). “Simple proton spectroscopic imaging.” In: *Radiology* 153.1, pp. 189–194. ISSN: 0033-8419, 1527-1315.
- DuBois, S. G. et al. (July 2022). “High-Risk and Relapsed Neuroblastoma: Toward More Cures and Better Outcomes”. In: *American Society of Clinical Oncology Educational Book* 42, pp. 768–780. ISSN: 1548-8748. DOI: 10.1200/EDBK_349783. URL: https://ascopubs.org/doi/full/10.1200/EDBK_349783 (visited on 01/31/2023).

- Dumba, M. et al. (2015). “Neuroblastoma and nephroblastoma: a radiological review”. In: *Cancer Imaging* 15.1. ISSN: 1740-5025. DOI: 10.1186/s40644-015-0040-6.
- Eisenhauer, E. et al. (2009). “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)”. In: *European Journal of Cancer* 45.2, pp. 228–247. ISSN: 0959-8049. DOI: 10.1016/j.ejca.2008.10.026.
- Even-Sapir, E. et al. (2001). “The New Technology of Combined Transmission and Emission Tomography in Evaluation of Endocrine Neoplasms”. In: *Journal of Nuclear Medicine* 42.7, pp. 998–1004. ISSN: 0161-5505, 2159-662X.
- Fedorov, A. et al. (2012). “3D Slicer as an Image Computing Platform for the Quantitative Imaging Network”. In: *Magnetic resonance imaging* 30.9, pp. 1323–1341. ISSN: 0730-725X. DOI: 10.1016/j.mri.2012.05.001. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466397/>.
- Fendler, W. P. et al. (2013). “High ^{123}I -MIBG uptake in neuroblastic tumours indicates unfavourable histopathology”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 40.11, pp. 1701–1710. ISSN: 1619-7089. DOI: 10.1007/s00259-013-2491-y.
- Gallamini, A. et al. (2009). “Interim positron emission tomography scan in Hodgkin lymphoma: definitions, interpretation rules, and clinical validation”. In: *Leukemia & Lymphoma* 50.11, pp. 1761–1764. ISSN: 1029-2403. DOI: 10.3109/10428190903308072.
- Garner, E. F. and Beierle, E. A. (Jan. 2016). “Cancer Stem Cells and Their Interaction with the Tumor Microenvironment in Neuroblastoma”. In: *Cancers* 8.1, p. 5. ISSN: 2072-6694. DOI: 10.3390/cancers8010005. URL: <https://www.mdpi.com/2072-6694/8/1/5>.
- George, R. E. et al. (2012). “Tumor histology during induction therapy in patients with high-risk neuroblastoma”. In: *Pediatric Blood & Cancer* 59.3, pp. 506–510. ISSN: 1545-5017. DOI: 10.1002/pbc.24013.
- Goetz, C. et al. (2008). “SPECT Low-Field MRI System for Small-Animal Imaging”. In: *Journal of Nuclear Medicine* 49.1, pp. 88–93. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.107.044313.
- Gregory, R. A. et al. (2017). “Objective comparison of lesion detectability in low and medium-energy collimator iodine-123 mIBG images using a channelized Hotelling observer”. In: *Physics in Medicine & Biology* 62.1, p. 17. ISSN: 0031-

9155. DOI: 10.1088/1361-6560/62/1/17. URL: <http://stacks.iop.org/0031-9155/62/i=1/a=17>.
- Ha, S. et al. (2010). “Development of a new RF coil and γ -ray radiation shielding assembly for improved MR image quality in SPECT/MRI”. In: *Physics in Medicine and Biology* 55.9, p. 2495. ISSN: 0031-9155. DOI: 10.1088/0031-9155/55/9/005.
- Hamamura, M. J. et al. (2010). “Development of an MR-compatible SPECT system (MRSPECT) for simultaneous data acquisition”. In: *Physics in Medicine and Biology* 55.6, p. 1563. ISSN: 0031-9155. DOI: 10.1088/0031-9155/55/6/002.
- Harris, C. R. et al. (2020). “Array programming with NumPy”. In: *Nature* 585.7825, pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: <https://www.nature.com/articles/s41586-020-2649-2>.
- Hasenclever, D. et al. (2014). “qPET - a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 41.7, pp. 1301–1308. ISSN: 1619-7089. DOI: 10.1007/s00259-014-2715-9.
- Hoesl, M. et al. (2022). *An AI-based algorithm for continuous Hounsfield units in the pelvis and brain – with syngo.via RT Image Suite (VB60)*. White Paper. Erlangen, Germany: Siemens Healthcare GmbH.
- Hofman, M. S. and Hicks, R. J. (Oct. 18, 2016). “How We Read Oncologic FDG PET/CT”. In: *Cancer Imaging* 16.1, p. 35. ISSN: 1470-7330. DOI: 10.1186/s40644-016-0091-3. URL: <https://doi.org/10.1186/s40644-016-0091-3>.
- Horger, M. and Bares, R. (2006). “The role of single-photon emission computed tomography/computed tomography in benign and malignant bone disease”. In: *Seminars in Nuclear Medicine* 36.4, pp. 286–294. ISSN: 0001-2998. DOI: 10.1053/j.semnuclmed.2006.05.001.
- Hudson, H. M. and Larkin, R. S. (1994). “Accelerated image reconstruction using ordered subsets of projection data”. In: *IEEE transactions on medical imaging* 13.4, pp. 601–609. ISSN: 0278-0062. DOI: 10.1109/42.363108.
- Hunter, J. D. (2007). “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science & Engineering* 9.3, pp. 90–95. ISSN: 1558-366X. DOI: 10.1109/MCSE.2007.55.

- Inoue, Y. et al. (2003). “Effect of collimator choice on quantitative assessment of cardiac iodine 123 MIBG uptake”. In: *Journal of Nuclear Cardiology: Official Publication of the American Society of Nuclear Cardiology* 10.6, pp. 623–632. ISSN: 1071-3581. DOI: 10.1016/s1071-3581(03)00652-4.
- Jacobsson, H. et al. (2007). “Different concentrations of I-123 MIBG and In-111 pentetreotide in the two main liver lobes in children: persisting regional functional differences after birth?” In: *Clinical Nuclear Medicine* 32.1, pp. 24–28. ISSN: 0363-9762. DOI: 10.1097/01.rlu.0000249592.95945.e4.
- Joshi, V. V. et al. (1992). “Age-linked prognostic categorization based on a new histologic grading system of neuroblastomas. A clinicopathologic study of 211 cases from the Pediatric Oncology Group”. In: *Cancer* 69.8, pp. 2197–2211. ISSN: 0008-543X. DOI: 10.1002/1097-0142(19920415)69:8<2197::aid-cncr2820690829>3.0.co;2-x.
- Kapoor, V. et al. (2004). “An Introduction to PET-CT Imaging”. In: *RadioGraphics* 24.2, pp. 523–543. ISSN: 0271-5333. DOI: 10.1148/rg.242025724.
- Kayano, D. and Kinuya, S. (Aug. 2018). “Current Consensus on I-131 MIBG Therapy”. In: *Nuclear Medicine and Molecular Imaging* 52.4, pp. 254–265. ISSN: 1869-3474. DOI: 10.1007/s13139-018-0523-z. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6066492/>.
- Keereman, V. et al. (2013). “Challenges and current methods for attenuation correction in PET/MR”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 26.1, pp. 81–98. ISSN: 0968-5243, 1352-8661. DOI: 10.1007/s10334-012-0334-7.
- Kinahan, P. E. et al. (1998). “Attenuation correction for a combined 3D PET/CT scanner”. In: *Medical Physics* 25.10, pp. 2046–2053. ISSN: 2473-4209. DOI: 10.1118/1.598392. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1118/1.598392>.
- Kinahan, P. E. and Fletcher, J. W. (2010). “PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy”. In: *Seminars in ultrasound, CT, and MR* 31.6, pp. 496–505. ISSN: 0887-2171. DOI: 10.1053/j.sult.2010.10.001. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3026294/>.
- King, M. A. et al. (2016). “Design of a Multi-Pinhole Collimator for I-123 DaTscan Imaging on Dual-Headed SPECT Systems in Combination with a Fan-Beam

- Collimator”. In: *IEEE transactions on nuclear science* 63.1, pp. 90–97. ISSN: 0018-9499. DOI: 10.1109/TNS.2016.2515519. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4864598/>.
- Kluyver, T. et al. (2016). “Jupyter Notebooks – a publishing format for reproducible computational workflows”. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing*, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87. URL: <https://ebooks.iospress.nl/doi/10.3233/978-1-61499-649-1-87>.
- Kushner, B. H. (2004). “Neuroblastoma: A Disease Requiring a Multitude of Imaging Studies”. In: *Journal of Nuclear Medicine* 45.7, pp. 1172–1188. ISSN: 0161-5505, 2159-662X.
- Ladenstein, R. et al. (2011). “The prognostic value of semi-quantitative 123I MIBG scintigraphy at diagnosis in high risk neuroblastoma: validation of the SIOOPEN score method”. In: 43rd Congress of the International Society of Paediatric Oncology (SIOP) 2011. Vol. 57, pp. 732–733.
- Ladenstein, R. et al. (2017). “Busulfan and melphalan versus carboplatin, etoposide, and melphalan as high-dose chemotherapy for high-risk neuroblastoma (HR-NBL1/SIOOPEN): an international, randomised, multi-arm, open-label, phase 3 trial”. In: *The Lancet Oncology* 18.4, pp. 500–514. ISSN: 1470-2045, 1474-5488. DOI: 10.1016/S1470-2045(17)30070-0. URL: [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(17\)30070-0/abstract](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(17)30070-0/abstract).
- Lee, M.-Y. et al. (2022). “Assessment of the Prognostic value of PERCIST 1.0 and Simplified PERCIST (“SPERCIST”) FDG PET in Patients with Advanced Melanoma Treated with Immune Check-point Inhibitors”. In: *Journal of Nuclear Medicine* 63 (supplement 2), pp. 2493–2493. ISSN: 0161-5505, 2159-662X.
- Lewington, V. et al. (2009). “Development of a semi-quantitative I -123 mIBG reporting method in high risk neuroblastoma”. In: 41st Congress of the International Society of Paediatric Oncology (SIOP) 2009. Vol. 53, p. 808.
- Manohar, K. et al. (2012). “Prognostic value of quantitative parameters derived on initial staging 18F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with high-grade non-Hodgkin’s lymphoma”. In: *Nuclear Medicine Communications* 33.9, pp. 974–981. ISSN: 1473-5628. DOI: 10.1097/MNM.0b013e32835673ec.

- Marshall, H. R. et al. (2011). “A comparison of MR-based attenuation correction in PET versus SPECT”. In: *Physics in Medicine and Biology* 56.14, p. 4613. ISSN: 0031-9155. DOI: 10.1088/0031-9155/56/14/024.
- Martinez-Möller, A. et al. (2009). “Tissue Classification as a Potential Approach for Attenuation Correction in Whole-Body PET/MRI: Evaluation with PET/CT Data”. In: *Journal of Nuclear Medicine* 50.4, pp. 520–526. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.108.054726.
- Matthay, K. K. et al. (2010). “Criteria for evaluation of disease extent by 123I-metaiodobenzylguanidine scans in neuroblastoma: a report for the International Neuroblastoma Risk Group (INRG) Task Force”. In: *British Journal of Cancer* 102.9, pp. 1319–1326. ISSN: 0007-0920. DOI: 10.1038/sj.bjc.6605621.
- Matthay, K. K. et al. (2003). “Correlation of Early Metastatic Response by 123I-Metaiodobenzylguanidine Scintigraphy With Overall Response and Event-Free Survival in Stage IV Neuroblastoma”. In: *Journal of Clinical Oncology* 21.13, pp. 2486–2491. ISSN: 0732-183X, 1527-7755. DOI: 10.1200/JCO.2003.09.122.
- Matthay, K. K. et al. (Nov. 10, 2016). “Neuroblastoma”. In: *Nature Reviews Disease Primers* 2.1, pp. 1–21. ISSN: 2056-676X. DOI: 10.1038/nrdp.2016.78. URL: <https://www.nature.com/articles/nrdp201678>.
- McKinney, W. (2010). “Data Structures for Statistical Computing in Python”. In: Python in Science Conference. Austin, Texas, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a. URL: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
- Meignan, M. et al. (2009). “Report on the First International Workshop on Interim-PET-Scan in Lymphoma”. In: *Leukemia & Lymphoma* 50.8, pp. 1257–1260. ISSN: 1029-2403. DOI: 10.1080/10428190903040048.
- Messina, J. A. et al. (2006). “Evaluation of semi-quantitative scoring system for metaiodobenzylguanidine (mIBG) scans in patients with relapsed neuroblastoma”. In: *Pediatric Blood & Cancer* 47.7, pp. 865–874. ISSN: 1545-5017. DOI: 10.1002/pbc.20777.
- Mueller, W. P. et al. (2013). “Nuclear medicine and multimodality imaging of pediatric neuroblastoma”. In: *Pediatric Radiology* 43.4, pp. 418–427. ISSN: 0301-0449, 1432-1998. DOI: 10.1007/s00247-012-2512-1.
- Nakazawa, A. (2021). “Biological categories of neuroblastoma based on the international neuroblastoma pathology classification for treatment stratification”.

- In: *Pathology International* 71.4, pp. 232–244. ISSN: 1440-1827. DOI: 10.1111/pin.13085. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/pin.13085>.
- NEMA (2001). *NEMA Standards Publication NU 2-2001: Performance Measurements of Positron Emission Tomographs*. NU 2-2001. Rosslyn, VA: National Electrical Manufacturers Association.
- (2012). *Standards Publication NU 1-2012: Performance Measurements of Gamma Cameras*. NU 1-2012. Rosslyn, VA: National Electrical Manufacturers Association.
- O, J. H. and Wahl, R. L. (2018). “PERCIST in Perspective”. In: *Nuclear Medicine and Molecular Imaging* 52.1, pp. 1–4. ISSN: 1869-3474. DOI: 10.1007/s13139-017-0507-4. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5777964/>.
- O, J. H. et al. (2016). “Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0”. In: *Radiology* 280.2, pp. 576–584. ISSN: 0033-8419. DOI: 10.1148/radiol.2016142043.
- Olivier, P. et al. (2003). “Guidelines for radioiodinated MIBG scintigraphy in children”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 30.5, B45–B50. ISSN: 1619-7089. DOI: 10.1007/s00259-003-1138-9.
- Paquet, N. et al. (2004). “Within-Patient Variability of 18F-FDG: Standardized Uptake Values in Normal Tissues”. In: *Journal of Nuclear Medicine* 45.5, pp. 784–788. ISSN: 0161-5505, 2159-662X. URL: <https://jnm.snmjournals.org/content/45/5/784> (visited on 12/29/2022).
- Peters, A. M. et al. (2011). “Estimation of lean body mass in children”. In: *British Journal of Anaesthesia* 106.5, pp. 719–723. ISSN: 0007-0912. DOI: 10.1093/bja/aer057. URL: <https://doi.org/10.1093/bja/aer057>.
- Pfannenbergl, A. C. et al. (2003). “Benefit of anatomical-functional image fusion in the diagnostic work-up of neuroendocrine neoplasms”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 30.6, pp. 835–843. ISSN: 1619-7070. DOI: 10.1007/s00259-003-1160-y.
- Pfluger, T. et al. (2003). “Integrated Imaging Using MRI and ¹²³I-Metaiodobenzylguanidine Scintigraphy to Improve Sensitivity and Specificity in the Diagnosis of Pediatric Neuroblastoma”. In: *American Journal of Roentgenology*

- 181.4, pp. 1115–1124. ISSN: 0361-803X, 1546-3141. DOI: 10.2214/ajr.181.4.1811115.
- Richardson, M. K. and Keuck, G. (June 2022). “The revolutionary developmental biology of Wilhelm His, Sr.” In: *Biological Reviews of the Cambridge Philosophical Society* 97.3, pp. 1131–1160. ISSN: 1464-7931. DOI: 10.1111/brv.12834. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9304566/>.
- Riedl, C. C. et al. (2017). “Comparison of FDG-PET/CT and contrast-enhanced CT for monitoring therapy response in patients with metastatic breast cancer”. In: *European journal of nuclear medicine and molecular imaging* 44.9, pp. 1428–1437. ISSN: 1619-7070. DOI: 10.1007/s00259-017-3703-7. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5526620/>.
- Ritt, P. et al. (May 1, 2011). “Absolute quantification in SPECT”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 38.1, pp. 69–77. ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-011-1770-8. URL: <http://link.springer.com/article/10.1007/s00259-011-1770-8>.
- Rossum, G. van (1995). *Python tutorial*. Technical Report CS-R9526. Amsterdam: Centrum voor Wiskunde en Informatica (CWI).
- Rozovsky, K. et al. (2008). “Added Value of SPECT/CT for Correlation of MIBG Scintigraphy and Diagnostic CT in Neuroblastoma and Pheochromocytoma”. In: *American Journal of Roentgenology* 190.4, pp. 1085–1090. ISSN: 0361-803X. DOI: 10.2214/AJR.07.2107.
- Samim, A. et al. (Dec. 12, 2022). “[18F]mFBG PET-CT for detection and localisation of neuroblastoma: a prospective pilot study”. In: *European Journal of Nuclear Medicine and Molecular Imaging*. ISSN: 1619-7089. DOI: 10.1007/s00259-022-06063-6.
- Schulthess, G. K. von et al. (2006). “Integrated PET/CT: Current Applications and Future Directions”. In: *Radiology* 238.2, pp. 405–422. ISSN: 0033-8419. DOI: 10.1148/radiol.2382041977.
- Schulz, V. et al. (2010). “Automatic, three-segment, MR-based attenuation correction for whole-body PET/MR data”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 38.1, pp. 138–152. ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-010-1603-1.

- Sharp, S. E. et al. (2011). “Pediatrics: Diagnosis of Neuroblastoma”. In: *Seminars in Nuclear Medicine* 41.5, pp. 345–353. ISSN: 0001-2998. DOI: 10.1053/j.semnuclmed.2011.05.001.
- Shimada, H. et al. (2001). “International neuroblastoma pathology classification for prognostic evaluation of patients with peripheral neuroblastic tumors: a report from the Children’s Cancer Group”. In: *Cancer* 92.9, pp. 2451–2461. ISSN: 0008-543X.
- Shimada, H. and Ikegaki, N. (2019). “Chapter 1 - Neuroblastoma Pathology and Classification for Precision Prognosis and Therapy Stratification”. In: *Neuroblastoma*. Ed. by S. K. Ray. Academic Press, pp. 1–22. ISBN: 978-0-12-812005-7. DOI: 10.1016/B978-0-12-812005-7.00001-1. URL: <https://www.sciencedirect.com/science/article/pii/B9780128120057000011>.
- Shimada, H. et al. (1999). “The International Neuroblastoma Pathology Classification (the Shimada system)”. In: *Cancer* 86.2, pp. 364–372. ISSN: 1097-0142. DOI: 10.1002/(SICI)1097-0142(19990715)86:2<364::AID-CNCR21>3.0.CO;2-7. URL: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0142\(19990715\)86:2%3C364::AID-CNCR21%3E3.0.CO;2-7/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0142(19990715)86:2%3C364::AID-CNCR21%3E3.0.CO;2-7/abstract).
- Simões-Costa, M. and Bronner, M. E. (2013). “Insights into neural crest development and evolution from genomic analysis”. In: *Genome Research* 23.7, p. 1069. DOI: 10.1101/gr.157586.113. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3698500/>.
- Skougaard, K. et al. (2013). “Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab”. In: *Journal of Nuclear Medicine* 54.7, pp. 1026–1031. ISSN: 1535-5667. DOI: 10.2967/jnumed.112.111757.
- Snay, E. R. et al. (2011). “Improved Quality of Pediatric 123I-MIBG Images with Medium-Energy Collimators”. In: *Journal of Nuclear Medicine Technology* 39.2, pp. 100–104. ISSN: 0091-4916, 1535-5675. DOI: 10.2967/jnmt.110.080309. URL: <http://tech.snmjournals.org/content/39/2/100>.
- Steinberg, J. et al. (2010). “Three-region MRI-based whole-body attenuation correction for automated PET reconstruction”. In: *Nuclear Medicine and Biology* 37.2, pp. 227–235. ISSN: 0969-8051. DOI: 10.1016/j.nucmedbio.2009.11.002.
- Sugawara, Y. et al. (1999). “Reevaluation of the Standardized Uptake Value for FDG: Variations with Body Weight and Methods for Correction”. In: *Radiol-*

- ogy* 213.2, pp. 521–525. ISSN: 0033-8419. DOI: 10.1148/radiology.213.2.r99nv37521. URL: <https://pubs.rsna.org/doi/10.1148/radiology.213.2.r99nv37521>.
- Sundquist, F. et al. (2022). “A Phase II Trial of a Personalized, Dose-Intense Administration Schedule of 177Lutetium-DOTATATE in Children With Primary Refractory or Relapsed High-Risk Neuroblastoma–LuDO-N”. In: *Frontiers in Pediatrics* 10. ISSN: 2296-2360. URL: <https://www.frontiersin.org/articles/10.3389/fped.2022.836230>.
- Taggart, D. R. et al. (2011). “Prognostic Value of the Stage 4S Metastatic Pattern and Tumor Biology in Patients With Metastatic Neuroblastoma Diagnosed Between Birth and 18 Months of Age”. In: *Journal of Clinical Oncology* 29.33, pp. 4358–4364. ISSN: 0732-183X. DOI: 10.1200/JCO.2011.35.9570. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3221520/>.
- Therasse, P. et al. (2000). “New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada”. In: *Journal of the National Cancer Institute* 92.3, pp. 205–216. ISSN: 0027-8874. DOI: 10.1093/jnci/92.3.205.
- UKHSA (Feb. 2021). *Children, teenagers and young adults UK cancer statistics report 2021*. UK Health Security Agency. URL: <http://www.ncin.org.uk/view?rid=4272>.
- Virtanen, P. et al. (2020). “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17.3, pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: <https://www.nature.com/articles/s41592-019-0686-2>.
- Wagenknecht, G. et al. (2013). “MRI for attenuation correction in PET: methods and challenges”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 26.1, pp. 99–113. ISSN: 0968-5243, 1352-8661. DOI: 10.1007/s10334-012-0353-4.
- Wahl, R. L. et al. (2009). “From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors”. In: *Journal of Nuclear Medicine* 50 Suppl 1, 122S–50S. ISSN: 0161-5505. DOI: 10.2967/jnumed.108.057307.

- Wallace, H. J. et al. (2014). "Simulation of SPECT/MR attenuation correction demonstrates potential for uptake quantification in Neuroblastoma". In: *Pediatric Blood & Cancer* 61, S336–S336.
- Wallace, H. J. et al. (2016). "Impact of SPECT/MR fusion on scoring of 123I-mIBG scintigraphy in neuroblastoma". In: *Nuclear Medicine Communications* 37.5, pp. 553–553.
- Woods, A.-M. et al. (2017). "Dual-isotope subtraction SPECT-CT in parathyroid localization". In: *Nuclear Medicine Communications* 38.12, pp. 1047–1054. ISSN: 1473-5628. DOI: 10.1097/MNM.0000000000000765.
- Yanagawa, M. et al. (2012). "Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors". In: *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* 53.6, pp. 872–880. ISSN: 1535-5667. DOI: 10.2967/jnumed.111.098699.
- Yanik, G. A. et al. (2013). "Semiquantitative mIBG Scoring as a Prognostic Indicator in Patients with Stage 4 Neuroblastoma: A Report from the Children's Oncology Group". In: *Journal of Nuclear Medicine* 54.4, pp. 541–548. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.112.112334.
- Yankeelov, T. E. et al. (2012). "Simultaneous PET–MRI in oncology: a solution looking for a problem?" In: *Magnetic Resonance Imaging* 30.9, pp. 1342–1356. ISSN: 0730-725X. DOI: 10.1016/j.mri.2012.06.001. URL: <http://www.sciencedirect.com/science/article/pii/S0730725X12001993>.
- Ziai, D. et al. (2013). "Therapy response evaluation with FDG-PET/CT in small cell lung cancer: a prognostic and comparison study of the PERCIST and EORTC criteria". In: *Cancer Imaging: The Official Publication of the International Cancer Imaging Society* 13, pp. 73–80. ISSN: 1470-7330. DOI: 10.1102/1470-7330.2013.0008.