



Gemmell, Alastair J. (2026) *Quantitative SPECT-CT and textural analysis for neuro-endocrine tumours*. PhD thesis.

<https://theses.gla.ac.uk/86259/>

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 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

Quantitative SPECT-CT and Textural Analysis for Neuro-Endocrine Tumours

Alastair J. Gemmell
B.Sc(Hons), M.Sc



**University
of Glasgow**

Submitted in fulfilment of the requirements for
the Degree of Doctor of Philosophy
19th of March 2026

School of Medicine, Dentistry and Nursing
College of Medical, Veterinary and Life Sciences
University of Glasgow

NHS Greater Glasgow and Clyde
Department of Clinical Physics and Bioengineering

Copyright by A.J. Gemmell. All Rights Reserved.

Abstract

Background

Quantification is an important emerging technique in nuclear medicine imaging. It may provide additional clinical information to add to visual interpretation, whilst reducing inter-operator variability. Technological developments and commercial solutions have recently increased access to quantitative Single Photon Emission Tomography (SPECT). There are few studies optimising acquisition and reconstruction parameters and only a limited number of clinical applications have been investigated. The growth in imaging of Neuroendocrine Tumours (NETs), both for diagnosis and following molecular radiotherapy, highlights Somatostatin Receptor Imaging (SRI) as an ideal candidate for applying quantification.

Aims

The thesis sought to assess the accuracy and longitudinal stability of a commercial quantification package for SPECT, Siemens Broad Quantification ("BroadQuant") for two radionuclides, ^{99m}Tc and ^{177}Lu . A further aim was to apply quantification to SPECT imaging of NETs, for the integration of quantification into clinical practice.

Methods

A series of experiments were designed to measure the accuracy and reproducibility of each stage of the BroadQuant calibration process. Uncertainties were calculated for each step and combined to provide an overall uncertainty on the calibration factor. A range of Quality Control tests were evaluated to determine how best to monitor the longitudinal stability of the calibration. A phantom that mimicked inhomogeneous lesions was used to identify robust textural analysis features for further study in clinical applications.

The use of SPECT quantification for imaging of NETs was investigated for a number of clinical applications. The impact of Somatostatin Analogue medication on quantitative SPECT measurement of ^{99m}Tc -EDDA/HYNIC-TOC was studied, as well as whether a Tumour Sink Effect (TSE) affected uptake. The use of quantitative metrics and textural analysis

features as an alternative to the visual assessment of Krenning Score in ^{99m}Tc -EDDA/HYNIC-TOC SPECT was examined. Post-therapy SPECT imaging following Peptide Receptor Radionuclide Therapy (PRRT) was analysed, to determine the predictive value of quantitative metrics. Statistical harmonisation methods were applied to quantitative data from two gamma cameras with the aim of expansion of quantitative SPECT to multiple sites for future studies.

Results

The calibration process was comprehensively examined, providing a number of recommendations for the user, and estimating an uncertainty in the calibration factor of 1.37% for ^{99m}Tc and 2.81% for ^{177}Lu respectively. Longitudinal measurements of calibration accuracy determined that stability could be successfully measured with Shewhart charts monitoring Standardised Uptake Value (SUV), with a suitable stable period for ^{99m}Tc calibration assessed as being up to 9 months.

For patients on Somatostatin Analogue medication, a reduction in physiological uptake of ^{99m}Tc -EDDA/HYNIC-TOC was shown, that was not reflected in pathological uptake. No evidence was found for the presence of a TSE for ^{99m}Tc -EDDA/HYNIC-TOC. A quantitative Krenning Score was found to be more robust than visual interpretation of Krenning Score, with a comparable level of accuracy when compared to expert assessment.

Following feature reduction, a range of textural analysis features were investigated in the context of post-PRRT SPECT imaging. Sphericity was determined to be predictive both of overall survival and for patient-reported symptom response.

Finally, statistical harmonisation using Location-Scale adjustment was found to improve the harmonisation between gamma cameras for multiple quantitative metrics, demonstrating a potential tool for combining data from multiple cameras for clinical studies.

Conclusion

This thesis addresses the key aspects of the SPECT quantification process, from the calibration of the gamma camera through to clinical applications. It demonstrates that SPECT quantification can be an accurate and robust tool. The use of quantification and textural analysis has been shown to have positive clinical value when applied to the imaging of neuroendocrine tumours.

Contents

Abstract	i
List of Tables	vii
List of Figures	x
Acknowledgements	xiv
Declaration	xvi
Publications	xvii
Abbreviations	xviii
Part 1: Introduction and background	1
1 Aims of the thesis	3
2 Background to Quantification in nuclear medicine	5
2.1 Introduction	5
2.2 Gamma camera: physics and technology	7
2.3 Quantification of SPECT images	12
2.4 Relevance to the thesis	16
3 Review of Textural Analysis and Radiomics	17
3.1 Introduction	17
3.2 Technical factors of textural analysis	18
3.3 Definitions of Textural Analysis features	19

3.4	Validation of Textural Analysis features	21
3.5	Clinical studies utilising textural analysis	24
3.6	Relevance to the thesis	25
4	Neuroendocrine Tumours and the role of nuclear medicine	26
4.1	Background to Neuroendocrine Tumours	26
4.2	Imaging of NETs	26
4.3	Treatment of NETs	29
4.4	Relevance to the thesis	37
5	Introduction to harmonisation of quantitative data	38
5.1	Introduction	38
5.2	Quantitative standards in nuclear medicine	38
5.3	Approaches to harmonisation	40
5.4	Relevance to the thesis	44
Part 2: Quantitative SPECT-CT calibration		45
6	Calibration of Siemens BroadQuant	46
6.1	Radionuclide calibrator	46
6.2	Point Source Sensitivity calibration	48
6.3	Volume Sensitivity calibration	59
6.4	Overall uncertainty in the calibration process	61
6.5	Conclusion and relevance for the thesis	67
7	Assessing stability and accuracy of quantitative calibrations	69
7.1	Introduction	69
7.2	Acquisition and Reconstruction parameters	70
7.3	Accuracy of the calibration	80
7.4	Longitudinal stability	88
7.5	Correlation of SUV measurements with other metrics	96
7.6	Conclusion and relevance for the thesis	104
8	Robustness of textural analysis features in quantitative ^{99m}Tc and ^{177}Lu SPECT-CT phantom acquisitions	106
8.1	Background	106
8.2	Materials and Methods	108
8.3	Results	111
8.4	Discussion	112
8.5	Conclusion	118
8.6	Relevance for the thesis	118

Part 3: Clinical quantification for Neuro-Endocrine Tumours	119
9 Quantitative Uptake in ^{99m}Tc -EDDA/HYNIC-TOC Somatostatin Receptor Imaging – the effect of Long-Acting Release Somatostatin Analogue Therapy	120
9.1 Context of published paper	120
9.2 <i>Content removed due to copyright restrictions</i>	121
9.3 Relevance for the thesis	121
10 Comparison of quantitative Krenning Scores with visual assessment in ^{99m}Tc -EDDA/HYNIC-TOC SPECT-CT	122
10.1 Context of published paper	122
10.2 <i>Content removed due to copyright restrictions</i>	122
10.3 Relevance for the thesis	122
11 Textural Analysis and comparison to visual Krenning Scores	125
11.1 Introduction	125
11.2 Materials and Methods	125
11.3 Results	126
11.4 Discussion	134
11.5 Conclusion and relevance for the thesis	136
12 The Tumour Sink Effect in ^{99m}Tc -EDDA/HYNIC-TOC quantitative SPECT . . .	138
12.1 Introduction	138
12.2 Materials and Methods	138
12.3 Results	139
12.4 Discussion	140
12.5 Conclusion and relevance for the thesis	146
13 Post-therapy quantitative SPECT imaging as a predictor of PRRT response .	147
13.1 Introduction	147
13.2 Materials and Methods	148
13.3 Results	150
13.4 Discussion	165
13.5 Conclusion and relevance for the thesis	167
14 Patient-reported Quality of Life metrics following PRRT and the predictive value of SPECT quantification	168
14.1 Introduction	168
14.2 Materials and Methods	169
14.3 Results	172
14.4 Discussion	180
14.5 Conclusion and relevance for the thesis	182

15	Harmonisation of clinical quantitative data	184
15.1	Introduction	184
15.2	Materials and Methods	184
15.3	Results	186
15.4	Discussion	188
15.5	Conclusion and relevance for the thesis	192
Part 4: Summary and Conclusions		193
16	Summary of the thesis	194
16.1	Findings on calibration and quantification	194
16.2	Findings on clinical applications	195
17	Future work	197
17.1	Further work on calibration and quantification	197
17.2	Further work on clinical applications	197
18	Conclusion	200
Appendices		201
A	Optimisation of reconstruction parameters for Siemens BroadQuant	202
B	Supplementary materials, Chapter 8	206
C	List of Textural Analysis features used in LifeX v6.3	213
D	Detailed results, Chapter 15	216
Bibliography		223

List of Tables

2.1	Accuracy of SPECT quantification using Siemens cameras	15
6.1	Measured difference between point source activity and syringe activity	49
6.2	Measured difference between point source activity at central dipper position and at the edges	50
6.3	Repeatability of point source activity measurements, with & without jig	51
6.4	Repeatability of syringe activity measurements	52
6.5	Repeatability of Point Source Sensitivity, static setup	53
6.6	Repeatability of Point Source Sensitivity, reset of setup	53
6.7	^{99m} Tc sensitivity measured by GCAT methodology	53
6.8	Reproducibility of Point Source Sensitivity measured on gamma camera during the calibration process, for two different periods	54
6.9	CoV for point sources with increasing volume	55
6.10	Acquisition parameters for Volume Calibration workflow.	60
6.11	Repeatability of Volume Sensitivity calibration	61
6.12	Reproducibility of Volume Sensitivity calibration. Expected and measured concentration are decay-corrected to preparation time.	61
6.13	Uncertainties for activity measurements in the radionuclide calibrator for both ^{99m} Tc and ¹⁷⁷ Lu.	63
6.14	Uncertainties for the point source calibration procedure for both ^{99m} Tc and ¹⁷⁷ Lu.	63
6.15	Uncertainties for the volume sensitivity calibration procedure for both ^{99m} Tc and ¹⁷⁷ Lu.	66
6.16	Combined uncertainties for the BroadQuant calibration process for both ^{99m} Tc and ¹⁷⁷ Lu.	66
7.1	Activities added to uniform phantom for reconstruction testing	73
7.2	Percentage of SUV measurements within 5% and 10% for each of ^{99m} Tc and ¹⁷⁷ Lu.	80
7.3	Logistic models fitted to Recovery Coefficients - assessment of best fit	84
7.4	criteria for remedial action on regular QC testing for a Shewhart control chart .	90
7.5	Summary of longitudinal SUV measurements per calibration	93
7.6	longitudinal point source sensitivity measurements with ^{99m} Tc, per calibration .	96

7.7 Occasions where ^{99m}Tc Point Source sensitivity breached either ± 2 or ± 3 standard deviations	96
8.1 Non-converging features following volume dependency testing	112
8.2 Average VOI size after application of threshold	113
8.3 Radiomics features with CoV $> 10\%$ for ^{99m}Tc	113
8.4 Radiomics features with CoV $> 10\%$ for ^{177}Lu	114
8.5 Remaining features following reduction	115
11.1 Univariate analysis using ordinal logistic regression	131
11.2 Multivariable analysis, using backward step-wise analysis	134
11.3 Multivariable analysis, using forward step-wise analysis	134
12.1 Correlation of body habitus measurements with splenic uptake	145
13.1 Demographics for patients undergoing PRRT and quantitative SPECT imaging	151
13.2 Comparison of median Overall Survival (OS) following PRRT	151
13.3 Results for univariate Cox survival analysis, based on 1st fraction imaging . .	156
13.4 Comparison of models by the Akaike Information Criterion, based on 1st fraction imaging	157
13.5 Results for univariate Cox survival analysis, based on differences between 1st and 4th fraction imaging	157
13.6 Comparison of models by the Akaike Information Criterion, based on differences between 1st and 4th fraction imaging	158
13.7 Survival rates for 1st fraction post-PRRT imaging risk factors	158
13.8 Survival rates for changes in post-PRRT between 1st and 4th fractions	159
14.1 Patient-Reported Quality of Life metrics for the EORTC QLQ-C30 questionnaire	170
14.2 Patient-Reported Quality of Life metrics for the EORTC QLQ-GI.NET21 questionnaire	171
14.3 Baseline HR-QoL results before starting PRRT	172
14.4 Features with significant p-values based on the patient-reported response to PRRT	178
14.5 Cutpoints for features established as predictive for changes in patient-reported outcomes following PRRT	179
14.6 Comparison of cutpoints established with QoL data and the cutpoint from OS data	180
15.1 Patient demographics by hospital site	185
15.2 Acquisition parameters for each hospital site	185
15.3 Reconstruction parameters for each hospital site	185
15.4 Summary of harmonisation effects on p-values for L-S method, without covariates and Hospital A as reference site	188
C.1 Abbreviation and description of textural analysis features	215

D.1	p-values for harmonised data from Hospitals A and B, for physiological liver uptake.	217
D.2	p-values for harmonised data from Hospitals A and B, for physiological spleen uptake	218
D.3	p-values for harmonised data from Hospitals A and B, for physiological blood pool	219
D.4	p-values for harmonised data from Hospitals A and B, for liver metastases . .	220
D.5	p-values for harmonised data from Hospitals A and B, for nodal metastases .	221
D.6	p-values for harmonised data from Hospitals A and B, for small bowel primaries	222

List of Figures

0.1	Number of publications by year on SPECT quantification	2
2.1	A Siemens Intevo Bold SPECT-CT gamma camera	6
3.1	Simulated distributions for skewness & kurtosis	20
3.2	Diagram of Grey Level values in an image, with calculation of Run Length . . .	21
4.1	Planar ¹¹¹ In-Octreoscan scan	28
4.2	Fused axial slice of ^{99m} Tc-EDDA/HYNIC-TOC SPECT-CT scan	29
5.1	Recovery Coefficient curve for FDG PET with EARL accreditation limits	39
6.1	Time difference (in minutes) between networked gamma camera and radionuclide calibrator	48
6.2	Siemens point source container with ruler for scale	49
6.3	Calibrator dipper	51
6.4	Positioning jig to allow reproducible positioning of point source within calibrator dipper	52
6.5	Point Source sensitivity measurements, relative to max. volume, with gauze .	56
6.6	Point Source sensitivity measurements, relative to max. volume, without gauze	57
6.7	Linearity of ^{99m} Tc Point Source sensitivity measurements with decaying activity	58
6.8	Linearity of ¹⁷⁷ Lu Point Source sensitivity measurements	59
6.9	CT image of a uniform phantom used for calibration with profile	65
7.1	Reconstructed axial slices of 208 keV and 113 keV energy windows for ¹⁷⁷ Lu NEMA phantom	71
7.2	Maximum Intensity Projection of 208 keV and 113 keV energy windows for ¹⁷⁷ Lu-DOTATATE post-therapy SPECT images	71
7.3	COV measurements for different radionuclides with Siemens Optimal Recon presets	74
7.4	COV measurements for different radionuclides, replicating Siemens Optimal Recon presets manually	75
7.5	NEMA IEC Body phantom	76
7.6	Combined RMSD results for determining optimal quantitative reconstruction parameters	79
7.7	Repeated measurements of ^{99m} Tc and ¹⁷⁷ Lu SUV	81

7.8	Distribution of ^{177}Lu -Chloride in a phantom. Adsorption to the walls of the phantom can be seen after 264 hours	82
7.9	Measured Recovery Coefficients for both $^{99\text{m}}\text{Tc}$ and ^{177}Lu with 60 mm sphere	83
7.10	Testing repeatability of $^{99\text{m}}\text{Tc}$ Recovery Coefficients across consecutive phantom acquisitions and multiple fills.	85
7.11	Testing reproducibility of $^{99\text{m}}\text{Tc}$ Recovery Coefficients across multiple phantom fills.	86
7.12	Comparison of ^{177}Lu Recovery Coefficients with varying time per view	87
7.13	Comparison of measured Recovery Coefficient for $^{99\text{m}}\text{Tc}$ in a NEMA IEC phantom with results from Siemens White Paper	88
7.14	Longitudinal measurements of SUV_{mean} for repeated $^{99\text{m}}\text{Tc}$ calibrations	91
7.15	Longitudinal measurements of SUV_{mean} for repeated ^{177}Lu calibrations	92
7.16	Longitudinal measurements of point source sensitivity for repeated $^{99\text{m}}\text{Tc}$ calibrations	95
7.17	Correlation of $^{99\text{m}}\text{Tc}$ SUV measurements with SUV measurements of other radionuclides	97
7.18	Correlation of $^{99\text{m}}\text{Tc}$ SUV measurements with point source sensitivity measurements	98
7.19	Correlation of $^{99\text{m}}\text{Tc}$ SUV measurements with extrinsic uniformity of ^{57}Co flood source	100
7.20	Correlation of $^{99\text{m}}\text{Tc}$ SUV with $^{99\text{m}}\text{Tc}$ intrinsic uniformity	101
7.21	Comparison of longitudinal $^{99\text{m}}\text{Tc}$ SUV measurements with ^{57}Co energy peak deviation, for detector head 1	102
7.22	Comparison of longitudinal $^{99\text{m}}\text{Tc}$ SUV measurements with ^{57}Co energy peak deviation, for detector head 2	103
8.1	Arrangement of activity in Revolver inserts	109
8.2	Top-down view of unfilled phantom with 2 Revolver inserts in place	109
8.3	Examples of (top) positive correlation; (middle) random distribution; and (bottom) convergence with increasing volume	111
8.4	Axial slices of (left) $^{99\text{m}}\text{Tc}$ and (right) ^{177}Lu Revolver phantoms	113
8.5	Distribution of CoV results for $^{99\text{m}}\text{Tc}$ by syringe size	116
11.1	Correlation matrix for quantitative and textural features, using Spearman's Correlation Coefficient	127
11.2	Intra-observer variability for textural analysis features	129
11.3	Inter-observer variability for textural analysis features	130
11.4	Confusion matrix of predicted versus Expert Consensus Krenning Score	132
11.5	Distribution of the normalised values for the reduced textural analysis features	133
11.6	Distribution of ECS Krenning Scores by dataset	136

12.1 Splenic SUV_{mean} for LAR-SSA naive patients, avid or no disease	140
12.2 Splenic SUV_{mean} for LAR-SSA naive patients, for no disease, solitary site or widespread disease	141
12.3 Splenic SUV_{mean} for LAR-SSA naive patients, based on location of disease . .	142
12.4 Splenic SUV_{mean} for patients on LAR-SSA, based on location of disease . . .	143
12.5 Correlation of the summed Tumour Volume with splenic SUV_{mean}	144
12.6 Correlation of the summed TLSRE with splenic SUV_{mean}	145
13.1 Kaplan-Meier Survival curve, differentiated by tumour type	152
13.2 Kaplan-Meier Survival curves for clinical factors	153
13.3 Comparison of SUV_{mean} between PRRT fractions 1 and 4	154
13.4 The distribution of paired SUV_{mean} between PRRT fractions 1 and 4	155
13.5 Changes in active volume of tumours between PRRT fractions 1 and 4	160
13.6 Correlation matrix for quantitative & textural analysis features from ^{177}Lu - DOTATATE images.	161
13.7 Flow diagram of process of combining risk from three features	162
13.8 Kaplan-Meier curves for Overall Survival stratified by relevant features on first fraction imaging	163
13.9 Kaplan-Meier curves for Overall Survival stratified by changes in GLCM Correlation, 1st vs 4th fraction imaging	164
14.1 Correlation of features from 1st fraction imaging with baseline QLQ-C30 responses	173
14.2 Correlation of features from 1st fraction imaging with baseline QLQ-GI.NET21 responses	174
14.3 Changes in HR-QoL from QLQ-C30 questionnaire, 6 months after starting PRRT	175
14.4 Changes in HR-QoL from GI.NET21 questionnaire, 6 months after starting PRRT	176
14.5 Changes in Endocrine Disease symptoms, quantified by NGLDM Coarseness	177
15.1 Distribution of values for two different clinical sites for GLCM Correlation, before and after harmonisation	187
15.2 Changes in p-value following harmonisation for each feature.	189
A.1 Recovery Coefficients for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu	204
A.2 Signal to Noise Ratio for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu	204
A.3 Measured lung error for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu	205
B.1 Volume dependence of textural analysis parameters for ^{99m}Tc	207

B.2 Volume dependence of textural analysis parameters for ^{177}Lu	210
---	-----

Acknowledgements

If each of my words were a drop of water, you would see through them and glimpse what I feel: gratitude, acknowledgement.

Octavio Paz
Nobel Lecture, 1990

The research and writing of this thesis started 6 years ago: the thoughts of focussing a potential PhD on quantitative SPECT began developing four years prior. However, the skills, knowledge and experience required to produce this thesis came from a lifetime of learning from many different people. There are certain people who I must specifically thank, as it would have been difficult, if not impossible, to get here without their help.

Firstly, my thanks go to my colleagues in NHS Tayside who introduced me to nuclear medicine. Clare Monahan, Dr Glen Gardner and Ed Hockaday, as well as Dr Ewan Eadie, my compatriot in learning on the job. I'm grateful to the technologists of Ninewells and PRI for putting up with me as I found my feet. My fellow trainees and training scheme survivors in the lunchtime group were invaluable for showing the way through training. I owe a great deal to the late Dr Norman Kennedy, for hiring me onto the training scheme, and teaching me not only about nuclear medicine but about being a scientist and a leader.

South Glasgow was the ideal next step for me, with great colleagues, and the endless work for the new hospital gave me a chance to spread my wings. My thanks to Dr Alice Nicol for entrusting me with so much. Working with Dr Jen Dennis and Ross Williamson and the rest of the team, particularly at the Victoria sites, were some of the best moments of my career.

At Gartnavel I've enjoyed the variety and challenges that PET and therapies bring, as well as a wonderful group of co-workers. My thanks to all the technologists who have tolerated without complaint my requests for recording more information for quantitative SPECT. To the physicists who have covered the daily grind as I disappeared down the PhD rabbit-hole: Dr

Colin Brown, Dr Caroline Findlay, Dr Mary-Frances Dempsey, Dr Scott Smith and Dr Clare McKeown - I owe you all innumerable empire biscuits!

Specifically for this PhD thesis, there are a number of people who deserve thanks. My NHS supervisors Dr Sandy Small and Dr Colin Brown have steered me through with challenging conversations, pinpoint advice and support when it seemed the road ahead was endless. To my University of Glasgow supervisor Professor Surajit Ray, thank you for your advice, and apologies for steering in a different direction to what we originally envisaged. Thank you to the medics who took part in the observer study in Chapter 10, and in particular Professor Dave Colville for patiently answering many questions on neuroendocrine disease and its treatments. The Clinical Nurse Specialists in Glasgow and Edinburgh, Irene Wotherspoon, Olivia Beattie and Katie Gibson were so helpful with patient questionnaires and tracking down missing survival data. I have to acknowledge the NHS GG&C staff bursary for supporting my initial years of study, as well as the NES HSSE programme for additional funding. I thank the anonymous patients over the years who have allowed their data to be used for this research.

I wish to conclude with some thanks for those who have supported me outside of work. A list of the ways my parents, Derek and Judith Gemmell, have helped me would be longer than this thesis; suffice to say I am forever grateful for all they have done for me, and also everything they continue to do for my family.

My final acknowledgement, though it is far too poor a word for what they deserve, are to my wife Swenja and daughter Charlotte. A source of unmatched positivity, Swenja has supported me through the whole six years (and pre-PhD too), in ways too numerous to count. Charlotte has brightened my life in a million different ways, and I look forward to every minute I spend with you both.

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.

Alastair James Gemmell

19th of March 2026

Publications

Work produced for this degree has been published in peer-reviewed journals. The contents are included in this thesis as part of a submission by alternative format (approved by the Higher Degrees Committee on 19/05/25). Permission has been sought from, and provided by, the respective publishers to reproduce this work in the print version of the thesis. The references to the published papers are given below.

Chapter 8:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, “Robustness of textural analysis features in quantitative ^{99m}Tc and ^{177}Lu SPECT-CT phantom acquisitions,” *EJNMMI Physics*, vol. 12, no. 1, p. 40, Apr. 2025, doi: 10.1186/s40658-025-00749-0.

Chapter 9:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, “Quantitative uptake in ^{99m}Tc -EDDA/HYNIC-TOC somatostatin receptor imaging – the effect of long-acting release somatostatin analogue therapy,” *Nuclear Medicine Communications*, vol. 44, no. 11, pp. 944–952, Nov. 2023, doi: 10.1097/MNM.0000000000001746.

Chapter 10:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, “Comparison of quantitative Krenning Scores with visual assessment in ^{99m}Tc -EDDA/HYNIC-TOC,” *Nuclear Medicine Communications*, vol. 46, no. 6, pp. 515–522, Jun. 2025, doi: 10.1097/MNM.0000000000001967.

Several of the chapters have been the subject of presentations or posters at European Association of Nuclear Medicine or British Nuclear Medicine Society conferences.

Abbreviations

Notation	Description	First Appearance
ADS	Acquire During Step	8
AIC	Akaike Information Criterion	23
ALT	Aminotransferase	32
BMI	Body Mass Index	139
BSA	Body Surface Area	139
CF	Calibration Factor	46
CFOV	Central Field of View	99
CgA	Chromogranin A	32
CoV	Coefficient of Variation	50
DEXA	Dual-Energy X-ray Absorptiometry	90
EANM	European Association of Nuclear Medicine	3
ECS	Expert Consensus Score	126
EDTA	Ethylenediaminetetraacetic acid	81
ENETS	European Neuroendocrine Tumour Society	123
EORTC	European Organisation for Research and Treatment in Cancer	168
FBP	Filtered Back Projection	8
FoV	Field of View	3
FWHM	Full Width Half Maximum	11
GAN	Generative Adversarial Networks	43
GFR	Glomerular Filtration Rate	31
HCRC	Hot Contrast Recovery Coefficient	74
HR	Hazard Ratio	157
HR-QoL	Health-Related Quality of Life	168
HU	Hounsfield Unit	64
IBI	Inflammation-based Index	34

Notation	Description	First Appearance
IBSI	Image Biomarker Standardisation Initiative	17
ICC	Intra-class Correlation Coefficient	126
IPEM	Institute of Physics and Engineering in Medicine	54
KS	Krenning Score	35
L-S	Location-Scale	184
LAC	Linear Attenuation Coefficient	9
LAR-SSA	Long-acting Release Somatostatin Analogue	29
LBM	Lean Body Mass	139
LEHR	Low Energy High Resolution	7
LET	Linear Energy Transfer	30
MELP	Medium Energy Low Penetration	8
MID	clinically Minimal Important Difference	171
MIM	Medical Image Merge	12
MIRD	Medical Internal Radiation Dose committee	70
MRT	Molecular Radiotherapy	30
MVLS	College of Medical, Veterinary & Life Sciences	4
NECs	Neuroendocrine Carcinomas	26
NEMA	National Electrical Manufacturers Association	15
NENs	Neoplasms of the Neuroendocrine system	26
NETs	Neuroendocrine Tumours	2
NIST	National Institute of Standards and Technology	13
NPL	National Physical Laboratory	46
NSE	Neuron-specific Enolase	34
OS	Overall Survival	31
OSCG	Ordered Subset Conjugate Gradient	9
OSEM	Ordered Subset Expectation Maximisation	9
PD	Progressive Disease	32
PET	Positron Emission Tomography	1
PFS	Progression-Free Survival	31
PPQ	PRRT Prediction Quotient	35
PR	Partial Response	32
PROs	Patient-Reported Outcomes	168
PRRT	Peptide Receptor Radionuclide Therapy	4
PSF	Point Spread Function	9
PSMA	Prostate-Specific Membrane Antigen	24
PVC	Partial Volume Correction	11

Notation	Description	First Appearance
PVE	Partial Volume Effect	11
QC	Quality Control	3
RC	Recovery Coefficient	11
RMSD	Root Mean Square Deviation	78
ROI	Region of Interest	75
SNR	Signal to Noise Ratio	78
SPECT	Single Photon Emission Computed Tomography	1
SRETV	Somatostatin Receptor-Expressing Tumour Volume	35
SRI	Somatostatin Receptor Imaging	i
SSA	Somatostatin Analogue	29
SSTR	Somatostatin Receptors	26
SUV	Standardised Uptake Value	ii
TLSRE	Total Lesion Somatostatin Receptor Expression	35
TSE	Tumour Sink Effect	138
UFOV	Useful Field of View	99
VOI	Volume of Interest	60
VSF	Volume Sensitivity Factor	60

Part 1:

Introduction and background

...when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

William Thomson, Baron Kelvin of Largs
Lecture on "Electrical Units of Measurement", 1883

Diagnostic nuclear medicine began as a quantitative modality: early studies in the 1930s and 1940s involved in-vitro counting of biological samples taken from patients who had previously been administered a radioactive substance [1]. In-vivo methodologies were soon implemented, measuring the emissions from a patient directly - initially with basic Geiger counters. Methods improved with the use of detectors consisting of scintillation crystals and photo-multiplier tubes [2]. As systems capable of imaging the distribution of radionuclides within the human body were introduced [3], the focus on quantification waned, as optimisation of image quality became the primary concern.

Visual interpretation of nuclear medicine images has continued to be the predominant approach into the 21st century. Applications of quantification in Positron Emission Tomography (PET) demonstrating clinical utility (for example in treatment response [4–6]) re-awakened interest in quantification, though there remain difficulties in the technical aspects of producing quantitative data from Single Photon Emission Computed Tomography (SPECT) imaging [7].

Increasing integration of diagnostic imaging with therapeutic radionuclide treatments (often termed "theragnostics" [8, 9]), along with work showing the advantages of individualised patient dosimetry [10] has increased demand for quantitative SPECT. This heightened interest in the benefits of SPECT quantification is demonstrated by the increasing numbers of publications on quantitative SPECT through the 2000s (Figure 0.1). It must be ensured,

however, not only that quantitative SPECT imaging is accurate and reproducible, but also provides clinical insight and benefit over visual interpretation.

This thesis will attempt to answer questions around accuracy and reproducibility of quantitative SPECT (focussing on one manufacturer's implementation), while introducing and advocating for clinical applications, specifically in the field of Neuroendocrine Tumours (NETs).

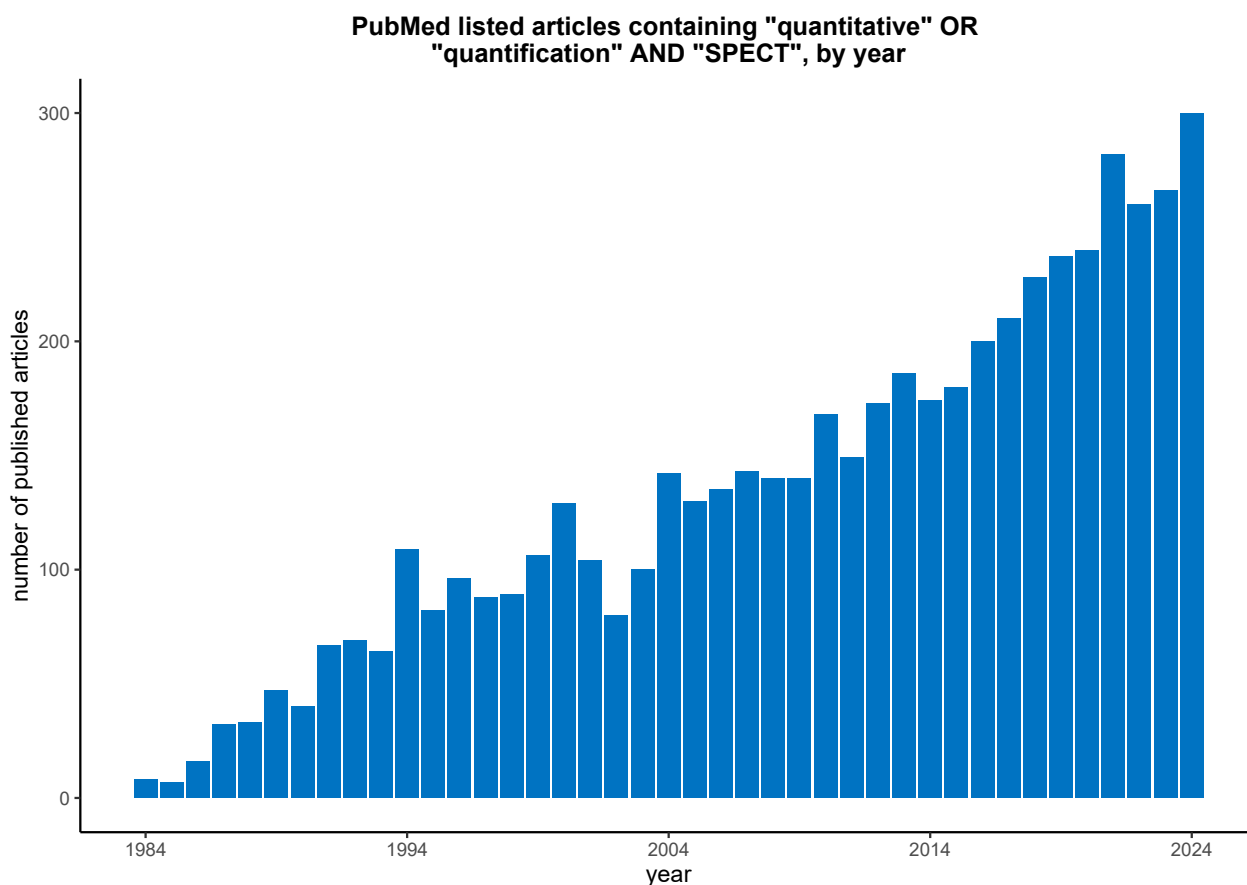


Figure 0.1: Number of publications by year on SPECT Quantification. Results are from a PubMed search (ran 24/04/25) for articles containing the phrases "quantitative" OR "quantification" AND "SPECT".

This introductory part will describe the format of this thesis, as well as the intended aims (Chapter 1), before describing in detail the technical and clinical backgrounds to the use of quantification in nuclear medicine, in particular with application to Neuroendocrine Tumours (NETs).

Chapter 1

Aims of the thesis

Increased clinical interest in quantitative data from nuclear medicine modalities provides opportunities for innovative research, and this thesis will aim to contribute to the knowledge base: specifically in the area of quantitative SPECT calibration and assessment of accuracy, as well as exploring avenues for clinical utility with a focus on uses in Neuroendocrine Tumours (NETs).

This thesis will attempt to meet two overarching objectives arising from a desire to improve the accessibility, usability and clinical value of quantitative SPECT:

- Critically evaluate the accuracy and longitudinal stability of quantitative SPECT calibration using the Siemens Healthineers software tools for Broad Quantification (hereafter referred to as "BroadQuant"), available on the Siemens Intevo family of gamma cameras. It will study the BroadQuant calibration process, assessing reproducibility and accuracy, before evaluating potential Quality Control (QC) tests to determine longitudinal stability.
- Apply quantitative SPECT to clinical applications and demonstrate clinical utility, particularly in the field of SRI for NETs.

Technical objectives will be met by an in-depth testing of the BroadQuant calibration process, investigating each step of the calibration process as outlined by the manufacturer. This work will be described in detail in Part 2. Testing of the quantitative accuracy of the calibration will provide useful information as to the overall accuracy of quantitative SPECT, building on work in the literature which has tested other quantification packages. Repeated testing of accuracy over prolonged periods will be used to assess the longitudinal stability of such quantification packages, aiming to produce advice for other users as to suitable testing frequencies, which has been identified as a gap in the knowledge base in the guidance issued by the European Association of Nuclear Medicine (EANM) [11].

A lack of validated clinical uses for quantitative SPECT has been identified as a major limitation in expanding its routine adoption [12, 13]. Part 3 of the thesis will focus on applications in SRI; this emphasis has been driven by recent developments, namely the introduction of ^{99m}Tc -EDDA/HYNIC-TOC and multiple Field of View (FoV) SPECT acquisitions, which allow for whole-body SPECT-CT scanning; and the clinical approval of

Peptide Receptor Radionuclide Therapy (PRRT) by regulatory bodies. Further motivation for this focus on NETs comes from the results of a local survey on the use of quantitative SPECT, where radiologists and physicians identified the use of quantitative SPECT for ^{99m}Tc -EDDA/HYNIC-TOC imaging as one of their key interests for future development.

A number of clinical projects will be described in Part 3 where quantitative SPECT is applied to SRI – firstly identifying differences in quantitative uptake based on patient's medication status (an important precursor to ensure the quantitative values can be assigned a clinical meaning) before developing a quantitative alternative to the use of visual scoring, with the aim of reducing inter-rater variability. The application of quantification to post-therapy imaging will aim to provide predictive information on the outcomes of patients undergoing PRRT. These projects will be enhanced by the application of textural analysis, which at this point in time has had relatively limited usage in quantitative SPECT. It is intended that these projects will demonstrate new quantitative metrics that will provide additional clinical information over standard measures currently in use. In order to improve the prospects of these clinical results being generalisable and applied to other quantitative SPECT-CT systems, the potential for harmonisation of quantitative data will be investigated.

Ethics and research approval was sought for the clinical projects contained in this thesis through NHS Greater Glasgow and Clyde Safe Haven (with agreement of the University of Glasgow College of Medical, Veterinary & Life Sciences (MVLS) Research Regulation & Compliance Officer), and approval granted under references GSH22NE001 and GSH24RA002.

The final part of the thesis (Part 4) will summarise the findings of the thesis, detailing their novel nature and how they fit in to the existing literature, as well as suggesting ideas for future work.

Chapter 2

Background to Quantification in nuclear medicine

2.1 Introduction

The first use of a radionuclide in medicine dates to the turn of the 20th Century, where shortly after the discovery of radium in 1898 it was used for the treatment of lupus [14]. The development of rectilinear scanners [15] in the early 1950s and then, notably, the (Anger) gamma camera [3] led to the production of images of the distribution of radiopharmaceuticals in patients. The introduction of ^{99m}Tc -labelled radiopharmaceuticals provided new clinical tests [16, 17], and led to a rapid expansion in the clinical role of nuclear medicine imaging.

SPECT was introduced by Jaszczak and others in the late 1970s [18] and allowed display of the 3-dimensional distribution of radiopharmaceuticals within a patient (itself an important advance for improving image interpretation). However, the ability to translate this to quantitative values measuring the radiopharmaceutical uptake was limited, due to the challenges of correcting for photon attenuation and scatter [19], the limited spatial resolution of SPECT systems [20] and the relatively poor outputs from reconstruction by Filtered Back Projection [21].

The quantitative aspect of nuclear medicine was rarely utilised over the next 30 years. It took until PET became widely adopted for quantification to again become a useful clinical aspect of nuclear medicine imaging [22, 23]. PET represented a new frontier where more accurate quantification was possible in the clinical setting – the biological advantages inherent to PET radiopharmaceuticals (particularly ^{18}F FDG [24]) and the introduction of commercial systems fitted with integrated CT for attenuation correction meant quantification was no longer restricted to research facilities.

A new generation of SPECT systems (Figure 2.1) became commercially available at the start of the 21st Century, featuring integrated CT for attenuation correction [25]. Concurrently, there were also improvements in image processing and correction algorithms [26]. Together with a renewed interest in clinical quantification (encouraged by its use in PET), this led to a revitalisation of the study of quantification in SPECT [12]. A number of studies demonstrated that improved corrections for attenuation and scatter, along with enhanced statistical reconstruction techniques allowed accurate measurement of radionuclide activity in phantoms [27–29], and investigations began into a range of clinical applications [30]. With



Figure 2.1: Example of a modern SPECT-CT gamma camera: a Siemens Intevo Bold (Siemens Healthineers, Erlangen Germany).

these hybrid modalities becoming fairly ubiquitous in modern nuclear medicine departments, henceforth all references to SPECT or PET in this thesis will be inclusive of the the relevant hybrid modalities SPECT-CT or PET-CT.

These quantitative studies relied on a series of calibration and processing steps which required intensive user input and advanced scientific and technical knowledge to perform correctly. While invaluable in proving the concept and applicability of quantitative analysis of SPECT, the difficulties associated with performing such processes routinely means that quantification has not yet been widely adopted. As reported in a 2019 survey of the UK nuclear medicine community [31]) 67% of respondents did not yet perform SUV or Bq/ml quantification for SPECT. A possible route to more widespread availability and use of quantification is the introduction of commercial quantification packages which require fewer resources to implement. This focus on accessibility has encouraged a “black box” format, where user interaction is minimised (though protection of commercially sensitive details is undoubtedly also a contributing factor). Limited access to the processes involved in calibration, acquisition and processing may frustrate the efforts of the user to fully understand the methods being applied and limit the value or transferability of clinical findings from such packages. Validation of these packages has been performed in several studies (to ensure accurate results are produced), though their chosen focus means they have necessarily limited their scope to a superficial examination of the packages themselves [32–35]. One aim of this thesis is to

delve in greater detail into each stage of the quantification process to better characterise the quantification package available within the Nuclear Medicine department at Gartnavel General Hospital, Glasgow; namely Siemens BroadQuant [36]. The following sections will focus on the improvements to acquisition and reconstruction processes that have enabled more accurate quantification of SPECT in recent years.

2.2 Gamma camera: physics and technology

The basics of the technology involved in scintillation crystal gamma cameras have changed little since the improvements made by Anger: the reader is referred to several helpful resources covering the components and physics of gamma cameras [7, 37]. Though not part of the experimental work of this thesis, semiconductor-based gamma cameras have the potential to dramatically improve quantitative accuracy over scintillation crystal gamma cameras (for some radionuclides, at least). While these semiconductor cameras are gradually appearing in clinical departments, they are still rare and more work is required to properly evaluate their use in quantification, though there are interesting results published in the literature [38, 39]. All further discussions raised in this thesis will be concerned solely with the standard scintillation crystal gamma camera. Subsequent sections will outline the technical choices made that will be applied throughout the work of the thesis.

2.2.1 Collimator choice

As collimators rely on the physical attenuation properties of dense material to absorb gamma photons on non-perpendicular paths, the technology involved in their construction has changed little. The introduction of cast collimators, replacing the previous method of lead-foil construction [40] is the only significant change since the days of Anger. Converging and pinhole collimators remain limited to specialist applications, and though some manufacturers have promoted parallel-hole collimators with preferential characteristics for reduced scatter or ultra-high resolution [41] they have not been widely adopted. The most important change regarding collimators and quantification has been improved modelling of their performance in the reconstruction process, which will be covered in the discussion of resolution recovery.

The choice of collimator for imaging a particular radionuclide may have implications for quantitative accuracy as well as for image quality. Low Energy High Resolution (LEHR) collimators are usually designed for imaging of $^{99\text{m}}\text{Tc}$ and its 140 keV emission, with little variation between manufacturers in performance, and guidance is available as to the minimum performance standard required for quantitative imaging [42]. The choice of collimator for quantitative imaging of other radionuclides is more nuanced, as some radionuclides will have multiple energy peaks (e.g. ^{177}Lu), or high energy emissions that can produce scatter and

affect measurements (e.g. ^{123}I) which necessitates striking a balance between low energy and medium energy collimator options [43, 44]. The medium energy collimator provided by Siemens, termed Medium Energy Low Penetration (MELP) has been shown to have acceptable sensitivity and resolution for use with ^{177}Lu [44], and will be used in much of the experimental work in this thesis, along with LEHR for $^{99\text{m}}\text{Tc}$.

2.2.2 Orbit type

SPECT acquisition parameter choices for orbit type have generally been limited to a continuous orbit (camera heads move at a constant speed around the patient or phantom; data is acquired throughout) or a “step-and-shoot” orbit (camera head pauses, acquires data, then moves to the next angular position). Early studies determined that continuous acquisition provided greater sensitivity, while resolution was improved with step-and-shoot [45]. The introduction of body contouring to SPECT acquisitions (where infrared or piezoelectric sensors are used to guide the camera head as close as possible to the patient without collision) meant step-and-shoot has been the preferred orbit in most clinical situations for many years. Siemens have recently introduced an option of an orbit that combines continuous with step-and-shoot; termed Acquire During Step (ADS). Siemens ADS mode uses a fixed acquisition time at the defined ‘steps’ in the angular orbit, but in addition acquires counts during travel between steps – these counts are binned into the next projection for reconstruction purposes. Local unpublished work found that ADS allowed for a reduction in total acquisition time whilst maintaining sensitivity when compared with step-and-shoot on a Siemens Bold camera: as has been demonstrated for other manufacturers [46]. A small amount of blurring (consistent with loss of resolution) may be introduced with the use of ADS; however this has been found to be limited and with no effect on visual interpretation [46, 47]. This thesis will investigate quantitative accuracy during ADS acquisitions.

2.2.3 Reconstruction algorithms

One of the first clinically applied methods for reconstruction of tomographic images was Filtered Back Projection (FBP) – the mathematics are covered in detail in many textbooks, but several summaries of its application to SPECT reconstruction demonstrate how it provides only limited image quality and is suboptimal if attempting to correct for physical processes that degrade the image such as attenuation and scatter [48, 49].

Iterative reconstruction allows for the physical properties inherent in radionuclide imaging to be built into the reconstruction process [21] – it has been shown that such modelling of the camera system has benefits over uncorrected reconstructions [50]. Commercial iterative reconstruction packages have for many years mainly provided algorithms using expectation

maximisation - Siemens offer as standard their Ordered Subset Expectation Maximisation (OSEM) algorithm, called Flash3D, which included an early application of resolution recovery with a generic Point Spread Function (PSF) model [51]. Although Flash3D has not been classified by Siemens as capable of quantitative reconstruction, it has been independently proven to be capable of quantitative measurement [52, 53], and other manufacturers use OSEM for their quantification packages [54]. Siemens released their xSPECT package in 2013 [55]. As well as improvements to system modelling and other corrections (discussed in further detail below), the package provided a new reconstruction algorithm based on an Ordered Subset Conjugate Gradient (OSCG) methodology [56]. OSCG uses an additive update term which can affect convergence for a point source [55], rather than the multiplicative updating of OSEM. The application of a merit function (in this case Mighell's Chi square) means OSCG can converge faster to the known activity value than an alternative OSEM algorithm [57], with an effect more pronounced as sphere size increases [55]. The noise pattern in OSCG is different from that seen with OSEM and tends to be more "speckled" than the smoother appearances of noise in OSEM. The use of OSCG as a reconstruction algorithm for quantitative SPECT will form part of the evaluation of the BroadQuant package in this thesis.

2.2.4 Attenuation Correction

Photons emitted deep in phantoms or patients are more likely to be attenuated as they must traverse a greater distance to reach the gamma camera crystal than photons emitted near the surface. Without correction, this results in images with "hot" rims and "cold" centres, limiting both image quality and quantitative accuracy.

Early methods of correction assumed a uniform density in the object [40], limiting the suitability to phantoms and clinical brain studies. Accurately defining the density along the path of the photon can allow for a more accurate correction but requires measurement of the density of the object. Historically, the use of a long-lived radionuclide as a transmission source provided a means to measure the density of a patient but required lengthening the time spent by the patient on the camera bed, and with low count density produced a poor-quality attenuation map [58]. Images from a stand-alone CT scanner could allow for the use of X-ray photons to map the density of the patient, which could be converted into a Linear Attenuation Coefficient (LAC) based on the expected energy of the radionuclide in use [58, 59]. However, spatial registration of the two images made this a difficult task. The introduction of the first hybrid SPECT-CT systems allowed for a CT scan (of a duration of the order of minutes) [25] to be appended to the SPECT acquisition. The sequential imaging, typically with minimal movement of the patient allows for improved spatial registration. Modern SPECT systems now have diagnostic quality multi-slice CT and the CT scans may take only seconds.

The LAC is energy dependent and can be taken from published values of measurements in water (for example, Siemens use 0.1585 cm^{-1} for $^{99\text{m}}\text{Tc}$ [60]). An attenuation map (μ -map) is created from the Hounsfield Unit values of the CT and corrected with the energy-appropriate LAC. The measured LAC is minimally affected by changes in the CT parameters (kV or mAs) [61] and whilst measured LAC in a phantom may deviate slightly from theoretical values, this results in errors of around 2% in the quantitative result for lung and bone tissue for ^{177}Lu [61]. The Siemens OSCG algorithm uses the native matrix CT for AC: as SPECT data is acquired in a 256×256 matrix, it is then translated to the CT coordinate system for reconstruction [51].

2.2.5 Scatter Correction

Compton scattering of gamma photons produces some of the largest sources of error when reconstructing a nuclear medicine image, with up to 50% of the counts acquired in a SPECT scan being due to scatter [40].

Correction for scatter can be incorporated into the LAC when calculating attenuation [19], though the most common methodology currently used for correcting for scatter is the calculation of a scatter fraction using energy windows that border the primary energy window. Either triple energy window (one on each side of the photopeak window) or dual energy window (a single scatter window below the photopeak) methods can be adopted. Siemens apply the method reported by Ichihara et al, which showed that for $^{99\text{m}}\text{Tc}$ a single lower window was suitable [62].

Modelling of the scatter process has been adopted for many applications, originally with simple models [63], but latterly with more advanced models utilising Monte Carlo modelling. Although such models may give good results and are now available commercially in packages that can provide corrections in a timely manner [64], they will not be discussed further in this thesis as they are not utilised in the Siemens BroadQuant package.

2.2.6 Resolution Recovery and System Modelling

Siemens have employed a number of features to improve the system modelling in the OSCG reconstruction process, with the aim of reducing noise and/or improving resolution [51]:

- Measurements of the deflection of the gantry during the SPECT acquisition
- Modelling of the collimator, including the angle of the holes.
- A PSF measurement for $^{99\text{m}}\text{Tc}$ across the entire field of view.
- Decay correction for each acquisition step.
- Attenuation correction during forward and back-projection steps.

The system modelling now available has been shown by Siemens to offer improved resolution when compared with their previous premium reconstruction method with Flash3D and its implementation of system modelling [55].

For 3rd party software manufacturers without direct access to the proprietary technology in the gamma cameras, it is more difficult to provide detailed system modelling to allow for resolution recovery. One approach is to use Monte Carlo simulations to approximate the imaging process, which can be accomplished using the limited data made available by the manufacturer; for example using the collimator parameters to provide enhanced corrections through better scatter modelling [64]. However, this thesis will focus on the BroadQuant package and the corrections used therein.

2.2.7 Partial Volume Effect

The accurate imaging of objects smaller than three times the Full Width Half Maximum (FWHM), where FWHM is measured from the profile of a line source, is impacted by the Partial Volume Effect (PVE) [65]. Counts originating from a particular location are averaged out amongst neighbouring pixels, creating a blurring effect on the visual appearances. The quantitative measurement of activity present in small objects is also affected.

As the dose rate from electrons often emitted from therapeutic radiopharmaceuticals is heavily dependent on the accuracy of the measured activity within a VOI [66], for proper application of dosimetry it is considered necessary to apply a Partial Volume Correction (PVC) to account for PVE. For images that are being assessed only for quantitative analysis, this may be of lesser import.

An example of the practical application of PVC for ^{177}Lu tested a correction using Recovery Coefficient (RC)s derived from spherical phantom inserts which showed an improved absolute measurement error in non-spherical phantom 'lesions' from 26% to 10% [67]. However, the study acknowledges that due to a high target:background ratio of 28:1, spill-in to the lesions did not have a major impact on the PVE, potentially a significant deviation from the clinical situation. The use of specialised anthropomorphic phantoms to produce non-spherical RCs has been broached [67], but the heterogeneous nature of lesion size and shape in clinical situations makes application of the most appropriate RC difficult.

A large number of variables will affect the applicability of a RC curve, such as target:background concentration, object size and shape, and the acquisition and reconstruction parameters of a specific centre [68]. These restrictions, and other practical considerations such as availability of commercial software with capacity to apply corrections to images, have limited the application of PVC in the clinical setting. How these limitations relate specifically to the work of this thesis will be discussed in Section 7.2.7.

2.3 Quantification of SPECT images

Once reconstruction has been performed, the quantitative data must be extracted from the resultant image. This section will discuss the software packages that allow for quantitative analysis, as well as the quantitative metrics that may be used in clinical interpretation.

2.3.1 Vendor-supplied packages for SPECT Quantification

The technical advances that have made SPECT quantification a viable approach (discussed in section 2.2) have generally originated in research-led environments and have struggled to gain favour for routine implementation in clinical departments. This has been due to factors such as the requirement for additional specialised processing or analysis software [28, 69] and the additional time and scientific expertise needed to calibrate the gamma camera and process the acquired data.

For quantitative SPECT to gain widespread adoption, there needs to be suitable applications for the clinical setting (a topic discussed in Part 3) and easily applied techniques for calibration, acquisition, and processing of data. This development would enable smaller departments and those with limited scientific resources to implement quantification in routine clinical practice. Such developments can most easily be provided by the manufacturers and suppliers to the nuclear medicine field, as most clinical departments cannot support development of individualised software solutions. This approach would mirror the route to quantification that occurred in PET: from research applications to software add-ons that were made available as purchase options from the major manufacturers to the current situation where all commercial PET systems come with quantitative calibration and analysis as standard.

The first major steps to expanding access to quantitative SPECT were produced by third-party software vendors, such as Hermes Medical Solutions [32] and Medical Image Merge (MIM) Software [70]. While these packages did help automate the reconstruction process and provided guidance on the calibration of a system, expert scientific input was still required to gauge the accuracy of the calibration and for extracting the required system parameters to provide accurate system modelling for advanced reconstruction algorithms. The cost that departments incur in purchasing such additional software may also act as a deterrent to adoption.

Developments from the gamma camera manufacturers themselves are likely to stimulate wider access to quantification; the advantages of a manufacturer-provided quantification solution over a third-party offering include a calibration process that is seamlessly integrated into the acquisition and processing software of the system and imbedded system modelling, aided by factory testing and the inclusion of commercially sensitive details of the cameras construction. Siemens made the first steps with their xSPECT Quant package [71]. Its use of

National Institute of Standards and Technology (NIST) traceable calibration sources should make the calibration process simple enough for inclusion in the routine QC workflow of a department. The reconstruction (using Siemens proprietary xRecon algorithm, a version of OSCG) can be integrated into current reconstruction workflows without significant user input. However, xSPECT Quant is only available for a small number of radionuclides: originally for ^{99m}Tc , it has now been extended to include ^{111}In , ^{123}I and ^{177}Lu but requires additional sealed sources [71]. The cost of replacement of these calibration sources after they decay below useable levels is a significant outlay for a department. A separate package available from Siemens is the BroadQuant software [36], which allows the user to calibrate for any radionuclide in a two-step process, involving measuring both point source sensitivity and volume sensitivity. BroadQuant will be used extensively in the forthcoming sections of this thesis.

2.3.2 Quantitative metrics

The Standardised Uptake Value (SUV) has found common usage in PET as a (presumed) patient-independent measure of quantified uptake (Equation 2.1). Mainly used with [^{18}F]FDG, the modelling behind the SUV being a reproducible and repeatable value relies on the physiological non-reversibility of the two-compartment model in which the FDG is phosphorylated within the cell [72]. It also relies on [^{18}F]FDG uptake having plateaued within the organ or tumour being studied at the time of imaging [24]. These assumptions on the kinetics of glucose transport and metabolism are known to be dubious and take no account of the micro-environment of the tumour or the many physiological variables inherent in tissue, particularly pathological tissue [24].

$$\text{SUV} = \frac{\text{activity concentration in VOI} \times \text{patient weight}}{\text{administered activity}} \quad (2.1)$$

As an easily relatable metric, many studies have adopted the use of SUV in SPECT-CT quantification [73–75]. The assumption of the same model restrictions as for [^{18}F]FDG may not be appropriate for all radiopharmaceuticals, and when attempting to image at the plateau stage of distribution, a greater focus on strict imaging timepoints than is generally used for SPECT imaging may need to be considered.

There are also questions regarding which particular SUV metric (and which corrections) would be appropriate in SPECT, though this should properly be considered on a per-radiopharmaceutical basis, as there is no guarantee that one answer will be right for all. For [^{18}F]FDG PET there has been discussion that SUV should be normalised to lean body mass rather than weight, as fat may constitute a significant portion of a patient's weight, but the amount of [^{18}F]FDG distributed in fat is negligible [76]. It is not currently clear what

normalisation would be appropriate for the various SPECT radiopharmaceuticals, given differing pharmacological distribution to the various tissue types. The choice of how the activity is measured is also important: value in the maximum voxel (SUV_{max}), the mean of the voxels in the VOI (SUV_{mean}) and the mean of a smaller volume centred on the maximum voxel (SUV_{peak}) are all in common use. While SUV_{max} is generally the chosen metric in clinical practice for PET, studies have shown that SUV_{peak} may be more reproducible, especially in low-volume lesions [77] and demonstrate less bias, particularly in low-count circumstances [78]. This could have important considerations for SPECT quantification, precisely as SPECT is a lower count technology than PET, and in smaller lesions the PVE will be more pronounced in SPECT than in PET due to the poorer resolution [12].

Moving beyond SUV, other imaging biomarkers obtained from statistical and textural analysis of images have been assessed for their usefulness, mainly in PET [79–81]. The lack of standardised quantified data for SPECT has limited the investigation of these biomarkers in SPECT imaging so far, with only a small number of articles published in this area [82–84]. Textural analysis and its application in SPECT will be covered further in Chapter 3.

2.3.3 Accuracy of SPECT Quantification

Determining the accuracy of quantitative SPECT in a clinical setting is challenging, as biological distribution and excretion are unique to each patient and will also vary with different radiopharmaceuticals. It is therefore difficult to accurately calculate the true activity in a particular organ at a particular time. Dynamic PET imaging has allowed for absolute quantification with the use of arterial blood sampling and kinetic modelling to approach the ground truth, but such studies are extremely complex and time-consuming, with SPECT imaging not well suited to dynamic acquisition. The lack of studies on the kinetic profiles of SPECT radiopharmaceuticals also mean dynamic imaging is not a realistic approach to apply to SPECT quantification.

One study found an innovative solution to the challenge of measuring accuracy by imaging the activity concentration of ^{99m}Tc -diphosphonate in the patient's bladder, before obtaining a urine sample from the patient, which could then be measured on a gamma counter to provide the true value [29]. This resulted in an average measured accuracy of within 1.1% across 16 patients (standard error of 8.4%) [29]. The uncertainties in this process arise from the fact that the concentration of radiopharmaceutical in the bladder will change between imaging and the time of urine sampling, as the radiopharmaceutical is excreted through the kidneys into the bladder. A similar study in ^{177}Lu -DOTATATE patients gave an accuracy of 3.9% (albeit with a large range of approximately $\pm 20\%$) [85].

As an alternative to in vivo patient measurements, many studies have used phantoms filled with known concentrations of a radiopharmaceutical. Phantoms may be anthropomorphic in nature, but more commonly are standard phantoms used in gamma camera or PET QC and owned by many nuclear medicine departments, such as the Jaszczak phantom or the National Electrical Manufacturers Association (NEMA) IEC phantom.

Studies have been conducted with many of the most common diagnostic and therapeutic radionuclides and with a range of different gamma cameras and software. Differences in acquisition and reconstruction protocols (and potentially in the underlying performance of gamma cameras) mean that study results may not be directly comparable, but most studies have shown acceptable results for accurate quantification of a uniform source of activity large enough to avoid partial volume effects. Results from some relevant studies performed on Siemens gamma cameras are shown in Table 2.1.

Authors	Radionuclide	Phantom type	Reconstruction	Reported accuracy
Armstrong and Hoffmann [33]	^{99m}Tc	NEMA IEC	OSCG (48i, 1s)	1% in background
Gnesin et al. [34]	^{99m}Tc	NEMA IEC	OSCG (16i, 4s)	<2.5% in background
Ryu et al. [65]	^{99m}Tc	NEMA IEC	Flash3D (4i 6s)	2.7% in background
Ryu et al. [65]	^{177}Lu	NEMA IEC	Flash3D (4i 6s)	6.5% in background
Tran-Gia and Lassmann [57]	^{177}Lu	Uniform phantom	OSCG (48i, 1s)	3.3% in background

Table 2.1: Accuracy of SPECT quantification using Siemens cameras as reported by studies in the literature (i=iterations, s=subsets).

Use of the OSCG algorithm has been shown to provide better quantitative accuracy for ^{177}Lu compared to Flash3D [57] across differing geometries and phantom materials [61]. The noise in the image has been reported to be higher for OSCG compared to Flash3D, but benefits from reduced spill-out of counts from high-activity objects [57]. Quantitative accuracy of both ^{99m}Tc and ^{177}Lu will be investigated extensively in Chapter 7, to define preferred acquisition and reconstruction parameters for clinical studies.

2.4 Relevance to the thesis

This chapter has introduced many of the key technologies and corrections that have made quantitative SPECT a possibility. The work in Part 2 of this thesis will evaluate the application of the quantitative calibration process for Siemens BroadQuant (Chapter 6) and assess the accuracy and stability of the calibrations (Chapter 7), before quantitative SPECT is applied clinically for imaging of NETs in Part 3.

Chapter 3

Review of Textural Analysis and Radiomics

3.1 Introduction

One method to extract further information from images than can be discerned with the naked eye is textural analysis. Mathematical processes are used to classify image features based on the values of pixels, and their relationships to neighbouring pixels. Some of the first applications of these techniques were to aerial photography to differentiate features in landscape, such as the work by Haralick et al [86] where many of the commonly used features were mathematically defined.

When applied to large tranches of medical imaging data, the use of textural analysis features is often termed “radiomics” [87] and aims to utilise the large quantities of information contained in each clinical image. The features extracted from the images may be combined with clinical information or other biological markers ("biomarkers"), and used to predict response to a treatment, or a patient's likelihood of survival. The literature in radiomics is well supplied with review articles (one review article puts it at 30% of the articles recently published on radiomics [88]); the sheer volume of opinions will naturally lead to discrepancies in recommendations that will propagate into a variation in practice in studies. However, as will be discussed later, standardisation is especially important in the field of textural analysis and radiomics. The Image Biomarker Standardisation Initiative (IBSI) provides a framework [89], while application to nuclear medicine should be aided by recent EANM guidelines [90].

The use of quantitative images is intrinsic to the textural analysis process, as pixel values must be standardised so as to be comparable between repeat imaging and for imaging on different equipment. Hence PET was the first nuclear medicine modality where textural analysis was studied in great detail [91, 92]. SPECT has generally lagged in both technical and clinical implementation of textural analysis, due to limited access to quantitative images as well as the poorer resolution and sensitivity which introduces greater variability in the features [93].

There are some quantitative measurements that can be taken from images before the processing required for textural analysis; these generally fall into the categories of intensity measurements (for example SUV) and shape or volume. Whilst shape-related features such as compactness or surface area may be less dependent on acquisition or reconstruction

parameters [94], SUV_{max} is well-known to display variability which can affect accuracy [95, 96].

3.2 Technical factors of textural analysis

Before the textural analysis features can be measured, the images concerned must be prepared to allow the features to be extracted in an accurate and reproducible manner [91, 97]. This thesis will draw heavily on settings that have been confirmed to work with [^{18}F]FDG PET images, specifically with the LifeX analysis software (Institut Curie, Paris, France) [88, 98] – though adaptations will be made to account for the specifics of SPECT imaging and the radiopharmaceuticals being imaged: details on the specified settings and parameters will be included in the methodology sections of the relevant chapters.

3.2.1 Segmentation

The delineation of areas of interest (usually tumour or disease sites showing increased uptake of a radiopharmaceutical) can be performed in multiple ways. Manual segmentation by an expert operator can be time-consuming and has the disadvantage of being susceptible to both intra-operator and inter-operator variance. Semi-automated methods reduce variance but require the setting of a seed point which can introduce variability, and a threshold level (usually based on a maximum value) to define the end of the volume, which assumes a clear boundary that is unlikely to be present in clinical conditions (especially in PET and SPECT imaging where Partial Volume Effect affects the boundaries). The preferred threshold level will also be modality, disease, and even anatomy-specific; multiple studies in PET have recommended different threshold levels, though values around the 40% to 43% threshold have been shown to provide the least inter-observer variability [99]. Fully automated segmentation methods may be more reproducible (eg Fuzzy Local Adaptive Bayesian or Deep Learning methods [100, 101]) but often require large training sets and expert observer input to labelling and clarifying if areas of uptake are physiological or pathological [102].

3.2.2 Spatial resampling

Image processing algorithms assume isotropic voxels when calculating textural analysis features (i.e. assumption is that the "nearest neighbour" is the same distance away in all dimensions). Therefore, an important initial step is spatial resampling, achieved by interpolation of the data (for which there are multiple methods available, see [88]). For example, where a voxel may have a larger dimension in the z-direction, it could be resampled to match the size of the voxel in the x and y-dimensions (work in this thesis will use isotropic voxels of $1.95 \times 1.95 \times 1.95$ mm). The voxel size and method of resampling will have an effect

on the values of many features (specific examples include GLCM Energy and compactness [103]), so it is important to define the choice when commencing a study and be cognisant when comparing to other studies published in the literature, though there has been little consensus to define a standardised voxel size to this point.

3.2.3 Intensity discretisation

As there are large number of grey levels in a PET or SPECT image, the pixel values must be discretised before further analysis can take place. Normalisation should preferably be done on an absolute basis (i.e. with fixed min and max bounds for all studies), rather than on a relative basis (using min and max of the individual lesions) [104]. Utilising a fixed bin width (rather than a fixed number of bins) is most common in PET [80], primarily as it reduces the correlation of features with volume [105]. The preferred bin width has no agreed convention [97], though suggestions for a sensible value include 0.3 SUV/bin [88].

3.2.4 Volume size

Different minimal sizes have been suggested for segmented volumes in PET to ensure that extracted features are reproducible. A high minimum value of 45 ml was suggested by Brooks and Grisby [106], though focussed work on the LifeX platform has suggested a minimum number of 64 voxels is more appropriate [88]. When applied to the SPECT imaging detailed in this thesis, and in particular the 256×256 matrix used, this equates to pixel size of 1.95 mm and volumes around 0.5 ml. This issue will be discussed further in Chapter 8.

3.3 Definitions of Textural Analysis features

Textural analysis features are often broken down into three classes (though 2nd and 3rd order features may be grouped as higher-order features [107]):

- 1st order features – these express how the intensity of voxels is distributed in the image, using global heterogeneity measures of the volume in question determined from histogram analysis. As such they are measures of frequency with no information on the spatial distribution within a volume (and hence sometimes not included as true “textural” features [88]). Some examples of 1st order features include skewness (describing the asymmetry of the histogram distribution) and kurtosis (representing the frequency of outliers deviating from the mean), as demonstrated with simulated data in Figure 3.1. An example of the potential range of normalised values can be seen for a practical implementation in Figure 11.5.

- 2nd order features – examine the spatial distribution of voxel values in relation to neighbouring voxels. They provide local information relative to a given voxel. Examples include Grey Level Co-occurrence Matrix (GLCM), where the number of pairs of neighbouring voxels with the same grey level is recorded in a matrix. Various summary statistics can be derived from the matrix, such as the homogeneity.
- 3rd order features – give measurements of trends in the voxel values, for example the maximum number of voxels in a row with the same value. They are measures of regional patterns in a volume. An example of higher-order feature would be Grey Level Run Length Matrix (GLRLM). As demonstrated in Figure 3.2, the value of each voxel is compared with the neighbouring voxel (this can happen in 2 or 3 dimensions). If the neighbouring voxel is of a different grey level, the run length is equal only to one. If the grey level is the same, the next voxel is also compared, to determine if the run length is 2 or more. Once the overall length of the run is determined, it is added to a matrix containing the total number of run lengths per each grey level. This matrix can then be normalised for volume size, and summarised in a multitude of ways, such as how strongly short or long runs are favoured (e.g. Short Run Emphasis, GLRLM SRE).

Simulated distributions demonstrating possible patterns for selected 1st order features

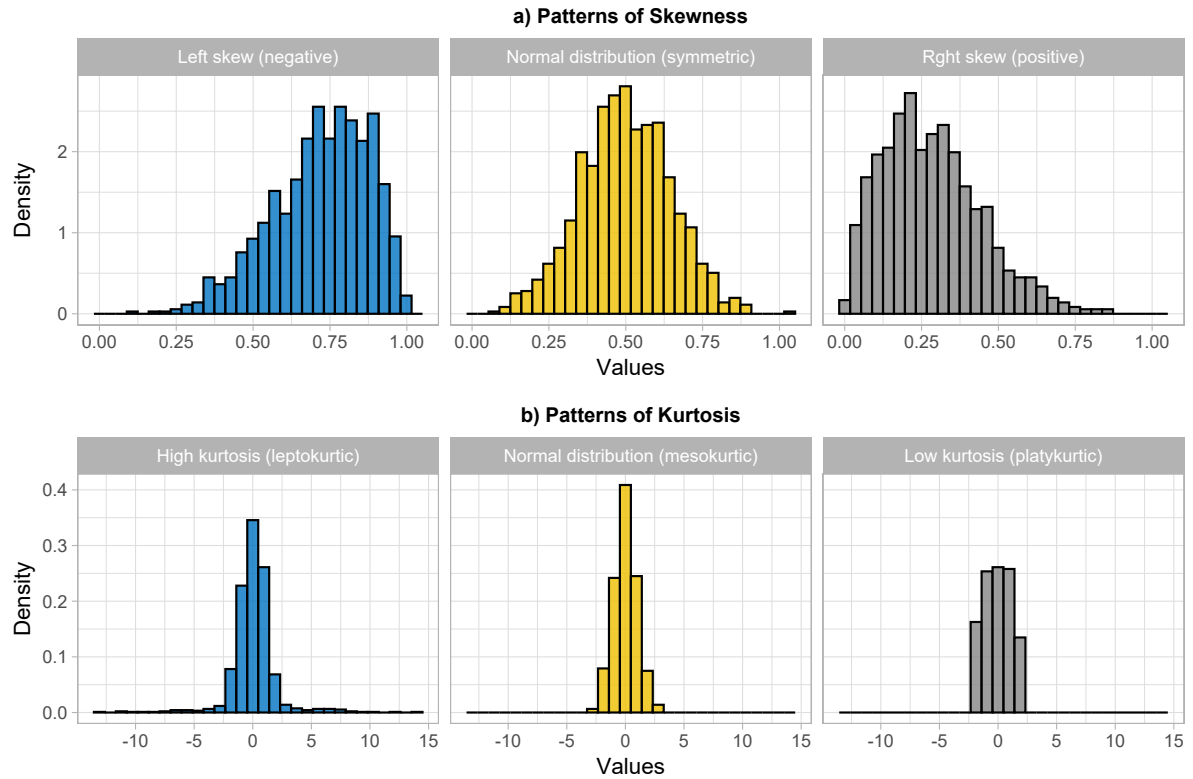


Figure 3.1: Simulation data demonstrating potential patterns of histogram distribution for a). different patterns of skewness; and b). different patterns of kurtosis.

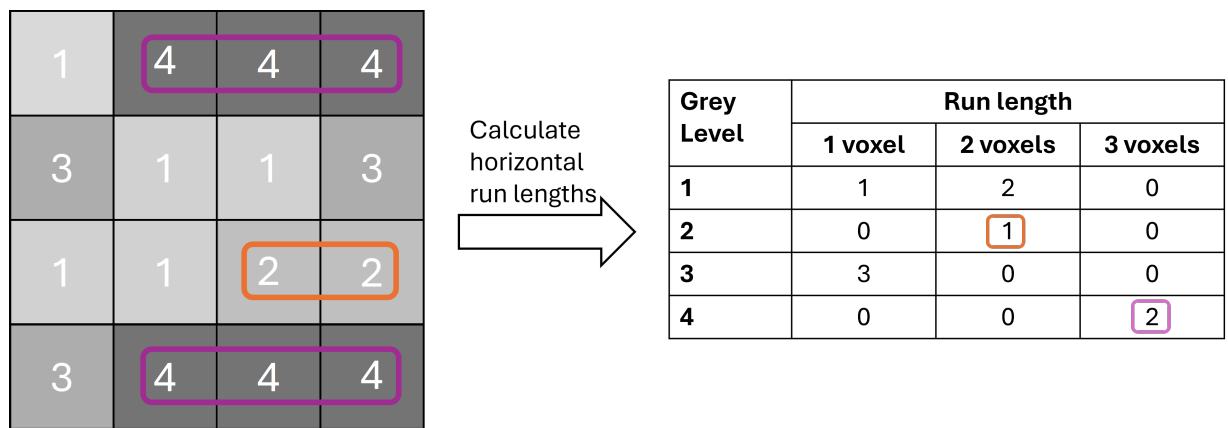


Figure 3.2: Diagram demonstrating an example of Grey Level values in image pixels, with calculation of Grey Level Run Length and conversion to a matrix, prior to summation.

Once textural analysis features have been calculated, their values can be combined with clinical data and used to define a model predicting a clinical outcome. This could be, for example, to determine a patient's response to a drug or the time to progression or death following a diagnosis. A model can take many forms, from simple linear regression to machine learning or deep learning models [81] – but should relate to physiological or pathological features. The challenges of model selection (such as interpretability [108] and generalisation [90]) go beyond the scope of this thesis.

3.4 Validation of Textural Analysis features

A significant issue in applications of textural analysis to nuclear medicine so far is that studies have not proven to be reproducible outwith the centre where they were conducted, and prognostic markers have varied from one study to the next [109–111]. Perhaps the definitive declaration on this effect was made by Pencina et al, who stated that “*A biomarker with great potential for improvement in risk prediction in one study can prove practically useless if it cannot be reliably measured elsewhere*” [112]. This lack of reproducible results can potentially be explained by two separate conditions: model over-fitting and the inclusion of non-robust features. That is not to say that small-scale or single-centre studies have no part to play: the testing of hypotheses as to which features may prove valuable, and the study of less-common cancers are important in identifying areas for further multi-centre study, and should continue to be shared amongst the nuclear medicine community, providing suitable caveats are given as to their limitations.

3.4.1 Model over-fitting

Model over-fitting, sometimes referred to as “*the curse of dimensionality*” [90], can occur if a large number of features are used to create a model, particularly with small sample

sizes. While it can be attractive to test as many features as can be calculated, it can lead to situations where the number of features exceeds the sample size, and result in incorrect classification of features as predictive [88]. The user can of course incorporate more features without causing dimensionality issues by increasing their sample size. However, this is often a difficult task to achieve in single-centre nuclear medicine studies. There is no definitive answer as to the “correct” number of features with respect to sample size [97], though a widely used ‘rule of thumb’ expecting ten events per feature has been deemed overly conservative for logistic and cox regression [113]. Examples of the ratios of data points to features seen in the PET Textural Analysis literature cover a wide range, including 1:1 [79], 3:1 [114], 5:1 [115] and 10:1 [116].

3.4.2 Feature robustness

It is important before beginning clinical studies to define a set of features that will prove technically robust across as many differing circumstances as possible [117, 118]. The main issues to consider are:

- Feature simplicity: An ideal feature is one which has a relatable formulation for the problem in question, and thus can be easily explained. For molecular imaging, and textural analysis of tumours, volume would perhaps be the most obvious candidate. Removing features that may be predictive, but also correlate strongly with a simple feature, helps with the problem of dimensionality. Several studies have demonstrated that tumour volume can be strongly correlated with textural analysis features or models that were identified as predictive [111, 115, 119], and hence the list of features can be reduced, with a preference for volume as the simplest feature.
- Reproducibility: features that exhibit high test-retest variability are unlikely to be suitable for building textural analysis models, as results are unlikely to be repeatable, or validated using external data [97]. This has proven to be a challenge in PET, with some features showing high variability between sites [120].
- Harmonisation between centres: this is one focus of the recent guidelines for radiomics in nuclear medicine [90], though it does not recommend optimised acquisition or reconstruction parameters outwith meeting EARL harmonisation requirements for PET [121]. Studies have shown that for some features a significant difference will still exist between harmonised systems [122], even if meeting EARL criteria.
- Other factors that influence the robustness of textural analysis features and models include image segmentation methods [123], imbalanced data sets and reuse of development data during the validation step [81].

The EANM guidelines suggest phantom studies may be used to identify features affected by a dependence on volume, or that suffer from poor reproducibility. The capability of physical phantoms to provide a suitably accurate representation of the physiological and pathological uptake of a radiopharmaceutical in a patient is limited. Some studies have chosen to avoid these limitations by using software simulations to represent the theoretical uptake of radiopharmaceuticals in patients [124, 125]. The applicability of such simulations to real life may be questioned. Ideally, such simulations will be bolstered by a physical experiment replicating the patient as closely as possible. Optimising the structure of a physical phantom may take several forms: the use of radioactivity mixed with gels can avoid the ‘cold’ walls seen with glass or plastic spheres [126]; 3D printing of lesions can provide irregular shapes that better replicate tumour lesions [127] when compared to the more common spherical lesions found in the commercially available phantoms; and at the greatest extent, 3D printing of a complex multi-organ phantom can provide the closest replication of actual patient anatomy [128].

While the structure of phantoms is improving to better represent the anatomy of patients, the functional nature of both patients and the radiopharmaceutical distribution process is more difficult to replicate: there are several methods of replicating some physiological processes such as respiratory [129, 130] or cardiac [131] motion. However, the heterogeneous patterns of radiopharmaceutical distribution that vary with time are a challenge of a different order. In particular, tumours may show inhomogeneous distribution of radiopharmaceuticals. Evidence in single-centre studies demonstrates this can predict outcomes, suggesting that this inhomogeneity is clinically relevant [132] (though the concept of image inhomogeneity being a practical representation of biological inhomogeneity is controversial [133]). Inhomogeneity can be difficult to test in phantoms that generally consist of containers uniformly filled with water mixed with the radionuclide and alternative phantom structures are required [93]. A method to better test inhomogeneity with a phantom is investigate in Chapter 8.

3.4.3 Model selection

The EANM guidelines discuss how the performance of different radiomics models can be analysed, referencing that a radiomic model should also be compared to the simpler clinical parameters (such as SUV) and also to the interpretation of clinical experts [90]. While many potential tools can be used to compare different models, from simple sensitivity and specificity measurements, to more complex statistical tests, all have strengths and weaknesses [88]. This thesis will utilise the Akaike Information Criterion (AIC) where applicable [88, 134].

The AIC is used to compare two or more models, with the lower AIC value determining the model which best represents the process, by ‘losing’ the least amount of information in the

act of modelling. Whilst it scores the models on goodness of fit to the data, it also penalises the use of a larger number of factors, hence reducing the issue of overfitting of data [134]. AIC has previously been applied in radiomics studies using SPECT or PET data [135, 136], confirming its suitability for the data used in this thesis. The adoption of AIC into many statistical analysis packages (e.g. survival analysis in R [137]) allowed for direct application to the data gathered and an adaptation for small sample size, AICc, makes it particularly relevant for the challenges faced with quantitative SPECT data.

3.5 Clinical studies utilising textural analysis

3.5.1 PET imaging studies

The main focus of research using textural analysis has been centred around [^{18}F]FDG [132], with studies into various cancers such as lung cancer [104, 116] and thyroid cancer [79]. As studies progressed, comparison of prognostic features between cancer types showed that some features were specific only to certain tumour types [138, 139]. Recent studies have highlighted uses for textural analysis with other radiopharmaceuticals, including ^{68}Ga -DOTA-peptides [140–142] and Prostate-Specific Membrane Antigen (PSMA) ligands [143, 144]. The rise in molecular radiotherapy and its combination with diagnostic imaging (“theragnostics”) mean that features predictive of response to expensive therapies using matched radiopharmaceuticals is of increasing importance [145, 146].

3.5.2 SPECT imaging studies

Oncological applications are perhaps the area where we expect most utilisation of textural analysis in SPECT, following the pattern of PET studies. However, at present there have only been a limited number of studies performed. The paper by Wetz et al [84] investigated a clinically relevant radiopharmaceutical ([^{111}In]In-DTPA-Octreotide) and applies textural analysis to a relevant clinical problem, namely using a diagnostic scan to predict a patient’s response to therapy, where it shows that asphericity can be a better predictor of response than current semi-quantitative measures. The study is limited in its applicability due to [^{111}In]In-DTPA-Octreotide being superseded by superior radiopharmaceuticals (namely $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC for SPECT and various ^{68}Ga -DOTA-peptides for PET) and the studies focus only on asphericity and not other textural features. A model utilising [$^{99\text{m}}\text{Tc}$]Tc-Sestamibi SPECT textural features proved more accurate in identifying renal oncocytomas from benign renal masses than visual interpretation only, when combined with the current gold standard of histopathology, potentially reducing unnecessary surgery [147]. De-risking treatment for paediatric neuroblastoma patients was a potential use of a radiomic model derived from

[^{123}I]-metaiodobenzylguanidine imaging, demonstrating a higher success rate than visual analysis using a standardised semi-quantitative score for predicting event-free survival [148].

Rahmim et al [83] demonstrated that several textural features in [^{123}I]-loflupane imaging correlated with clinical outcomes for Parkinsonism patients, though the study did not compare to current semi-quantitative methods such as comparison to normal database and therefore the additional value of textural analysis is unclear. A more recent study using both textural analysis of striatal uptake and deep learning showed strong predictive value for staging of Parkinson's Disease [149].

Cardiac imaging has a significant share of the textural analysis studies in the SPECT field, including phantom [150] and clinical [151] studies on reproducibility. Clinical studies have looked at risk stratification in coronary artery disease [152] and cardiomyopathy [153]. The addition of clinical factors to textural analysis features is highlighted as an important feature in improving the accuracy of a multi-variable model in work on coronary artery calcification by Ashrafinia et al [154].

There have been recent innovative studies utilising SPECT radiomics for indications that differ from the standard cardiac or oncology applications. For example, for diagnosing Grave's ophthalmopathy a radiomic model using [$^{99\text{m}}\text{Tc}$]Tc-DTPA SPECT-CT was more accurate than MRI [155]. Radiomic models of [$^{99\text{m}}\text{Tc}$]Tc-MAA SPECT-CT out-performed radiomic features from CT alone when diagnosing pulmonary perfusion deficiency in lung subsegments [156].

3.6 Relevance to the thesis

Textural analysis has potential advantages for evaluating the 'hidden' data contained within medical images. Use of textural analysis in PET has shown it may have clinical benefits, in terms of response or survival prediction. The growing trend to study textural analysis in SPECT suggests that an increasing number of uses will be discovered. After appropriate multi-centre validation there is a hope that SPECT radiomics will have a real clinical impact. This thesis aims to incorporate textural analysis into quantitative SPECT imaging in a rigorous, reproducible manner, by identifying robust features before implementing them in single-centre clinical studies using SRI.

Chapter 4

Neuroendocrine Tumours and the role of nuclear medicine

4.1 Background to Neuroendocrine Tumours

The neuroendocrine system is involved in longer-term control of regulation of bodily processes via the secretion of hormones. Neuroendocrine cells can be found either clustered in the endocrine organs (e.g. thyroid, pituitary, etc) or scattered diffusely through other organs (e.g. epithelial lining of the intestine) [157]. Neoplasms of the Neuroendocrine system (NENs) have an incidence of 8.6 per 100,000 in the UK [158]. Increasing prevalence of NENs in the developed world can partially be explained by more sensitive diagnostic tests and increased knowledge leading to earlier detection, though it is suspected there are underlying environmental factors also driving the rise [159]. If these neoplasms are well differentiated, they are known as Neuroendocrine Tumours (NETs). De-differentiated, aggressive neoplasms are known as Neuroendocrine Carcinomas (NECs) and make up around 10-20% of incidences. The majority of NETs (between 60-70% [160]) arise in the gastroenteropancreatic system (GEP-NETs). They are also commonly found in the lungs [161] (where they may be termed lung carcinoids), but can theoretically arise anywhere in the body where neuroendocrine cells are present.

4.2 Imaging of NETs

The hormone somatostatin is produced in the pancreas and gastro-intestinal system and is involved in the regulation of the endocrine system [157]. Many well-differentiated NETs over-express Somatostatin Receptors (SSTR) sites [162]. Five SSTR sub-types have been identified in humans (SSTR1 - SSTR5) [163], which demonstrate variability in expression at different locations in the neuroendocrine system, and varying biological functions [164]. Expression in the majority of tumours is dominated by SSTR2, followed by SSTR1, with SSTR5 being the least common [164].

Imaging of radiopharmaceuticals with an affinity for certain SSTR sub-types can be used to image well-differentiated NETs, referred to as Somatostatin Receptor Imaging (SRI). SRI is known to offer improved sensitivity over structural imaging, particularly for metastatic disease

[165]. A recent meta-analysis found 23 out of 24 studies demonstrated that PET imaging with SSTR-peptides was useful in guiding patient management, over and above structural imaging alone [166]. This is reflected in the professional guidelines, which recommend the use of SRI in tumour staging and re-staging [165] and for determining eligibility for PRRT [167]. Generally, SRI radiopharmaceuticals are targeted to have the highest affinity for SSTR2 [168, 169], due to its dominant expression in tumours. In-vitro testing has suggested there are differences in the rate of internalisation of chemicals attaching to the receptor sub-types, which has implications for the suitability of targeting radiopharmaceuticals [163].

The most commonly-used SRI today are of the agonist class, with one of the first examples used being ^{111}In -Octreoscan with planar gamma camera imaging (Figure 4.1), and later using SPECT [170]. The introduction of $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC (Tektrotyd – POLATOM; Otwock, Poland [171]) results in reduced radiation dose and improved image quality due to the use of LEHR collimators (Figure 4.2). The improvement in image quality (as well as increased use of SPECT-CT) has improved the patient experience, allowing for a one-day imaging protocol without laxative, rather than the multi-day protocol with laxative required for ^{111}In -Octreoscan [170], as abdominal uptake can be more clearly identified as tumour-related or bowel excretion, rather than relying on comparison of 24- and 48-hour images to demonstrate clearance by bowel excretion. Direct comparison of the two radiopharmaceuticals demonstrated that $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC showed greater target:background uptake, higher sensitivity and improved overall diagnostic efficacy [172, 173]. For PET imaging, there are a choice of pharmaceuticals now widely used, with improved SSTR affinity, sensitivity and accuracy over $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC or ^{111}In -Octreoscan, namely ^{68}Ga -labelled DOTATATE, DOTANOC and DOTATOC [169, 174–176], collectively referred to as ^{68}Ga -DOTA peptides. Much recent work has been focused on the development of SSTR antagonists [168]. These radiopharmaceuticals demonstrate increased binding to SSTR2, reduced disassociation and with improved target:background ratio over the agonists [177]. Examples of some antagonist radiopharmaceuticals currently at early stages of workup include ^{18}F AlF-NOTA-LM3 [178], ^{68}Ga -DATA5m-LM4 [179] and ^{68}Ga -NODAGA-LM3 [180]. Beyond SRI, bombesin analogues may have a role in targeting different peptide receptors specific to certain tumour types [168], and specifically for non-GEP NETs.

NECs are unlikely to over-express somatostatin receptors due to their highly de-differentiated nature and hence often do not show increased uptake on SRI. ^{18}F FDG PET imaging is more suitable for NECs, with their higher proliferation requiring high glucose metabolism – though the often-assumed binary relationship between ^{18}F FDG uptake and SRI uptake (where a tumour will show high uptake for either ^{18}F FDG or SRI, dependent on the aggressiveness and differentiation of the tumour) and their correlation with tumour grade is not always

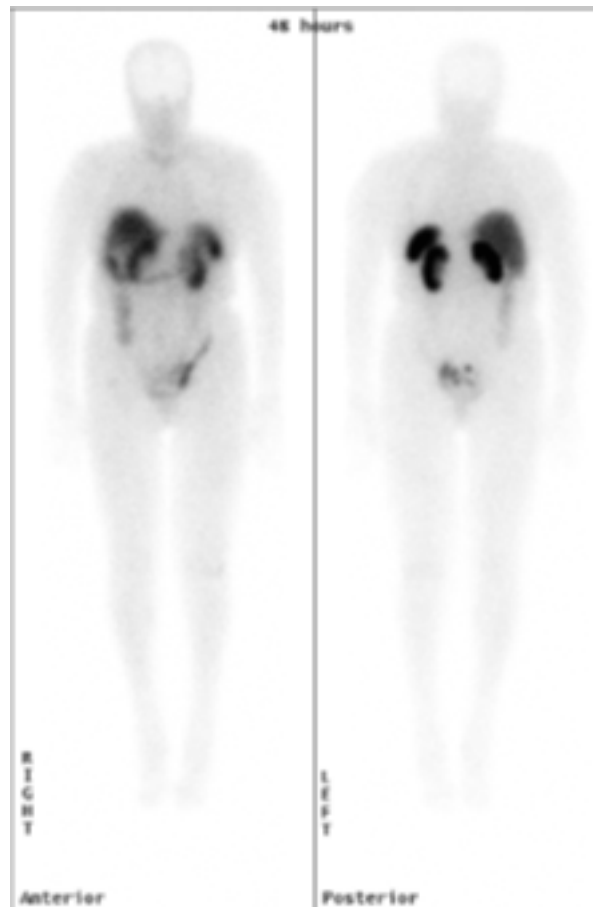


Figure 4.1: ^{111}In -Octreoscan planar whole-body scan, acquired at 48 hours post-administration. It demonstrates only normal physiological uptake.

straightforward [181]. Another agent that may have a role in the imaging process is ^{18}F -DOPA, for NENs with variable SSTR expression [182].

There is a confirmed role for functional imaging in NETs, both with SRI and FDG PET: guidelines recommend their use in the staging of confirmed NETs, in the identification of occult primary, and in determining suitability for therapies that target somatostatin receptor sites [165, 182]. There is also a defined scope in monitoring response to therapy [168]; the use of SRI as a prognostic marker, or as a predictor of response to therapy will be discussed in more detail in Section 4.3.4 of this chapter.

The varied field of potential SRI radiopharmaceuticals means there is a continual flow of publications examining the pharmacology, chemistry and diagnostic accuracy of the most recent promising candidates. However, this understandable focus on the cutting edge of imaging and radiopharmaceuticals does leave gaps where the utility of SRI in routine clinical service is neglected, and those without access to the most up-to-date technology or radiopharmaceuticals are somewhat neglected. This thesis will focus on an SRI radiopharmaceutical in $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC that should be widely accessible due to its

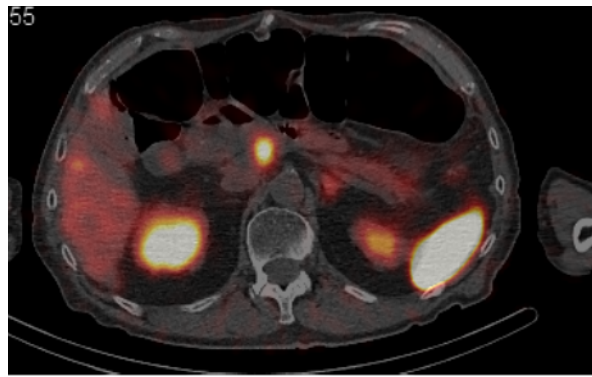


Figure 4.2: ^{99m}Tc -EDDA/HYNIC-TOC SPECT-CT scan, fused axial slice showing increased uptake in pancreatic lesion and physiological uptake in spleen, kidneys and left adrenal gland.

commercial availability, the ease of distribution and reconstitution due to the use of ^{99m}Tc as its radionuclide, and the ubiquity of gamma cameras in both the developed and developing world.

4.3 Treatment of NETs

As per the European Consensus guidelines [183], treatment pathways for NETs depend on the presence of metastases; surgery is the first-line option for patients without metastatic spread. This emphasises the importance of SRI in the patient workup for identifying metastatic spread, while it is also required in determining eligibility for systemic therapies which also target SSTR over-expression. Both “cold” somatostatin analogues and PRRT require SRI to show increased uptake at all tumour sites, otherwise other therapies such as the targeted drugs Sunitinib or Everolimus [183], or chemotherapy [162] should be considered. In some patients there may be consideration of liver embolisation or even liver transplant [183].

4.3.1 Somatostatin analogues

Somatostatin is a hormone secreted from cells of the central nervous system and gastrointestinal tract [162], consisting of a chain of amino acids [163]. Due to the short half-life of natural somatostatin in blood, it was unsuitable for use in therapy (to block somatostatin receptor sites) or for use in imaging. Therefore, a range of Somatostatin Analogue (SSA)s were developed [163]. An inhibitive effect on NETs growth has been proven in patients treated with the SSA octreotide [184], though the various SSA have often been found to stabilise a patient's condition, rather than reduce tumour burden [185]. Nevertheless, consensus guidelines recommends SSA therapy as a common first line treatment for metastatic NETs [183], proving particularly useful for symptom control. The development of Long-acting Release Somatostatin Analogue (LAR-SSA) has made delivery of therapy to patients easier, as it allows for monthly administration via an intra-muscular

injection, with the SSA encapsulated in a dissolving polymer delivering a steady dosage of the SSA to the patient [186].

4.3.2 Peptide Receptor Radionuclide Therapy

Pharmaceuticals that have high affinity for somatostatin receptor sites can be labelled with beta or alpha emitting radionuclides, to be used for Molecular Radiotherapy (MRT) to treat NETs, as they overexpress somatostatin receptors. PRRT was originally performed with high activities of ^{111}In Octreoscan, where the highly targeted nature of the treatment, due to the high affinity for the somatostatin receptor, showed initially promising results [187]. However ^{111}In is not an ideal radiopharmaceutical for therapeutic use due to relying on Auger and conversion electrons to cause cell damage [187]; these have a very limited range [188], which will limit the cells that will receive a killing dose. This limited efficacy has been shown in the lack of trials showing reduction in tumour size for ^{111}In -Octreoscan therapy patients [189].

These disadvantages prompted work on developing radiopharmaceuticals labelled with either ^{90}Y or ^{177}Lu , which are beta-emitting radionuclides with few or no gamma emissions (0% for ^{90}Y [190] and 10.6% intensity for the 208 keV peak of ^{177}Lu [191]). The differing ranges of the beta emissions (max range 2 mm for ^{177}Lu and 12 mm for ^{90}Y [192]) result in a different Linear Energy Transfer (LET) for ^{177}Lu and ^{90}Y . This means that for smaller tumours ^{177}Lu is the optimal choice of radiopharmaceutical, whereas for larger tumours ^{90}Y will provide better cell-killing properties as the inter-cell crossfire effect will play a larger role [193]. This effect involves the beta emissions from cells that have internalised the therapeutic radiopharmaceutical delivering cell-killing doses to nearby cells that are not expressing receptors sites. Some centres have used a mix of beta emitters to try and optimise the dose delivered to tumour, targeting both large and small lesions [194].

Peptides labelled with ^{90}Y have significant downsides that have driven the move towards ^{177}Lu -labelled PRRT. The practicalities of handling ^{90}Y , with its longer beta range, make it less appealing from the point of view of staff radiation exposure and in the event of an accidental spill. Post-therapy dosimetry of ^{90}Y requires either (poor-quality) bremsstrahlung imaging with SPECT, or the use of PET imaging, where available [195]. A high renal dose has been observed in multiple dosimetric studies of ^{90}Y -DOTATOC [196], which can be a significant factor in limiting the overall activity able to be administered. Long-term follow-up also suggests that Overall Survival (OS) for patients treated with ^{90}Y -DOTATOC is lower than those treated with ^{177}Lu -DOTATATE [197].

The limitations of other radionuclides have meant many centres now concentrate on ^{177}Lu -labeled PRRT therapies, with regulatory approval for ^{177}Lu -DOTATATE following from the results of the Phase 3 NETTER-1 trial [198]. The radiopharmaceutical ^{177}Lu -DOTATATE

(Advanced Accelerator Applications, France; trade name Lutathera) is delivered in four fractionated treatments of fixed activity, 8 weeks apart [199], with fractionation allowing recovery of organs at risk. The main concerns are renal function, liver function and both white blood cell and platelet count, which all must be reviewed between each fraction of therapy. Renal toxicity can be mostly avoided through concurrent administration of amino acids, which limits the tubular re-absorption of the radiopharmaceutical [200]. Long-lasting renal damage is only reported in patients who had restricted renal function before PRRT [201]. As a precaution, the patient's Glomerular Filtration Rate (GFR) should measure >60% of age-corrected normal lower limit [167]. Due to the 6.6 day half-life of ^{177}Lu [202] there is a "continuous superfractionation" effect [203], as dose is delivered to the tumour site at a slower rate than external beam radiotherapy. The initial results of the NETTER-1 trial showed improved Progression-Free Survival (PFS) over LAR-SSA alone, as well as higher radiological response [198]. However, the 5-year review of NETTER-1, did not show a statistically significant difference in Overall Survival (OS) against LAR-SSA, though the median OS was 11.7 months longer [204]. Following the preliminary PFS results of the NETTER-1 trial, approval from various regulatory bodies followed, including the Scottish Medicine Consortium approving the use of ^{177}Lu -DOTATATE in clinical practice in Scotland in July 2018 [205].

The direction of future developments in PRRT is focussed towards the use of alpha-emitting therapies, following the theory that improved dose distributions and increased cell apoptosis [206] will result in beneficial outcomes for patients. This may be of particular benefit in cases with extensive skeletal metastases, where the number of fractions of ^{177}Lu administered may have to be limited by the impact of haematological toxicity [207], though large-scale studies have found that the long-term risk is not significantly correlated with bone metastases [208]. The most progress has been made with ^{225}Ac -labelled treatments [209], but ^{212}Pb -labelled compounds have also shown promise [210]. The review article by Kunikowska and Krolicki details some of the early pre-clinical work done with a variety of alpha emitters in NETs mouse models [211]. PRRT is generally considered as a 2nd line treatment at the earliest, after 1st line treatment with SSA [201], but further developments such as new clinical trials like NETTER-3, or the success of alpha therapy trials may result in an earlier position in the patient pathway [212].

4.3.3 Measuring response to treatment

Early guidelines for PRRT recommended follow-up with both blood tests and structural imaging (CT or MRI), but also monitoring of symptomatic response [213]. This is recommended at 3 months following the end of therapy, then at 6 months and recurring every 6 months thereafter, preferably for lifelong monitoring.

To aid interpretation of imaging reports, recommendations for standardised reporting metrics have been made [214]. For CT and MRI the RECIST 1.1 criteria [215] are recommended, particularly in clinical trials [165]. This criteria assess up to 5 target lesions against a baseline of pre-therapeutic structural imaging, with Partial Response (PR) requiring at least 30% decrease in the summed diameter of lesions, and Progressive Disease (PD) defined as a 20% increase in diameter, or new lesions [215]. This has been reported to require a much larger volume increase to demonstrate PD than previous structural imaging criteria [216], and the ability of individual reporters to select concordant target lesions (when NETs may demonstrate innumerable metastases), can affect inter-operator agreement [217]. The prospect of a transient 'flare' response affecting response measurement using structural imaging has been shown, reporting that 9% of patients later found to have Stable Disease had a >10% increase in lesion size at 6 weeks post-PRRT, which was reversed in half of these patients at a 6 month scan [218]. CT also struggles to accurately measure response on bone-predominant disease, where SRI may be of more benefit [219].

For SRI, various scoring systems have been proposed including the Krenning Score [214], SSTR-RADS [220], NETPET [221] and the PERCIST criteria [4] (though the latter has only been validated in [^{18}F]FDG, and not in SRI [216]). Certain issues may affect response assessment with SRI, such as concurrent SSA treatment affecting interpretation, or the uncertainty as to whether reduced uptake compared to baseline scan corresponds to reduced disease or de-differentiated disease. There is also the potential for tumour sink phenomenon to distort response assessment, where small or low-avidity lesions that were previously hidden on pre-therapeutic diagnostic imaging due to exceedingly high uptake in tumour lesions are unmasked once the more avid lesions have been treated by PRRT [219]. It has been shown that SRI may be more sensitive than structural imaging in detecting new lesions [44].

Monitoring response through standard blood tests is potentially the ideal solution due to the tests being more widely accessible and (relatively) inexpensive compared to imaging. Several blood parameters have been studied: for example Aminotransferase (ALT) and Chromogranin A (CgA). It has been shown that such markers can be unreliable immediately following therapy however, possibly due to radiation-induced inflammation [218]. CgA levels have been found to be increased after the 1st fraction in 29% of patients; only at 12 weeks post final fraction could CgA marker differentiate between PD and non-PD (though other studies have shown that CgA failed to correlate with response at all [222]). A permanent increase in ALT was seen in 43% of PD patients, compared to only 9% with objective response or stable disease [218]. A more complicated method looking at gene transcription has been proposed; the NETest [222]. A quarterly blood test that looks at 51 genes related to NETs, it has been shown to correlate with response measured by structural imaging - a study in an independent

cohort found that NETest levels decreased from baseline by an average 37% in responders to PRRT, while non-responders saw an average increase of 76%, with an overall accuracy of 90% compared to RECIST [223]. The use of structural imaging as the gold standard in this comparison does perhaps limit the applicability of the results, as the limitations of structural imaging in response monitoring are discussed above.

Functional imaging will still have a useful role to play in monitoring treatment response, as tests studying genes will not be widely accessible for the foreseeable future due to the advanced technical requirements and costs of the processes. With functional (hybrid) imaging, it is also possible to observe the changing structure and function of individual lesions, providing the opportunity to target surgery where tumour burden is reduced, and to diagnose potential complications following therapy, such as bowel obstruction [224]. Additionally, multi-timepoint functional imaging allows for dosimetry of the molecular radiotherapy treatment. This thesis will attempt to expand on the utility of post-therapy imaging, particularly focusing on the SPECT-CT imaging of ^{177}Lu at 24 hours post PRRT administration.

4.3.4 Predicting response to treatment

There are a range of clinical, biological and imaging indicators that have been studied to assess if they can predict a NET patient's response to the varying treatments available. The value of predictive markers would be to stratify patients into those likely and those unlikely to respond to a specific therapy – potentially saving expensive but ineffectual treatment and quickly moving the less likely to respond onto other therapies. This is especially important for PRRT where success rate in achieving response have been reported as quite low (interim analysis in the NETTER-1 trial showed only 18% of patients with complete or partial response, as measured by RECIST criteria [225]) - though the therapy has had more success in halting further progression of the disease). Discussion in the remainder of this chapter will focus on a limited range of markers that have been tested solely in relation to PRRT. For the enquiring reader, the review article by Barriuso et al [226] details many more tissue and gene biomarkers that are of interest to researchers. To clarify terminology, it is important to highlight that many of the studies referenced below do not have two arms (treatment/placebo or treatment/alternative treatment – though some such as the NETTER-1 trial do meet this criteria [204]); therefore they cannot strictly be termed predictive biomarkers [227]. To avoid using language that contradicts that used in the original published studies, the use of predictive/prognostic will be maintained as in the relevant article.

Response predictors for PRRT is a muddled field, with almost every marker having studies with contradictory findings on their predictive (or prognostic) value. It is proposed that such inconsistency may be due to the heterogeneity implicit in the grouping of NETs as a single tumour type [226, 228]. Where so much variety exists in terms of cell biology it may not

be useful to classify all under one banner. A notable consequence of this inconsistency is large variances in reported OS based on primary tumour site: for example, pancreatic NETs OS at 71 months, compared to bronchial NETs at 52 months [216]. A meta-analysis and review noted that most studies looking at imaging as a predictor of PRRT response did not perform subgroup analysis based on the primary tumour location [166], potentially limiting the applicability of their results. Another issue may be measurement at differing timepoints in the follow-up process, with evidence that many markers from blood tests may show transient increases in the immediate aftermath of PRRT (a ‘flare’ response) [218]. Similar findings, described as pseudo-progression due to temporary increase in lesion size on CT or MRI, have also been described for structural imaging [216, 218].

One clinical factor to show prognostic value in several studies has been liver tumour burden measured on structural imaging. Three studies correlated poorer survival following PRRT with greater metastatic involvement of the liver. However, the cut-off applied varied between studies (e.g. $\geq 25\%$ [229], $>50\%$ [230], and defined as “moderate” or “extensive” [189]). Re-analysis of the NETTER-1 trial data found no significant difference in PFS in the Lutathera arm between low, moderate or high disease burden measured on CT or MRI [231]. There was however a correlation on the LAR-SSA treatment arm. This may show that the crossfire effect of radiation-induced cell damage [232] increases the effectiveness of PRRT over other treatments. A potential limitation of these studies is that pre-therapy liver burden was measured on CT or MRI, and not on SRI; structural imaging may not correspond to the SSTR-positive tumour burden that is treated by PRRT.

Ki-67 index is a measure of proliferation in tumour cells taken from biopsied tissue and acts as a corollary of aggressiveness [233]. It is used as an important component in the WHO classification of NENs into Grades 1, 2 or 3 [234]. Ki-67 proliferation index has been shown to be inversely correlated to SSTR uptake [235]. There are contradictory reports as to its predictive value, possibly due to inter- and intra-tumour heterogeneity [216]. A study of 74 patients following PRRT found Ki-67 $>10\%$ predicted lower PFS [229], though this was not supported in the NETTER-1 trial [225].

The predictive nature of circulating biomarkers, assessed via blood sampling, has been examined for several different candidates. A study of midgut only NETs found both baseline CgA ≥ 600 ng/mL and patients with serotine-hypersecreting tumours had shorter PFS [236]. Neuron-specific Enolase (NSE) was found in one study to predict longer OS when < 15 ng/mL at baseline, with the proviso that it wasn’t significantly different in non-pancreatic tumours [229]. This finding was supported in a study of midgut tumours where NSE did not predict PFS [236]. A suggested combination of biomarkers to create a new metric is the Inflammation-based Index (IBI) which combines Albumin and C-reactive Protein levels in a 3-point score [237]. A pilot study [237] and an independent assessment in 89 patients

undergoing PRRT [238] showed the baseline IBI was predictive of OS. However, a critique of this method was there was no way to determine if the IBI was merely highlighting patients with more aggressive tumours, and their corresponding poorer prognosis [239], rather than predicting their response to therapy (though this issue affects nearly all studies evaluating predictive biomarkers outwith the randomised clinical trial setting). Preliminary work using biomarkers that test for tumour DNA or RNA in circulating blood (sometimes termed “liquid biopsies” [216]) have shown promise as a pre-treatment predictor of response – with the PRRT Prediction Quotient (PPQ) of prime interest. This biomarker combines the tumour grade, growth factor and metabolism gene expressions assessed in baseline blood [240]. A study evaluated the predictive value of baseline PPQ on 67 patients undergoing PRRT and found an overall predictive accuracy of 96%, with those measuring $PPQ_{negative}$ were 24 times more likely than $PPQ_{positive}$ patients to suffer progression or death [223]. Studies of genetic abnormalities in tumour cells may be a future area of study, though initial results have so far not identified a relationship between genetic alterations in biopsied tissue and PRRT response [223].

The most widely adopted semi-quantitative scoring system for SRI has been the Krenning Score (KS): originally developed for planar ^{111}In -Octreoscan it uses a visual comparison of uptake in tumour lesions compared to various areas of physiological uptake. Though not validated in other imaging techniques [239], it has been used with PET SRI [241] and indeed is referenced in ENETS guidelines [214]. The NETTER-1 study used Krenning Score on planar ^{111}In -Octreoscan as inclusion criteria, and the final published results found no significant difference in median OS based on baseline Krenning Score [204].

The most obvious quantitative imaging parameter for initial assessment of predictive value is SUV, which has familiarity due to its use in $[^{18}F]FDG$ PET (though the potential limitations of SUV for radiopharmaceuticals other than $[^{18}F]FDG$ are discussed in Section 2.3.2). Several studies for the different ^{68}Ga -DOTA peptides have shown predictive value for PRRT response, with various cut-offs proposed to select responders: for example $SUV_{max} \geq 13.0$ [242] and ≥ 16.4 ([243]. Other studies have supported the predictive value of SUV_{max} without determining a threshold value [244, 245]. In opposition, there have also been studies that found no significant predictive value for SUV_{max} [141, 246, 247]. Given the known effect of image noise on reproducibility of SUV_{max} [78] and the differing PET scanners and reconstructions used, it is perhaps not surprising that different thresholds or even contradictory results are reported.

There has been recent interest in quantitative measurements of the overall tumour burden on pre-therapy PET SRI as a predictive metric. Borrowing the concept of Metabolic Tumour Volume from $[^{18}F]FDG$ PET, a metric termed Somatostatin Receptor-Expressing Tumour Volume (SRETV) can be defined. This then allows Total Lesion Somatostatin Receptor

Expression (TLSRE) to be calculated (comparable to Total Lesion Glycolysis), by multiplying the SRETV by the SUV_{mean} in the lesion. SRETV or TLSRE may then be summed over all lesions in the scan, denoted by $\Sigma SRETV$ or $\Sigma TLSRE$. Three studies have shown $\Sigma SRETV$ to be a predictor of either PFS or OS under univariate analysis [238, 246, 248], though each with different cut-off criteria, and this has not always shown predictive value under multivariate analysis. For $\Sigma TLSRE$, a significant association with PFS ($p < 0.001$) at $\Sigma TLSRE \geq 306.8$ was found [248], though a different study identified a cut-off at $\Sigma TLSRE \geq 146.48$ with a p-value of 0.056 [246]. One study also found the change in $\Sigma SRETV$ from baseline to post-therapy ^{68}Ga -DOTATOC was predictive of PFS [249]. Methodology for segmentation of lesions to give reproducible SRETV measurements is still under study [250] and the time requirements of manual delineation of every lesion makes such techniques unsuitable for clinical practice, though semi- and fully-automated software packages to assist with total body delineation of lesions are becoming available [251, 252].

Though use of radiomics or textural analysis has been increasing in SRI in recent years [142], there has been a limited number of studies attempting to identify features that predict PRRT response. One study that examined the asphericity of lesions on ^{111}In -Octreoscan SPECT-CT found it to have greater predictive value for response than Krenning Score or tumour to liver ratio [84]. Almost all other studies focused on pre-therapy ^{68}Ga SRI. A study evaluating 64 patients assessed tumour heterogeneity in a visual manner: variation in Krenning Score within the lesion was used to classify a lesion as hetero- or homogeneous, and patients classified as <50% or >50% of their lesions showing heterogeneity [253]. This study found those patients classified as having heterogeneous lesions had significantly lower overall survival. A study of ^{68}Ga -DOTATOC using PET-MRI examined a very limited number of textural features but found no significance in predicting treatment response [247]. Entropy was recorded to correlate with longer OS, but not statistically significant longer PFS over a value of 6.7 [254]. Skewness and kurtosis were identified as showing significantly different values in responders and non-responders in one study, but it also commented on the disappointing sensitivity and specificity results using optimised cut-offs [255]. Other studies have also found skewness [141, 245, 256] and kurtosis [141, 256] to be predictors of PRRT response. Variations in the features chosen for analysis and use of multiple scanners (without mention of harmonisation steps) highlights the lack of standardisation specific to the process of textural analysis, which will affect the reproducibility of many studies [117, 216, 257]. As such, the use of textural analysis in the routine clinical setting as a predictor of response to PRRT is still some distance away.

The use of post-PRRT imaging of the gamma photon component of ^{177}Lu -DOTATATE has also been investigated, though the small number of studies have been limited in sample size or follow-up timescale [258, 259]. One study of 90 patients evaluated response at 3

months post-treatment and found the single time-point absorbed dose corresponded with significant changes in OS [260]. Following on from the above-mentioned study on asphericity in ^{111}In -Octreoscan, the same group studied the same feature in 33 patients who underwent post first fraction ^{177}Lu -DOTATATE imaging, and found higher asphericity correlated with poorer response on a per-lesion basis [261]. A study examining the Krenning Score on planar post-therapy ^{177}Lu imaging found significantly longer PFS and OS for those scoring Krenning 4 versus Krenning 3 [262]. It can be seen from the summary of the literature above that identification of biomarkers that can consistently and accurately predict a patient's likely response to therapy for NETs is a worthwhile pursuit. Future developments in blood and gene testing may be of value in identifying accurate markers, though this thesis will concentrate on identifying the additional value that SRI may provide over current uses of functional imaging.

4.4 Relevance to the thesis

There have been several significant developments in the field of NENs, leading to both an increase in diagnoses, but also greater treatment and monitoring options. The approval of ^{177}Lu -DOTATATE for PRRT in many countries has refocused attention on the uses of SRI in diagnosis, staging and predicting and monitoring response to treatment. However, emphasis is placed on the use of PET radiopharmaceuticals, and little on SPECT-CT imaging. This focus is demonstrated by the very limited number of studies that have so far investigated the role of post-PRRT ^{177}Lu imaging in predicting response to PRRT, and no studies have been found for pre-PRRT quantitative $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC imaging. One of the aims of Part 3 of this thesis is to further explore imaging biomarkers in these use cases, particularly with quantitative measurements and textural analysis with the aim of identifying suitable biomarker candidates for multi-centre study.

Chapter 5

Introduction to harmonisation of quantitative data

5.1 Introduction

Quantification of images for SPECT involves converting data, gathered in the raw form of counts, into units describing the concentration of radiopharmaceutical. This can then be normalised to account for differences in patient size and administered activity. The quantified data produced by a calibrated SPECT system (or indeed for quantified data in other fields of nuclear medicine, such as PET), should be able to be compared directly to any other system calibrated to the same standard. However, the lack of agreed-upon standards for acquisition and reconstruction can limit the utility of data comparison, as can differences in performance inherent even between systems of the same make and model [263, 264]. Such differences hinder the replication of results and can discourage the use of quantitative SPECT in multi-centre trials, where aggregation of quantitative data is required to provide robust evidence for the introduction of new clinical practices and therapies. Even for [^{18}F]FDG PET, where standards for calibration and quantification are well established, variability on inter-camera measurements has been found to be higher than intra-camera variability [265, 266].

This section will discuss how quantitative data in nuclear medicine compares between systems and evaluates potential methodologies for improving inter-system agreement and how they may be applied to quantitative SPECT.

5.2 Quantitative standards in nuclear medicine

As the focus of quantification in nuclear medicine has been on [^{18}F]FDG PET imaging for many years, it was here that standards were first developed and are most widely adopted. The EANM programme, under the branding of EARL, has run accreditation services for [^{18}F]FDG since 2010. The accreditation process requires acquisition of a uniform phantom, to assess any bias in SUV, as well as acquisition of the NEMA IEC phantom, from which RCs are measured. Both results are compared to specifications derived from multiple scanners [267]. An example of a RC curve for [^{18}F]FDG PET and the appropriate EARL standard limits are shown in Figure 5.1. The EARL standards have recently been updated to EARL-2, in order to incorporate the improved performance of the new generation of digital PET scanners

[268], although the continued presence of non-digital scanners in clinical practice means that the reconstructions on more advanced scanners are often intentionally limited to meet the older standards.

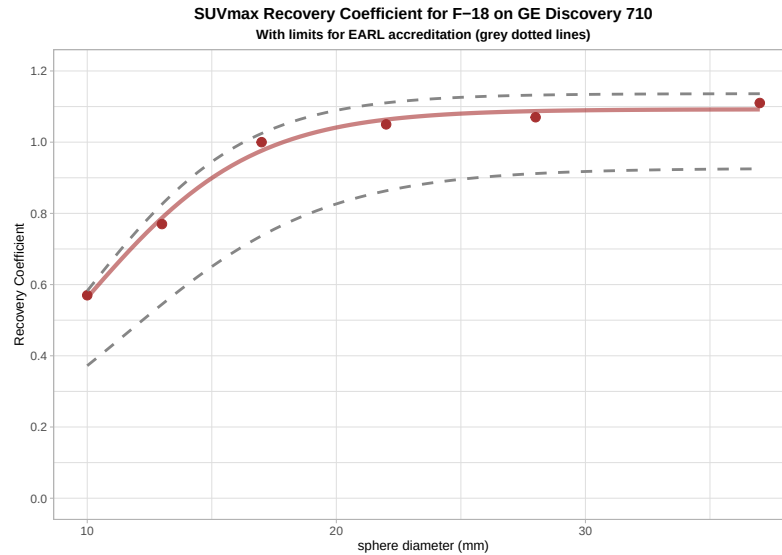


Figure 5.1: Recovery Coefficient curve (solid line) calculated from NEMA IEC Body phantom acquisition for [^{18}F]FDG PET on GE Discovery 710 PET scanner, with EARL accreditation limits (dotted line).

Several other professional bodies have introduced their own approaches to accreditation, often tied to regimented protocols and guidelines, with the aim of improving standardisation, particularly in the domain of clinical trials. A summary of some of these harmonisation strategies is given in Akamatsu et al [269].

Recent advances by manufacturers to provide on-board quantification for SPECT have expanded the availability of quantification beyond a few research-focussed centres and accentuated the need for harmonisation of acquisition and reconstruction. This approach will be imperative if quantitative SPECT-CT is ever to be adopted for clinical trials, as has been the case for the use of [^{18}F]FDG PET in clinical trials. The more limited patient numbers for many indications in SPECT (when compared to oncological imaging using [^{18}F]FDG in PET, for example) mean that multi-site data aggregation will be of even greater importance for demonstrating the clinical utility of quantitative SPECT, which is needed for widespread acceptance and adoption into clinical practice.

There are also technical factors that inhibit SPECT harmonisation: for example, the measurement of RC in NEMA phantoms has been shown to be less reproducible in SPECT, with dependency on sphere position [33] and even on the relative position of spheres to others [270]: the poorer resolution in SPECT will result in proportionally more spill-in from nearby spheres. The non-standard nature of how vendors interpret quantitative data in DICOM format will also raise issues, as with inconsistent calculation of scan durations [270], or that some manufacturers treat quantitative SPECT data as if being from a PET modality.

Recent publication of guidelines for quantitative SPECT-CT [11] are a welcome advance in promoting its use to the community. However, the availability of guidelines does not guarantee that any data gathered while following these standards will be comparable across systems. There are several possible solutions to improve harmonisation:

- Widespread adoption of vendor-neutral quantification software
- A centralised accreditation service for SPECT-CT
- Use of quantitative features with high reproducibility and low inter-scanner variability
- Post-acquisition data harmonisation using additional smoothing filters
- Post-acquisition data harmonisation using statistical tools

5.3 Approaches to harmonisation

5.3.1 Vendor-neutral quantification software

Multi-centre phantom studies have shown variation between quantitative SPECT results, where different manufacturer systems have been used [271, 272]. Reconstruction of data from different manufacturer's gamma cameras using a vendor-neutral reconstruction software solution may reduce the variability expected to be caused by manufacturer-specific reconstruction algorithms [54, 70, 272, 273]. However, such software is expensive, and it is unlikely the same package would be adopted by every centre. For single-camera or single-manufacturer centres, there would be little motivation for doing so: experience in [^{18}F]FDG PET clinical trials has shown that possession of specific software is rarely, if ever, an exclusionary factor.

An open-licence (freeware) software solution could perhaps result in more widespread adoption, but such a package does not exist at present, and the requirement to obtain CE marking and/or FDA approval before use in a medical environment means one is unlikely to be forthcoming. This means that vendor-neutral quantification software is unlikely to be the panacea to the problem raised.

5.3.2 SPECT-CT Quantification accreditation service

One detriment of accreditation for harmonisation that has arisen in [^{18}F]FDG PET is the requirement to include as many centres as possible: specifically, the range of possible camera technologies that is the inevitable result. A principle of broad inclusion has often meant that accreditation standards apply the 'lowest common denominator' theory, limiting the application of newer technology, for example, digital PET. This is slowly changing with the

introduction of the EARL-2 standard, but occurring many years after the first implementation of digital PET. Slow evolution of standards has meant that image quality for patients in clinical trials is often deliberately reduced from the best achievable image quality.

Establishment of such a service for quantitative SPECT will require a critical mass of centres willing to participate, plus an onus for centres to demonstrate accreditation (most likely for participation in multi-centre clinical trials). Until such requirements exist, those wishing to improve harmonisation of SPECT quantification between multiple cameras may have to look to other solutions. The upcoming launch of ^{177}Lu accreditation by EARL [274] will provide an interesting view as to how easily accreditation may be adopted in SPECT, and the popularity of an accreditation process with cost implications for SPECT users.

5.3.3 Use of low-variability quantitative features

There has been limited work on the repeatability and inter-scanner variability of quantitative features in SPECT-CT, but it may be expected that findings from PET could be considered as an exemplar.

Use of SUV_{peak} has been mooted to provide benefits in terms of standardisation compared to the previously universal use of SUV_{max} : there are multiple methodologies to measure SUV_{peak} , but in general it involves taking the average SUV of voxels within a sphere of specified diameter encircling the pixel with the highest SUV [275]. Studies in ^{18}F FDG PET have determined SUV_{peak} to be more reproducible than SUV_{max} in small volume lesions [77] and in higher noise images [78]. However, it has been found that SUV_{peak} provides poorer harmonisation results than prospective harmonisation of scanner acquisition and reconstruction parameters [276]. Local work for quantification of $^{99\text{m}}\text{Tc}$ across seven gamma cameras found no significant improvement in SUV range for phantom study between SUV_{peak} and SUV_{max} [277].

The use of SUVs normalised to patient-specific background uptake has found favour in some circles [278–280], for example, Tumour to Blood Pool ratios. While accounting for calibration biases, such methods fail to normalise for differences in camera image quality or reconstruction algorithms [276].

5.3.4 Additional filtering of images post-acquisition

A method to achieve harmonisation on images that have already been acquired, potentially without meeting guidelines on standardisation, is to apply smoothing filters post-hoc. This approach has been adopted as part of the move to digital PET systems, with the expectation of a step-change in image quality and concurrent large differences in measured SUV requiring a process to allow comparison of patient measurements against their previous

scans. Lasnon et al. suggested the use of a Gaussian filter with a large FWHM following an advanced reconstruction could match performance against a less-technologically advanced reconstruction [281] for PET imaging.

Siemens have introduced a proprietary program (EQ-PET) that can incorporate the calculation of amended SUVs in PET from an appropriately filtered image [282] into the routine reporting workflow. It requires a phantom acquisition on the high-performance scanner, for which the RC curve is matched to a reference RC curve by exploring the application of different filters. Once the appropriate filter has been found, it can be saved and then applied to future patient images to obtain harmonised SUVs. The benefit of this software is that the reporter only observes the higher-quality images, and it is only when measurements are made that the translated SUVs are displayed. The software has been tested satisfactorily for [^{18}F]FDG PET in a multi-centre study [283] and shows it is possible to achieve harmonisation of images acquired on digital and non-digital systems.

At present there is not a similar software solution for quantitative SPECT. Indeed, it presumably requires the definition of a suitable reference RC curve, which by definition requires the establishment of a quantitative SPECT accreditation programme. This methodology also suffers from the “lowest common denominator” approach of reducing to a broad standard.

5.3.5 Statistical harmonisation

The problem of combining data from multiple centres (or acquired using different equipment/parameters) has been examined in fields outside molecular imaging; one such field is laboratory medicine, where differences between ‘batches’ of samples are important when combining datasets or even following changes to a chemical reagent at the same centre [284].

Statistical methods developed for this issue originally required large numbers of samples in each batch, reducing the utility. The Location-Scale model (L-S) has been commonly used in laboratory medicine, particularly in genomic analysis. The L-S model looks at two parameters for each batch of data, the location and the scale (or variance) and then adjusts each data point to try and minimise the differences between the batches. For example, for a normal distribution the location may be taken to be the mean value, and the scale to be the standard deviation. A full derivation of the L-S model is discussed in Karvanen [285].

One such model was implemented in the R package “ComBat” by Johnson et al [284], which also extended the L-S to include an empirical Bayes method, to address its limitations at small sample sizes, where the presence of outliers affected robustness. The Bayesian method assumes that estimates of the requirements needed to harmonise one feature (originally a

gene, but here applied to an imaging feature such as SUV) can be applied to subsequent related features. This estimate then forms a Bayesian prior that is applied to the L-S model, with the aim of reducing the effect that outliers would have on location and scale if applying the L-S model alone [284]. An explanation of how the statistical formulation is applied to features from imaging data is given in Orlhac et al [286].

These techniques have been adapted for use with imaging data in R packages such as NeuroCombat [287] and applied successfully to harmonise quantitative imaging data post-acquisition [286, 288, 289]. Applied in PET, the effects of variability due to acquisition in differing centres and due to varying reconstruction parameters were successfully normalised and allowed the combining of datasets [286]. Other studies have used modified versions of ComBat in attempt to improve downsides of the original algorithm; for example, the base version of ComBat shifts data based on the overall mean of the feature from the combined centres – this could render the harmonised feature meaningless if, for example it created a feature with negative values that do not relate to real-world conditions. A modification to shift data based on a chosen reference ‘batch’ (or centre) was proposed by Stein et al [290], but a study applying this modification to radiomic data from MRI, PET and CT found no improvement to performance over the base ComBat package [288]. The same study also tested a modification to provide bootstrapping and Monte Carlo modelling of the estimated Bayesian parameters. This adapted algorithm, termed B-ComBat, did reduce variability in the harmonised data. However, the improvement was small [288].

The above studies successfully implementing ComBat in PET suggests that its application in quantitative SPECT would be possible. The use of a reference centre would have the advantage of being able to harmonise to the most advanced equipment, rather than requiring matching to the lowest common denominator.

An in-depth discussion of alternative statistical techniques is beyond the scope of this thesis, but there are some promising techniques that are beginning to be investigated in medical imaging. Generative Adversarial Networks (GAN) can utilise a learned prior from a teaching dataset, to regularise a given image [291]: an advantage of this is that the process occurs in the image space rather than with measured values, whilst Convolutional Neural Networks produce synthetic images that are easier for visual interpretation [292]. Domain Adaptation, a machine learning technique that has efficacy in smaller datasets that are common in imaging, has been tested for use in functional MRI and found to improve harmonisation [293]. Within the field of image processing for radiomics, there are a multitude of approaches that have been tested, many with deep learning models [294]. Chapter 15 of the thesis will focus on investigating how a statistical harmonisation technique (namely, the L-S and Bayesian models) could be applied to SPECT quantification.

5.4 Relevance to the thesis

While individual departments may face challenges with technical implementation of quantitative SPECT, or its integration into clinical workflows, it appears clear that wider challenges exist to embed quantitative SPECT fully into the nuclear medicine toolbox. A requirement for validation of clinically useful features necessitates a programme of harmonisation to ensure data can be successfully combined for multi-centre trials. This chapter has outlined some potential routes to achieve harmonised data: which methods will be most widely adopted by the SPECT community has yet to be determined. This thesis will focus on the statistical harmonisation methodology, as the most easily applicable method currently for SPECT.

Part 2:

Quantitative SPECT-CT calibration

Success is a science; if you have the conditions, you get the result.

Oscar Wilde

Letter to Marie Prescott, The New York Herald, 12th August 1883

In order to achieve accurate and precise quantification of clinical images, a solid foundation is needed: this comprises an accurate calibration of the imaging system and traceability of measurements made on adjunct equipment. For quantification-enabled systems the manufacturer should provide instructions on the calibration process as well as the required frequency. A full analysis of the methods and results of the calibration process for Siemens BroadQuant are reported in Chapter 6, detailing any required amendments to the manufacturer's instructions and providing suggestions for producing a robust calibration.

It is incumbent on the operator to perform QC to ensure a SPECT system meets an acceptable level of performance. Guidelines on QC for Nuclear Medicine have been available for many years [295] and are still relevant to many of the performance characteristics that will affect quantification. Methods have been developed to more precisely test the quantitative performance of gamma cameras, often using tests developed for quantitative PET [65]. Chapter 7 will attempt to demonstrate a number of these tests for multiple radionuclides, as well as assessing the duration for which calibrations may remain valid.

The final chapter of this Part (Chapter 8) will explore how a user may apply their newly-validated calibrated images to move beyond the limited quantitative features available with Siemens software, expanding into the greater depths of textural analysis. It will aim to identify which textural analysis features provide consistent results through the use of phantom studies, reducing the range of features to a workable subset before they are applied to clinical studies in Part 3.

Chapter 6

Calibration of Siemens BroadQuant

Whilst gamma cameras are the central focus of a nuclear medicine department's Quality Assurance programme, there are several pieces of adjunct equipment which are essential to the production of accurate and precise results for clinical tests, particularly when quantification is involved. As a result, rigorous calibration and QC checks of the equipment are vitally important. Despite this, there remain both systematic and random sources of error that may impact the calculated Calibration Factor (CF). It is therefore important to quantify the uncertainty in the calibration process - each component step has an uncertainty (denoted u) associated with it, and can be combined using addition in quadrature as demonstrated in EANM guidelines on uncertainty analysis [296].

This chapter aims to examine each step of the calibration process, in order to identify any refinements to the manufacturer's protocol to improve accuracy or reduce variability. It will then define the uncertainty for each step, before combining the uncertainties into a single value for each of the two radionuclides tested (^{99m}Tc and ^{177}Lu).

6.1 Radionuclide calibrator

6.1.1 Calibration against national standard

The radionuclide calibrator used in the Nuclear Medicine department at Gartnavel General Hospital is a Capintec CRC-15R (Capintec, Inc. Pittsburgh, USA). It is classified as a field instrument and can be referenced back to the national standard instrument at the UK's National Physical Laboratory (NPL) through a chain of calibrations, such that the accuracy of the measured value for the radionuclides used in routine clinical practice should be $\pm 5\%$ [297]. Annual calibrations are carried out to determine the required calibration factors for a number of radionuclides.

As ^{177}Lu cannot be sourced in a suitable and cost-effective manner for transport to NPL, the calibrator was calibrated against a reference source that was provided as part of a clinical trial and included an appropriate certificate of calibration. An NPL cross-calibration exercise in 2023 for ^{177}Lu showed the calibrator within this department, using the calibration factor used for the work in this thesis, was within 1.23% of the NPL approved value. A study of 17 hospitals in Europe in 2012 showed that 81% could measure ^{177}Lu to within $\pm 5\%$ of the

reference activity determined by NPL [298]. Although impurities of $^{177\text{m}}\text{Lu}$ (half-life = 160 days) are present in small quantities in ^{177}Lu radiopharmaceuticals, the percentage impurity is minuscule until many weeks after the reference date [299] and does not require correction to the calibrator factor to ensure accurate measurement [298].

6.1.2 Calibrator clock synchronicity

As quantification requires accurate timing of activity measurement and scan time, it is important that clocks throughout the department are synchronised – the literature reports differences of up to 10 minutes can occur in clinical departments [300]. Recording the time to the nearest minute will result in an error of less than 0.1% for $^{99\text{m}}\text{Tc}$ [297] and smaller still for other commonly-used radionuclides for SPECT, which have longer half-lives than $^{99\text{m}}\text{Tc}$. Modern gamma cameras generally allow synchronisation through a Network Time Protocol server. However, there rarely such facility for most radionuclide calibrators which tend to be of older provenance and lack the security protocols to allow them to be safely networked. EANM guidelines for QC of radionuclide calibrators suggest a daily check that the calibrator clock is synchronised with others in the department [295]. This approach has been adopted locally, but with an action level of 3 minutes. To assess the likelihood of the calibrator time drifting from the clock time of the gamma camera (assumed to be the “correct” time, as synchronised with the server) a frequent check was made of the calibrator time and adjusted to return it to synchronicity with the gamma camera.

Over the course of approximately two years, 70 measurements were made of the calibrator and gamma camera clock times. The average interval between measurements was 18 days. An adjustment of 1 minute or less was required 79% of the time (i.e. error of less than 0.1% for $^{99\text{m}}\text{Tc}$) and exceeded 3 minutes on only one occasion. The required alteration of 12 minutes on that occasion was over 6 standard deviations away from the mean change. This anomaly suggests an outside issue such as a power cut (the calibrator is not equipped with an Uninterruptible Power Supply), though there were no indications other items of equipment in the department were affected around the same date. The distribution of the deviation from the correct time is shown in Figure 6.1.

The average drift of the calibrator clock away from the time recorded on the gamma camera is calculated at approximately 27 seconds per week. To exceed the local action level of ± 3 minutes would take approximately seven weeks. Remaining within a stricter tolerance of ± 1 minute would require a correction every two weeks on average. Therefore, a QC protocol involving weekly checks of calibrator time should be satisfactory to maintain accuracy for quantitative acquisition.

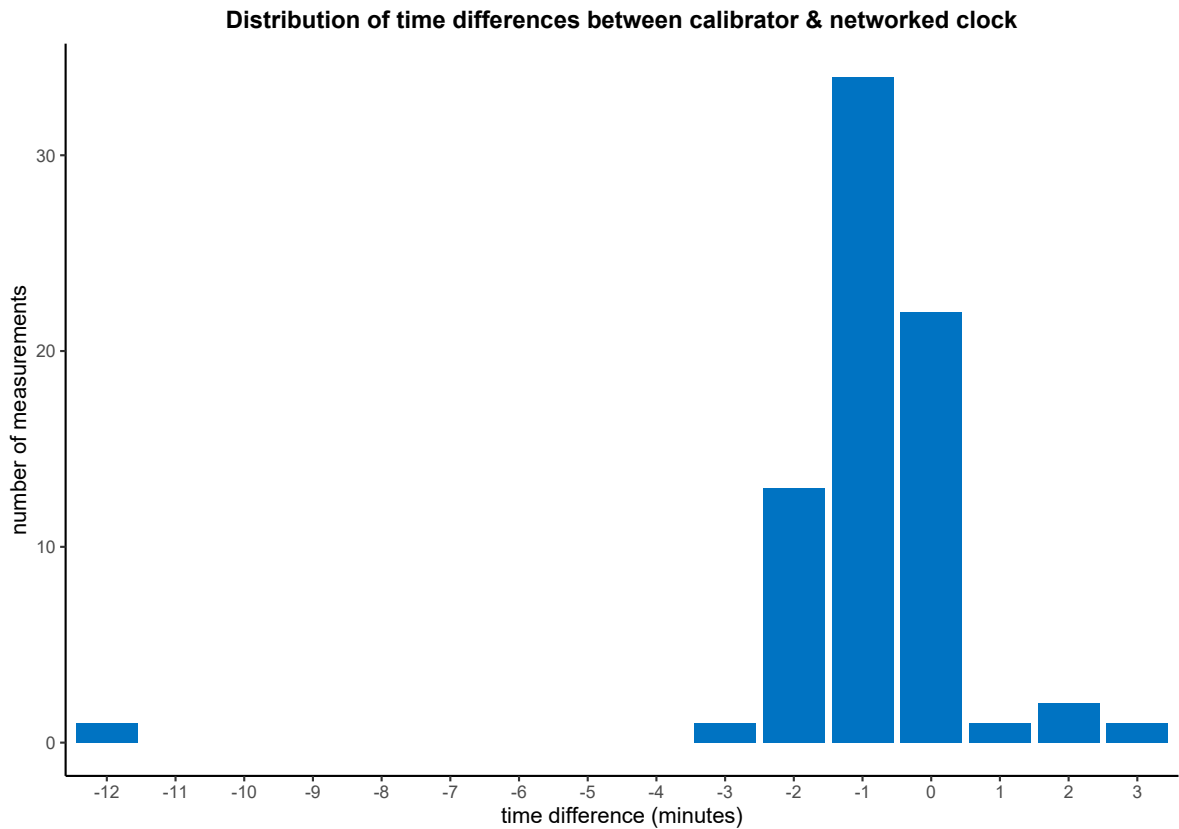


Figure 6.1: Measured time difference (in minutes) between a gamma camera (connected to Network Time Protocol server) and a radionuclide calibrator from repeated measurements over a two year time period.

6.2 Point Source Sensitivity calibration

The Sensitivity Calibration step for BroadQuant Calibration utilises a Siemens point source container (Figure 6.2), containing a known activity of the radionuclide to be calibrated. The point source is positioned between the collimated camera heads at a fixed distance and an image of 5 million counts is acquired. The counts per second can then be divided by the activity of the point source to calculate the sensitivity.

$$\text{Point Source Sensitivity} = \frac{\text{counts per second}}{\text{measured activity}} \quad (6.1)$$

The container has a cylindrical body with a conical bottom and total capacity of around 1.5 ml. As the point source has to be positioned horizontally for the calibration procedure (attached to the metal arm extending from the camera bed), typically the liquid would spread throughout the body of the container, resulting in a poorer approximation of a point source being presented to the detectors. To help create a reproducible geometry, Siemens recommend for the container to have an amount of absorbent material (usually cotton wool or gauze) placed in it, which will remain fixed in the conical part of the container. The manufacturer recommends this calibration is performed monthly [36].

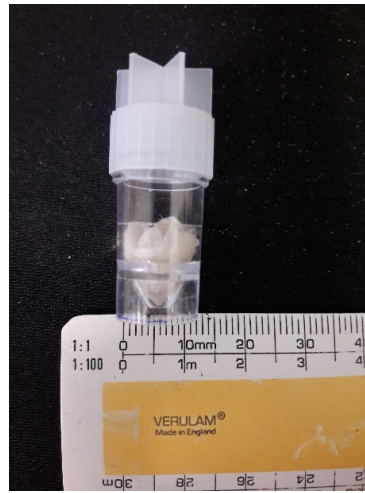


Figure 6.2: Siemens point source container with absorbent material, used to hold radionuclide during point source calibration. Pictured with ruler for scale.

6.2.1 Point Source activity measurement

Direct versus indirect activity measurement

For the gamma camera sensitivity to be calculated from a measurement of the point source, the activity present in the point source must be known, through measurement in a radionuclide calibrator traceable to the national standard. Measurement of the activity could either be through directly measuring the point source container in the calibrator (positioned in the sample holder, or "dipper": Figure 6.3) or indirectly, by measuring the syringe used to prepare the point source before and after the activity is dispensed into the point source container. The Siemens white paper on quantification [36] recommends the direct measurement of the point source.

To determine if the direct method of measuring the point source is valid, the activity measured in the point source can be compared to the syringe activity before and after preparation of the point source, to determine if the same radionuclide calibrator calibration factor can be used for the point source container. Measurements were made during the preparation of ^{99m}Tc and ^{177}Lu point sources (Table 6.1).

Radionuclide	number of measurements	mean difference
^{99m}Tc	23	-1.94%
^{177}Lu	8	-1.13%

Table 6.1: Measured difference between point source activity and syringe activity (including residual correction).

For both radionuclides, every activity measured in the point source was lower than the respective activity calculated from the syringe measurements. This could be attributable to the change in geometry: the syringe is measured hanging from the top of the dipper and the

calibration for the radionuclide has been validated for this positioning, whilst the point source must sit in the base of the dipper. These results suggest that to achieve optimal accuracy in the quantification process, a separate calibration factor would be required for measuring the point source.

Activity measurement - positioning of source

The calibrator dipper is designed to provide reproducible geometry in the calibrator chamber for measurements made of a vial or syringe. Although Siemens recommend direct measurement of the point source in the dipper to obtain the activity to be entered in the calibration process, the dipper does not provide reproducible positioning of the source container. The container is also prone to toppling over when placed directly in the dipper, which can lead to spread of the liquid through the container and hence disrupting the intended point-like nature of the source.

Measurements were made for both ^{99m}Tc and ^{177}Lu point sources at different locations within the base of the dipper (marked A-E on the right image in Figure 6.3, with the central position, E, taken as the reference point). The source was removed from the calibrator and repositioned at each of the five positions, with four repeat measurements at each position. The overall experiment was performed four times for ^{99m}Tc and twice for ^{177}Lu . The difference between the measured activity at the edges and the central reference measurement was calculated for each position, with the mean and maximum difference reported in Table 6.2. This data was also used to calculate the overall repeatability of activity measurements with the point source, using the Coefficient of Variation (CoV) with the results shown in Table 6.3.

$$\text{Coefficient of Variation, CoV(\%)} = \frac{\text{standard deviation}}{\text{mean}} \times 100 \quad (6.2)$$

Radionuclide	Mean difference	max difference
^{99m}Tc	0.61%	1.24%
^{177}Lu	0.57%	1.51%

Table 6.2: Measured difference between point source activity at central dipper position and at the edges (as illustrated in Figure 6.3).

The mean difference for the activity measured away from the central position is quite low for both radionuclides, the maximum measured difference is not inconsequential, at up to 1.5% for ^{177}Lu . Although both radionuclides have a CoV within the tolerance for repeatability defined by NPL for a field instrument of $\pm 1\%$, it was felt that the results could be improved upon.

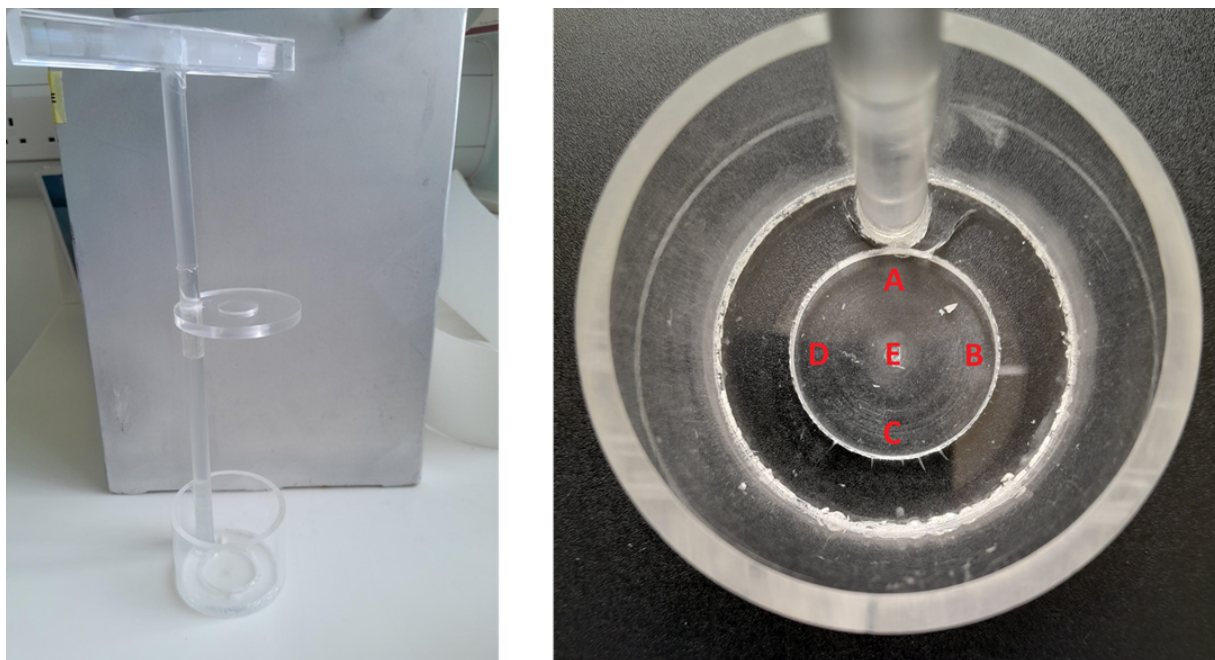


Figure 6.3: Left: Dipper used to lower sources into the radionuclide calibrator. Right: Diagram indicating the positions used at the base of the calibrator dipper to make measurements with point source container

A jig was designed to hold the point source container in a reproducible position within the calibrator dipper (Figure 6.4). The design was based on the holder for the sealed point source used in the xSPECT Quant calibration procedure and was manufactured locally. The final product was manufactured from polymethyl methacrylate (PMMA) at less than 3 mm thickness, which given the 0.17 cm^{-1} attenuation coefficient of PMMA at 140 keV [301], will result in minimal attenuation of the activity in the point source.

Radionuclide	Mean CoV without jig	Mean CoV with jig
$^{99\text{m}}\text{Tc}$	0.65%	0.17%
^{177}Lu	0.70%	0.23%

Table 6.3: Repeatability of activity measurements with point source, with and without use of a positioning jig.

Twelve repeated measurements were made of both $^{99\text{m}}\text{Tc}$ and ^{177}Lu point sources, resetting the jig and source in the dipper each time. The CoV were calculated and the experiment repeated (four times for $^{99\text{m}}\text{Tc}$ and twice for ^{177}Lu) before the overall mean CoV was calculated (Table 6.3). This showed the CoV could be reduced by almost two thirds with the use of the positioning jig.

The findings above suggest that if required to measure the point source container directly in the calibrator, it is recommended that a method to allow for reproducible positioning of the container is found and that an individual calibrator factor is determined for each radionuclide used, based on the NPL method for determining sub-calibrations.



Figure 6.4: Positioning jig (designed and manufactured for this thesis) to allow reproducible positioning of point source within calibrator dipper.

However, the EANM guidelines for quantitative SPECT recommend that the clinical setup for calibrator measurement should be replicated for the calibration and verification steps [11]. As the clinical application measures activity in a syringe, the repeatability of the activity measurement in a syringe was tested (Table 6.4).

Radionuclide	Mean CoV
^{99m}Tc	0.08%
^{177}Lu	0.19%

Table 6.4: Repeatability of measurements of activity in syringe.

The results in Table 6.4, when compared to Table 6.3, show the measurement activity was more repeatable when measured in the syringe than in the point source container. Measurement in syringe also replicates the measurements made for patient administrations, hence measurement of the syringe before and after dispensing to the point source container will be the chosen method of activity measurement for point source calibration in this thesis.

6.2.2 Repeatability of Point Source Sensitivity

Testing of repeatability was done using two methodologies; both used repeated measurements of the same source of ^{99m}Tc : the first method used consecutive measurements without adjustment to the set-up of the gamma camera, and the second method involved removing the source from the gamma camera, returning the support arm to the bed and resetting the gamma camera to its default positioning. This second method accounts for small differences in camera head position, extension of the support arm, and rotation of the point source that will be present in circumstances where setup is changed between acquisitions.

For each set of measurements, the mean and standard deviation of the point source sensitivity was determined, allowing CoV to be calculated (Tables 6.5 and 6.6).

number of measurements	H1 Sensitivity (cps/MBq)	H1 CoV	H2 Sensitivity (cps/MBq)	H2 CoV
8	84.44 ± 0.09	0.11%	85.21 ± 0.08	0.10%
8	83.26 ± 0.09	0.11%	83.73 ± 0.09	0.11%
8	82.81 ± 0.04	0.04%	82.71 ± 0.08	0.08%
7	84.17 ± 0.05	0.06%	83.86 ± 0.05	0.06%
8	83.06 ± 0.05	0.06%	83.39 ± 0.04	0.04%
Mean	83.55 ± 0.07	0.08%	83.78 ± 0.07	0.08%

Table 6.5: Repeatability of Point Source Sensitivity measured on gamma camera during the calibration process, using a static camera setup.

number of measurements	H1 Sensitivity (cps/MBq)	H1 CoV	H2 Sensitivity (cps/MBq)	H2 CoV
8	85.10 ± 0.16	0.19%	85.43 ± 0.16	0.10%
8	84.71 ± 0.23	0.27%	85.00 ± 0.11	0.11%
7	82.63 ± 0.11	0.13%	82.99 ± 0.09	0.08%
Mean	84.15 ± 0.19	0.23%	84.47 ± 0.14	0.16%

Table 6.6: Repeatability of Point Source Sensitivity measured on gamma camera during the calibration process, where camera setup was reset between measurements

Point source sensitivity measurements show high repeatability when equipment is not adjusted between acquisitions – with a slightly higher variability if the equipment is reset (albeit still of a relatively small magnitude). This is likely due to minor differences in positioning of the camera heads, positioning of the extending arm that holds the point source, and potentially due to movement of the liquid in the point source as it is repositioned.

Differences in sensitivity between the two detectors are noted to be minimal (on average 0.3%). For comparison, acceptance testing of this gamma camera in 2019 showed H2 to be 0.5% more sensitive for ^{99m}Tc than H1 (Table 6.7). The point source sensitivity does use a different method than was used at acceptance testing, where an extended source was positioned at a fixed distance from each detector (following the protocol outlined by the Gamma Camera Assessment Team of the Medical Devices Agency [302]).

	H1 Sensitivity (cps/MBq)	H2 Sensitivity (cps/MBq)	Relative difference
Acceptance testing 2019	88.72	89.17	0.5%

Table 6.7: Camera sensitivity for ^{99m}Tc measured by planar source at 10 cm (using Gamma Camera Assessment Team methodology) as part of acceptance testing of the camera.

A difference in sensitivity between the two detectors of this magnitude is to be expected (for example, Institute of Physics and Engineering in Medicine (IPeM) Report 111 suggest that the relative sensitivity of multiple detectors for SPECT imaging should be within $\pm 10\%$ [303], without affecting visual interpretation). The calibration process should adapt the resulting quantitative results for any difference between the camera heads.

6.2.3 Reproducibility of Point Source Sensitivity

Using ^{99m}Tc -Pertechnetate, sensitivity was measured over a period of 11 consecutive days, with the process repeated over a 12 day span, each time making 9 measurements (Table 6.8). Due to the relatively short half-life, a new ^{99m}Tc source was dispensed each day. The resultant CoV is significantly higher than either measurement of repeatability, suggesting that some amount of changing camera performance with time could be impacting the result (though uncertainty in the calibrator measurements will also have an effect).

	number of measurements	H1 CoV	H2 CoV	mean activity
11-day period	9	1.19%	1.15%	139.3 ± 10.0
12-day period	9	1.67%	1.51%	141.8 ± 11.5
mean CoV		1.43%	1.33%	

Table 6.8: Reproducibility of Point Source Sensitivity measured on gamma camera during the calibration process, for two different periods, each consisting of 9 measurements.

A study looking at short-term reproducibility of a ^{57}Co point source sensitivity examined 35 measurements over one month and found a maximum CoV of 0.34% [35]. As it re-used the same source (which due to the sealed nature of the source would be expected to have less variability in terms of source activity and positioning), the variability could be almost wholly attributed to day-to-day variations in camera performance. Comparison with the repeatability and reproducibility measurements performed here would therefore suggest that there are other factors affecting reproducibility in BroadQuant calibration measurements, outside of the expected fluctuations in camera performance, syringe activity measurements and setup repeatability: potential causes will be investigated further in the following Sections 6.2.4 and 6.2.5.

6.2.4 Effect of Point Source Volume

A variable that can be controlled by the user is the volume of liquid that can be used in the point source. The point source container can hold a volume of up to 1.55 ml (Siemens do not provide a recommendation on volume). However, the user manuals for gamma camera QC do recommend the use of gauze or cotton wool to absorb the liquid and keep it located in the conical 'point' of the container.

The user will be reliant on the activity concentrations supplied to their department, particularly for radionuclides supplied directly from the manufacturer (e.g. ^{177}Lu and other products that are not eluted locally from generators, as with $^{99\text{m}}\text{Tc}$). If the user can obtain activity of sufficient concentration to meet Siemens recommended activity within a small volume, it raises the question: does the volume used have an influence on the measured sensitivity and should the user dilute higher concentrations of radionuclides?

A Siemens point source container was prepared with sufficient gauze to absorb the liquid to be added – as the source is placed horizontally between the camera heads, the volume of liquid will no longer sit naturally in the ‘point’ of the container. The mass of the container was measured on a balance (which regular QC testing shows to be accurate to within ± 0.03 g). A source of $^{99\text{m}}\text{Tc}$ -Pertechnetate with high concentration was obtained and a small volume drawn into a syringe and measured in the radionuclide calibrator. After adding the activity to the point source, the syringe residue was measured, allowing the activity in the point source to be calculated. The point source container was re-weighed before a sensitivity calibration measurement was made on the gamma camera with LEHR collimators and the calculated sensitivity (in cps/MBq) was recorded for each detector. The sensitivity measurement was repeated after incremental volumes of water were added to the point source (between 0.1 and 0.2 ml increments), diluting the activity already present, without adding additional absorbent material. The added volume was calculated by weighing the point source before and after (assuming that 1 g of water = 1 ml of water).

For analysis, the measured sensitivity at each volume was compared to the measured sensitivity at the maximum volume (as variance of the sensitivity was lowest at the highest volumes) and the deviation plotted. The experiment was repeated on three separate occasions and the average CoV taken at different ranges of volumes, to determine if there is an optimal volume that reduces variability in the calibration process (Table 6.9). The experiment was also repeated without gauze in the point source.

	all volumes	>0.3 ml only	>0.4 ml only	>0.5 ml only
COV with gauze	1.35%	0.84%	0.72%	0.58%
	1.29%	0.66%	0.45%	0.46%
	1.47%	1.03%	0.80%	0.80%
mean COV with gauze	1.37%	0.84%	0.66%	0.61%
COV without gauze	1.01%	1.04%	1.06%	0.98%
<i>Sensitivity with gauze</i>	<i>83.74</i>	<i>83.16</i>	<i>83.01</i>	<i>82.96</i>
<i>Sensitivity without gauze</i>	<i>83.09</i>	<i>82.58</i>	<i>82.23</i>	<i>81.87</i>

Table 6.9: Measurements of Coefficient of Variation for point sources with increasing volume, with and without the use of gauze in the container. The mean sensitivities are given in italics.

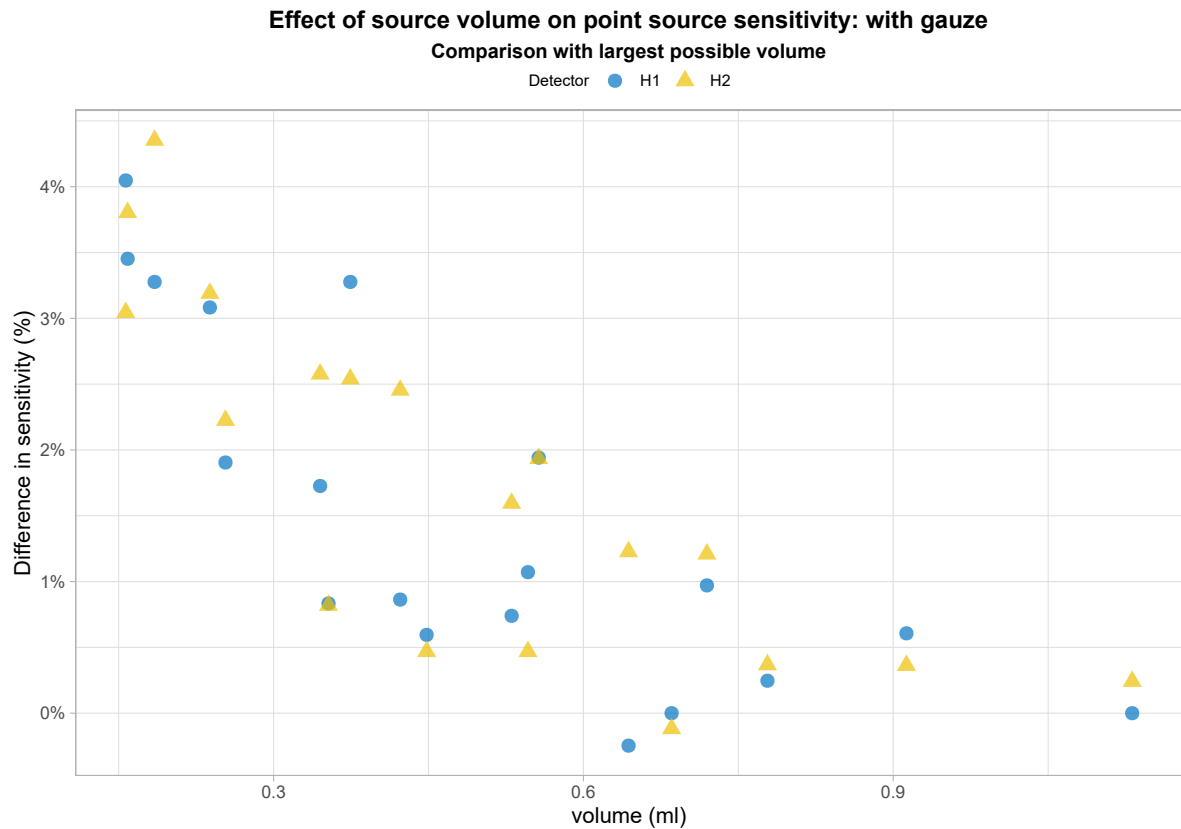


Figure 6.5: Point Source sensitivity measurements, relative to maximum volume (1.5 ml). Measurements are with gauze included in the point source container.

Figure 6.5 demonstrates that the measured sensitivity can vary with volume. This variation (as measured by CoV, shown in Table 6.9) is greater than would be expected based on the variation from repeated measurements with the same source (see Table 6.5).

Without gauze in the container, the CoV showed greater consistency over the range of volumes (Table 6.9 and Figure 6.6). However, this came at the expense of higher CoVs at all volumes than that achieved at >0.4 ml with the gauze.

The conclusion of this section is that to reduce the variability in the sensitivity calibration process, the optimal procedure is to construct the point source using gauze to absorb the liquid, ensure the total volume of liquid in the point source is > 0.4 ml, and maintain a consistent volume at each calibration. These propositions were implemented in all point source sensitivity measurements in the remainder of this thesis.

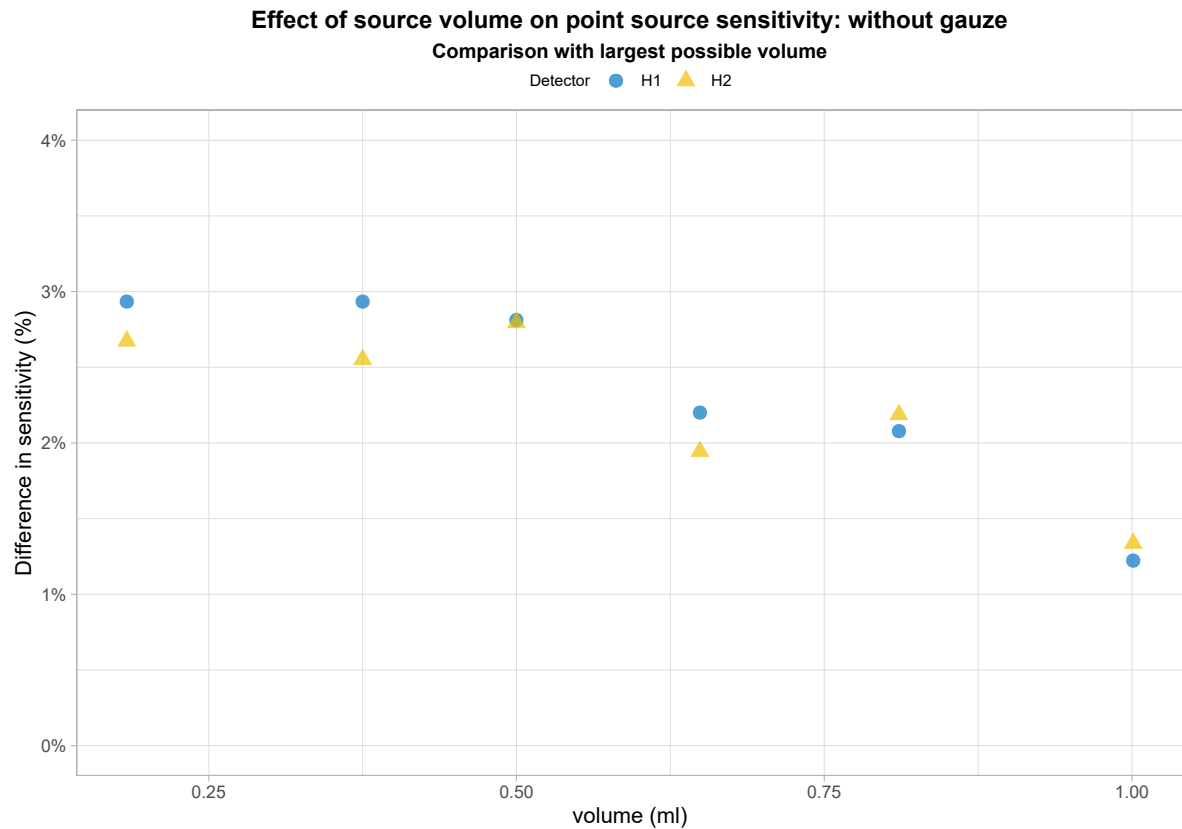


Figure 6.6: Point Source sensitivity measurements, relative to maximum volume (1.5 ml). Measurements are without gauze.

6.2.5 Linearity of point source sensitivity

The Siemens work instruction for the point source sensitivity calibration [304] details that the required activity should be between 37 and 185 MBq for the majority of radionuclides (the exception being ^{177}Lu where the required range is 370 to 1110 MBq), in order to produce the required count rate of between 5 and 30 thousand counts per second (kcps).

For modern gamma cameras such as the Siemens Intevo Bold, such count rates remain well within the linear range: for example, at acceptance testing of this camera the difference between expected and measured count rates for $^{99\text{m}}\text{Tc}$ in this range was on average 0.36%. Given expected clinical count rates averaged 5.4 kcps (maximum 8.5 kcps) for $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC and 8.7 kcps (maximum 35.2 kcps) for ^{177}Lu -DOTATATE imaged at 24 hours, there is not expected to be significant dead-time effects for clinical acquisitions.

The only radionuclide where it was possible to obtain activity with concentration $> 250 \text{ MBq/ml}$ in order to test the higher end of the suggested activity range was $^{99\text{m}}\text{Tc}$. Due to the half-life of $^{99\text{m}}\text{Tc}$, it was not possible to test one single source across the full range of activities. Therefore two sources were used, such that the activities overlapped.

A Siemens point source container was filled with a known activity of ^{99m}Tc -Pertechnetate and sensitivity measurements were made on the camera at least once every 2 hours, in line with NPL recommendations for the decaying-source method of testing linearity of ^{99m}Tc for radionuclide calibrators [297]. As the measurements had to be arranged around routine clinical imaging, the measurements for the higher activity source will feature a higher variation, due to the resetting of the camera between measurements (as per repeatability results in Table 6.6).

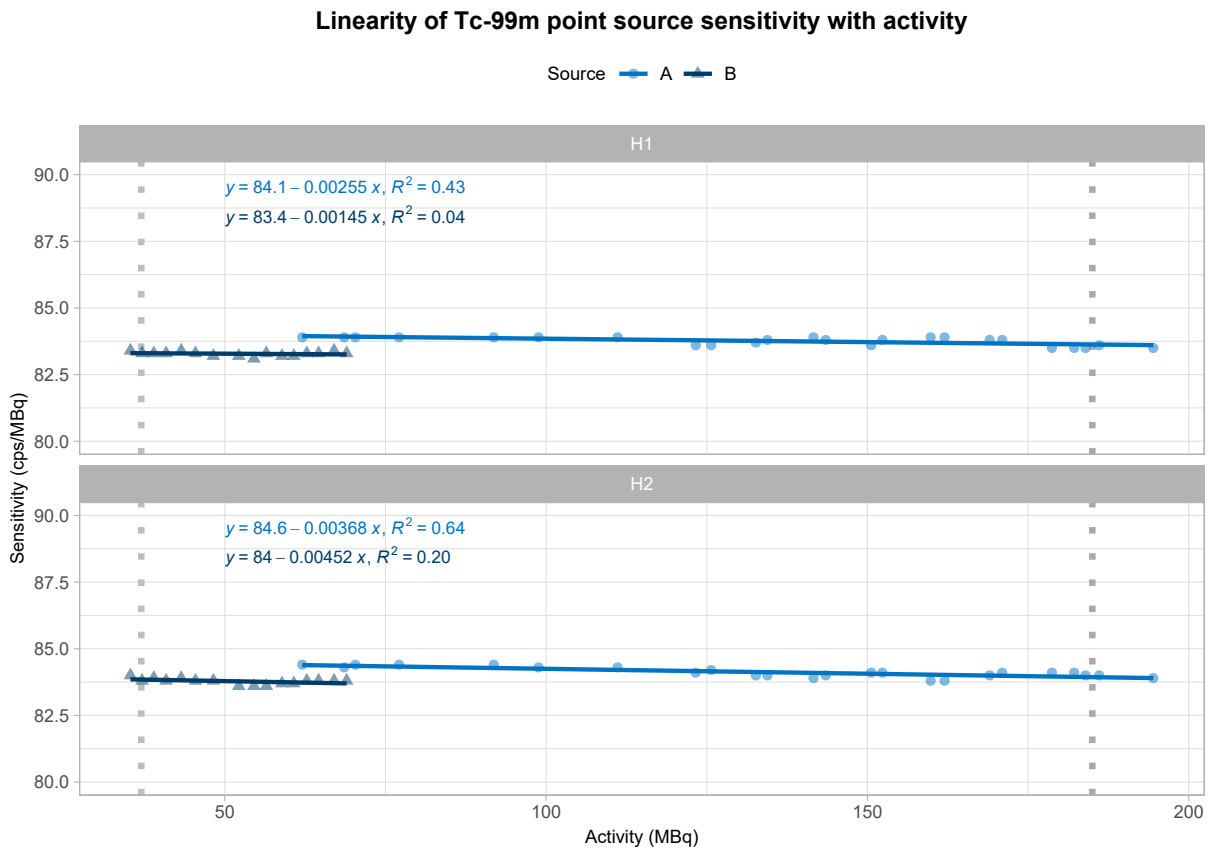


Figure 6.7: Linearity of ^{99m}Tc Point Source sensitivity measurements with decaying activity. Two separate sources (A and B) were used to cover the full range of suitable activities suggested by the manufacturer (range indicated by grey dotted lines).

Figure 6.7 demonstrates the sensitivity measurements for both sources: the difference between the average sensitivity measurement for the two sources was less than 2 standard deviations, as determined from the repeatability experiments (Section 6.2.3), and thus it can be assumed the two sources were not significantly different and can be assessed as if one source was used. The calculated gradient across the activity range for both of the ^{99m}Tc sources is also displayed in Figure 6.7: if the maximum gradient shown was extended to the minimum and maximum of the required activities, the estimated sensitivities would be 84.9 and 86.2 cps/MBq respectively (i.e. a change in sensitivity of 0.009 cps/MBq per 1 MBq increase). Taking the mid-point of this potential sensitivity range (85.55 cps/MBq), both the minimum and maximum estimated sensitivities would be within one standard deviation

($85.55 \pm 1.43\%$ is equivalent to 85.55 ± 1.22 cps/MBq) as assessed by the reproducibility experiment. We can therefore consider ^{99m}Tc sensitivity to be linear over the range of activities recommended by Siemens for use in the calibration step. The CoV of the sensitivity measurements was 0.22%, and consistent with the repeatability results in Table 6.6.

For ^{177}Lu linearity, it was only possible to obtain a concentration high enough to allow partial coverage of the recommended range (maximum activity obtained was 501 MBq, with Siemens recommended upper limit being 1110 MBq). Measurements were made once per day for fourteen days, except for weekends (Figure 6.8). A similar process as detailed in the last paragraph for comparing the possible range of sensitivity measurements suggested by the linearity gradient, with the uncertainty from reproducibility measurements, was performed for ^{177}Lu . This demonstrated that the estimated sensitivities would be within 1.03 standard deviations of the data from the reproducibility measurements, and it can therefore be considered that ^{177}Lu sensitivity is linear over the limited range tested.

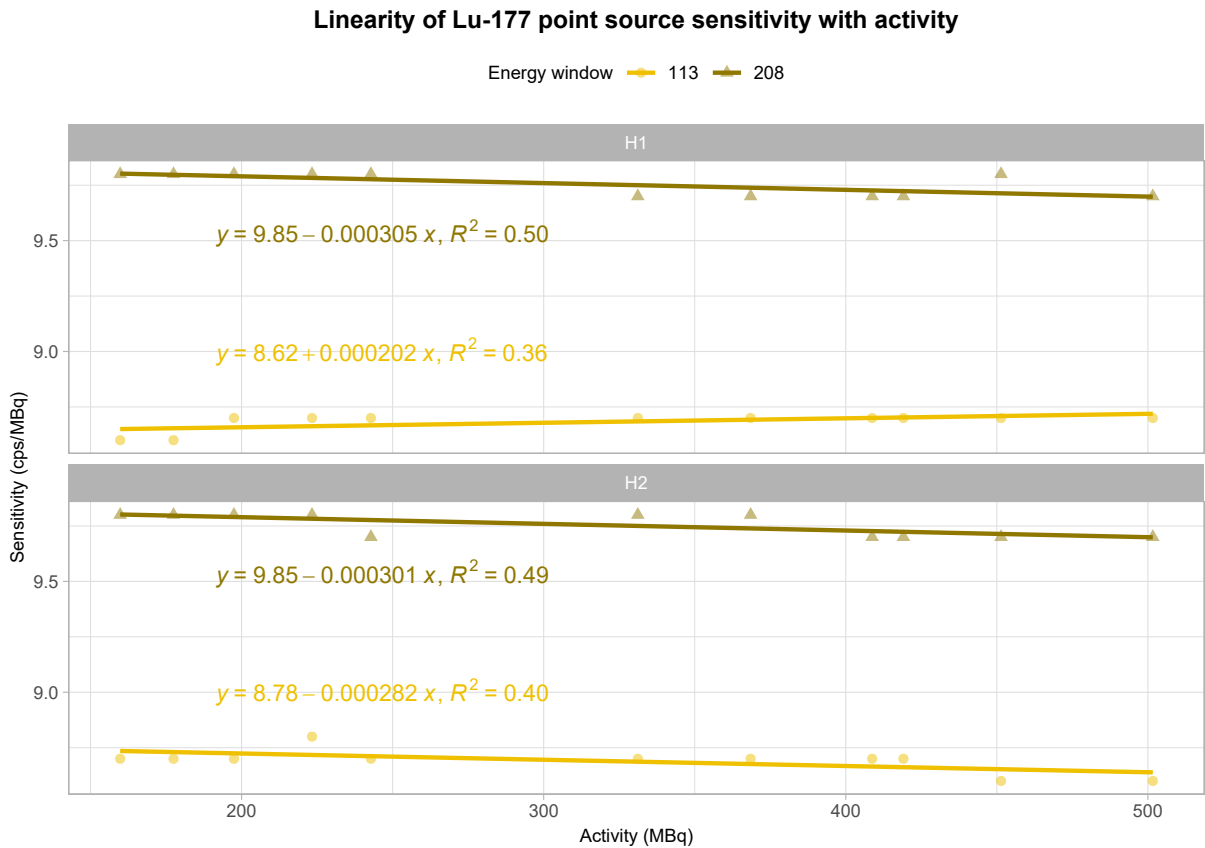


Figure 6.8: Linearity of ^{177}Lu Point Source sensitivity measurements using the decaying activity method.

6.3 Volume Sensitivity calibration

After the Point Source calibration, the BroadQuant calibration process requires a SPECT scan of a uniform phantom filled with a known activity of the radionuclide to be calibrated.

The required count rate is between 5 and 30 kcps [304]. The calibration workflow uses acquisition and reconstruction parameters specified by the manufacturer, that may differ from those used in clinical scans. Once the image has been reconstructed a Volume of Interest (VOI) is positioned by the user in the centre of the uniform phantom and the measured concentration is divided by the known activity concentration in the phantom to calculate the Volume Sensitivity Factor (VSF).

$$\text{Volume Sensitivity Factor, VSF} = \frac{\text{Measured activity concentration in phantom}}{\text{Known activity concentration in phantom}} \quad (6.3)$$

6.3.1 Repeatability of Volume Sensitivity calibration

A uniform cylindrical phantom was filled with a known amount of $^{99\text{m}}\text{Tc}$ -Pertechnetate and thoroughly mixed. The phantom was imaged with the acquisition parameters used in the Volume Calibration workflow (Table 6.10), supplied by the manufacturer for performing the BroadQuant calibration. After SPECT and CT acquisitions were performed, the camera was reset and the scan repeated, for five total acquisitions. Each acquisition was reconstructed with the OSCG algorithm using parameters set by Siemens for the calibration step. A 12 cm diameter spherical VOI was placed in the centre of the phantom, avoiding the edges. This VOI was copied to all acquisitions and the Volume Sensitivity Factor for each VOI was recorded. The Coefficient of Variation was calculated to examine the repeatability of the calibration acquisitions (Table 6.11).

parameter	setting
matrix	256×256
zoom	1.00
energy window	140 keV, width 15%
counts	100 Kcts in 1st view
orbit	non-circular, Step & Shoot
number of views	64

Table 6.10: Acquisition parameters for Volume Calibration workflow.

6.3.2 Reproducibility of Volume Sensitivity calibration

Five separate fills were made of uniform phantoms, each of which was scanned and reconstructed using the manufacturer's parameters for calibration.

Expected concentration (kBq/ml)	Measured concentration (kBq/ml) ¹	Volume Sensitivity Factor
62.6	68.7	1.10
62.6	69.6	1.11
62.6	68.7	1.10
62.6	69.3	1.11
62.6	69.8	1.11
	mean	1.11 ± 0.01
	CoV	0.71%

¹ The measured concentrations are displayed without sensitivity correction.

Table 6.11: Repeatability of Volume Sensitivity calibration. Expected and measured concentration are decay-corrected to preparation time.

days following Point Source calibration	expected concentration (kBq/ml)	measured concentration (kBq/ml) ¹	Volume Sensitivity Factor
0	49.2	54.1	1.10
1	62.6	68.7	1.10
3	59.1	65.6	1.11
8	47.4	52.0	1.10
9	49.7	55.3	1.11
		mean	1.10 ± 0.01
		CoV	0.65%

¹ The measured concentrations are displayed without sensitivity correction.

Table 6.12: Reproducibility of Volume Sensitivity calibration. Expected and measured concentration are decay-corrected to preparation time.

The CoV shown in Table 6.12 is similar to that calculated for repeatability and provides confidence that the VSF is not greatly affected by small changes in camera performance over relatively short time frames.

6.4 Overall uncertainty in the calibration process

Uncertainty in the gamma camera calibration process has been noted to be one of the major factors in overall uncertainty in patient dosimetry measurements [66]. The uncertainty in the Siemens BroadQuant calibration process is a combination of uncertainty from several components, which are added in quadrature using Equation 6.4 (where $u_{a,b,c}$ are the individual uncertainties from the components comprising the overall measurement).

$$\text{Combined Uncertainty, } u^2 = u_a^2 + u_b^2 + u_c^2 + \dots \quad (6.4)$$

The individual uncertainties, and the combined uncertainties for the various stages in the calibration process are detailed in the discussion of each uncertainty in Sections 6.4.1 to 6.4.3.

6.4.1 Uncertainty for radionuclide calibrator activity measurement

For ^{99m}Tc , the standard uncertainty in a measurement from a calibrator with traceability to the National Physical Laboratories standard can be assumed to be in the order of 2.0% [297]. This typical uncertainty is sufficient for the process of calculating a local uncertainty, as per EANM guidelines [296]. However, this site utilises a Fidelis Secondary Standard instrument for the purposes of cross-calibration and traceability of the radionuclide calibrators to NPL reference values. This allows for a more accurate calibration than to field instrument standards, with the uncertainty calculated as a combination of the Fidelis calibration factor, the CoV for the reference assay by the Fidelis instrument and the CoV of multiple measurements in the radionuclide calibrator (when the calibration factor is matched to the Fidelis).

For ^{177}Lu , NPL do not provide a similar typical uncertainty, and the local radionuclide calibrator had previously not been directly calibrated with a vial also measured at NPL. Instead, a local calibration factor had been derived from reference vials produced by the manufacturer of the clinical radiopharmaceutical ^{177}Lu -DOTATATE. However, a national inter-comparison exercise with NPL utilising a vial previously calibrated at NPL allowed for a traceable calibration factor to be obtained; this had a reported uncertainty of 1.01%.

Regular checks of agreement of clock time between the calibrator and gamma camera should maintain a difference of less than minute – hence the contribution to uncertainty in any ^{99m}Tc measurement would be <0.1% [297]. For ^{177}Lu and any other longer-lived radionuclide, this would be much smaller and inconsequential (in the order of <0.0001%).

The combined standard uncertainties in the radionuclide calibrator activity measurements are given in Table 6.13. For ^{99m}Tc the standard uncertainty of 1.53%, which is comparable to the activity measurement uncertainty measured by Scott, et al of 1.31% [305]. For ^{177}Lu , the standard uncertainty of 1.06% is of the same magnitude as the ^{177}Lu uncertainty reported by Gustafsson et al of 2% [66]. These uncertainties will be incorporated into the calculations for both point source and volume calibrations.

6.4.2 Uncertainty for Point Source Sensitivity calibration

It is recommended, based on the work in Section 6.2.4, to use a consistent volume greater than 0.4 ml when preparing point sources. This approach would mean there should be minimal additional uncertainty in the point source sensitivity measurement due to the volume in the point source. The use of gauze is also recommended. As the sensitivity appears to be

Calibrator uncertainty components	Measured uncertainty
For ^{99m}Tc	
Secondary standard uncertainty on factor	0.90%
Secondary standard CoV	0.03%
Field instrument CoV	0.13%
Clock	0.10%
Syringe measurement CoV	0.08%
Combined ^{99m}Tc calibrator uncertainty	0.92%
For ^{177}Lu	
NPL inter-comparison source uncertainty	1.01%
Field instrument CoV	0.25%
Clock	<0.0001%
Syringe measurement CoV	0.19%
Combined ^{177}Lu calibrator uncertainty	1.06%

Table 6.13: Uncertainties for activity measurements in the radionuclide calibrator for both ^{99m}Tc and ^{177}Lu .

linear for both radionuclides within the recommended activity limits proposed by Siemens, activities within those ranges should be used to further minimise uncertainties.

For the combination of uncertainties (Table 6.14), the uncertainty in the measured activity will be as defined in Section 6.4.1. The CoV of repeated identical measurements should be used to determine the uncertainty in the measured counts [296, 306] for the calibration process. For both radionuclides, the repeatability measurements will be used (for ^{177}Lu , point source measurements have a sensitivity for each of the two energy peaks; here the poorer CoV was chosen, as a worst-case example).

Point source uncertainty components	Measured uncertainty
For ^{99m}Tc	
Repeatability CoV	0.23%
Calibrator uncertainty (Table 6.13)	0.92%
Combined ^{99m}Tc point source uncertainty	0.95%
For ^{177}Lu	
Repeatability CoV	0.63%
Calibrator uncertainty (Table 6.13)	1.06%
Combined ^{177}Lu point source uncertainty	1.23%

Table 6.14: Uncertainties for the point source calibration procedure for both ^{99m}Tc and ^{177}Lu .

6.4.3 Uncertainty for Volume Sensitivity calibration

The volume sensitivity measurement will have uncertainty from both the activity measurement and the acquired counts measurement, similar to the point source uncertainty. However, there may also be contributions from uncertainties in the measurement of the volume of the phantom and the CT attenuation correction process. As repeatability measurements with ^{177}Lu were not possible during the course of this thesis, the CoV of all ^{177}Lu calibration factor measurements was used instead. The final contributing uncertainties are noted in Table 6.15.

Patient weighing scales

Before administration of the radiopharmaceutical, patients who are to undergo SPECT imaging have their height and weight measured on a set of Seca weighing scales (Seca UK, Birmingham, United Kingdom), which are tested on an annual basis by a local calibration service. The scales measure in kilograms to two decimal places, and the most recent testing showed no deviation from nominal value across a range of masses. Therefore, the only uncertainty in the measured weight will be due to truncation and can be calculated as per Equation 6.5 [307] (where the constant 12 relates to the variance of the rectangular probability distribution):

$$u(\text{Scales}) = \frac{0.01}{\sqrt{12}} = 0.29\% \quad (6.5)$$

The phantom was also measured using the patient scales, therefore this uncertainty will be applied to the calibration process. The volume of the phantom could also be measured by directly measuring the liquid required to fill the phantom: a colleague made repeated measurements of the volume of a NEMA IQ phantom with a CoV of 0.09% (S. Williams, personal communication, 20th January 2023). Given the patient scales will be used in clinical practice, the larger uncertainty from a direct measurement on the scales will be used for the overall analysis.

CT density data for attenuation correction

As CT images are used to correct for attenuation, the process of measuring the density of tissue (in a patient, or water in a uniform phantom) and converting it to a LAC that is specific to the radionuclide being investigated could result in uncertainty in the resulting calibration.

Gustafsson et al modelled the uncertainties in this process by both random sampling of noise in the Hounsfield Unit (HU) measurements, as well using simulations to add uncertainty to the conversion to LAC for ^{177}Lu [66]. However, the simplified composition of a uniform phantom (i.e. consisting solely of water with a thin casing of Perspex) versus a patient's

habitus and the highly-filtered CT reconstruction used for Attenuation Correction means there is likely to be little variance in the HU measurements of the phantom (see Figure 6.9). The EANM guidelines suggest that scatter and attenuation correction are sufficiently complex to preclude individual analysis of their uncertainties, and their effects will be seen nonetheless in measurement of RCs [296]; as such, this effect will be ignored during this stage of uncertainty calculation.

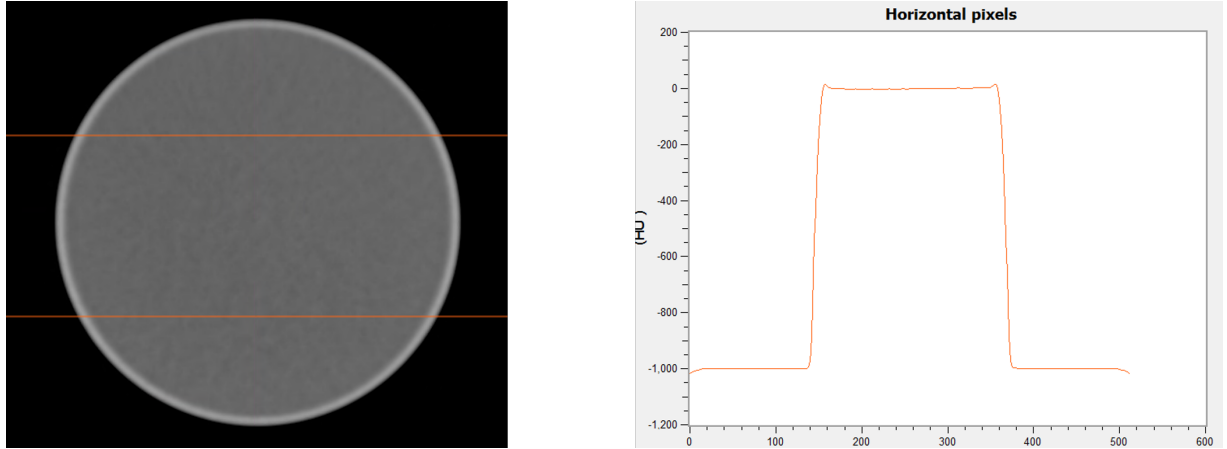


Figure 6.9: CT image of a uniform phantom used for calibration (left) with a profile displaying minimal variation of Hounsfield Units (right).

Voxelisation

The uncertainty on positioning of a VOI, $u_{\text{Voxelisation}}$ drawn on a clinical image (e.g. segmenting a tumour), of a given diameter d relative to the SPECT matrix can be calculated based on the pixel size of the SPECT image [296]. This uncertainty is caused by the inclusion or exclusion of voxels on the edges of the VOI, interpolating what should be a spherical VOI onto a matrix with cubic voxels. However, for this phantom analysis where the VOI is placed centrally within a uniform phantom, avoiding edge effects, then the voxels included or excluded should have similar activity concentrations. Hence, the contribution to uncertainty will be negligible and will be discounted for the purposes of this study.

6.4.4 Combination of uncertainties

labelcombined unc There is not sufficient information from the manufacturer to determine how the two stages of calibration (Point Source Sensitivity and Volume Sensitivity) are combined in order to mathematically define exactly how their uncertainties should be combined. As a compromise, it is assumed both calibrations contribute equally to the process of calibrating the system for quantification, and hence uncertainties will be added in quadrature (Table 6.16). For $^{99\text{m}}\text{Tc}$, the combined standard uncertainty is 1.53% and for ^{177}Lu , the combined standard uncertainty is 2.26%.

Volume sensitivity uncertainty components	Measured uncertainty
For ^{99m}Tc	
Weighing scales	0.29%
Repeatability CoV	0.71%
Calibrator uncertainty (Table 6.13)	0.92%
Combined ^{99m}Tc volume sensitivity uncertainty	1.20%
For ^{177}Lu	
Weighing scales	0.29%
Repeatability CoV	0.63%
Calibrator uncertainty (Table 6.13)	1.06%
Combined ^{177}Lu volume sensitivity uncertainty	1.23%

Table 6.15: Uncertainties for the volume sensitivity calibration procedure for both ^{99m}Tc and ^{177}Lu .

Combined calibration uncertainty components	Measured uncertainty
For ^{99m}Tc	
Point source sensitivity uncertainty (Table 6.14)	0.95%
Volume sensitivity uncertainty (Table 6.15)	1.20%
Combined ^{99m}Tc BroadQuant calibration uncertainty	1.53%
For ^{177}Lu	
Point source sensitivity uncertainty (Table 6.14)	1.23%
Volume sensitivity uncertainty (Table 6.15)	1.89%
Combined ^{177}Lu BroadQuant calibration uncertainty	2.26%

Table 6.16: Combined uncertainties for the BroadQuant calibration process for both ^{99m}Tc and ^{177}Lu .

There are several studies that have investigated uncertainties in the calibration process for quantitative SPECT. Generally, they were conducted using centre-specific calibration protocols, whereas this study utilised the Siemens BroadQuant package. There are also small differences in the quantities included in the uncertainty budget on a study-by-study basis, due to differences in equipment, national regulations on metrology and the availability of traceable standards.

For ^{99m}Tc , Willowson et al reported an uncertainty of 2.7% in a sensitivity calibration [28], which although not directly comparable, is higher than the uncertainty calculated with ^{99m}Tc in this thesis. The use of a uniform phantom for calibration was calculated to have an uncertainty of 8% by Zeintl et al [29]. However, this evaluation included a presumed 5% uncertainty due to the radionuclide calibrator activity measurement.

A recent publication by NPL on uncertainty in filling a phantom with radioactivity reported an expected uncertainty for ^{177}Lu in a uniform phantom of 1.5% [308]. This can be compared to the measured uncertainty of 1.06% for activity measurement in this study (though this

omits some phantom-specific uncertainties that are included in the NPL uncertainty budget). Including the gamma camera contribution of uncertainty into the calibration process, a study of ^{177}Lu calculated uncertainties of 3.1% and 3.2% for point source and uniform phantom calibrations respectively [309]. However, these were dominated by a 3.0% uncertainty in the activity measurement [309]. Activity measurement was also reported to be the dominant factor (at up to 4%) by a multi-centre study on harmonising ^{177}Lu calibration in the Netherlands [273]. A similar pan-European study reported the uncertainty component due to activity measurements to be between 2.83% and 4.25% in three out of four centres; only one centre reported an uncertainty contribution of 0.52% [310]. A large study focussing specifically on accuracy of radionuclide calibrators found 24% of calibrators were outside the recommended accuracy of $\pm 5\%$ [311]. Such discrepancy in accuracy results in a large uncertainty contribution, which explains the higher uncertainties for ^{177}Lu than found in this study, where accurate inter-comparison of activity with an NPL reference source results in a lower uncertainty contribution.

6.5 Conclusion and relevance for the thesis

This chapter has investigated the calibration process for Siemens BroadQuant by evaluating the accuracy and reproducibility of the measurements that comprise the calibration. It has demonstrated that for the two radionuclides tested, $^{99\text{m}}\text{Tc}$ and ^{177}Lu , the measurements are robust, reproducible and have minimal uncertainties. Testing demonstrated several areas where the user can improve the accuracy of the calibration process and the following recommendations are made for performing the BroadQuant calibration process:

- Activity measurement for calibration point sources should be assayed in a syringe, rather than a point source container. This aids repeatability and more closely replicates the activity measurement of patient studies.
- If direct measurement of a point source container in the calibrator is preferred, a method to replicate positioning (such as the jig in Figure 6.4) should be used.
- The point source for sensitivity calibrations should use gauze to absorb the liquid and contain at least 0.4 ml of volume, in order to reduce variability. The volume used should remain consistent across all calibrations and checks.
- Following the recommendations above, an overall standard uncertainty of 1.53% for $^{99\text{m}}\text{Tc}$ and of 2.26% for ^{177}Lu can be applied to the BroadQuant Calibration Factor.

The accurate measurement of activity has a large effect on the uncertainty in the measured Calibration Factor: it highlights the importance of having traceability in the activity measurement. The UK metrology service provided by NPL includes traceability for ^{177}Lu ,

that may not be available for all departments. Hence the uncertainties calculated above are lower than reported elsewhere for ^{177}Lu (where a larger uncertainty in the activity measurement has to be assumed). As new radionuclides are introduced, particularly therapeutic radiopharmaceuticals that have gamma emissions for SPECT imaging with an increased demand for dosimetry, there will be a need for traceability of activity measurements to allow for accurate quantification.

The uncertainties inherent in the calibration process will be informative when assessing the accuracy of quantitative measurements in the following chapter, and will inform clinical measurements made in later chapters.

Chapter 7

Assessing stability and accuracy of quantitative calibrations

7.1 Introduction

After the quantitative calibration process for a SPECT gamma camera has been performed, it is imperative to assess the accuracy of a quantitative measurement, to ensure any potential error in the calibration process will not be applied to patient results. Furthermore, the principles of quality assurance require that continuing checks are made over the period of time when a particular calibration is applied, to confirm that measurements remain consistently accurate and that variations in camera performance do not require that the system is recalibrated.

Recent EANM guidelines on SPECT quantification provide the advice that a validation of the calibration should be performed with a uniform phantom and the intended clinical acquisition protocol (though not with the same filling of the uniform phantom used to produce the calibration) [11]. The guidelines also suggest measuring the accuracy of the quantitative measurement across differing volume sizes, commonly known as activity RCs. The RC are a common feature of the calibration process in PET, where a NEMA-specified phantom is most often used. This chapter will make such measurements as part of the larger goal of providing a holistic characterisation of the BroadQuant package, including comparing results to measurements made using BroadQuant by other studies in the literature.

Periodic checks of the system sensitivity are also recommended in the guidelines (here meaning the sensitivity used to derive the calibration factor for quantification), but the guidance leaves it to local scientists to define the frequency of such testing, due to variations in the procedures and recommendations between the different manufacturers [11]. For Siemens BroadQuant, the manufacturer recommends a monthly re-calibration of each radionuclide [36]. Such frequent re-calibration would have implications in terms of cost of the radionuclide, as well as staff radiation exposure and camera availability. The process of re-calibration is also an opportunity for an error to occur and result in an incorrect calibration, potentially affecting clinical results. Therefore, it would be beneficial if the existing calibration can be proven to remain accurate beyond the manufacturer's recommended lifetime, allowing for less frequent re-calibration. To date, there have not been suitable studies in the literature

examining the stability of a quantitative calibration over prolonged time periods. Hence, this thesis will aim to provide such information for Siemens BroadQuant, specifically for ^{99m}Tc and ^{177}Lu .

7.2 Acquisition and Reconstruction parameters

For some of the experiments detailed in this chapter, specific acquisition and reconstruction parameters were required (i.e. manufacturer-supplied parameters fixed for calibration acquisitions). Otherwise, consistent parameters across all experiments were used: the parameters are recorded here to reduce repetition in each experiment.

7.2.1 Collimator choice and Energy windows

The departmental standard energy window for ^{99m}Tc was used, consisting of a photopeak at 140 keV and window width of 15% (an immediately adjacent lower scatter window was also used, with width of 15%). The LEHR collimators were used.

For ^{177}Lu , imaging was performed with MELP collimators and the energy window consisted of two photopeaks at 113 keV and 208 keV, each with a $\pm 15\%$ energy window – this is in line with guidelines for ^{177}Lu quantification for radionuclide therapy [68]. It has been shown that there is a minimal contribution to the energy spectra from Bremsstrahlung in ^{177}Lu , with a maximum of 0.12% for 113 keV photopeak [312]. However, some reports conclude that the influence of Bremsstrahlung on the 113 keV peak (in conjunction with 208 keV downscatter and Compton scatter of 113 keV photons) makes the 113 keV peak unsuitable for quantification if using a single peak [309]. Some studies suggest using only the 208 keV peak [271]. The Medical Internal Radiation Dose committee (MIRD) have fewer concerns regarding this, and suggest that if the 113 keV energy window images are reconstructed separately and reviewed for image quality, then the lower window can be included in the reconstruction (though it must be noted this applies specifically to dosimetry, and has not been proven to be appropriate for textural analysis - for which there is currently no formal position to either include or exclude the 113 keV window). Figure 7.1 shows images from both 208 keV and 113 keV energy windows for a NEMA phantom, with Figure 7.2 showing likewise for patient images. Visual inspection of these images, by multiple Medical Physics Experts, resulted in the 113 keV window being deemed as providing comparable image quality to the 208 keV window. With the rationale that including the 113 keV window will increase the number of counts in the image and hence reduce uncertainty (as seen in Figure 7.12, where reduced acquisition time resulted in reduced reproducibility), both energy peaks will be included in the quantification work on ^{177}Lu in this thesis.

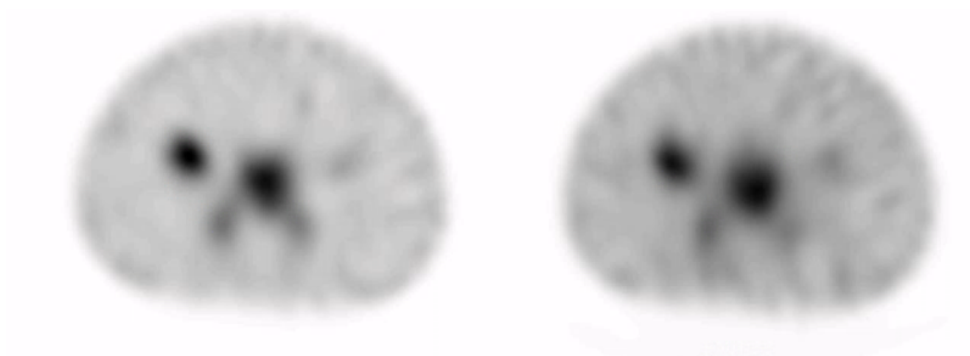


Figure 7.1: Reconstructed axial slices of 208 keV (left) and 113 keV (right) energy windows for ^{177}Lu NEMA phantom.

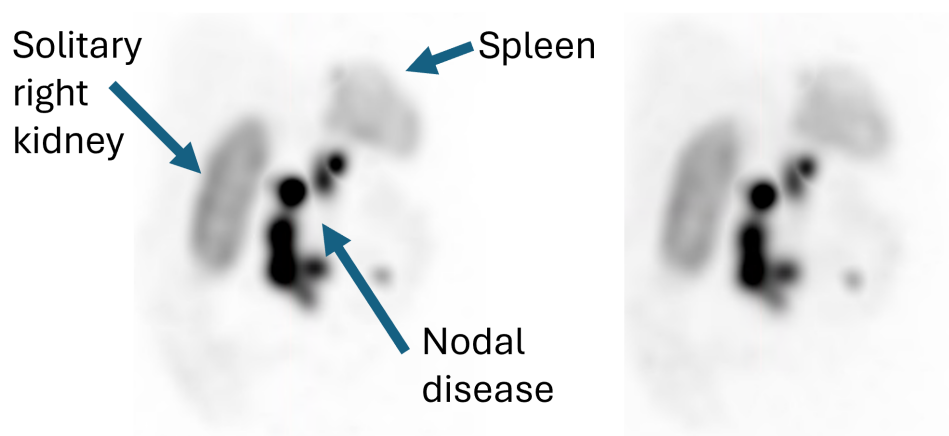


Figure 7.2: Maximum Intensity Projection of 208 keV (left) and 113 keV (right) energy windows for ^{177}Lu -DOTATATE post-therapy SPECT images.

Although the EANM/MIRD guidelines recommend a Triple Energy Window method for scatter correction of the 113 keV photopeak, the clinical workflow at this site includes only Dual Energy Window scatter correction, chosen originally to offer improvements in contrast and count statistics [313]. To retain consistency with our previous clinical imaging, this was maintained as the preference for scatter correction in the quantitative images.

7.2.2 Matrix

For the OSCG algorithm, Siemens recommend that a matrix size of 256 x 256 is used to obtain greater spatial accuracy in a matrix size closer to that acquired for CT, which is 512 x 512 [51]. This was adopted into local acquisition protocols.

7.2.3 Number of projections

Departmental protocols use 128 projections (i.e. 64 views per detector) as default. This relates to historically requiring $<3^\circ$ angular sampling for reconstruction using FBP. Recommendations for quantitative ^{177}Lu SPECT [68] details clinical studies using between 60-120 projections - as we are using a 256 x 256 matrix, the larger number of projections is beneficial to better match the data sampling of the matrix [48], hence 128 projections will be maintained.

7.2.4 Orbit type

Local unpublished work demonstrated that the time per view could be reduced while maintaining count density and quantitative accuracy using a body-contouring orbit and ADS, instead of the default Step and Shoot. This acquisition method was later validated by colleagues elsewhere for ^{177}Lu [314]. The use of ADS has become the departmental standard and implications for the extent of reducing scan time will be discussed later in this chapter.

7.2.5 Siemens default reconstruction settings

The Siemens default reconstruction options for the OSCG algorithm, termed “Best”, “Standard” and “Fast”, were investigated to determine their usefulness for this thesis and wider clinical application. Each of the three options has predetermined combinations of iterations and subsets. These vary based on the total counts acquired [315]. Initial testing assessed the noise, measured by CoV (Equation 6.2), in a uniform phantom (acquired for four different radionuclides; Table 7.1). Each phantom was reconstructed with the three Siemens default reconstruction options. The CoV is expected to increase as acquisition time decreases; a simple consequence of reduced acquisition time is reduced number of photons,

and hence higher noise in the image. This pattern can be seen for all three reconstruction options in Figure 7.3 for ^{111}In and $^{99\text{m}}\text{Tc}$, as total acquisition time is reduced (in this case, by reducing time per view and keeping number of views constant). However, an unforeseen property of the Siemens default reconstruction options was discovered. For both ^{131}I and ^{177}Lu an initial increase in CoV with reducing acquisition time is followed by a sharp decrease in CoV as time per view is reduced to 4 seconds. Upon investigation of the DICOM headers for these images, it was discovered that the Gaussian filter applied by the default reconstruction was varying with the time per view (or more strictly, varying with the number of counts acquired in the image); a smoother filter was applied as the number of counts in the image reduced. The experiment was repeated with manual selection of a fixed set of reconstruction parameters (Figure 7.4), confirming the unexpected CoV performance was due to the automatic adjustment of the filtering. This adjustment can be seen as an attempt by the manufacturer to maintain visual image quality for poorer count images, but it is not suitable for measuring quantitative accuracy. The inability to control the smoothness of the images when using the default reconstruction settings makes them unsuitable for the planned work of this thesis and therefore manual selection of the reconstruction parameters will be used throughout this thesis, and discussed below.

Radionuclide	Phantom activity (MBq)
$^{99\text{m}}\text{Tc}$	348
^{177}Lu	300
^{111}In	204
^{131}I	30

Table 7.1: Activities added to uniform phantom for Coefficient of Variation measurement using different reconstruction options.

7.2.6 NEMA IEC Body phantom

Once it was determined that manual selection of reconstruction parameters was required, it was necessary to test a range of parameters to determine the optimal combination of parameters and evaluate if the different radionuclides under consideration required different parameters. The NEMA IEC Body phantom (Data Spectrum, USA) [316] shown in Figure 7.5 is recommended by the NEMA NU 2-2018 standard [317] for testing of the performance of PET systems and provided a suitable phantom to test optimal quantitative SPECT parameters.

Most SPECT phantoms are constructed in a cylindrical shape, however the NEMA IEC Body phantom has a cross-section that better replicates the body habitus of a patient lying on the scanner bed. The cross-section approximates a Body Mass Index of 25 kg/m^2 [272]. A cylindrical insert mimics the density of lung tissue when filled with a mixture of polystyrene beads and water. Six hollow spheres (the NEMA NU 2-2018 standard for PET imaging

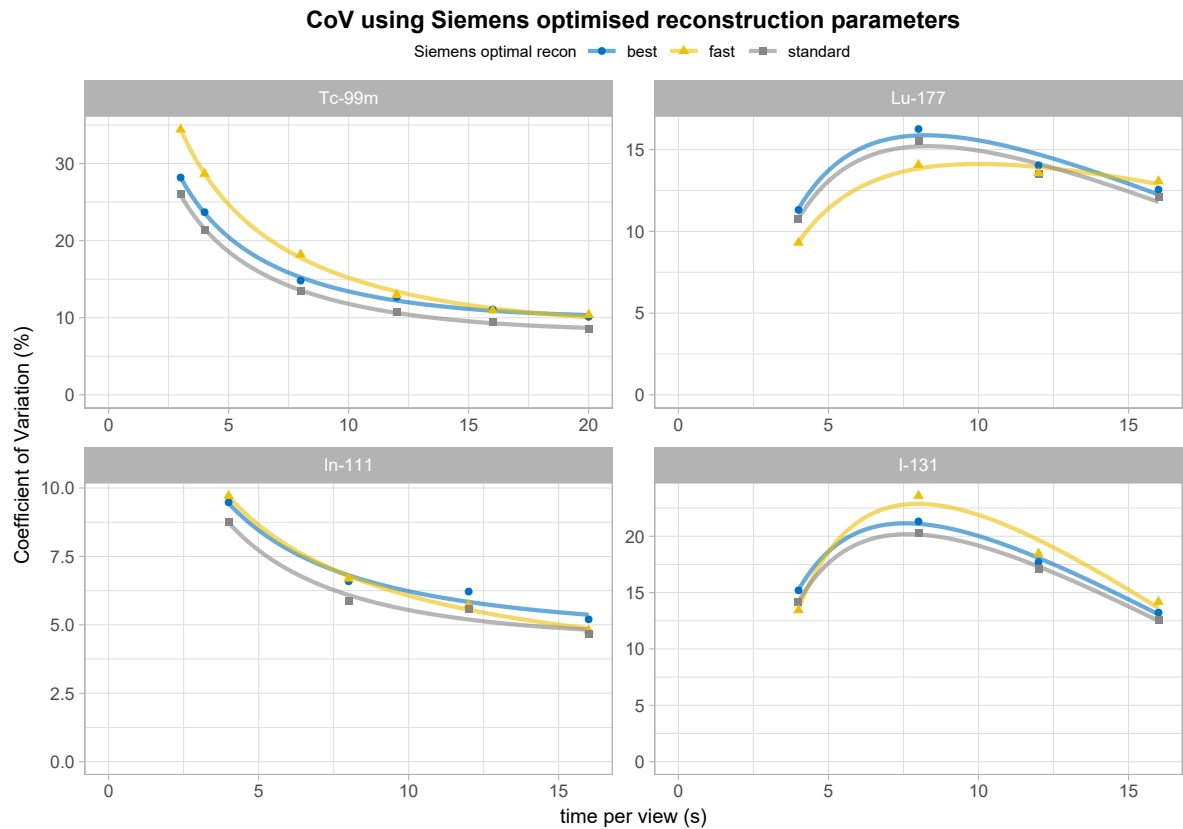


Figure 7.3: Coefficient of Variation measurements for different radionuclides with reducing time per view, using Siemens Optimal Recon presets.

requires spheres from 10 mm to 37 mm diameter) can be filled through ports in the top surface of the phantom. The NEMA NU 2-2018 standard recommends filling of all six spheres with radioactivity, whereas the previous standard (NEMA NU 2-2012) advises the two largest spheres were to be filled with ‘cold’ water for measuring cold contrast recovery. As SPECT has poorer spatial resolution than PET imaging, filling of all six spheres (with the option of larger spheres - see Section 7.3.1) was deemed necessary to measure the quantitative accuracy across a range of sizes resolvable by SPECT.

7.2.7 Measurement of Recovery Coefficients

The RC is a measure of how close a quantitative measurement is to reproducing the actual, known concentration of radionuclide in a phantom as a function of sphere size. For both PET and SPECT, the RC is a surrogate measure of the PVE.

There are two main methodologies in the literature for calculating RCs – although often used interchangeably, they provide different information. The Hot Contrast Recovery Coefficient (HCRC) is often used in PET imaging, and the formal definition is taken from the NEMA standards [317].

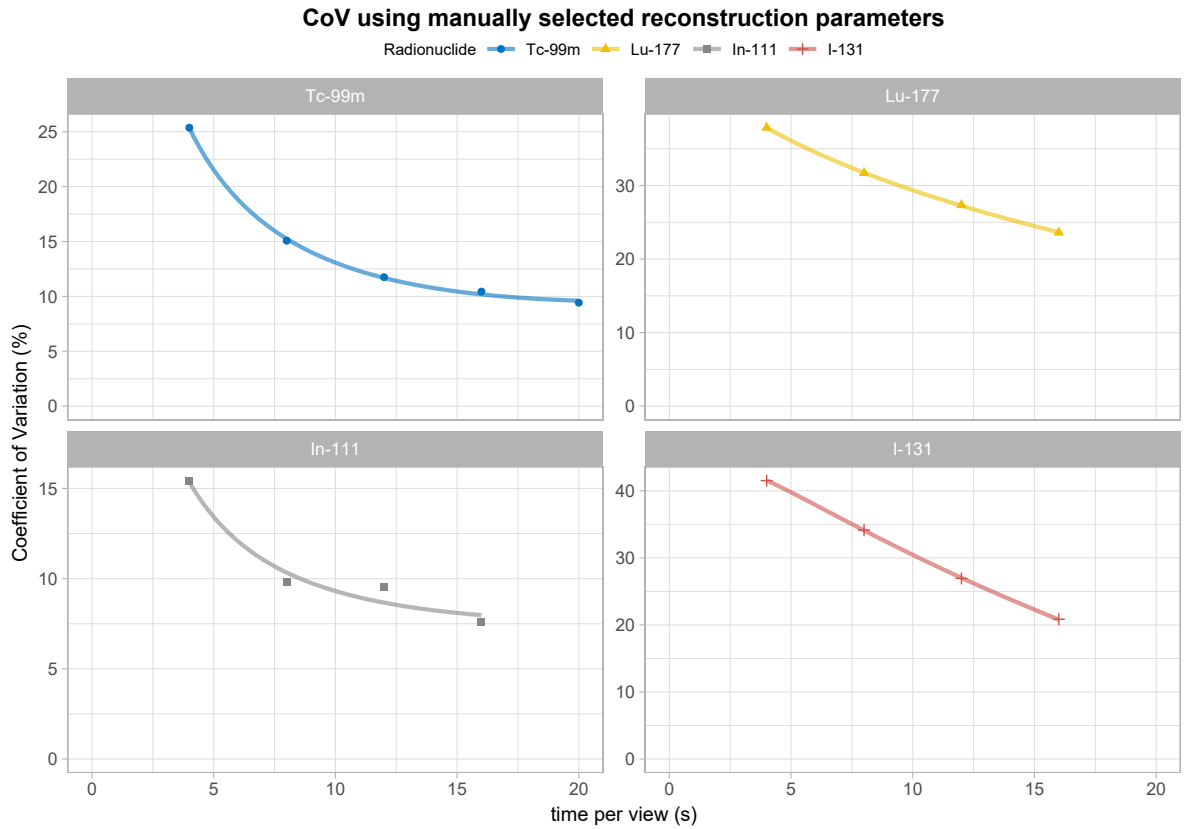


Figure 7.4: Coefficient of Variation measurements for different radionuclides with reducing time per view, replicating Siemens Optimal Recon presets but using manual selection of iterations & subsets in order to keep filtering constant.

$$\text{Hot Contrast Recovery Coefficient, HCRC}_j = \frac{\left(\frac{C_j}{C_{B,j}}\right) - 1}{\left(\frac{a_j}{a_B}\right) - 1} \times 100\% \quad (7.1)$$

Where C_j is the average counts in the Region of Interest (ROI) drawn for sphere j , $C_{B,j}$ is the average of the counts in the background ROI for sphere j , a_j is the known activity concentration in the spheres and a_B is the known activity concentration in the background of the phantom [317]. The HCRC measures how well differently sized lesions would be detectable against a background concentration of radioactivity: for example, liver metastases against physiological uptake in the liver in a [^{18}F]FDG PET scan.

A different definition, often used in quantitative SPECT, is concerned with how close a measured concentration in a sphere is to the known concentration of activity: i.e., a measure of how quantitative accuracy varies with lesion size. The definition in Gear et al [296] uses measured count rates; though more commonly it is defined by activity concentration, as in Peters et al [272]:

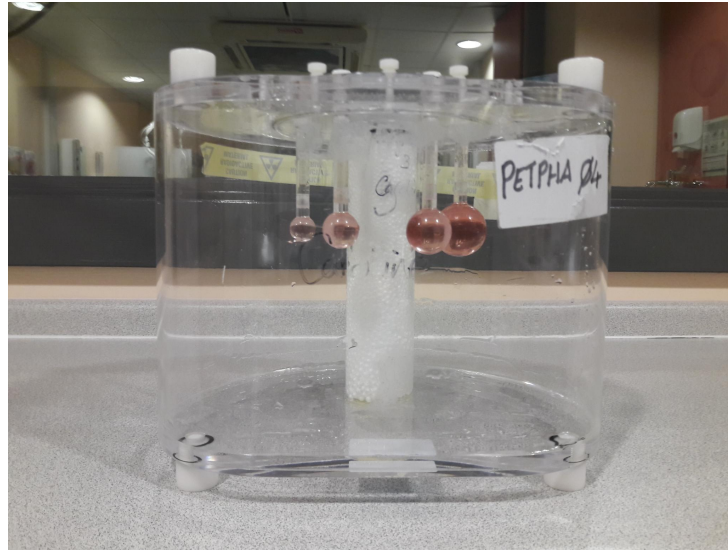


Figure 7.5: Example of a NEMA IEC Body phantom, with lung insert in-situ and spheres filled with red dye for visibility.

$$\text{Recovery Coefficient, } RC_j = \frac{a_j}{A} \quad (7.2)$$

Where RC_j is the Recovery Coefficient for a sphere j , a_j is the measured activity concentration in the image for a sphere j and A is the known activity concentration in the spheres.

Given the aims of this thesis are closely associated with quantitative accuracy, and lesion detectability is a different (though equally important) problem, the definition from Peters et al used in Equation 7.2 will be used henceforth when discussing RCs. The specific choice of metric to measure may differ depending on the intended application:

- RC_{mean} where the mean pixel value in the VOI is used
- RC_{max} where the maximum pixel value is used
- RC_{peak} is dependent on the software package; but for Hermes Hybrid Viewer "peak" is defined as the mean of the voxels in the "hottest" 1 cm^3 constrained in the demarcated VOI [318]. If the volume of the VOI is $< 1 \text{ cm}^3$, no value is given for peak.

For example, as suggested by Peters et al, RC_{max} may be preferred for evaluating treatment response where reproducibility between repeat measurements is important, whereas for dosimetry applications RC_{mean} is more appropriate, as the activity must be measured accurately [273]. For the purposes of this section, where the aim is to characterise the quantitative accuracy of the BroadQuant software, RC_{mean} is the most suitable metric.

One use for RC curves is to apply PVC [67]. The PVE may underestimate mean activity in a sphere of 40 mm diameter by around 20% for $^{99\text{m}}\text{Tc}$ [319] or 30% for ^{177}Lu [320] - results

in keeping with the findings of this thesis (Figure 7.9). However, no proposed method for PVE correction based on RC measurement in spheres in a phantom fully addresses the fundamental problems that occur when trying to apply such a correction to clinical data. These issues include:

- An assumption that lesions in a patient would be spherical, as with the fillable spheres in the NEMA or Jaszczak phantoms (though some studies have attempted to simulate lesions [67]).
- The “cold” walls of the sphere can alter “spill-out” of counts in a very different way to the physiological process
- Any corrections would only be applicable to the activity concentration to background ratio used in the phantom
- There is an assumption that resolution and sensitivity would not vary across the field of view.
- To date, the majority of major manufacturers of acquisition or processing software in Nuclear Medicine have not incorporated PVE correction through phantom-generated RC into their algorithms (with the notable exception of MiM Software). This means that applying a partial volume correction is a resource-intensive manual task, not easily integrated with image display and reporting systems.

Some of these challenges may be able to be overcome; for example, 3D printing of lesions could overcome the dependence on spherical objects [127]; the use of gel-based lesions can remove the impact of “cold” glass or plastic walls [126, 321]. However, some applications of PVC have been shown to result in activity overestimations of the same magnitude as the initial PVE [319]. The accumulation of these aforementioned difficulties (and the challenges and time-consuming nature of the potential solutions) limit accurate PVE correction to the preserve of the research application, and make it a particularly challenging goal for an average nuclear medicine department.

7.2.8 Optimised reconstruction parameters

Testing was performed using the NEMA IEC Body phantom described above, filled separately with ^{99m}Tc and ^{177}Lu . Acquisition parameters were as described above, while the reconstruction parameters (number of iterations, number of subsets and post-reconstruction filtering) were varied to test the response. A full description of the reconstruction parameters and the results is detailed in Appendix A, with a summary of the combined results presented here.

The three tested metrics (RC, Signal to Noise Ratio (SNR) and lung error), based on the recommendation of the EANM SPECT quantification guidelines [11], were combined in quadrature after applying the Root Mean Square Deviation (RMSD) from the measurement's expected value.

$$\text{Root Mean Square Deviation, RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Expected} - \text{Measured}_i)^2} \quad (7.3)$$

Where n is the number of measurements made, *Expected* is the value of a theoretically ideal reconstruction (i.e. for RC the *Expected* value is 1) and *Measured* is the actual measurement from the NEMA phantom.

The optimal reconstruction parameters for ^{99m}Tc were found to be 48 iterations and 1 subset (Figure 7.6), showing the lowest RMSD when combining the sum of the deviations for RC, the SNR and the error in the lung insert. This corresponds with the findings of studies using Siemens OSCG algorithm for ^{99m}Tc in the literature, one of which reported the highest SNR with 48 iterations and 1 subset and indiscernible differences in RCs [33], whilst another demonstrated the optimal recovery with 48 total iterations [34]. The measured lung error was in the range 2.4% - 5.2%, comparable to results reported elsewhere of <7.8% [34], accounting for differences in reconstruction subsets and filtering. As there were minimal differences between the results for different post-reconstruction filtering, a 5 mm Gaussian filter was chosen, firstly to match a recommendation of a FWHM approximately twice the pixel size [322] and also to match previous work by the author and other colleagues [323].

For ^{177}Lu , the choice of iterations had less impact on the RMSD result than the choice of filter, where 0 mm of post-filtering was optimal (Figure 7.6). The results for RC absolute error were given priority in selecting the preferred parameters (due to alignment with the aims of the thesis, prioritising quantitative accuracy) and hence 72 iterations with 1 subset (and 0 mm filter) were selected. A study of quantification of ^{177}Lu using Siemens OSCG reported that noise is minimised with the use of one subset and that recovery also varied with subsets [57]; they opted for 48 iterations, which in this optimisation work had similar results for RMSD as for 72 iterations.

**Optimal reconstruction parameters,
assessed by Root Mean Square Difference**

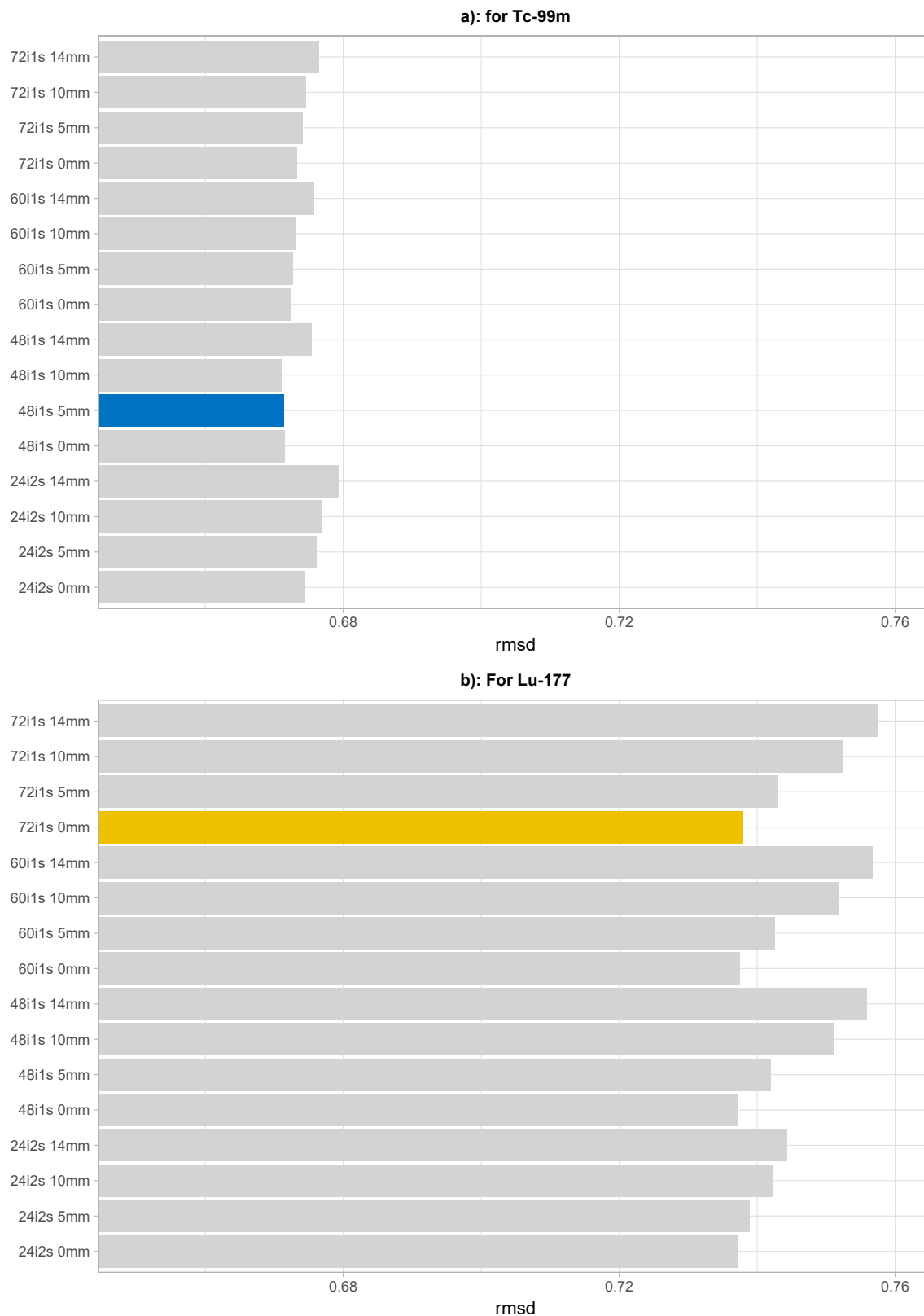


Figure 7.6: Root Mean Square Deviation (RMSD) of quantitative reconstruction parameters for a): ^{99m}Tc (preferred in blue) and b): ^{177}Lu (preferred in yellow). RMSD combines measurements of Recovery Coefficients, Signal to Noise Ratio and Lung Error, with full results in Appendix A.

7.3 Accuracy of the calibration

The simplest method of assessing accuracy of a quantitative calibration is measurement of activity concentration in a uniform phantom. It is recommended for verifying the accuracy of a calibration for quantitative SPECT [11]. EANM guidelines advise that the measured activity concentration should deviate by no more than $\pm 10\%$ from the expected activity, but that $\pm 5\%$ is possible with a sufficiently accurate radionuclide calibrator [324]. In everyday practice, SUV_{mean} is commonly used as it normalises the expected result to 1, making it easier to compare the accuracy across multiple fills of a phantom with different activity concentrations.

Different uniform phantoms were used during the course of this thesis, with volumes between 6244 ml and 7350 ml. The phantoms were filled with on average 319 MBq of ^{99m}Tc or 551 MBq of ^{177}Lu . The phantoms were scanned using an appropriate clinical protocol (generally 15-20 minute acquisition), before reconstruction with the optimised reconstruction protocols. A spherical VOI of 12 cm diameter (900 ml volume) was drawn on the centre of each phantom, and activity in Bq/ml and SUV_{mean} were recorded. A visual display of the distribution of these results for both radionuclides is shown in Figure 7.7, whilst a comparison with the recommended tolerance limits is made in Table 7.2. Both show that regular measurement within the tighter tolerance of $\pm 5\%$ is routinely achievable. The stability of these measurements, and hence of the underlying calibration, will be discussed further in Section 7.4.1, as will the potential cause of the outliers seen in Figure 7.7.

Radionuclide	Percentage of measurements within tolerance	
	$\pm 5\%$	$\pm 10\%$
^{99m}Tc (n=134)	97.8%	99.3%
^{177}Lu (n=22)	100%	100%

Table 7.2: Percentage of SUV measurements in a uniform phantom (multiple acquisitions over an extended time period) within 5% and 10% for each of ^{99m}Tc and ^{177}Lu .

7.3.1 Measured Recovery Coefficients

^{99m}Tc and ^{177}Lu were used in repeated fillings of the NEMA IEC Body phantom. The concentration of radionuclide (in kBq/ml, as measured pre-administration in the radionuclide calibrator) was targeted to match the measured ratio of mean tumour concentration to mean liver background concentration in ^{99m}Tc -EDDA/HYNIC-TOC patients imaged at this centre. This ratio was measured to average approximately 6:1 in a sample of 27 patient scans (note: these patients were on somatostatin analogue treatment, the implications of which will be discussed further in Chapter 9). Other studies in the literature have used a range of ^{99m}Tc activity concentrations in phantoms, including 4:1 (based on uptake in bone scans [53]), 8:1 [325], 10:1 (based on EARL guidance for ^{18}F FDG PET imaging [272]) and 11.5:1 [52].

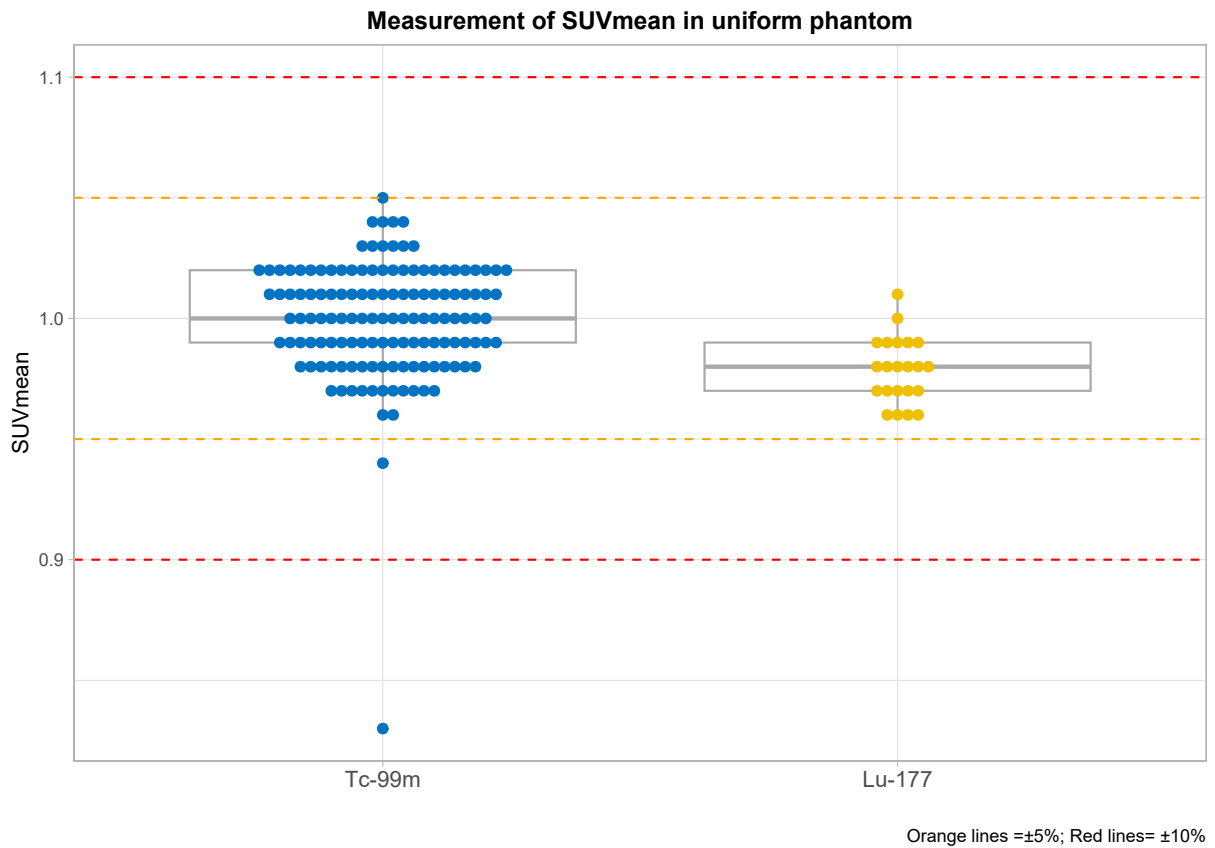


Figure 7.7: Repeated measurements of SUVs (across an extended time period) for both ^{99m}Tc and ^{177}Lu in uniform phantoms. Dotted lines denote 5% and 10% from the expected SUV=1.

For ^{177}Lu , it was also investigated whether the phantom may be subject to precipitation and/or adsorption: a phenomenon colloquially referred to as “sticking”, where the radionuclide adheres to the walls of the phantom. This has been reported for other radio-metals, notably ^{111}In and ^{90}Y [326]. Peters et al reported using an Ethylenediaminetetraacetic acid (EDTA) solution to prevent precipitation of ^{177}Lu in a phantom that was filled once and scanned multiple times over a 74 hour span [273]. Saturation of the plastic phantom with a chelator was not performed for the work carried out here, as local testing showed no evidence of adsorption at the imaging time used for measurements (generally <4 hours post-filling) – though images acquired at later time points did demonstrate this effect (Figure 7.8) and use of EDTA to prepare a phantom and reduce adsorption may be beneficial where imaging is not taking place within hours of filling. It is important to note that often the phantom was filled with ^{177}Lu -DOTATATE (making use of an unused vial ordered for a patient), rather than the ^{177}Lu -Chloride which was imaged in Figure 7.8, though again no evidence of “sticking” was noticed in any reconstructed images.

As the NEMA IEC Body phantom was originally designed for use in PET, the poorer resolution of SPECT (and hence expected larger impact of PVE) means that the maximum recovery coefficient may not be established using only the standard spheres (the largest sphere

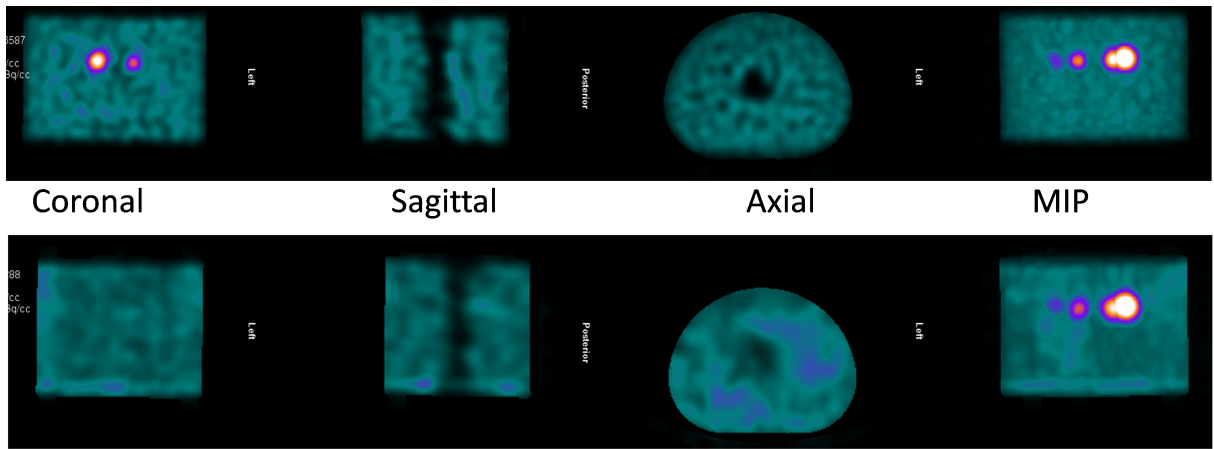


Figure 7.8: Images showing distribution of ^{177}Lu -Chloride four hours (top) and 264 hours (bottom) after filling. Adsorption to the walls of the phantom can be seen after 264 hours.

has a diameter of 37 mm). It was proposed by Bailey and Willowson [30] that a sphere diameter $> 3 \times \text{FWHM}$ of spatial resolution would be required before PVE losses cause $< 5\%$ underestimation of activity. Further work by the same group suggests the minimum sphere diameter required would be 37.8 mm for $^{99\text{m}}\text{Tc}$ and 64.8 mm for ^{177}Lu [65]. To evaluate this, a larger sphere of 60 mm diameter was positioned inside the phantom for further measurements, which required removing the lung insert.

Reconstruction was performed using Siemens xRecon workflow activity, using OSCG and the iterative reconstruction parameters as described in section 7.2.8 previously.

A VOI was preferred over the ROI used in NEMA documentation, as recent EANM guidelines on radiomics in Nuclear Medicine recommend that textural features should be calculated in 3D [90], and further work in this thesis will explore the use of textural features in phantom and clinical work. Definition of the VOI to measure the activity in the sphere was performed using a fixed spherical VOI based on the known physical diameters of the spheres, as taken from the manufacturer's manual [316] and positioned using the co-registered CT. This is one of the outlining methods suggested by Kaalep et al [121]: alternative methods include, for example an 80% VOI based on physical dimensions [327] (see Section 7.3.3 for discussion) or threshold techniques based on a percentage of measured activity [121, 272]. Testing of threshold techniques at the pre-analysis stage found it was often not possible to identify the smaller spheres against the background concentration: this situation was also identified by Kaalep et al as a limitation of the use of PET images for VOI placement - image noise can affect the location, and a limitation on acceptable background noise should be used if considering this approach [121].

For each of the radionuclides, the phantom was filled multiple times on separate days (to test reproducibility) and multiple acquisitions were performed of each filling (to test repeatability).

Figure 7.9 displays the mean RC values with error bars of ± 1 standard deviation for each sphere diameter.

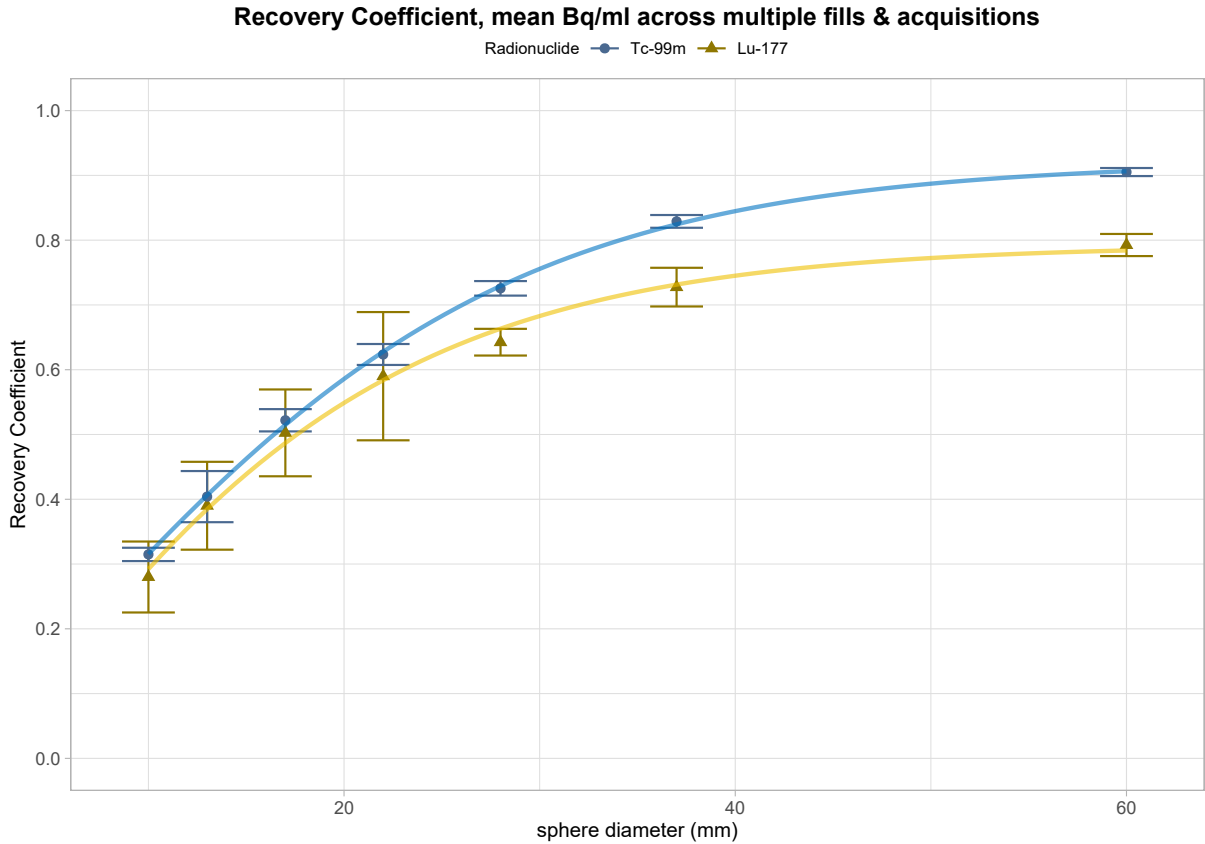


Figure 7.9: Measured Recovery Coefficients for both ^{99m}Tc and ^{177}Lu , including 60 mm sphere. Error bars are ± 1 standard deviation, from repeated acquisitions.

A curve was fitted to the data using a logistic model, as suggested by the EANM guidelines on dosimetry [296]. The guidelines propose a 2-parameter model, but Peters, et al propose a 3-parameter model on the basis of a better fit around the asymptote [273]. Note that Peters et al describe a good fit with the data by evaluating the Pearson R^2 . However, this is not the correct metric for assessing fit for non-linear functions [328], as it should only be applied to linear regression. Therefore for this work multiple models were tested with different numbers of parameters, with fit assessed by the standard error and goodness of fit (Table 7.3).

There is little difference between the performance of the models: since a 4-parameter model provided a high goodness-of-fit and low standard error for both radionuclide, this will be used for the remainder of the thesis.

There have been several other studies that have shown comparable RC curves for both ^{99m}Tc [54, 65, 272] and ^{177}Lu [65, 273]. It is noticeable in Figure 7.9 that the fitted curves for both radionuclides fall somewhat short of the ideal RC of 1, as seen in other studies [54, 65, 329], potentially due to PVE. For this work we have chosen to optimise the reconstruction based

Parameters in logistic model	Standard error		Goodness-of-fit	
	^{99m} Tc	¹⁷⁷ Lu	^{99m} Tc	¹⁷⁷ Lu
2	0.019	0.058	0.991	0.9
3	0.019	0.053	0.991	0.915
4	0.018	0.053	0.993	0.916
5	0.017	0.054	0.993	0.912

Table 7.3: Logistic models fitted to Recovery Coefficients by number of parameters - assessment of best fit using Standard error and Goodness-of-fit.

not only on RC, but also SNR and lung error. Hence, a balance was struck that may result in a reduced RC. The reasoning behind this choice is that we wish to analyse more metrics than just the mean activity concentration in both phantom and patient studies, including other formulations of SUV, and textural analysis features that may be affected by e.g. noise:

- Measurement of lung error is a test of both the spill out effect and image noise, evaluating their effect in a uniform area without activity, that therefore should be "cold" on SPECT imaging. The reference to the lung comes from the use of a section of the NEMA phantom containing water and polystyrene beads, to mimic the LAC of lung tissue in the patient; in [¹⁸F]FDG studies the lungs should generally show no uptake.
- The SNR is important in metrics heavily affected by image noise, such as SUV_{max} . This is a commonly-used metric in quantitative PET and SPECT clinical studies, so optimisation to provide a high SNR is of benefit to allow comparison to studies in the literature. The low count nature of many clinical SPECT images also benefits from optimising the SNR.

The shape of both curves does rise beyond the 37 mm sphere, and the 60 mm sphere had higher RC than the 37 mm sphere for both radionuclides, confirming that a minimum sphere diameter of 60 mm (or preferably larger) is required to achieve maximum recovery [330].

Testing of repeatability for ^{99m}Tc with consecutive acquisitions demonstrates almost complete overlap of the measured RCs (Figure 7.10) and provides confidence as to the repeatability of quantitative scans.

Reproducibility was also tested by comparing the average result between separate phantom fills carried out 6 days apart. This timespan means there is unlikely to be a significant change in camera performance between acquisitions (confirmed by the results of the regular QC testing). The results, as demonstrated in Figure 7.11, show close agreement of the fitted lines. The equations for the fitted RC line are used to define correction factors for Partial

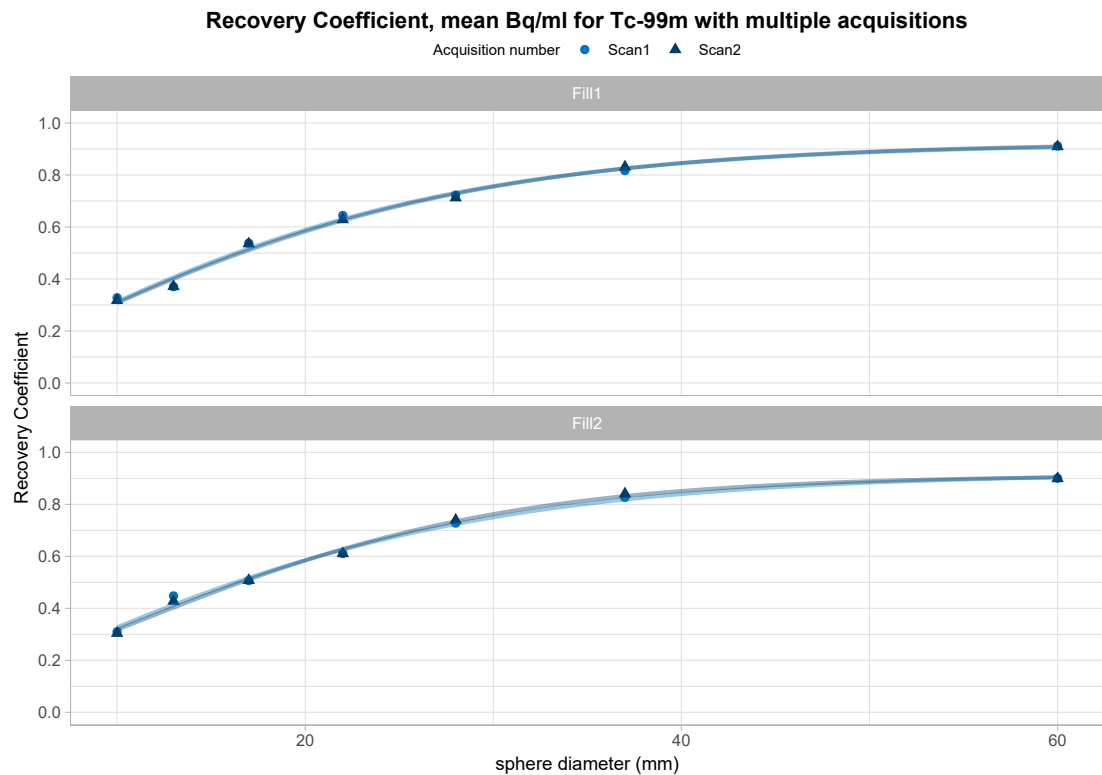


Figure 7.10: Testing repeatability of ^{99m}Tc Recovery Coefficients across consecutive phantom acquisitions and multiple fills.

Volume Effects [28], therefore the results from the different fills can be considered to be indistinguishable.

7.3.2 Time per view for ^{177}Lu acquisitions

To minimise the prospect of motion artefact during a long three field of view scan for post-PRRT patients, a time constraint is necessary on the potential duration of a patient's post-therapy scan at 24 hours (this also helps for timeous discharge from the in-patient ward). To determine if the resultant images would still be quantitatively accurate if acquisition time was reduced, multiple acquisitions were performed of the NEMA phantom at 5, 10 and 16 seconds per view. The results in Figure 7.12 show little difference between the scan speeds, especially at larger sphere sizes. The uncertainties do become larger for the smaller spheres at shorter time per view, though this is unlikely to greatly affect the overall quantitative accuracy, as many lesions will have a larger volume.

To maximise patient comfort, reduce possibility of motion artefact and streamline the clinical workflow, 5 seconds/view was selected for clinical scan protocols for ^{177}Lu -DOTATATE imaging at 24 hours. This is considerably shorter than the acquisition times noted by other centres in the literature: Ryden et al [331] reviewed 13 papers and listed a range of times per view from 10 seconds up to 80 seconds (though this included sites imaging at up to 192 hours

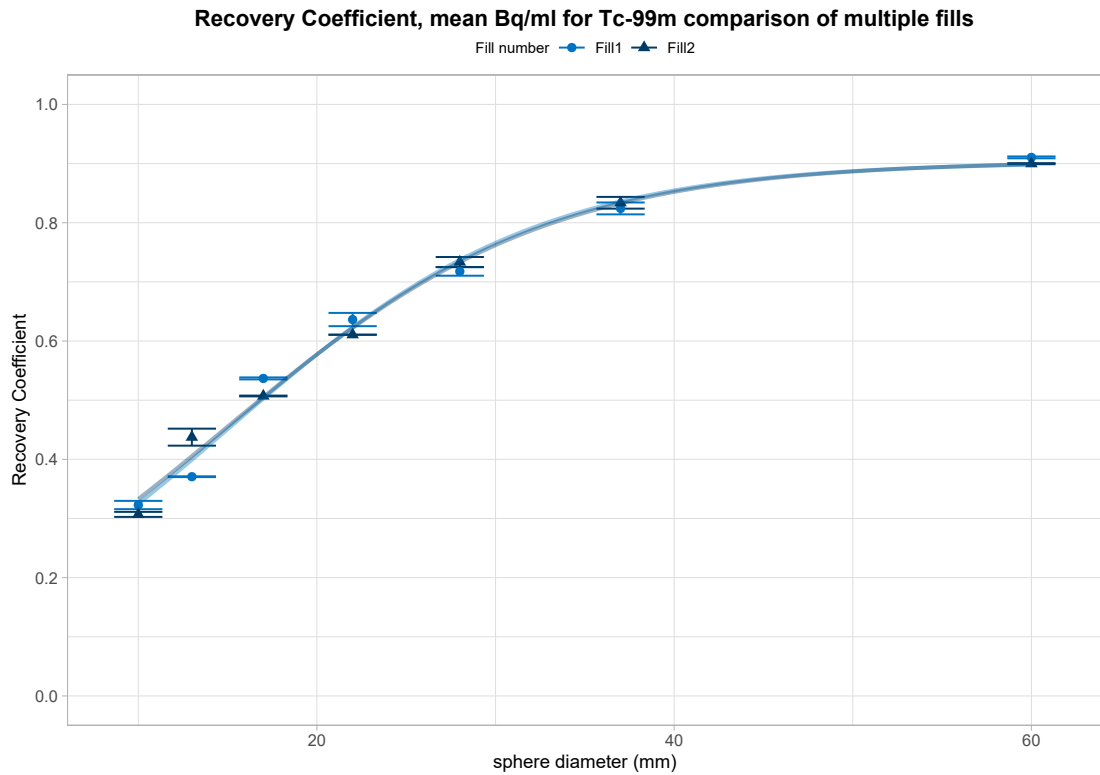


Figure 7.11: Testing reproducibility of ^{99m}Tc Recovery Coefficients across multiple phantom fills.

post-administration). The use of ADS rather than Step and Shoot has been shown to allow for reduced acquisition time whilst maintaining quantitative accuracy [46, 314] and hence will aid the process of reducing acquisition time. The results demonstrated in Figure 7.12 are in broad agreement with the results from a simulated reduction in projections (equivalent to a drop in time per view from 30 to 7.5 seconds) tested by Ryden et al [331].

7.3.3 Comparison to RC measured in Siemens White Paper

A comparison was made with the results offered by Siemens in their white paper validating the xSPECT quantification method [332], authored by Vija. There were differences in the Siemens methodology from that used in the majority of this thesis, therefore the method was adjusted where possible to make as direct a comparison as possible:

- Siemens used a Jaszczak phantom with fillable spheres that differ in size from those available with the NEMA IEC phantom. The phantom is also cylindrical, rather than the torso-mimicking shape of the NEMA phantom. This means the contoured orbit of the camera heads will differ between the phantoms, potentially being closer or further from the phantom at different points in the orbit. As relative sensitivity can vary with distance due to septal penetration from high energy photons [333], this could result in differences in measured activity. However, for ^{99m}Tc and LEHR collimators this effect is likely to be minimal.

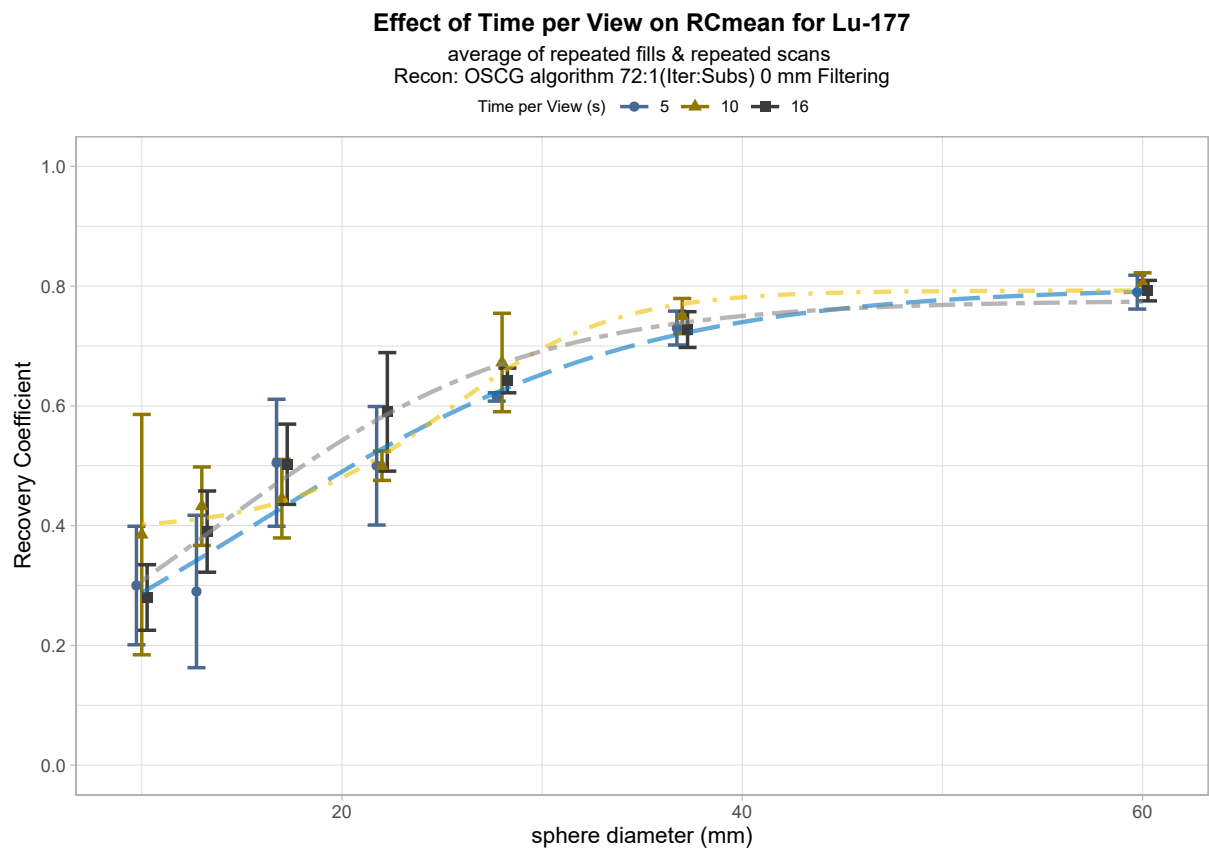


Figure 7.12: Comparison of ^{177}Lu Recovery Coefficient curves (dotted lines) for mean concentration values, with varying time per view.

- Rather than a VOI following the known physical dimensions of the spheres, Siemens chose to use a VOI with diameter 80% of the physical dimensions, which is claimed to follow the methodology used in Seret et al [327]. This paper used a phantom with hot and cold rods for measuring contrast, and the 80% ROI was only for the uniform section (for the hot and cold rods, the diameter of the ROIs was chosen as 100% or 50% of the physical diameter). These differences call into question the rationale for choosing 80% VOIs, which are not elaborated on by Vija. Nevertheless, the 80% VOI was used to analyse the NEMA phantom images for this section, in order to be more comparable with the Siemens White Paper results.
- Siemens opted for Step and Shoot acquisition, rather than ADS.
- 60 views/120 projections were used by Siemens, while this study opted for 64 views/128 projections, as this is the setting for local clinical protocols: the white paper showed only small differences in RC (less than 0.6% for the largest sphere) for a doubling of the angular sampling (i.e. 60 views to 120 views) – therefore the difference between 60 and 64 views is unlikely to be appreciable.

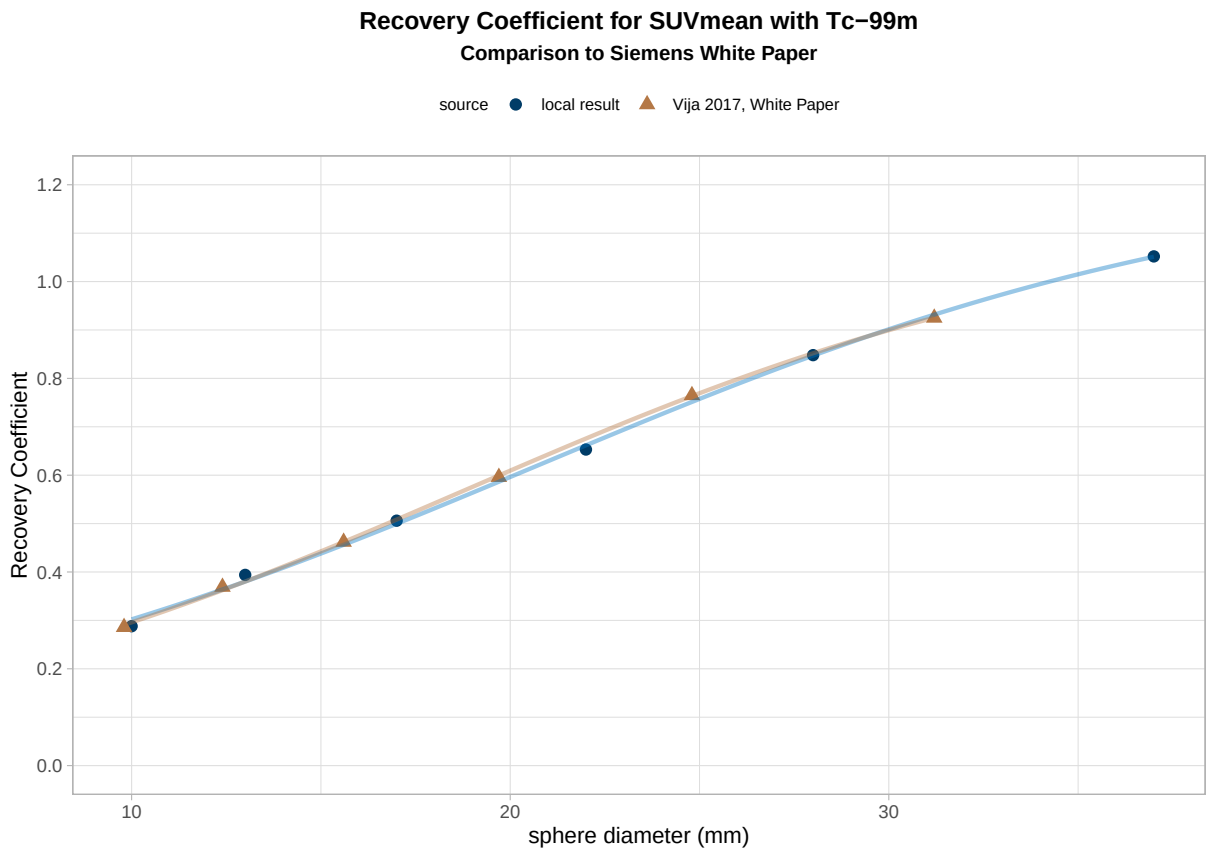


Figure 7.13: Comparison of measured Recovery Coefficient for ^{99m}Tc in a NEMA IEC phantom with results from Siemens White Paper (utilising a different phantom and sphere sizes).

The graph in Figure 7.13 demonstrates almost identical curves fitted to the Recovery Coefficients. As the spheres in the different phantoms have different physical sizes, a point-for-point comparison cannot be made; however it is clear that results are consistent between the White Paper and the local test.

7.4 Longitudinal stability

7.4.1 Stability of SUV measurements

The calibration process for Siemens BroadQuant, involving a point source sensitivity measurement and a volume sensitivity measurement using a uniform phantom, is described in detail in Chapter 6. The Siemens equipment manual and related white papers states that this point source sensitivity calibration procedure should be performed on a monthly basis [304], with only an initial volume sensitivity measurement required for any radionuclide and collimator combination. While the calibration process is relatively simple, it does require the user to obtain activity for each radionuclide they intend to calibrate – this may prove difficult for rarely-used or expensive radionuclides, especially where a high activity concentration is required for the point source measurement. There are also resource implications in terms of

camera time and availability of suitably-trained staff. Therefore, it would be useful to determine the minimum required frequency of calibration to maintain quantitative accuracy.

Regular QC tests are used to ensure that gamma camera performance remains within acceptable tolerances, with guidance on suitable measurements and frequencies given in long-standing guidelines [295, 303]. As quantitative SPECT is a relatively new technology, previous guidelines have not referenced how QC for quantification should be performed. The recent publication of EANM practice guidelines for quantitative SPECT provide some limited guidance; it advises a validation of the quantitative accuracy following a calibration, and assessment of sensitivity at a frequency to be determined by local physics support. Annual re-calibration is recommended, assuming there are no requirements for it to be re-calibrated sooner [11].

To determine a suitable frequency for checks after the first calibration of the quantitative SPECT system, reference was made to recent EANM guidelines for PET QC (released in 2022) that recommend a daily SUV measurement [324], which is commonly implemented using a ^{68}Ge sealed source phantom. Given the relatively infrequent use of quantitative SPECT in clinical practice and the lack of a suitable sealed, long-lived uniform radioactive source for SPECT, a daily frequency was determined to be impracticable. Lack of availability of other radionuclides means that only $^{99\text{m}}\text{Tc}$ was regularly available for filling both uniform phantoms (for SUV assessment) and point sources (for planar sensitivity checks). Regular measurements were made to assess the accuracy of the $^{99\text{m}}\text{Tc}$ calibration and assess for how long the calibration may retain its validity. As the guidelines recommend that re-calibration be performed at least annually, the stability was assessed over a number of periods of 8-11 months for $^{99\text{m}}\text{Tc}$. The calibration for ^{177}Lu was tracked over longer periods (up to 2 years) in order to obtain more measurements. As the system was not yet being used for quantitative clinical acquisitions of ^{177}Lu , exceeding the recommendations of the guidelines was deemed justified for the purposes of characterising the calibration process and its stability.

One method to assess longitudinal stability is with the use of process control charts. These statistical testing methods were developed for use in industry to provide information on manufacturing processes, particularly to track whether the level of quality has dropped from that required. A commonly used method is the Shewhart chart [334]. These have been studied for use in nuclear medicine, for radionuclide calibrators and also for monitoring gamma camera quality control on a longitudinal basis [335]. The rationale for using such charts in conjunction with quantifiable metrics is that they may be able to identify trends in performance that would not be identifiable on visual checks, and that remedial action can be taken before performance is significantly impaired [335].

The following methodology was used to evaluate the stability of the system through quantitative measurements. Acquisitions were performed using a water-filled uniform phantom: either a flanged Jaszczak phantom (Data Spectrum; Durham, North Carolina USA) with inserts removed, or a flangeless Esser phantom (Data Spectrum; Durham, North Carolina USA) with lid inserts left unfilled. Both phantoms were cylindrical. The volume of water contained in the phantom was 6244 and 7300 ml respectively. Activity was measured in the radionuclide calibrator and added to the phantom, with sufficient mixing before acquisition. A separate fill was used for the QC acquisitions than for the calibration process.

For each calibration of ^{99m}Tc , the first five measurements of SUV_{mean} were averaged to provide the “baseline” value for the process charts [336] (referenced elsewhere as target or reference value). Standard deviations were calculated from these readings and used to produce the tolerance limits. For the Shewhart chart, action criteria were taken from work on Dual-Energy X-ray Absorptiometry (DEXA) QC by Pearson and Cawte [337], with a small adjustment to make the criteria tighter. These criteria can be seen in Table 7.4.

Shewhart chart action criteria

Two consecutive points outside ± 2 standard deviations

One point outside ± 3 standard deviations

Four consecutive points outside ± 1 standard deviation, on one side of the target value

A difference of 4 standard deviations between consecutive points

Ten consecutive points on one side of the target value

Table 7.4: criteria for remedial action on regular QC testing for a Shewhart control chart; adapted from [337].

Retrospective analysis was performed for ^{99m}Tc over four periods of up to 11 months (before re-calibration was performed), using Shewhart charts to assess when quantitative accuracy may be sufficiently degraded to require re-calibration of the system (Figure 7.14).

The four ^{99m}Tc calibrations produced two failures on Shewhart analysis when using the action criteria defined in Table 7.4.

- The 2nd calibration failed with one point outside 3 standard deviations at 201 days post-calibration (SUV of 0.83, equivalent to -9.8 standard deviations from target value). However, this point correlated with an intrinsic calibration of the gamma camera performed immediately after a prolonged period without electrical power. This would be expected to affect the performance of the camera. A repeat of the intrinsic calibration brought camera performance to within normal limits and repeat measurement of the SUV the following day showed an acceptable result. The calibration remained within all tolerances for a further 82 days before it was re-calibrated.

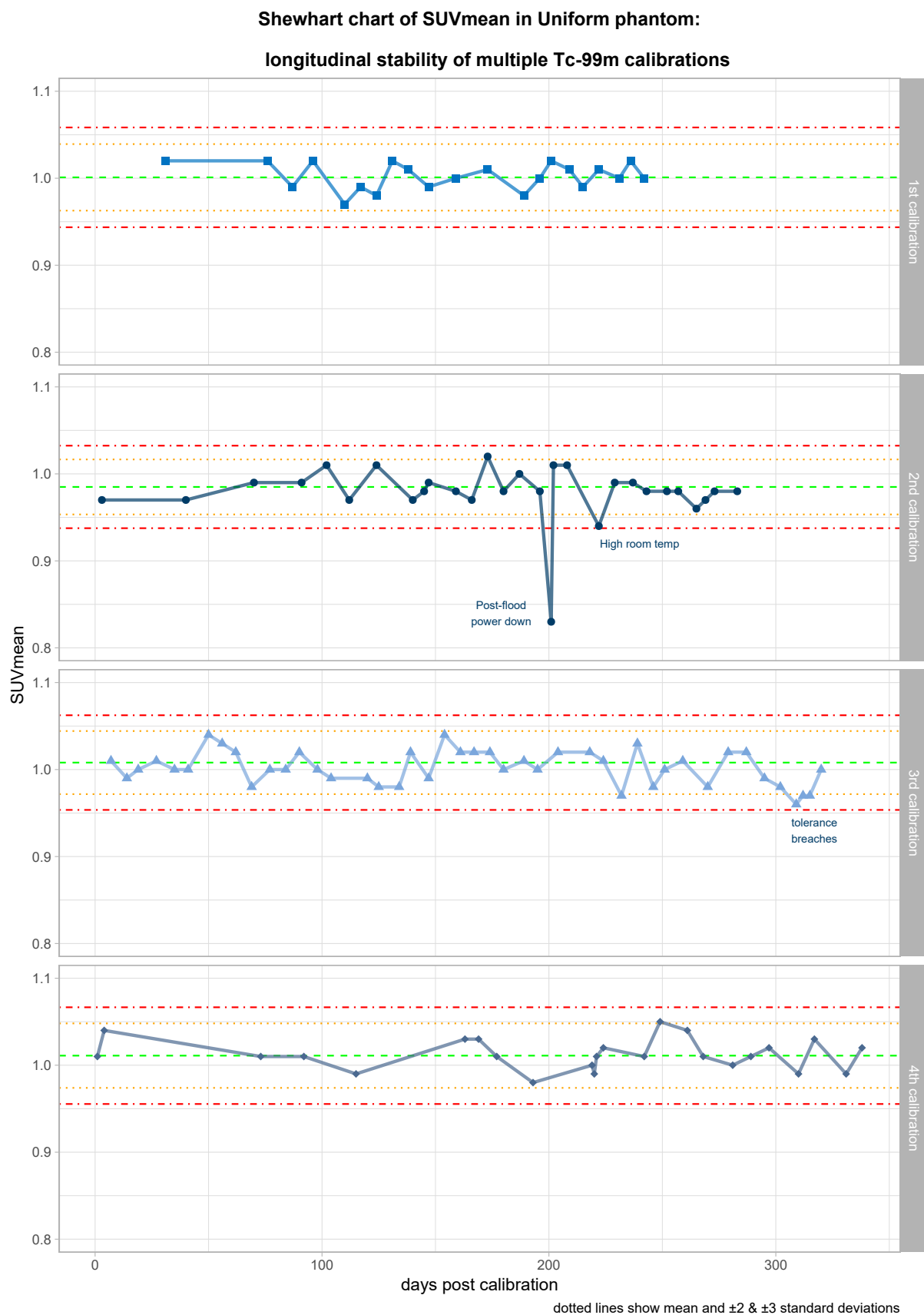


Figure 7.14: Longitudinal measurements of SUV_{mean} for repeated ^{99m}Tc calibrations. Points of interest (e.g. breaches in Shewhart tolerance criteria) are denoted by text at the relevant point in time.

- The 3rd calibration breached the tolerances at day 312, with a 2nd consecutive measurement at greater than 2 standard deviations under the target value (0.96 and 0.97 compared to target value of 1.008). A further measurement at 315 days would also have triggered a failure based on four consecutive measurements at greater than one standard deviation from the target value. There were no known issues with environment, camera performance or performance of the tests to ascribe these results to – as such it was deemed a “true” failure of the accuracy of the calibration.

Due to the limited availability of ^{177}Lu , the number of acquisitions possible within a suitable time frame was not adequate to calculate a practical mean reference value for setting Shewhart limits. Instead, a chart (Figure 7.15) was produced using limits of $\pm 5\%$ of expected activity concentration.

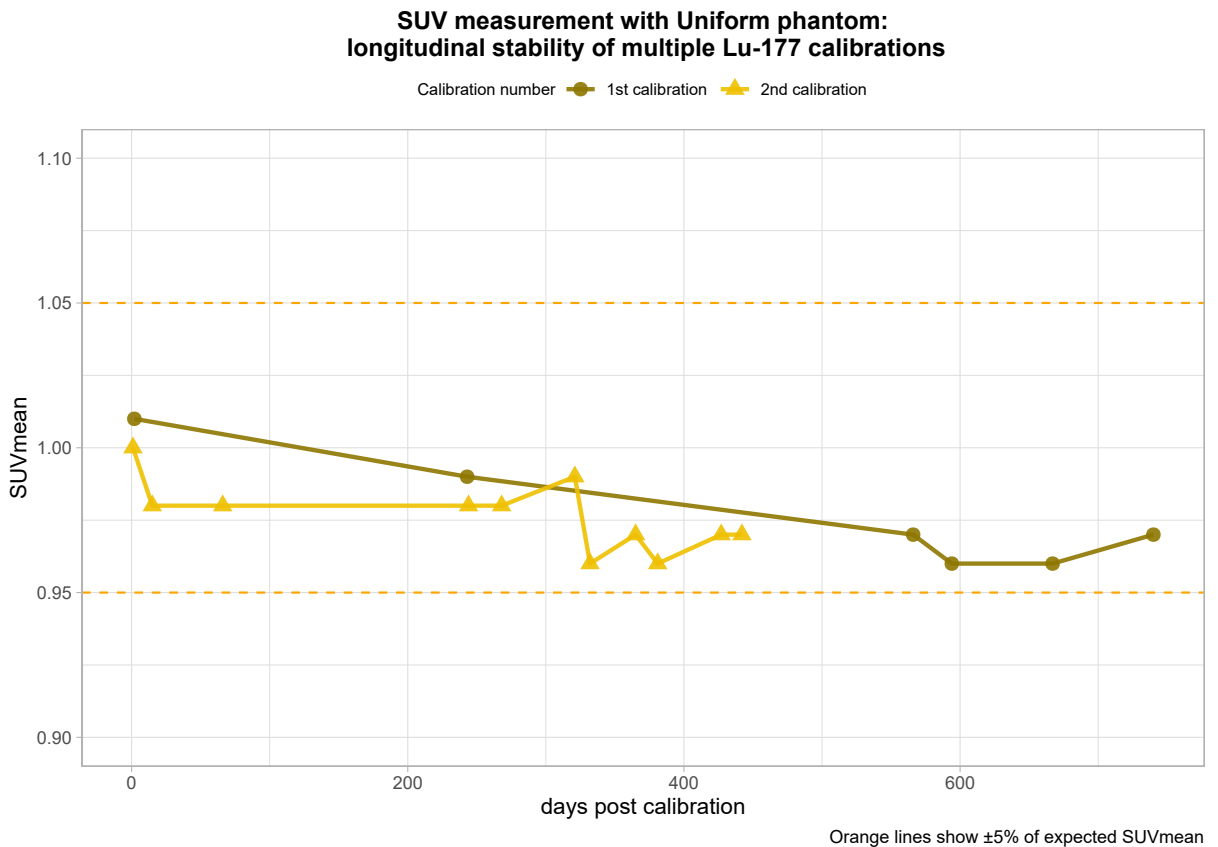


Figure 7.15: Longitudinal measurements of SUV_{mean} for two ^{177}Lu calibrations.

The ^{177}Lu SUV measurements all remained within $\pm 5\%$ of expected activity concentration throughout the lifespan of their respective calibrations. In PET QC guidelines, a difference of $\pm 5\%$ is defined as “achievable”, with $\pm 10\%$ as “acceptable” [324]. These results therefore meet the suggested tolerances (the relevant EANM quantitative SPECT guidelines do not give tolerances, suggesting it should be at the discretion of local physicists [11]).

Radionuclide	Calibration	n	Duration (days)	Mean SUV _{mean}	Standard deviation	CoV
^{99m} Tc	1st	21	242	1.00	0.015	1.54%
	2nd	29	283	0.98	0.017	1.78% ¹
	3rd	44	320	1.00	0.020	2.02%
	4th	14	242	1.01	0.017	1.78%
	Average					1.78%
¹⁷⁷ Lu	1st	6	740	0.98	0.019	2.01%
	2nd	9	381	0.98	0.013	1.24%
	Average					1.63%

¹ this calculation excludes the point at >3sd from expected activity due to an incorrect intrinsic calibration following power outage

Table 7.5: Summary of longitudinal SUV measurements per calibration for both ^{99m}Tc and ¹⁷⁷Lu. The number of measurements per calibration is shown in column titled "n".

To summarise the accuracy of the calibration longitudinally, the CoV was also used. This measure allows comparison between calibration periods (and radionuclides). These results are summarised in Table 7.5. The measured longitudinal CoVs (1.78% for ^{99m}Tc and 1.63% for ¹⁷⁷Lu) are similar for both radionuclides, and represent the repeatability of the SUV measurement process over prolonged periods of time, where changes in camera performance may occur: due to changes in environment (temperature, humidity, etc.) or drift in the electronics. Each calibration has an uncertainty related to the calibration process (and detailed in Section 6.16 and Table 6.16) which may explain why the average SUV_{mean} does not equal the expected value of 1 across each calibration.

A limited number of studies have attempted to monitor quantitative SPECT performance longitudinally. A study by McDougald et al [300] measured the required calibration factor for an accurate quantitative measurement for ^{99m}Tc on a Philips Precedence SPECT-CT camera. While these calibration factor values are not directly comparable to the SUVs or measured activity concentrations produced in this thesis, through the calculation of longitudinal CoVs it is possible to compare results. For 51 scans across an 18-month period (acquired weekly) they calculated a longitudinal CoV of 2.24% [300]. Comparison with the results in Table 7.5 shows a similar magnitude of variation, albeit their measurement period was longer than any evaluated for this project. Also, their result was based on only of one measurement period, rather than repeated periods, as for this study. Acquisition parameters were comparable, with the exception of matrix size (128 x 128 vs 256 x 256) and reconstruction algorithm was a proprietary Philips OSEM algorithm, and therefore not directly comparable to parameters used here for the OSCG algorithm.

Marin et al [338] measured the calibration factor for a Siemens Symbia T gamma camera for ¹⁷⁷Lu, with measurements at four timepoints, nine months apart. They calculated a mean

activity quantification error of -1.05% [338] for activity concentrations above 20 kBq/ml; a similar measurement from the data this thesis gave a mean error of -2.1% for ^{177}Lu . However, this thesis evaluated activity concentrations between 41 and 115 kBq/ml, whilst Marin et al used activity concentrations up to 933 kBq/ml. Visual inspection of Figure 6 in [338] suggests that for a similar activity concentration range, their average measured error may have been greater than that quoted, and in line with the results produced here. Differences in acquisition and reconstruction protocols (most notably in time per view, projection number and reconstruction algorithm) may explain some of the difference – the use of repeated acquisitions from each fill will also reduce the likely variation between measurements compared to the single acquisition per fill (as variance in the filling process and activity measurement will contribute to the quantification error [300]).

7.4.2 Stability of Point Source measurements

Regular measurements were also made of Point Source sensitivity for the individual gamma camera detectors. A $^{99\text{m}}\text{Tc}$ point source was prepared on each day a measurement was to be made, following the procedure outline in the manufacturer's instructions and as discussed in Chapter 6. Action levels were established based on the mean and standard deviation from early measurements, as discussed in 7.4.1, based on an average of the two detector heads (to allow for one set of tolerance limits to be produced).

Figure 7.16 demonstrates for the 2nd calibration that the point source sensitivity dipped at occasions when gamma camera performance was compromised by environmental factors (as did the SUVs: Figure 7.14). There were additional measurements where at least one of the detector heads were outside three standard deviations. Further interrogation of when these breaches occurred does not discern a clear pattern as to either a cause of the variation, or a good correlation with SUV measurements. For example, although one case where SUV measured 1.05 triggered for both detectors based on being outside three standard deviations, another SUV measurement of 0.94 showed both point source measurements comfortably within two standard deviations. Additionally, for the five measurements where both detectors were outside three standard deviations, only two had SUV measurements of note: the 1.05 measurements discussed in the last point, and the previously discussed instance of 0.83. The three remaining occasions corresponded to SUV measurements of 0.98, 1.01 and 1.101, well within expected tolerances for an ideal SUV of 1. Therefore it is challenging to construct a single measure that would provide certainty that the quantitative accuracy was declining; though point source measurements outside three standard deviations may suggest further investigation by SUV would be necessary.

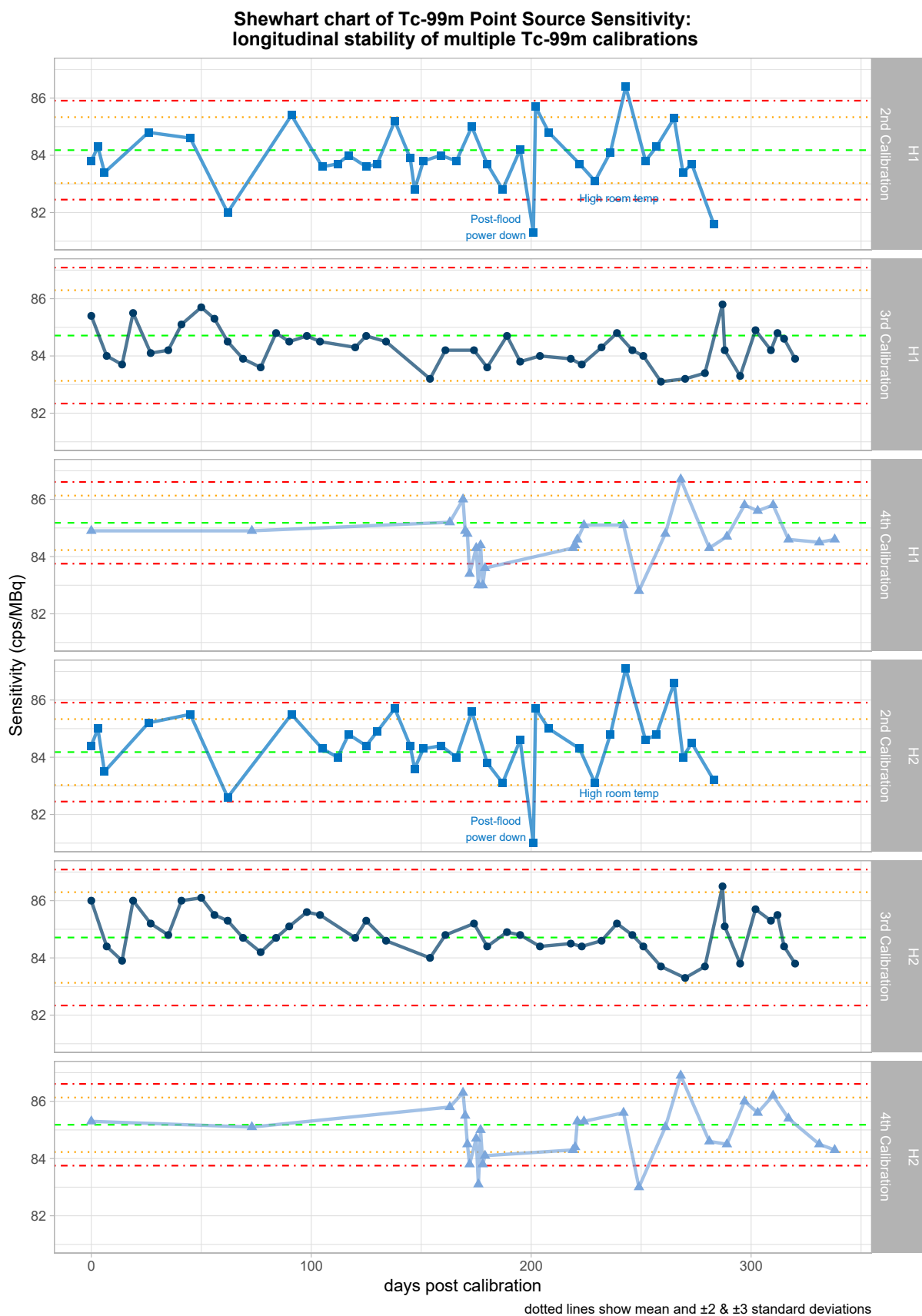


Figure 7.16: Longitudinal measurements of point source sensitivity for repeated ^{99m}Tc calibrations

Radionuclide	Calibration	n	Duration (days)	Mean Sensitivity (cps/MBq)	Standard deviation	CoV
H1	2nd	35	283	84.18	0.58	0.68%
	3rd	43	320	84.71	0.79	0.93%
	4th	19	261	84.39	0.86	1.02%
	Average					0.88%
H2	2nd	35	283	84.47	1.13	1.34%
	3rd	43	320	84.86	0.74	0.87%
	4th	19	261	84.74	0.89	1.06%
	Average					1.09%

Table 7.6: Summary of longitudinal point source sensitivity measurements with ^{99m}Tc , per calibration. The number of measurements per calibration is shown in column titled "n".

Detector	Calibration	Number of breaches	
		outside ± 2 s.d.	outside ± 3 s.d.
H1 only	2nd	3	2
	3rd	1	0
	4th	0	3
H2 only	2nd	4	1
	3rd	1	0
	4th	2	0
both detectors	2nd	5	2
	3rd	0	0
	4th	6	3

Table 7.7: Summary of occasions where ^{99m}Tc Point Source sensitivity breached either ± 2 or ± 3 standard deviations.

7.5 Correlation of SUV measurements with other metrics

Regular SUV measurements are not generally part of routine QC for nuclear medicine departments: while commonplace in PET QC, the lack of a suitable sealed source phantom limits the capacity for regular tests. Other regular QC tests are commonly used to assess the performance of the gamma camera. This section will investigate if these tests provide any useful information that could predict the quantitative accuracy of the gamma camera.

7.5.1 Correlation of uniform phantom measurements with different radionuclides

Due to the expense of some radionuclides, a regular check using a uniform phantom filled with several hundred MBq of the radionuclide may not be possible. As each radionuclide is calibrated individually for Siemens BroadQuant, it is not guaranteed that another radionuclide can be used to check that all calibrations remain valid. However, it could be assumed that

deteriorating performance of the gamma camera (for example, poor uniformity or a change in sensitivity) would have similar effects for all radionuclides. This assumption was tested by comparing the performance of a ^{99m}Tc check of SUV with measurements of SUV using other radionuclides that were performed within 7 days. Due to limited availability, only 17 ^{177}Lu measurements met this criteria. To provide further data, the comparison was also made with ^{123}I , for which 20 measurements were available.

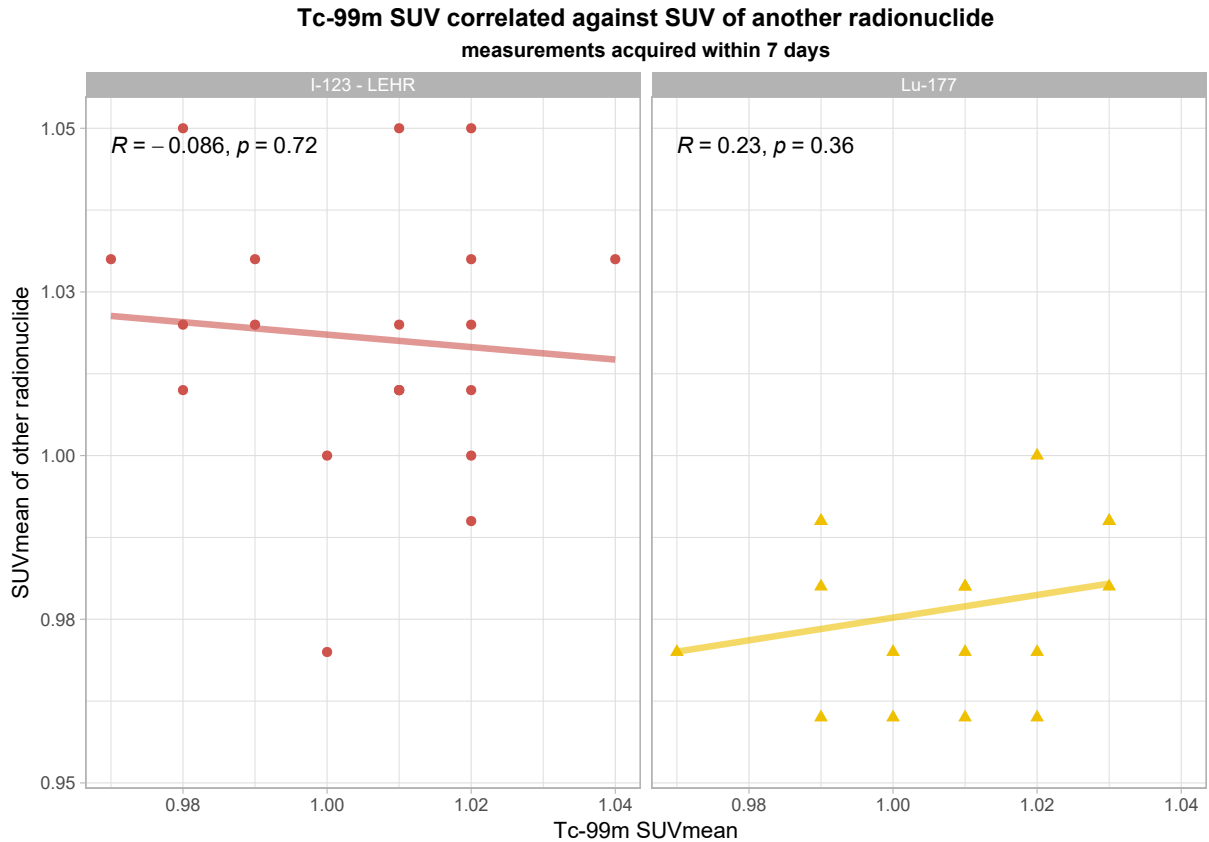


Figure 7.17: Correlation of ^{99m}Tc SUV measurements with SUV measurements of ^{123}I and ^{177}Lu .

There was no significant correlation between ^{99m}Tc SUV_{mean} and the SUV_{mean} results for either ^{177}Lu or ^{123}I (Figure 7.17). One caveat is that all results for each of the radionuclides were within expected performance (i.e. 1.00 ± 0.05), and so extremes of camera performance were not tested – it is hypothesised that an extreme deterioration in camera performance would be demonstrated by SUV measurement regardless of the radionuclide used (see Figure 7.14).

7.5.2 Correlation of point source and uniform phantom measurements

Making regular measurements of SUV using a uniform phantom has several potential drawbacks compared to a measurement made with a point source:

- The uniform phantom requires a higher activity to achieve the required count rate – this exposes staff to a higher radiation dose. It may also be difficult or expensive to source sufficient activity of some radionuclides (e.g. ^{177}Lu). By comparison, a point source may be easily accessible by using the residual from a patient dose.
- The acquisition time of a uniform source is longer – including CT acquisition and reconstruction time, the entire process may take 25-30 mins. By comparison, a point source sensitivity measurement may only take 5 minutes to acquire sufficient counts.

To test if a point source sensitivity measurement could act as a suitable surrogate for an SUV measurement, correlation between measurements acquired within 24 hours were made (to negate changes in performance of the gamma camera that may be expected with time). Correlation was tested using Pearson's correlation coefficient, with a significance level of $p < 0.05$. Pairs of measurements were tested for both $^{99\text{m}}\text{Tc}$ and ^{177}Lu ; due to less frequent availability of ^{177}Lu , only nine pairs of measurements were tested, compared to ninety-six for $^{99\text{m}}\text{Tc}$.

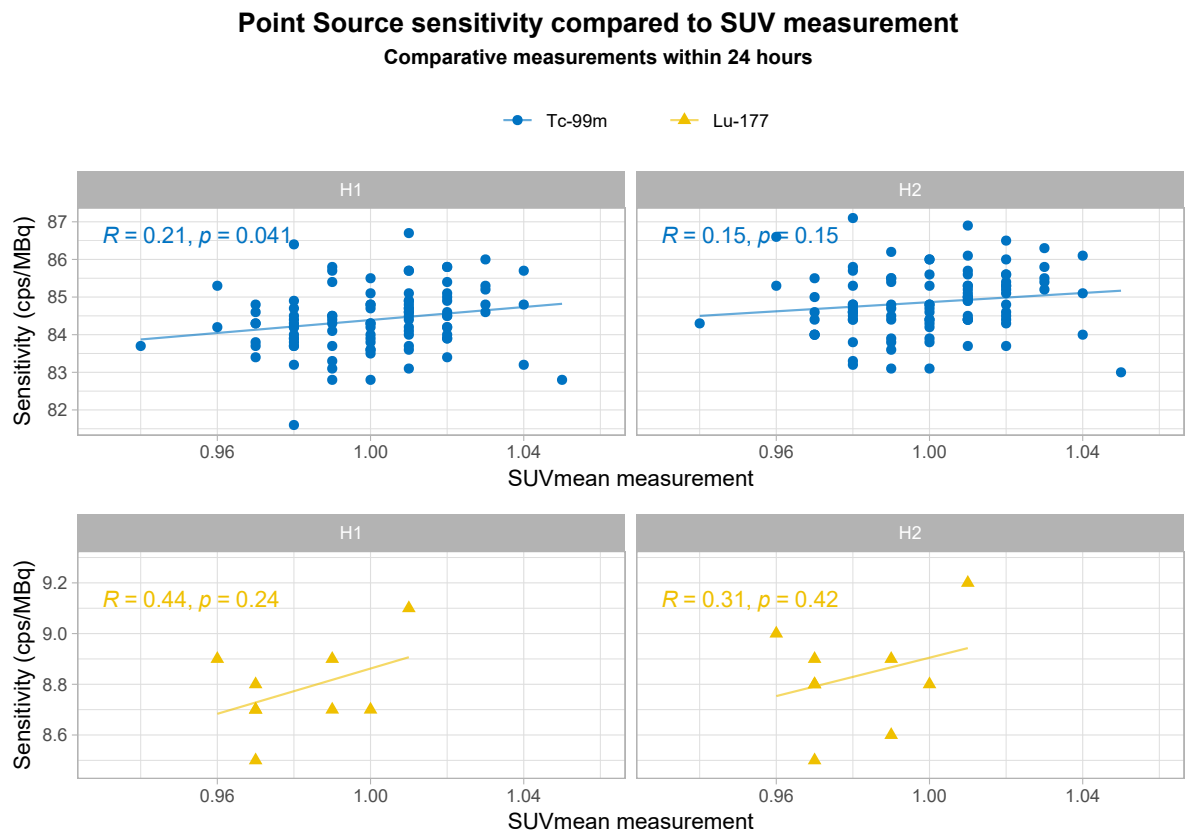


Figure 7.18: Correlation of $^{99\text{m}}\text{Tc}$ SUV measurements with point source sensitivity measurements for both $^{99\text{m}}\text{Tc}$ and ^{177}Lu .

Only the H1 measurements for $^{99\text{m}}\text{Tc}$ had a statistically significant correlation at $p=0.04$ (Figure 7.18). However, the Pearson correlation coefficient of $r=0.21$ shows this was only a

very weak correlation. For ^{177}Lu , there are similar trends, but again no significant correlation. Although we can conclude that a point source sensitivity measurement does not provide a suitable surrogate for directly measuring the SUV, as all measurements were during a period of relative stability of the SUV, it does not preclude that point source sensitivity may help with periods of extreme out-of-tolerance performance, as seen above in Figure 7.16.

7.5.3 Correlation of uniform phantom measurements and routine QC

A test that is already performed as part of routine QC within the department would be an ideal marker of quantitative performance: comparison of the SUV measured by the uniform phantom against other QC tests was performed in order to determine if any correlated well with SUV, or if any could be used to highlight where the SUV may be out of tolerance.

As the department's QC programme only requires a visual check of the uniform flood source each day, rather than a quantitative measurement, the only quantified daily test that can be compared to each SUV measurement is the comparison of the ^{57}Co peak to the expected energy peak of 122 keV. This is performed for both detectors, with a tolerance limit of $\pm 2\%$. This should highlight any gross abnormalities in the performance of the camera. For example, if the camera has been switched off overnight, the lack of continuous power will result in a variation in the energy peak.

Quantitative measurement of camera uniformity is performed weekly with a ^{57}Co uniform flood source, rotating between the four sets of collimators, resulting in a monthly check for each collimator. Three uniformity metrics are measured: Differential Uniformity, Integral Uniformity and Uniformity Index. Whilst measurements are made for both Useful Field of View (UFOV) and Central Field of View (CFOV), only the UFOV measurements will be considered here (as clinical SPECT-CT images, particularly Whole-Body SPECT-CT, usually will make use of the area covered in the UFOV). Only measurements acquired within 24 hours of comparable SUV measurement were utilised.

The equations used to determine uniformity are those applied in the Hermes QC utility [318], as the software used locally for QC analysis.

$$\text{Differential Uniformity (\%)} = \frac{D_L - D_S}{D_L + D_S} \times 100 \quad (7.4)$$

Where D_L is the largest difference between counts in five contiguous pixels, and D_S is the smallest difference.

$$\text{Integral Uniformity (\%)} = \frac{C_{\max} - C_{\min}}{C_{\max} + C_{\min}} \times 100 \quad (7.5)$$

Where C_{max} is the maximum counts per pixel in the image, and C_{min} is the minimum counts per pixel.

$$\text{Uniformity Index(\%)} = \frac{\sqrt{SD^2 - Mean}}{Mean} \quad (7.6)$$

Where SD is the standard deviation of counts in the UFOV and $Mean$ is the average counts per pixel in the UFOV.

The correlation graphs in Figure 7.19 show no significant correlation between the ^{99m}Tc SUV measurement and a ^{57}Co uniformity measurement (for any of the three uniformity metrics).

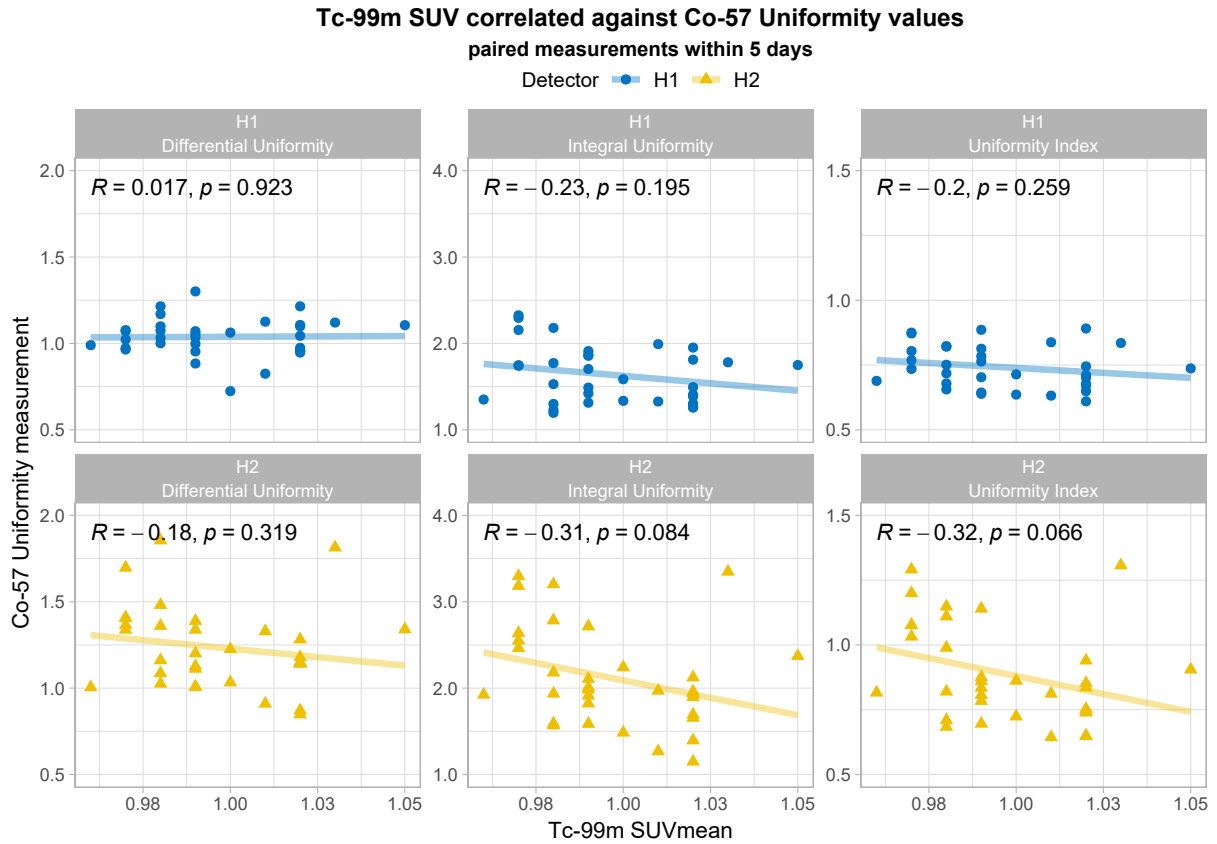


Figure 7.19: Correlation of ^{99m}Tc SUV measurements with extrinsic uniformity of ^{57}Co flood source using Differential Uniformity (left); Integral Uniformity (middle) and Uniformity Index (right).

An alternative method to measure uniformity requires the collimators to be removed and used a point source of activity rather than a uniform source. This intrinsic method avoids the need for purchasing a ^{57}Co flood source (or repeated filling of an unwieldy ^{99m}Tc flood phantom), but does not test the collimator performance (i.e. signal any damage to the collimators themselves). For the Siemens gamma camera, the manufacturer's recommended method uses a Siemens point source to assess uniformity intrinsically: it does not meet the usual definition of an intrinsic uniformity check, as the source is significantly closer than > 5.5 times the largest dimension of the crystal [339]. Instead Siemens have designed the protocol so

that the source is scanned between both detectors (using the source holder that is also used for extrinsic sensitivity measurements) and a geometric correction is then applied to correct for the source being a reduced distance from the detectors. Measurements are made on the camera on a weekly basis.

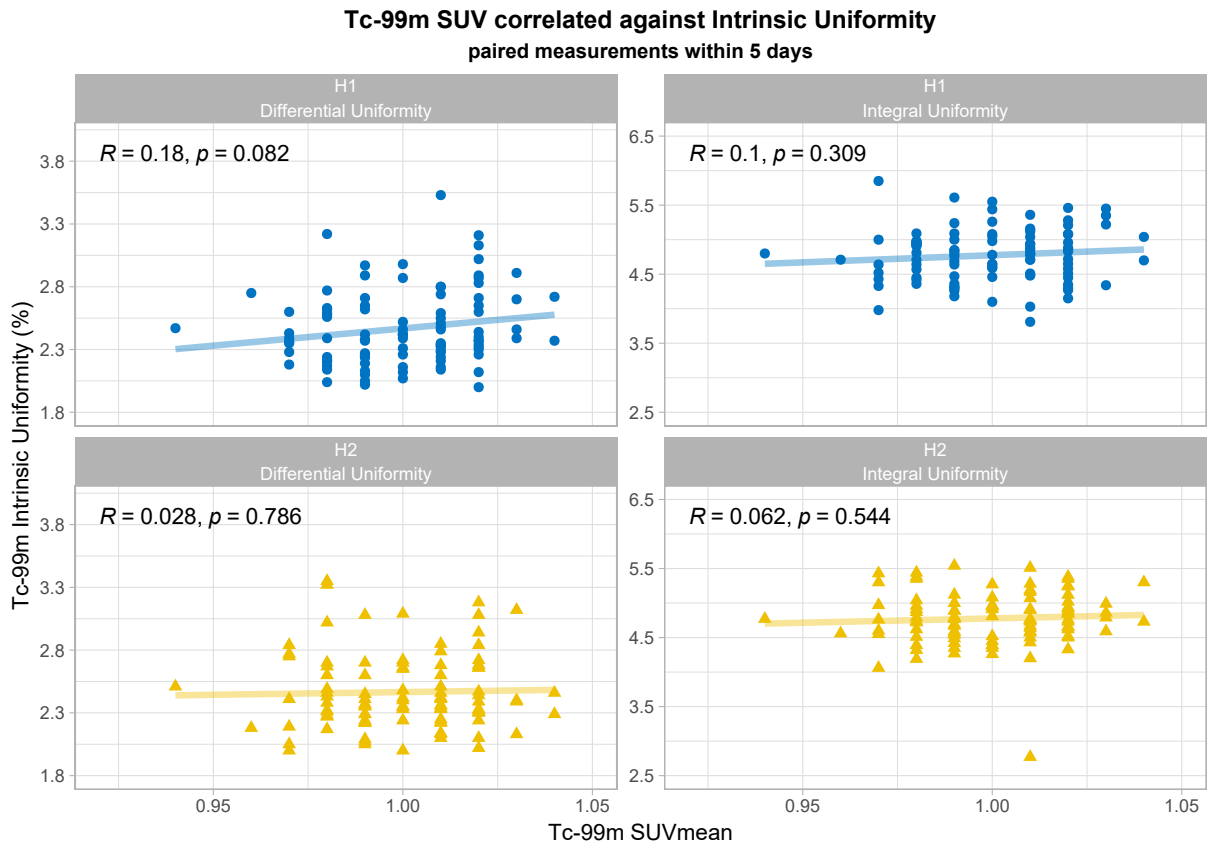


Figure 7.20: Correlation of ^{99m}Tc SUV measurements with intrinsic measurements with ^{99m}Tc point source of Differential Uniformity (left) and Integral Uniformity (right).

The results (Figure 7.20) show no significant correlation with any of the uniformity metrics. This finding suggests there will be no predictive value in a camera's uniformity measurement on the quantitative SPECT performance.

Another daily check made as routine practice is the position of the energy peak for each detector head. This test is performed using the ^{57}Co uniform flood source, and is to ensure that the photopeak of the radionuclide is correctly centred over the expected energy peak.

The results are shown in Figures 7.21 and 7.22 where the standard deviations for each measurement of the respective metric are plotted longitudinally, as shown as a Z score. Whilst this may demonstrate the limited use of the ^{57}Co energy peak for estimating an expected SUV measurement, it does suggest that in instances of extreme poor performance of the gamma camera (e.g. following a power outage or in times of rapid temperature or



Figure 7.21: Comparison of longitudinal ^{99m}Tc SUV measurements with ^{57}Co energy peak deviation, for detector head 1

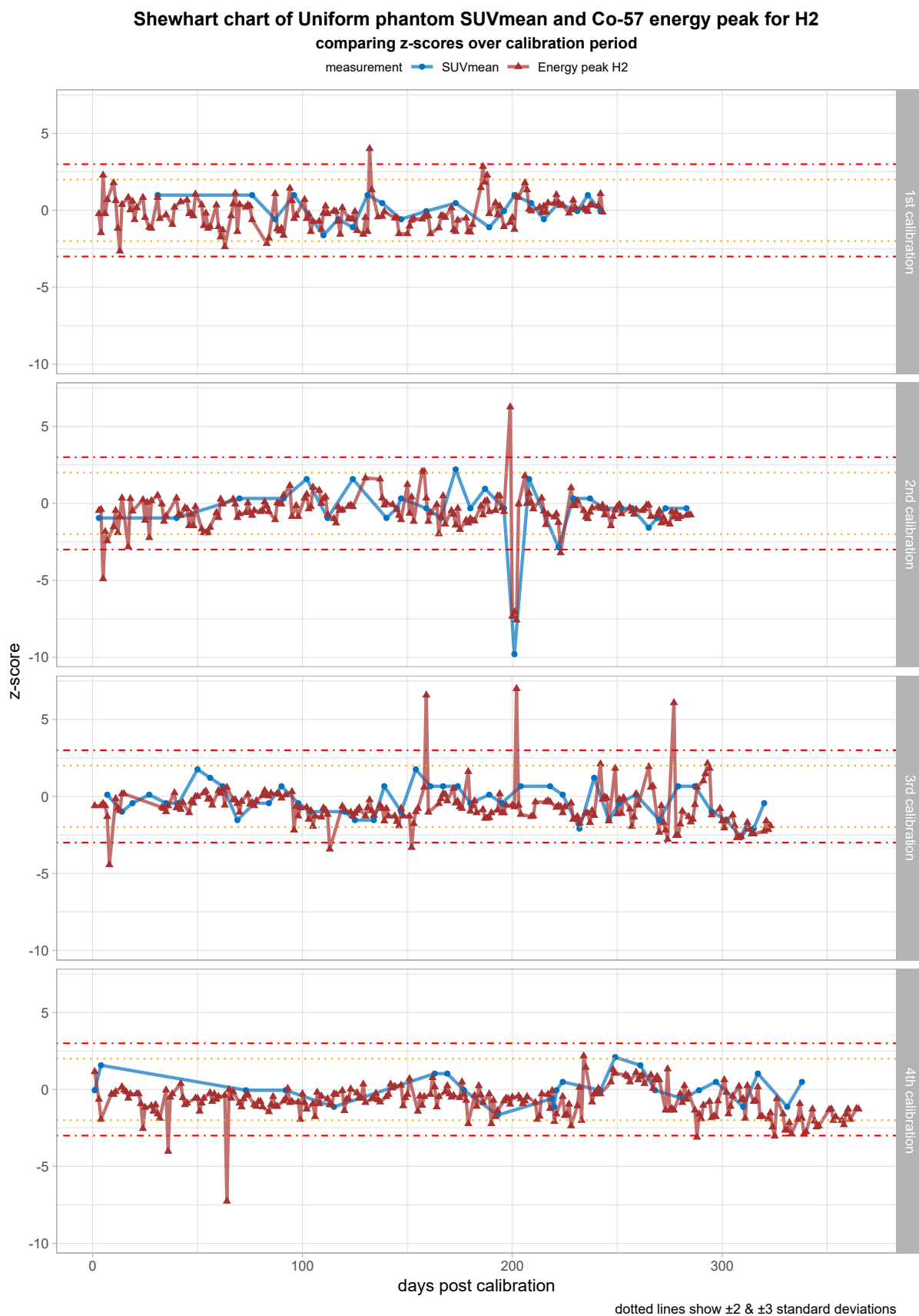


Figure 7.22: Comparison of longitudinal ^{99m}Tc SUV measurements with ^{57}Co energy peak deviation, for detector head 2

humidity changes) then an outlying energy peak measurement may at least act as a warning that SUV may be affected.

7.6 Conclusion and relevance for the thesis

This chapter has used phantom studies to identify optimal acquisition and reconstruction parameters to deliver accurate quantitative imaging, without excessively degrading the SNR or high levels of count spill-out. Though the specific parameters will have some dependency on the individual camera used, equivalent performance to that quoted by the manufacturer and other studies in the literature was achieved. Studies also investigated the longitudinal stability of the calibrations, and investigated QC tests that could be used as markers of potential poor quantitative performance.

The work presented here allows for some recommendations to be made regarding assessing performance of the Siemens BroadQuant calibrations:

- If the user is concerned with the quantitative accuracy of reconstructions produced using the Siemens xRecon workflow, the default recommended reconstructions algorithms (denoted as “Best”, “Standard” and “Fast”) should not be used. The adaptive filtering is dependent on number of counts in the acquired image and does not allow for consistent image processing, critical for quantitative accuracy.
- Calibration period can exceed the Siemens recommendation of monthly point source sensitivity calibrations without adversely affecting quantitative accuracy, with stability shown to extend up to 9 months. This avoids the requirement for more frequent re-calibration with associated radiopharmaceutical expenses, as well as the burden on camera and staff time. However, a single initial volume sensitivity calibration does not meet EANM recommendations [11], and is not supported by the findings of this thesis.
- Regular monitoring of quantitative accuracy through imaging of a uniform filled phantom is recommended for indicating when re-calibration is required. This QC check is less time-consuming than the re-calibration process and will identify deteriorating performance that may be hidden by regular re-calibration. Reducing re-calibration frequency also reduces the chance of a incorrect calibration factor being applied, if an error occurs in the calibration process.
- Although it was not found that an individual metric from routine gamma camera QC tests could accurately predict a declining trend in quantitative accuracy, several metrics (including point source sensitivity and energy peak deviation) can alert the user to sudden, gross changes in camera performance that may warrant further investigations,

as they could critically affect quantitative accuracy and require immediate corrective action and re-calibration.

- It is important that the user ensures the gamma camera is operating in an optimal condition before performing the quantitative calibration. Otherwise, there is a risk of an incorrect calibration perpetuating inaccurate quantitative measurements until re-calibration occurs. This detrimental effect can be avoided by performing rigorous QC testing before the calibration, and conducting a post-calibration quantitative measurement with a filled phantom independent of the one used for the calibration.

The work documented in this chapter provides confidence in the accuracy of quantitative measurements for the remainder of this thesis, and for the application of quantification to clinical SPECT-CT studies. It is important for analyses in multiple patients (and potentially repeat measurements in the same patient), that the quantitative metrics are reproducible (demonstrated in Figure 7.7 and Tables 7.2 and 7.5) and that repeated calibrations over long timeframes can maintain quantitative accuracy (as shown in Section 7.4.1). However, a significant PVE is present for both ^{99m}Tc and ^{177}Lu at small volumes (Figure 7.9) and will need to be considered in the clinical context of any measurements made in patient images.

Chapter 8

Robustness of textural analysis features in quantitative $^{99\text{m}}\text{Tc}$ and ^{177}Lu SPECT-CT phantom acquisitions

As discussed in Chapter 3, standard nuclear medicine phantoms generally consist of water-filled uniform cylinders, which are not useful for testing measures of heterogeneity. Heterogeneity is a key concept in the hypothesis that textural analysis can provide critical clinical information, as many studies have linked heterogeneity in tumour lesions to poorer prognosis [79, 138, 254]. Specialised anthropomorphic phantoms may provide a solution, but the time, expertise and expense involved in creating and using such advanced phantoms will necessarily limit their application in routine clinical settings. The recommendations from professional bodies must be able to be implemented by the largest possible cross-section of departments, hence the recent EANM guidelines on radiomics advocated the use of two simple phantoms. This chapter details the application of these phantoms to SPECT-CT imaging, with two radionuclides $^{99\text{m}}\text{Tc}$ and ^{177}Lu . This work was published in the *European Journal of Nuclear Medicine and Molecular Imaging Physics*, with the below citation and is reproduced with the permission of the publisher:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, “Robustness of textural analysis features in quantitative $^{99\text{m}}\text{Tc}$ and ^{177}Lu SPECT-CT phantom acquisitions,” *EJNMMI Physics*, vol. 12, no. 1, p.40, Apr. 2025, doi: 10.1186/s40658-025-00749-0.

The published article is replicated, formatted in line with the remainder of the thesis, in Sections 8.1 to 8.5 and the Supplementary material to the article is reproduced in Appendix B.

8.1 Background

In recent years the application of Textural Analysis has seen a rise across medical imaging, where it is often termed radiomics [340]. While prominent in CT and MRI, this trend has also been seen in Positron Emission Tomography (PET) imaging, where a number of papers have focussed on its use in measuring or predicting response in lung cancer [104, 341], Neuroendocrine Tumours [254, 257, 342], and a wide variety of other tumour types [79, 132,

138, 343]. The adoption into Single Photon Emission Computed Tomography (SPECT) has been more limited. Recently joint European Association of Nuclear Medicine (EANM) and Society of Nuclear Medicine and Molecular Imaging (SNMMI) guidelines on radiomics in Nuclear Medicine [90] have been published, covering the use of Textural Analysis in both PET-CT and SPECT-CT.

Unfortunately many clinical studies have proven not to be reproducible outside of the study centre; often the chosen Textural Analysis features or multi-factorial model will not provide the same prognostic or diagnostic benefit when applied to different patient populations or using images acquired with different parameters [109, 111]. A potential cause of this lack of reproducibility is the use of non-robust features.

Testing of quantitative accuracy in SPECT would ideally be performed with custom-built and/or anthropomorphic phantoms. However, these phantoms may require specialist 3D printing equipment or be prohibitively expensive for many centres [344]. The EANM and SNMMI guidelines on radiomics advocate the use of some simpler phantoms to test potential Textural Analysis features, recommending the use of the following:

- A uniform phantom that could consist of an empty Jaszczak phantom or similar cylindrical phantom: this can be used to test the dependency on volume of Textural Analysis features.
- A 'Revolver' phantom that can be constructed by binding together seven syringes and filling them with differing concentrations of a radionuclide to approximate an inhomogeneous lesion (based on the first demonstration of this phantom by Forgacs et al [345]). When placed in a phantom with background activity, repeated scans of this Revolver insert can be used to test the repeatability of Textural Analysis features when applied to lesions.

While the guidelines are mostly focussed on application to PET-CT, there is a suggestion that the same phantom experiments could also be applied to quantitative SPECT: there has been increased interest in quantitative SPECT for clinical applications in recent years [13], with more accurate corrections for scatter, attenuation and resolution recovery helping to improve the accuracy of quantitative SPECT [346]. Potential clinical uses have been investigated, notably for ^{177}Lu imaging and dosimetry in molecular radiotherapy [260, 261, 347], but also for a number of other radiopharmaceuticals [73, 74, 325]. However, many such quantitative SPECT studies have focussed on the more common quantitative parameters such as Standardised Uptake Values (SUV), with few SPECT studies directly investigating Textural Analysis. This work implements the use of phantom testing of Textural Analysis features for volume dependency and reproducibility in inhomogeneous lesions with quantitative SPECT

imaging of two different radiopharmaceuticals - ^{99m}Tc and ^{177}Lu , with the aim of identifying robust features for use in clinical studies.

8.2 Materials and Methods

8.2.1 Volume dependency phantom

A uniform cylindrical phantom of height 20 cm and diameter 22 cm was filled with water (total volume = 6244 ml). The activity added was 278 MBq for ^{99m}Tc and 670 MBq for ^{177}Lu , accounting for differences in photon emissions.

8.2.2 Revolver phantom

The supplementary material in the EANM/SNMMI guidelines [90] describes a methodology for the use of a Revolver phantom for PET imaging with ^{18}F . Given the differences in resolution between PET and SPECT imaging, it was hypothesised that the size of syringes suggested (3 ml) may be non-optimal, and partial volume effects may dominate over inhomogeneity. Therefore, for assessment of SPECT imaging multiple Revolver inserts were created, for 2.5 ml, 5 ml, and 10 ml syringes, each consisting of 7 syringes bound together in a circular arrangement. These were filled with intended syringe: background concentrations of 4:1, 8:1 and 16:1 (Figure 8.1). The filled Revolver inserts were attached to the lung insert of a National Electrical Manufacturers Association (NEMA) Image Quality phantom (Data Spectrum, NC USA) [316] and placed in the body of the phantom (Figure 8.2), which was then filled with background activity of 472 MBq for ^{99m}Tc and 386 MBq for ^{177}Lu . For both the ^{99m}Tc and ^{177}Lu filled phantoms, four consecutive scans were acquired using the acquisition parameters detailed below.

8.2.3 Image acquisition

For both phantoms, images were acquired on a Siemens Intevo Bold gamma camera (Siemens Healthineers, Erlangen, Germany) using Low Energy High Resolution (LEHR) collimators for ^{99m}Tc and Medium Energy Low Penetration (MELP) collimators for ^{177}Lu . The camera had previously been calibrated for quantitative acquisitions for both ^{99m}Tc and ^{177}Lu using the Siemens BroadQuant software. Acquisitions used Acquire During Step with 64 views (128 projections) and 16s/view. The energy window for ^{99m}Tc was 140 keV $\pm$ 15% with a lower scatter window width of 15% immediately adjacent to the primary window. The energy windows for ^{177}Lu were 113 keV $\pm$ 15% (with a lower scatter window width of 15% immediately adjacent of the 113 keV window), and a second primary window of 208 keV $\pm$ 15%. Quantitative reconstruction used the Ordered Subset Conjugate Gradient method with

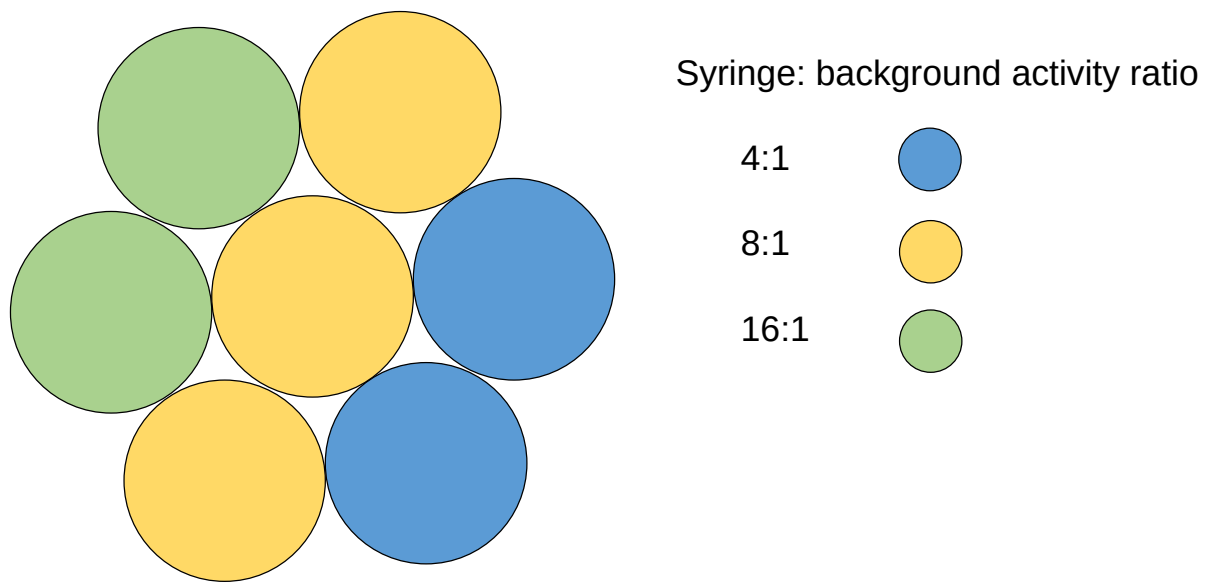


Figure 8.1: Arrangement of activity in Revolver inserts

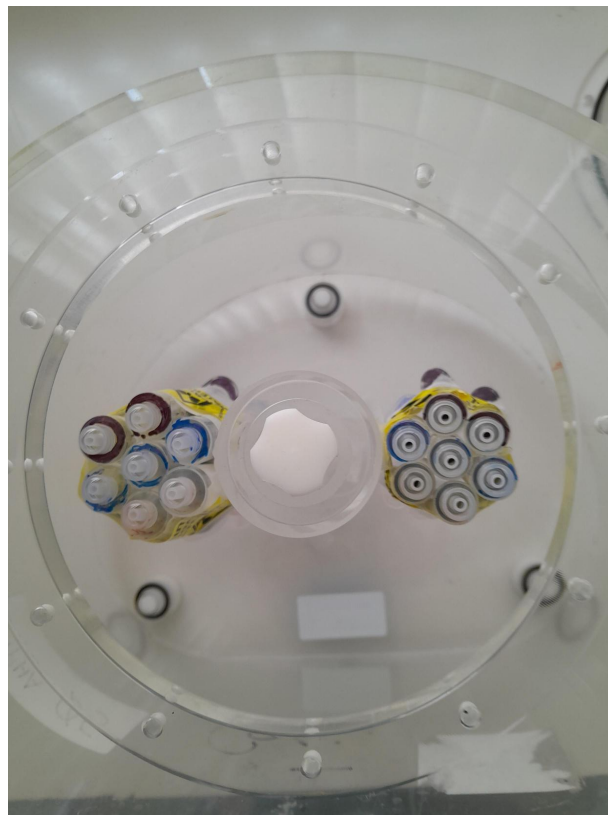


Figure 8.2: Top-down view of unfilled phantom with 2 Revolver inserts in place

parameters that had been optimised locally for quantitative accuracy (for ^{99m}Tc – 48 iterations, 1 subset and 5 mm filtering; and for ^{177}Lu - 72 iterations, 1 subset and 0mm filtering).

8.2.4 Feature extraction

Image analysis was performed using Lifex v6.3 [98] (www.lifexsoft.org): pixel size was 1.95 x 1.95 x 1.95 mm with discretisation to 64 bins (as recommended in [138] and [88]) and SUV range of 0-20 to give bin width of 0.3 SUV/bin (as per [88]). The choice of a fixed bin width and 2mm isotropic voxel size meets the recommendations of Hatt et al [92], and also showed the most reproducible features in one study of PET Textural Analysis [122].

For the uniform phantom, 29 spherical Volumes of Interest (VOIs) of increasing size from 0.9 ml up to 1021 ml (diameters from 10 mm to 122 mm, in 4 mm increments) were placed expanding from the centre of the phantom and avoiding the phantom edges. Sixty-seven Textural Analysis features compliant with the IBSI [89] (and applicable to SPECT imaging), were extracted. The measured values were correlated against VOI volume graphically for visual interpretation.

For the Revolver phantom, VOIs using a threshold SUV of 2.5 times the background was drawn separately over each Revolver insert, from which the 67 features were extracted. The Coefficient of Variation (CoV) was measured across the four scans for each feature (and each Revolver size and radionuclide) then compared to the suggested upper threshold of 10% from the supplementary material of the EANM/SNMMI radiomics guidelines [90].

8.2.5 Feature reduction

It is important that any feature selected for clinical analysis should behave similarly across a range of volumes, as tumour size will vary considerably between patients and diseases. The graphs of volume dependency for each feature (supplementary material) were visually inspected to determine those features that showed non-converging behaviour (as described by Forgacs et al [345]) and hence should be excluded from further analysis. Those features with a known or intrinsic correlation with volume (e.g. SUVs and shape-related features such as sphericity) but have been shown to have clinical value [84, 278] and are in common usage within the literature were not excluded.

Features must also be robust to repeat measurements (for example to allow confidence in assessing changes following treatment), therefore for the Revolver phantom experiment, those features that showed CoV >10% in any Revolver insert were excluded.

8.3 Results

8.3.1 Volume dependency

Graphs displaying the volume dependency against 67 features for both radionuclides (up to a VOI of 150 ml) are contained in the supplementary information. Forgacs et al demonstrated three typical patterns of volume dependency [345]: Figure 8.3 shows examples of these three patterns taken from this work. All converging features showed a volume dependency at volumes below approximately 25 ml (Table 8.1). Features that showed convergence with increasing volume were identified by visual inspection (supplementary material).

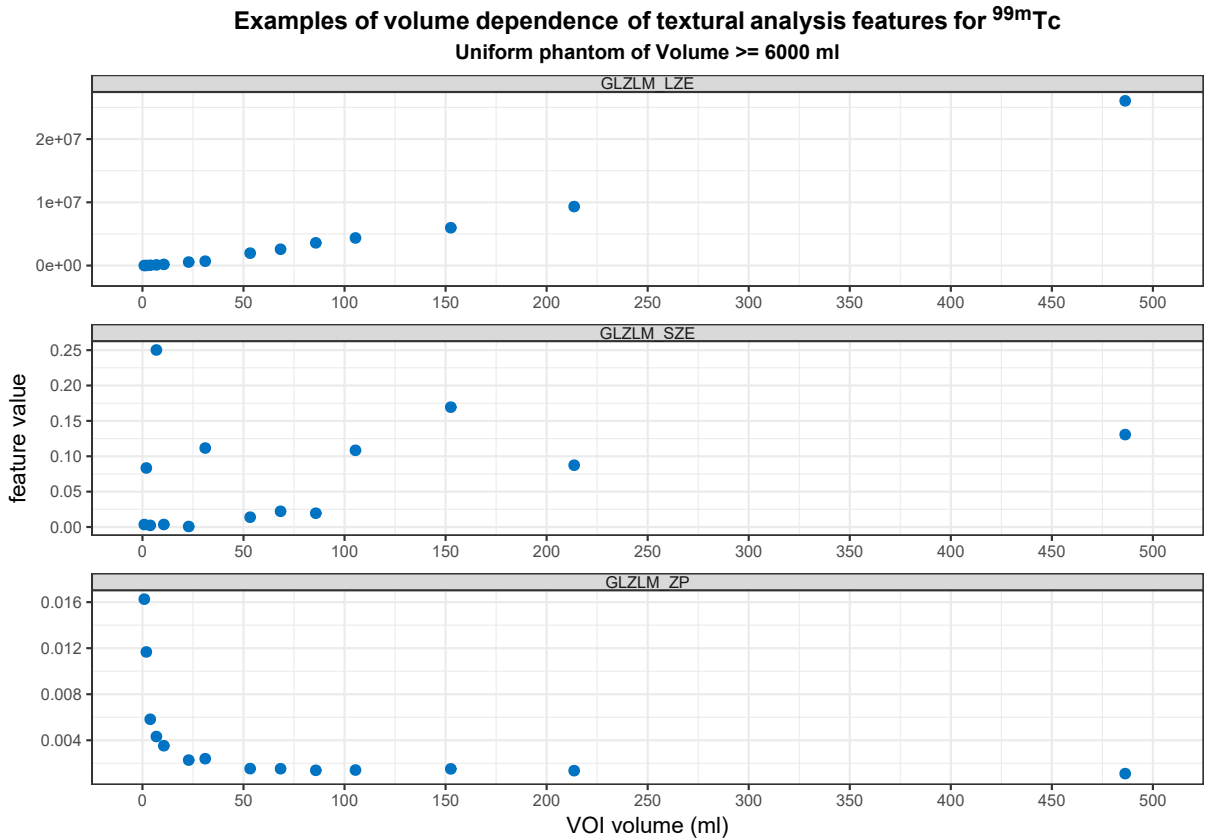


Figure 8.3: Examples of (top) positive correlation; (middle) random distribution; and (bottom) convergence with increasing volume

8.3.2 Reproducibility in inhomogeneous volume

An example slice from the reconstructed images is shown in Figure 8.4 for each of the radionuclides tested. The average volume of the VOI after application of thresholding is shown in Table 8.2.

The features that exceed the 10% CoV threshold suggested by the EANM/SNMMI radiomics guidelines are summarised in Table 8.3 and Table 8.4 for each Revolver insert, where multiple failures across the three Revolver sizes are emphasised.

^{99m} Tc	¹⁷⁷ Lu
SUVQ1	SUV Skewness
SUVQ2	SUV Kurtosis
SUVQ3	SUV Excess Kurtosis
Discretised SUVmean	Discretised SUVmin
Discretised SUVmax	Discretised SUVmax
Discretised TLSRE	Discretised SUVQ3
GLCM Homogeneity	Discretised SUV Skewness
GLCM Contrast	Discretised SUVpeak 0.5ml
GLCM Dissimilarity	Discretised SUVpeak 1ml
GLRLM LGRE	Discretised TLSRE
GLRLM HGRE	GLRLM HGRE
GLRLM SRLGE	GLRLM GLNU
GLRLM GLNU	GLRLM RLNU
GLRLM RLNU	NGLDM Busyness
NGLDM Contrast	GLZLM SZE
NGLDM Busyness	GLZLM LZE
GLZLM SZE	GLZLM LGZE
GLZLM LZE	GLZLM HGZE
GLZLM LGZE	GLZLM SZLGE
GLZLM HGZE	GLZLM SZHGE
GLZLM SZLGE	GLZLM LZLGE
GLZLM SZHGE	GLZLM LZHGE
GLZLM LZLGE	GLZLM GLNU
GLZLM LZHGE	GLZLM ZLNU
GLZLM GLNU	GLRLM LRE
GLZLM ZLNU	

Table 8.1: Non-converging features following volume dependency testing

8.3.3 Final feature reduction

The remaining features from the volume dependency and Revolver phantoms were combined to produce a list of remaining features: these features were suggestive for future exploration in clinical studies using data acquired and reconstructed with the local parameters as listed in Table 8.5: it consists of 39 features for ^{99m}Tc and 33 features for ¹⁷⁷Lu.

8.4 Discussion

With the recent publication of EANM guidelines for quantitative SPECT-CT [11], there is greater emphasis on the harmonisation of quantitative SPECT-CT; though further work is needed to harmonise acquisition and reconstruction parameters in the same way that EARL accreditation aids harmonisation for PET-CT [268]. This paper confirms that the

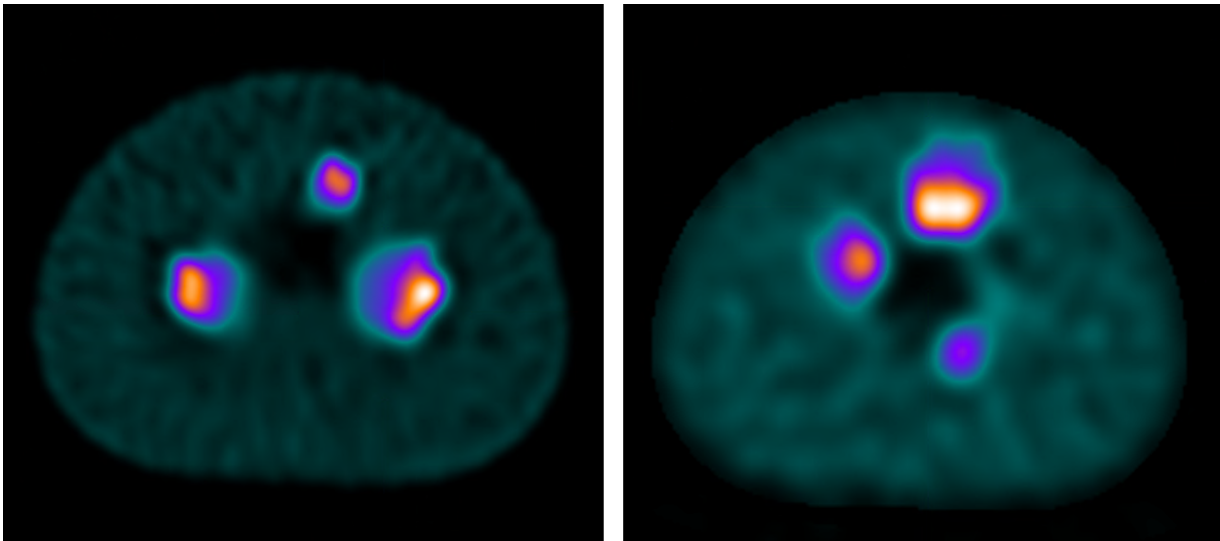


Figure 8.4: Axial slices of (left) ^{99m}Tc and (right) ^{177}Lu Revolver phantoms

Revolver insert (syringe size)	^{99m}Tc	^{177}Lu
2.5 ml	33.3 ± 0.5 ml	30.3 ± 1.3 ml
5 ml	66.3 ± 1.0 ml	64.0 ± 3.7 ml
10 ml	141.0 ± 1.1 ml	121.9 ± 6.7 ml

Table 8.2: Average VOI size after application of threshold

methodology outlined in EANM and SNMMI guidelines on radiomics [90] are applicable to SPECT, though the specific features selected should be confirmed by testing at other centres, due to differences in acquisition and processing parameters as well as inherent differences in camera performance that may affect reproducibility (as demonstrated in PET radiomics [117]).

For some features, a direct dependence on volume is to be expected (e.g. for surface area) although that in and of itself is not a reason for exclusion from further analysis, as it may be of clinical value. For example, two tumours may have the same volume, but an irregular-shaped lesion would have a greater surface area. This may have clinical implications: one study has proposed that the density of somatostatin receptors would be higher for such an

CoV >10% in 1 Revolver insert	CoV >10% in 2 Revolver inserts	CoV >10% in all 3 Revolver inserts
Discretised SUV Excess Kurtosis GLZLM LZE GLZLM LZLGE GLZLM SZHGE GLZLM SZLGE GLZLM ZLNU NGLDM Busyness	Discretised Histogram Excess Kurtosis SUV Excess Kurtosis	

Table 8.3: Radiomics features with CoV >10% for ^{99m}Tc

CoV >10% in 1 Revolver insert	CoV >10% in 2 Revolver inserts	CoV >10% in all 3 Revolver inserts
Compacity	Discretised Histogram Kurtosis	Discretised Histogram Excess Kurtosis
Discretised SUV Kurtosis	GLZLM SZE	Discretised SUV Excess Kurtosis
Discretised SUV Skewness	GLZLM SZHGE	GLCM Energy
Discretised SUVmin	GLZLM ZP	GLZLM GLNU
GLZCM Contrast	NGLDM Busyness	GLZLM HGZE
GLRLM LGRE	NGLDM Contrast	GLZLM LGZE
GLRLM LRLGE	SUVmin	GLZLM LZE
GLRLM SRLGE		GLZLM LZHGE
SUV Kurtosis		GLZLM LZLGE
SUV Skewness		GLZLM SZLGE
		GLZLM ZLNU
		SUV Excess Kurtosis

Table 8.4: Radiomics features with CoV >10% for ¹⁷⁷Lu

irregular-shaped tumour, therefore somatostatin receptor radiopharmaceuticals would show greater uptake, and hence a feature such as low sphericity could be predictive [84]. Some calculated features (such as those that combine SUV and volume, for example, Total Lesion Somatostatin Receptor Expression) are also specifically constructed so as to be dependent on volume and have been shown to have predictive value in patients [248].

As well as theoretical limitations of testing and excluding features due to volume dependence (as outlined in the point above), the practical implementation of testing with spherical VOIs in an image made up of rectangular voxels also has drawbacks. The VOI will not be truly spherical, hence affecting the measurement of some features, e.g. sphericity. The pixelation will be more prominent at lower volumes and may imply a volume dependence not truly inherent to the feature. Spherical lesions are also unlikely to be delineated in clinical studies.

A study on tumoural homogeneity [106] used a mathematical framework to calculate a minimum tumour size for accurate inhomogeneity measurement in PET of 45 ml (though dependent on their assumptions in imaging parameters). They propose that this limit is not just influenced by technical factors, but also the biology of the tumour, hypothesising that smaller tumours will not have sufficient size to allow for the potential biological variations (e.g. hypoxia, necrosis) that result in inhomogeneity to exist [106]. The limitations of their minimum volume calculation, as highlighted by Hatt et al [92] are that it was performed only for a single feature (GLCM entropy), fixed bin size quantisation was used and testing was only in 2D. It has been shown that using fixed bin width (used in this study) instead of fixed bin size is that it can reduce the correlation of features with volume [105]. A different study looking at both simulated spherical lesions and multiple cohorts of patient images found (for a limited set of features) that while a volume dependency did exist, the correlation was limited above volumes of 10 ml, and useful information could be extracted, potentially down to volumes greater than 3 ml – for example with Entropy, Dissimilarity, or Zone Percentage [138]. The work reported here did show similar results for these features, with reducing correlation with

^{99m} Tc	¹⁷⁷ Lu
SUVmin	SUVmin
SUVmean	SUVmean
SUVstd	SUVstd
SUVmax	SUVmax
SUV Skewness	SUVQ1
SUV Kurtosis	SUVQ2
SUVpeak 0.5ml	SUVQ3
SUVpeak 1ml	SUVpeak 0.5ml
TLSRE	SUVpeak 1ml
Discretised SUVmin	TLSRE
Discretised SUVstd	DISC SUVmean
Discretised SUVQ1	Discretised SUVstd
Discretised SUVQ2	Discretised SUVQ1
Discretised SUVQ3	Discretised SUVQ2
Discretised SUV Skewness	Discretised Histogram Skewness
Discretised SUV Kurtosis	Discretised Histogram Entropy log10
Discretised SUVpeak 0.5ml	Discretised Histogram Entropy log2
Discretised SUVpeak 1ml	Discretised Histogram Energy
Discretised Histogram Energy	Sphericity
Discretised Histogram Skewness	Surface Area
Discretised Histogram Kurtosis	GLCM Homogeneity
Discretised Histogram Entropy log10	GLCM Correlation
Discretised Histogram Entropy log2	GLCM Entropy log10
Sphericity	GLCM Entropy log2
Surface Area	GLCM Dissimilarity
Compacity	GLRLM SRE
GLCM Energy	GLRLM SRHGE
GLCM Correlation	GLRLM LRHGE
GLCM Entropy log10	GLRLM RP
GLCM Entropy log2	NGLDM Coarseness
GLRLM SRE	NGLDM Contrast
GLRLM LRE	GLZLM ZP
GLRLM SRHGE	Volume
GLRLM LRLGE	
GLRLM LRHGE	
GLRLM RP	
NGLDM Coarseness	
GLZLM ZP	
Volume	

Table 8.5: Remaining features following reduction

increasing volume for ^{177}Lu however, the results were not replicated for $^{99\text{m}}\text{Tc}$. The work here was done with spherical VOIs in a uniform phantom however and may not be directly comparable to segmented lesions in patients, as in the original study.

The importance of the number of voxels, rather than volume, is emphasised by Orlhac et al in their guide to implementing Textural Analysis in PET [88]; they suggest a minimum of 64 voxels is required in a VOI in order to calculate robust features (for the pixel size in this study that would correspond to a minimum volume size of 0.5 ml – smaller than the VOIs used in this study).

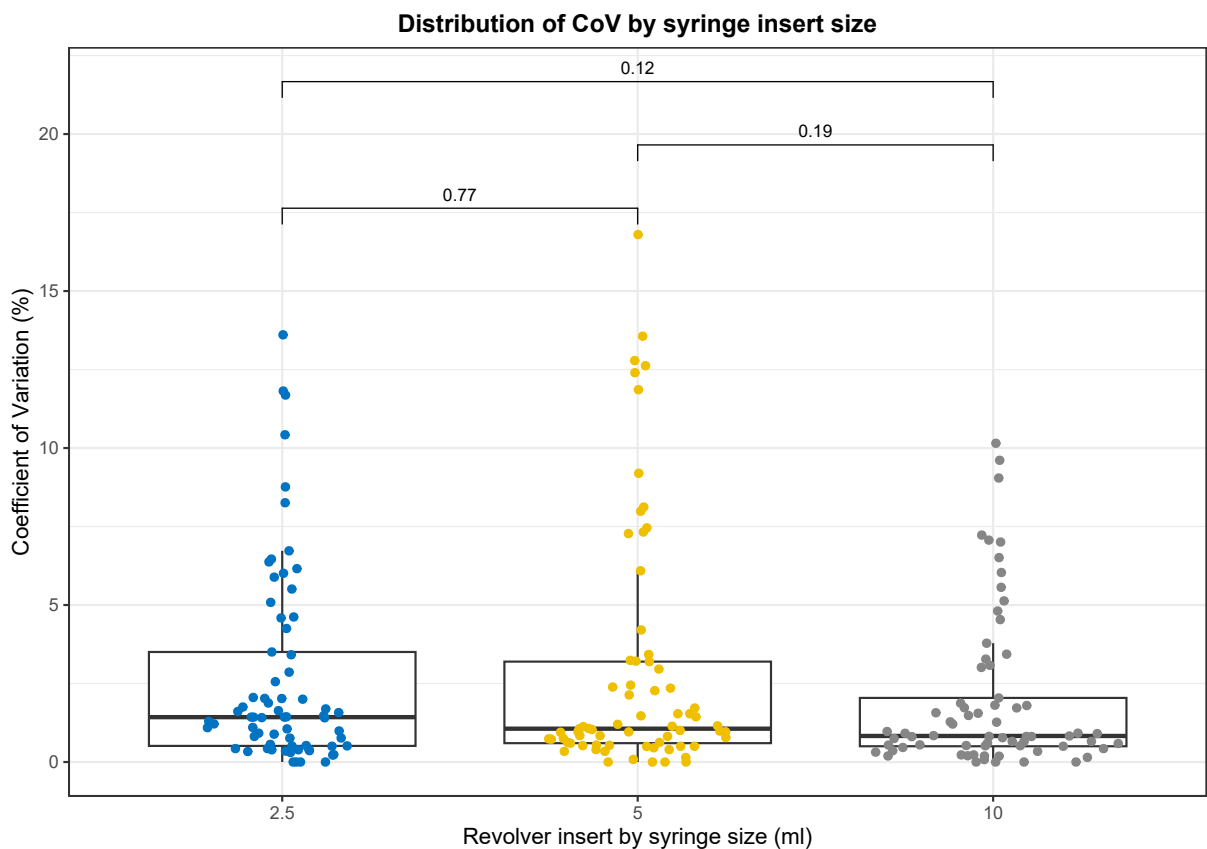


Figure 8.5: Distribution of CoV results for $^{99\text{m}}\text{Tc}$ by syringe size

An increasing size of the Revolver insert did not appear to decrease variability in the feature values: more features exceeded the CoV threshold of 10% for the 5 ml syringes than the 2.5 ml, for both radionuclides. There was no significant difference between the distributions of CoV results for any of the three Revolver insert sizes (Figure 8.5). The higher number of features exceeding CoV >10% for ^{177}Lu compared to $^{99\text{m}}\text{Tc}$ may be due to higher noise in the images, potentially caused by the lower frequency of gamma emissions from ^{177}Lu than $^{99\text{m}}\text{Tc}$, or due the differing number of iterations and post-reconstruction filtering applied in the optimised reconstructions for each radionuclide. However, results from uniform phantom acquisitions where both radionuclides were reconstructed with the same reconstruction

parameters showed that ^{177}Lu had a higher CoV than $^{99\text{m}}\text{Tc}$ (20.2% versus 14.3%, $p < 0.001$) and highlights the higher noise in ^{177}Lu images.

This study does not investigate the impact of segmentation on reproducibility of features: neither the impact of differing segmentation algorithms nor inter-operator variability. Experimental evaluation of the effect of different segmentation algorithms would be a significant undertaking, given the extensive variety of methods available – see Foster et al for a review of segmentation algorithms applied to PET [99]. This study utilises an adaptive threshold based on 2.5 times the background intensity, as prescribed in the Revolver phantom instructions in the EANM/SNMMI guidelines' supplementary information [90]. The advantages of this method are the simplicity of application using a variety of processing platforms, and that it is adaptable to changes in local background: an important factor in clinical studies. An issue for adaptive threshold segmentation can be the exclusion of necrotic areas of tumours, which may show reduced uptake centrally [91]. Inter-operator variability for segmentation is likely to be small for a SPECT phantom study using an adaptive threshold method; in clinical SPECT studies the variability in a segmented volume may be higher, due to difficulties in agreeing on the tumour geometry [348]. There is also likely to be inter-feature correlation (particularly between discretised and non-discretised versions of the same feature), however this is better tested in analysis of clinical images, where other highly correlated features can also be identified [341].

Phantom acquisitions may not provide a sufficient representation of the volumetric complexity and heterogeneity of lesions in the clinical setting – spherical inserts or, in this particular study, syringes do little to mimic the appearance of clinically relevant lesions. The use of 3D-printed lesions, designed from imaging of actual tumours would help phantom studies more closely replicate clinical imaging [127]. Factors such as patient movement, respiratory motion and physiological excretion are also not incorporated into most currently available phantoms. The identification of robust radiomic features by phantom study is just one step in the path to clinical translation, with significant issues identified in several review articles and editorials [81, 349].

It is clear from Table 8.3 and Table 8.4 that certain 'families' of feature were particularly prone to higher CoV and could be chosen to be excluded from further analyses (notably Excess Kurtosis and the family of GLZLM features). It is expected that many of the discretised features will show strong correlation with their non-discretised versions. However, the user must be careful not to remove features that may provide clinical information [81], hence for both radionuclides a separate list of remaining features has been produced that may be applied to local clinical studies on Textural Analysis, where more targeted feature reduction, based on clinical data, can be performed.

The clinical impact of radiomic features in Nuclear Medicine has been highlighted by a number of studies in the literature – mainly in PET, for example showing prognostic value for patient response to treatment [79, 350] or to non-invasively predict tumour metabolic phenotypes [116]. It has been suggested that a lack of reproducible and robust features has led to limited interest in radiomics for SPECT [349], with only a small number of published studies (for example in Parkinson’s Disease [83] and Neuro-Endocrine Tumours [84]). This study aims to promote the identification of robust radiomic features for SPECT imaging through the important initial steps of phantom studies, by adapting techniques recommended for PET imaging to SPECT. It is hoped that by validating the methodology as practicable and relevant to SPECT imaging, it will promote the use of phantom studies in other centres to select robust radiomic features for SPECT imaging and comparison of features across multiple centres will be encouraged to gauge compatibility, with the aim of an increased number of SPECT studies utilising radiomics clinically.

8.5 Conclusion

This study demonstrates that the quality assurance and feature reduction techniques using phantoms proposed in recent Textural Analysis guidelines for Nuclear Medicine can be applied to SPECT for ^{177}Lu and $^{99\text{m}}\text{Tc}$ radiopharmaceuticals. Features identified as having a strong dependence on volume or high CoV in inhomogeneous lesions may be removed from further analysis, but identification of such features will depend on local acquisition and reconstruction parameters and should be confirmed by other centres.

8.6 Relevance for the thesis

The work described above demonstrates that under the specific conditions of a phantom experiment, there are numerous textural analysis features that allow for reproducible measurements for both ^{177}Lu and $^{99\text{m}}\text{Tc}$. The features shown to be reproducible for the two radionuclides may differ due to differences in sensitivity (and hence total acquired counts and noise), as well as differences in spatial resolution.

If non-reproducible features are included, it increases the chance of inaccurate results and makes external validation more difficult. A reduced feature set also decreases the problem of dimensionality and allows for development of simpler models, both important factors when applying textural analysis to clinical data, which will be demonstrated in later chapters.

Part 3:

Clinical quantification for Neuro-Endocrine Tumours

The life and soul of science is its practical application.

William Thomson, Baron Kelvin of Largs
Lecture on "Electrical Units of Measurement", 1883

The work in the 2nd Part of this thesis described the calibration process and validation of quantitative SPECT. This validation is essential: it must be shown that any measurements made in clinical studies are sufficiently accurate and reproducible. The 3rd Part will hence describe how the metrics quantified in the 2nd Part, such as SUV and textural analysis features can be applied to imaging of NETs.

Chapters in the 3rd Part will demonstrate how quantitative SPECT can add value to the imaging process; for example, identifying changes to uptake caused by medication, or how visual assessment of uptake can be rendered more reproducible with quantitative metrics. There will also be a focus on how quantification may be used to predict patient outcomes from PRRT, either measured objectively or by patient-reported outcomes.

The final chapter in the 3rd Part will investigate the use of statistical tools to harmonise data between SPECT systems. The utility of any clinically relevant quantitative metrics determined in a study is limited to the specific centre and circumstances under which data was gathered, unless it can be validated independently. This has proven difficult for many quantitative markers in PET imaging, and is a potential issue for the growth of quantitative SPECT. Work to more closely harmonise results from different centres will benefit the validation process and help with widespread adoption of quantitative metrics.

Chapter 9

Quantitative Uptake in ^{99m}Tc -EDDA/HYNIC-TOC Somatostatin Receptor Imaging – the effect of Long-Acting Release Somatostatin Analogue Therapy

9.1 Context of published paper

The following chapter details an example of how quantification may provide added value to visual interpretation of images. A visual reporter is limited to assessing the uptake of radiopharmaceutical within a single patient they are currently viewing, which can be subject to a variety of influences (e.g. hydration status, time since administration, and most pertinently for the following work, interactions with concurrent medications). Often the information provided on the referral relating to a patient's medication is limited, while for patients from external health boards access to the appropriate hospital information systems is often not possible. Relying on the patient to provide an accurate history of their medication is problematic, with names and dates of medication often unclear. The ability of quantification to provide objective evidence as to a patient's medical status could be of value in clinical practice, especially if ratios of tumour uptake to physiological uptake are used to decide eligibility for therapy. In this chapter it is also shown that quantification could provide evidence to change the guidance from professional bodies on the optimal scheduling of patient's medications to maximise the effect of therapeutic administration of radiopharmaceuticals for neuroendocrine tumours.

The published work has been redacted from the electronic version of this thesis, in line with copyright restrictions. The original paper can be found under the citation:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, "Quantitative uptake in ^{99m}Tc -EDDA/HYNIC-TOC somatostatin receptor imaging – the effect of long-acting release somatostatin analogue therapy," *Nuclear Medicine Communications*, vol. 44, no. 11, pp. 944–952, Nov. 2023, doi:10.1097/MNM.0000000000001746.

9.2 *Content removed due to copyright restrictions*

9.3 Relevance for the thesis

This work demonstrates that quantitative analysis using SPECT-CT can provide important clinical information for NET patients and emphasises how factors like concomitant medication may affect quantitative measurements, sometimes in a counter-intuitive manner. Selection of patients for clinical studies must therefore account for these differences, suggesting analysis in two separate groups (on LAR-SSA treatment and LAR-SSA naive) is appropriate when utilising quantitative features.

The differences in uptake between the two groups (especially between tumour and physiological uptake), also raises important issues for visual interpretation of patient scans. This matter will be investigated further in the following chapter, specifically in the case of visual rating of uptake based on the Krenning Score.

Chapter 10

Comparison of quantitative Krenning Scores with visual assessment in ^{99m}Tc -EDDA/HYNIC-TOC SPECT-CT

10.1 Context of published paper

The previous chapter demonstrated that quantitative SPECT-CT could provide useful information in SRI, as well as highlighting that the use of such data must take into account factors such as tumour type and if the patient is taking somatostatin analogue treatment. For adoption into the clinical workflow, quantification must also show it meets or surpasses the utility of visual interpretation, as the current standard practice. For SRI an important aspect of reporting, which quantification may assist with, is determination of the Krenning Score. This five-point scale uses visual assessment of uptake in a site of suspected tumour compared to physiological uptake elsewhere, and is often used to determine eligibility for PRRT. The current practice for determining this score on a visual basis has several downsides associated with any visual analysis task performed by multiple operators and it is proposed that quantification may aid repeatability and standardisation: this will be investigated in the following work.

The published work has been redacted from the electronic version of this thesis, in line with copyright restrictions. The original paper can be found under the citation:

A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, "Comparison of quantitative Krenning Scores with visual assessment in ^{99m}Tc -EDDA/HYNIC-TOC SPECT-CT," *Nuclear Medicine Communications*, vol.46, no. 6, pp. 515–522, Jun. 2025, doi: 10.1097/MNM.0000000000001967.

10.2 *Content removed due to copyright restrictions*

10.3 Relevance for the thesis

This chapter demonstrates a potential clinical role for quantification, identifying metrics that could provide more reproducible scoring for determining eligibility for PRRT, whilst maintaining

the link to the Krenning Score that is familiar both to reporters and referrers. Quantification has the potential to act as confirmation to an expert reporter's opinion or as a training tool for less experienced reporters, particularly those working in centres without specialists in SRI.

As described in the Discussion of the published paper [351], the use of the Krenning Score in SPECT imaging where it has not been validated is a potential drawback to this study. Alternative scoring metrics for SRI do exist, for example the SSTR-RADS scale that has been developed for PET imaging [220]; however, this metric has not been tested in SPECT. The Krenning Score has more widespread adoption than SSTR-RADS, as seen by its promotion in the European Neuroendocrine Tumour Society (ENETS) standardised reporting guidelines [214]. Importantly, Krenning Score is often the criteria for determining eligibility for PRRT, adopted due to its use in the clinical trials used to prove the efficacy of ^{177}Lu -DOTATATE (for example, as part of the eligibility criteria for the NETTER-1 trial [225]). Adoption of a different metric than Krenning Score would require extensive re-training of reporting staff, as well as providing information on the new scoring system and how it relates to Krenning Score to the oncologists and Multi-Disciplinary Teams who care for the patients. The techniques applied in this chapter could easily be applied to SSTR-RADS, as it relies on a qualitative assessment of uptake, and used to derive equivalent quantitative scores.

The Discussion also considers the drawbacks of using an Expert Consensus as the gold standard, making the case as to why a quantitative metric may be preferred. However, fully proving the accuracy of quantitative metrics would require a prospective study comparing against an objective endpoint. Such trials are hard to design, especially in NETs where there are multiple treatment options, different imaging modalities (including multiple choice of radiopharmaceuticals) and progression of disease may only occur many years later [219]. A further limitation of this study is that it does not conform to optimal study design by testing against an external dataset consisting of data acquired at a different centre or analysed by different observers [90]. Therefore, the results shown above will need further validation before they could be judged applicable to wider use outwith our department. Further work on this topic may involve:

- Including patients who are receiving LAR-SSA treatment, as it has been shown that SUVs are significantly different for physiological uptake in the spleen and liver (see Chapter 9): as these are the organs mainly used for comparative purposes in the visual Krenning Score, the observer's classification of the Krenning Score could be altered.
- Incorporating additional observers from different centres to derive a wider consensus that could potentially expand the applicability of conclusions beyond this centre.
- Test the developed models (e.g. qK scores) with imaging data from other centres, to evaluate applicability when acquisition and reconstruction protocols are varied.

If the results from this study were validated at other centres, then it would display the potential for quantification to act as a training tool or to provide reassurance to reporters in centres without access to expert review.

Chapter 11

Textural Analysis and comparison to visual Krenning Scores

11.1 Introduction

The previous chapter investigated the use of some simple intensity-based quantitative metrics such as SUV as an alternative to visual scoring of ^{99m}Tc -EDDA/HYNIC-TOC using the Krenning Score. These metrics are commonly understood amongst the nuclear medicine community and benefit from being easily explained, an important consideration if used as part of a clinical decision making process. However, as outlined in Chapter 3, recent research on quantitative imaging has often focussed on radiomics, and the use of textural analysis to extract greater information from the visual data. It was therefore sought to expand the work of Chapter 10 by exploring the potential uses of textural analysis as an adjunct or replacement to visual interpretation of the Krenning Score.

11.2 Materials and Methods

This study utilised the same patient data as gathered for Chapter 10, with the imaging and image reconstruction protocols remaining unchanged from those in the "Materials and Methods" section of the published version [351].

11.2.1 Quantitative Analysis

Segmentation of lesions used the same data analysis performed for Chapter 10; the LifeX software (Lifex v6.3, www.lifexsoft.org) [98] was used to extract additional metrics to the SUVs used in Chapter 10. These consisted of intensity-derived features, histogram features and textural analysis features, numbering 67 in total. A full list of the available features and a description of each is given in Appendix C. Using the feature reduction work undertaken in Chapter 8, an initial list of 39 features (as shown in Table 8.1) were selected for further analysis, having proven robust in repeatability testing.

11.2.2 Feature reduction

Further feature reduction was performed to remove redundant features and those which exhibited poor intra- and inter-observer variability. To test features for redundancy, a Spearman's Rank Correlation test was run, and features with $|p| > 0.95$ flagged for review [154]. Variability was tested for using Intra-class Correlation Coefficient (ICC) for both intra- and inter-observer testing, as outlined in Chapter 10.

11.2.3 Univariate and multivariable analysis

After normalisation of the feature values, individual (univariate) logistic regressions were performed for the quantitative and textural analysis features using the MASS package in R [352]. The resultant predictive models were tested using the Comparison dataset, to assess their predictive value when compared to the Expert Consensus Score (ECS), using proportions of agreement and Fleiss' Kappa. For multivariable analysis, a penalised constrained continuation ratio model was used (R implementation: `glmnet` [353]) to identify the most appropriate models, for both forward and backward stepwise feature selection by optimising the AIC.

Results are compared against the ECS that was measured in Chapter 10, using percentage agreement, Kappa statistic and predicted Krenning Scores are displayed in a confusion matrix (where the predicted Krenning Score using the model created by univariate analysis is compared to the gold standard of the ECS).

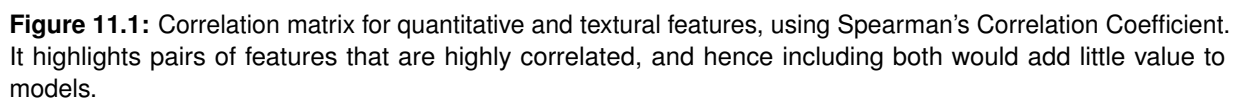
11.3 Results

11.3.1 Feature reduction

Spearman's correlation coefficients were produced for each pairing of features and are displayed in a correlation matrix in Figure 11.1.

For pairings with correlation above 0.95, it was considered if one of the features could be removed due to redundancy. Preference was given to clinically explainable features, non-discretised feature(s) and those with the highest variability (as measured by variance and CoV [97]).

- GLCM Entropy log2 was exactly correlated with GLCM Entropy log 10, and highly correlated with the two Discretised versions of entropy, as well as Discretised Histogram Energy – all but GLCM Entropy log2 were removed.



- GLCM Energy was highly correlated with Discretised Histogram Skewness and Discretised Histogram Kurtosis, which were both removed (It is interesting to note that the SUV and Histogram versions of skewness and kurtosis were not highly correlated).
- The majority of SUV features were highly correlated with each other (e.g. SUV_{mean} and SUV_{max} , $\rho=0.998$). However, these features are widely used in the literature and readily understood by clinicians, so were retained for further analysis. Less clinically intuitive SUV features, such as SUV_{std} and Discretised SUV_{Q1} , were also highly correlated and therefore removed. $SUV_{peak\ 0.5ml}$ had a higher variance than $SUV_{peak\ 1ml}$ (in line with the findings of Vanderhoek, et al [275]) and so $SUV_{peak\ 1ml}$ was removed.
- GLRLM LRHGE was strongly correlated with all SUV features, and so was removed.
- In the GLRLM family, three features were strongly correlated with each other: LRE, SRE and RP. GLRLM LRE had significantly higher variability, so was selected as the feature to retain.
- Several of the Shape features were strongly correlated: Volume, Surface Area and Compacity. Volume is a simple, explainable feature, and widely utilised in other

studies, and therefore was retained. As Surface Area has a potential clinically useful interpretation (see discussion below) it is also retained. Since Compacity is calculated using a formula comprising of Volume and Surface Area, this feature could be removed.

In total, 18 features were removed for redundancy due to high correlation with remaining features.

The ICC scores for intra-rater and inter-rater agreement for the textural analysis features are displayed in Figures 11.2 and 11.3. The qK metrics were included along with the textural analysis features, to check for their agreement. Each ICC was compared to the suggested strengths of agreement of Landis and Koch [354], which although designed for Kappa tests of agreement are suitable as conservative rankings for ICC (it being the most generalised case of Kappa). Those features with a mean ICC score <0.81 (lower than “Almost Perfect”) were removed: this threshold for removal is in line with the majority of studies using ICC for reproducibility in textural analysis that were included in the systematic review by Traverso et al [110]. Four features were identified for removal due to poor ICC levels of agreement: SUV Skewness and SUV Kurtosis (both intra- and inter-operator agreement were poor), Sphericity and NGLDM Coarseness (inter-operator agreement was poor). Following this manual reduction of features for redundancy and non-reproducibility, there remained 17 features that were subject to univariate regression analysis.

11.3.2 Univariate analysis

Table 11.1 displays the Odds Ratio and p-values for the univariate analysis of the reduced feature set. It also displays the levels of agreement with the ECS from the Comparison dataset.

Evidence for some of the ordinal regression models being inappropriate can be seen with the GLRLM LRLGE model, despite relatively good agreement and Kappa scores. A large standard error, non-existent odds ratio and a narrow distribution of the confusion matrix in Figure 11.4 (i.e. repeated prediction of Krenning =3) suggests that the model is not a good representation of the data. This is further demonstrated in boxplots of the distribution of normalised values for each feature (Figure 11.5), where GLRLM LRLGE values are concentrated at a small range for Krenning 2, 3 and 4 and show little possibility of discrimination. In contrast, features such as GLRLM SRHGE and the SUV metrics show distribution of values more amenable to discriminating between the ECS Krenning Scores. Overall, the univariate results do not show any features with better agreement than the quantitative Krenning Score established in Chapter 10.

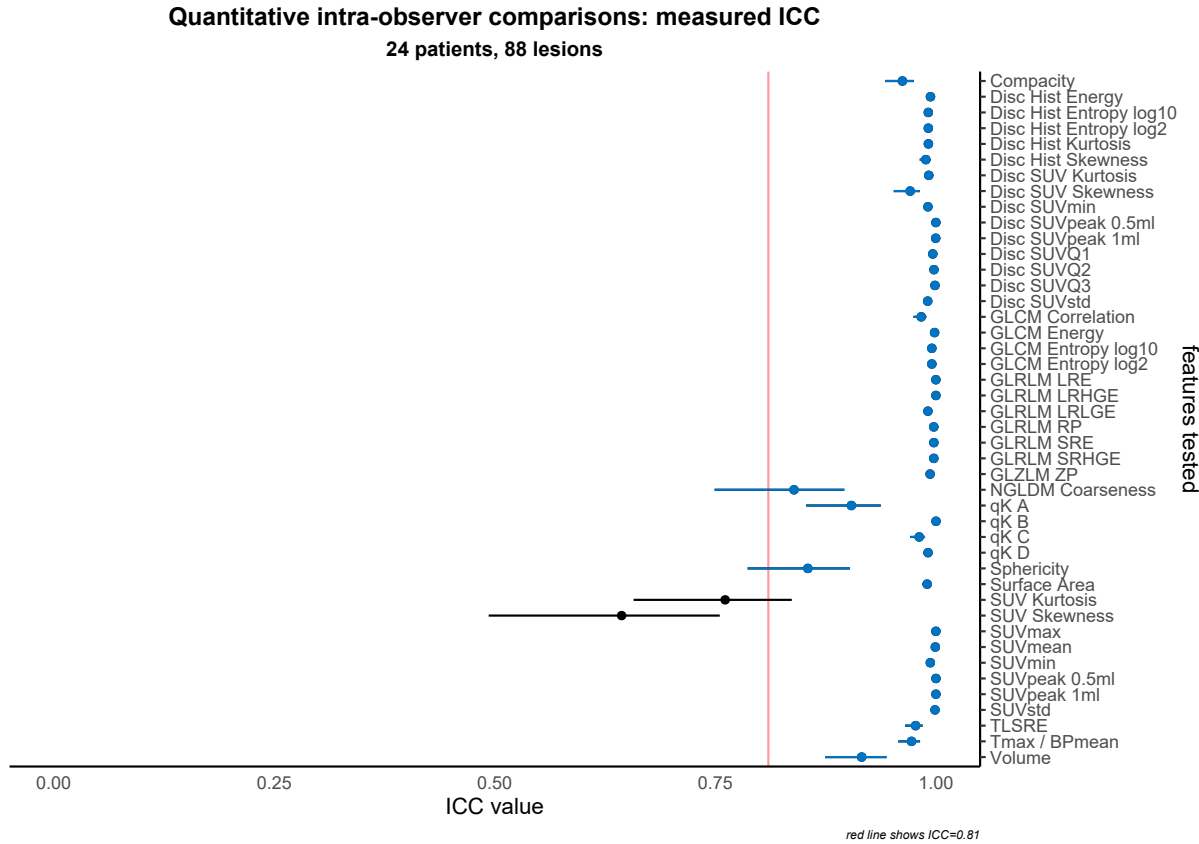


Figure 11.2: Intra-observer variability for textural analysis features, scored by Intra-Class Correlation Coefficient. Cut-off of ICC=0.8 highlighted in red.

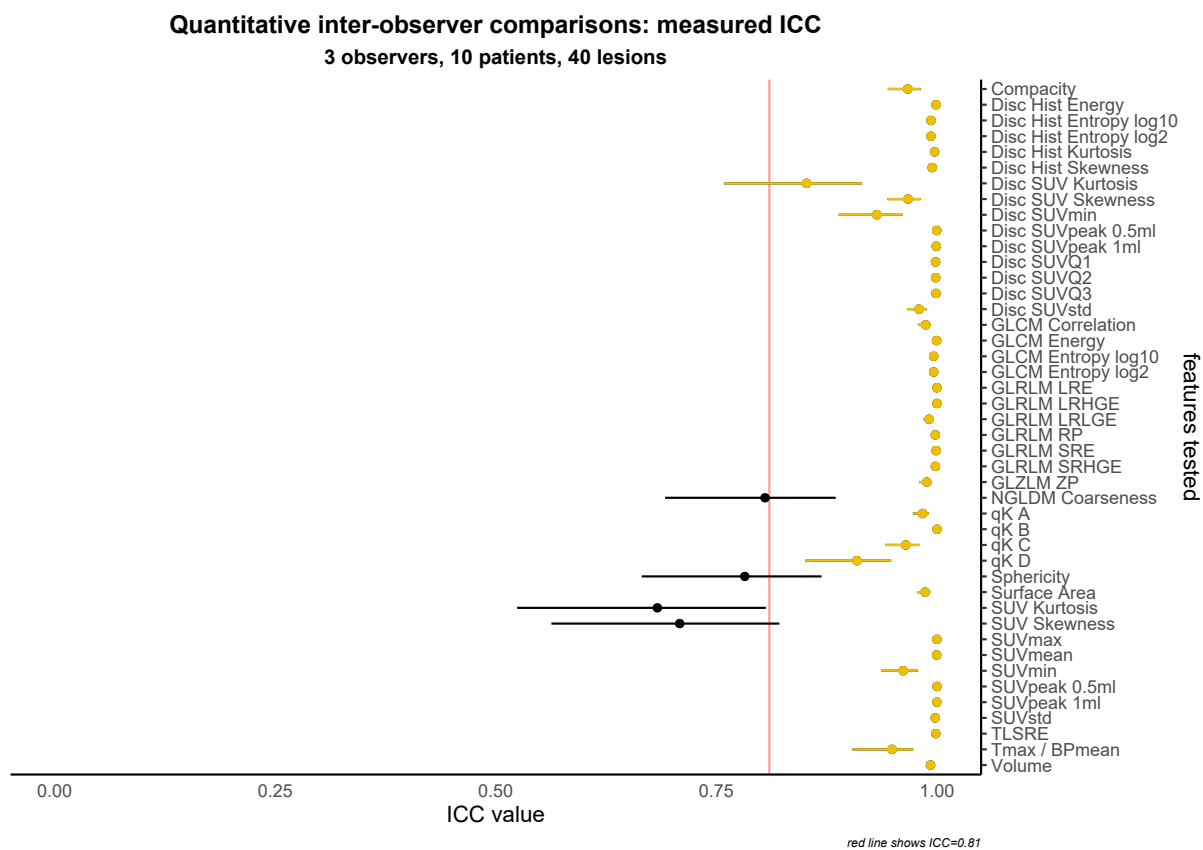


Figure 11.3: Inter-observer variability for textural analysis features, scored by Intra-Class Correlation Coefficient. Cut-off of ICC=0.8 highlighted in red.

Feature	Univariate analysis		Comparison with ECS	
	Odds Ratio (95% CI)	p-value	Agreement	Cohen's Kappa
SUVmin	2.39 (1.85-3.29)	<0.001	79.30%	0.565
SUVmean	2.08 (1.66-2.77)	<0.001	72.40%	0.506
SUVmax	1.59 (1.38 -1.91)	<0.001	75.90%	0.565
SUVpeak0.5	1.63 (1.40-1.97)	<0.001	79.30%	0.631
Disc SUV Skewness	0.31 (0.16-0.55)	<0.001	41.40%	-0.235
Disc SUV Kurtosis	1.64 (1.17-2.63)	0.018	62.10%	-0.122
Volume	1.01 (1.00-1.03)	0.036	69%	0
GLCM Energy	178 (10.9-5861)	<0.001	69%	0
GLCM Correlation	0.16 (0.01-1.90)	0.312	69%	-0.048
GLCM Entropy log2	1.1 (1.03-1.69)	0.405	69%	0
GLRLM LRE	1.15 (1.04-1.31)	0.015	69%	0
GLRLM SRHGE	1.01 (1.01-1.01)	<0.001	72.40%	0.558
GLRLM LRLGE	0 (0.00-0.00)	<0.001	79.30%	0.587
GLZLM ZP	134.28 (38.4-7177)	<0.001	58.60%	0.234
Tmax / BPmean	1.22 (1.14-1.33)	<0.001	72.40%	0.552
TLSRE ¹	-	-	-	-
Surface Area	2.9 (n/a) ²	0.015	10.30%	0

¹ Initial starting values could not be established, therefore regression failed

² Errors in regression meant confidence intervals could not be calculated

Table 11.1: Univariate analysis for a reduced set of textural analysis features, using ordinal logistic regression.

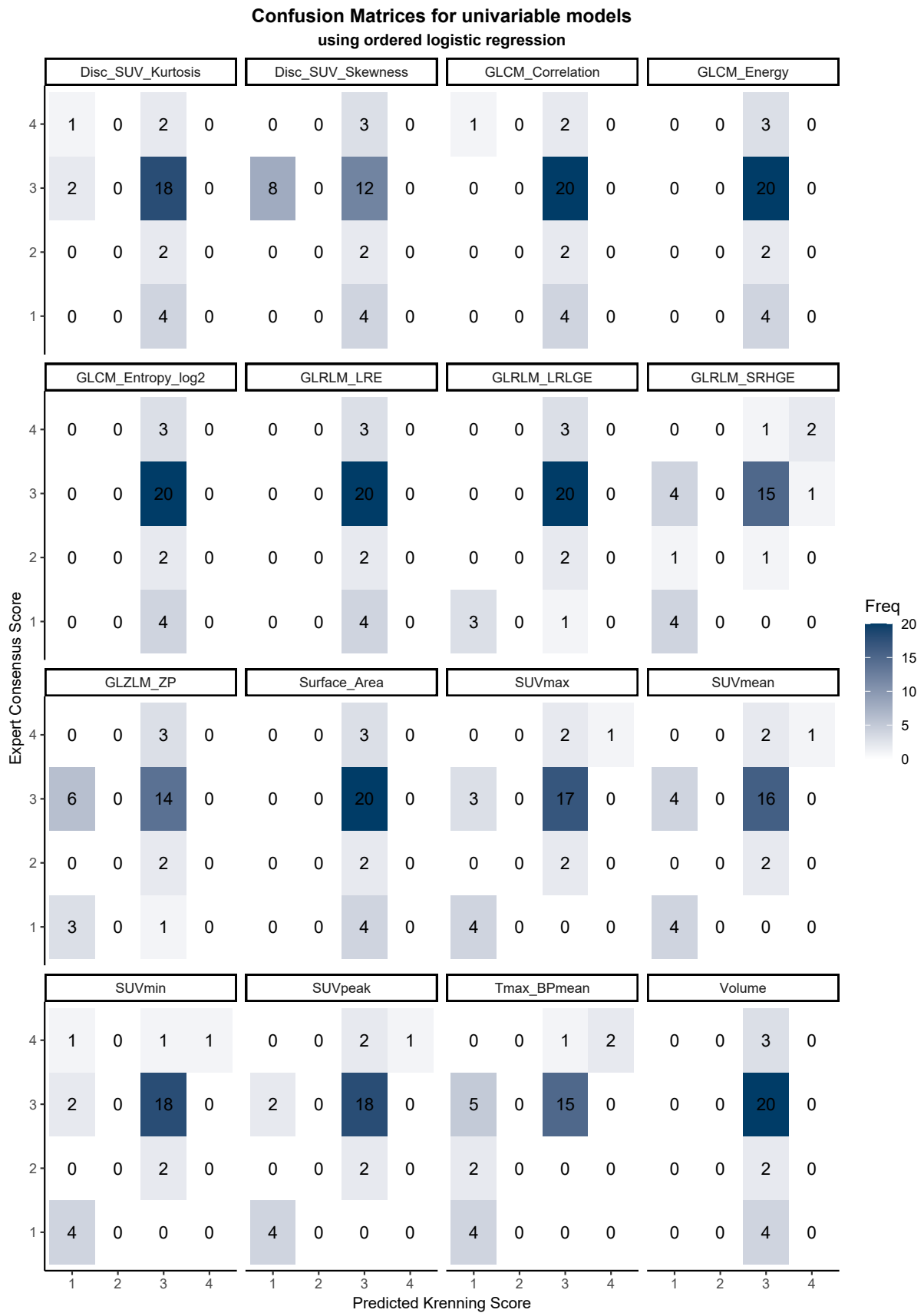


Figure 11.4: Confusion matrix comparing the predicted Krenning Scores by textural analysis against Expert Consensus Score.

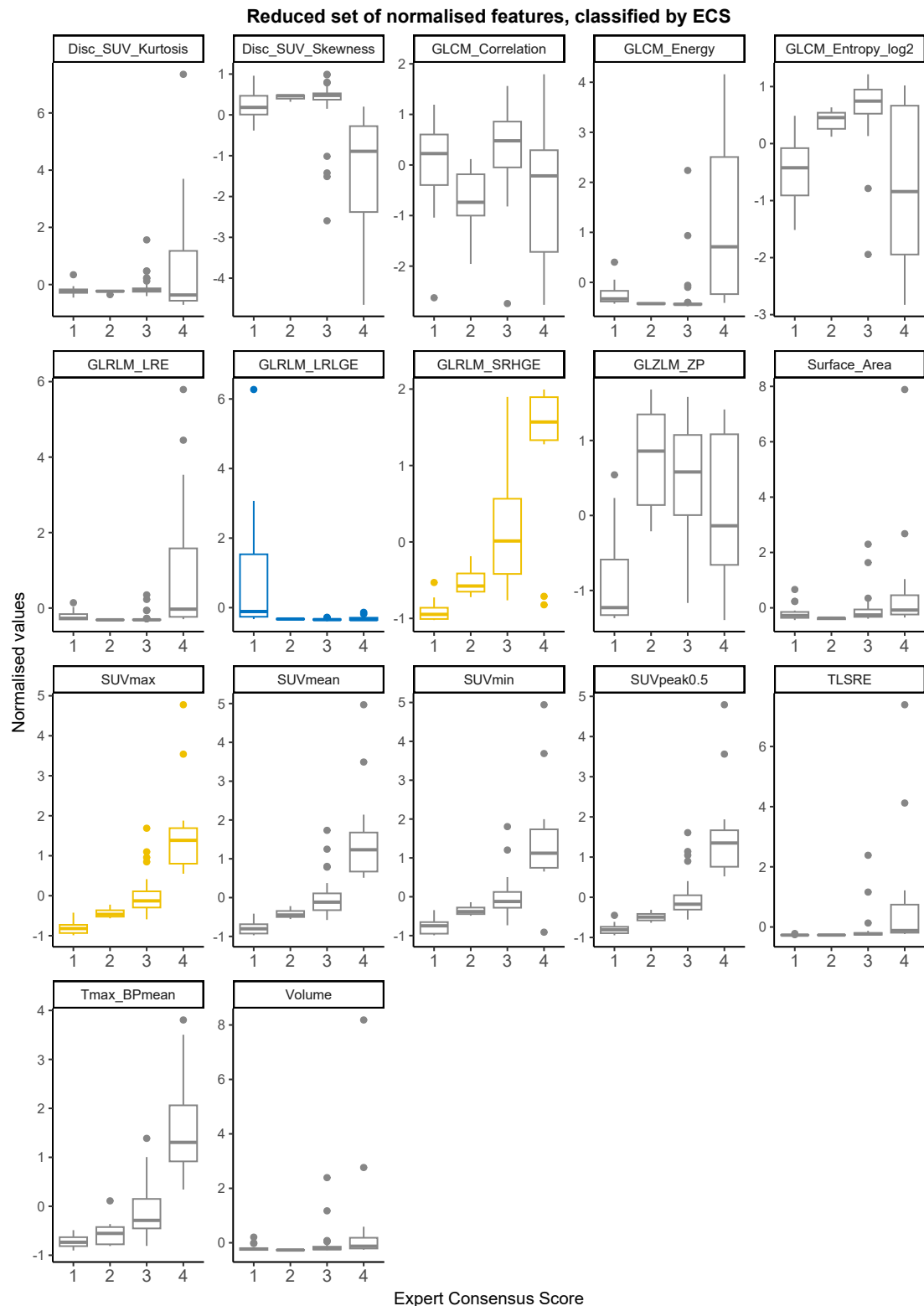


Figure 11.5: Distribution by Expert Consensus Score (ECS) of the normalised values for the reduced textural analysis features. Highlighted in blue is GLRLM LRLGE, which shows good agreement with ECS under univariate analysis, but distribution shows poor discrimination between the ECSs. Highlighted in yellow are GLRLM SRHGE and SUV_{max} , as examples of features where the distribution allows for good discrimination between ECS results.

11.3.3 Multivariable analysis

Results for the multivariable modelling are displayed in tables 11.2 and 11.3. Agreement with the ECS for both forward and backward step-wise feature selection are similar, though both models have features with low coefficients, suggesting they have limited importance in determining the predictions of the model. Results were again inferior to those found in Chapter 10.

Components of Model	Coefficient	Stepwise selection	Comparison with ECS	
			Agreement	Cohen's κ
Intercept	-15.89	Backward	75.90%	0.622
SUVpeak0.5	0.43			
Disc SUV Skewness	2.17			
GLCM Entropy log2	0.34			
GLRLM LRE	0.06			
GLRLM SRHGE	0.001			
GLZLM ZP	0.49			
Tmax / BPmean	0.04			
TLSRE	-0.0001			

Table 11.2: Results from multivariable analysis of the reduced feature set, showing only features identified as significant by the backward step-wise analysis.

Components of Model	Coefficient	Stepwise selection	Comparison with ECS	
			Agreement	Cohen's κ
Intercept	0.79	Forward	79.30%	0.587
SUVpeak0.5	-0.01			
GLRLM SRHGE	-0.001			

Table 11.3: Results from multivariable analysis of the reduced feature set, showing only features identified as significant by the forward step-wise analysis.

11.4 Discussion

Analysis of quantitative metrics in comparison to visual interpretation of Krenning Scores was expanded from the work performed with SUVs in Chapter 10 to include multiple quantitative and textural analysis features. Predictive performance for both univariate and multivariable models was poorer than for the quantitative Krenning Score derived in Chapter 10. The magnitude of some of the mis-classifications can be visualised in Figure 11.4, where predicted Krenning Score is plotted against the ECS in a confusion matrix. For example, for Discretised Skewness (1st row, 2nd column) a Krenning Score of 3 is predicted four times, when the ECS scored the lesion as Krenning Score 1. Conversely, 8 lesions were predicted by Discretised

Skewness to be a Krenning Score of 1, but by ECS were scored as Krenning Score 3. Both of these examples would be clinically significant, changing a patient's eligibility for PRRT.

One reason for the poor performance of some models could be that the distribution of data amongst the different levels of the Krenning Score was not uniform, as seen in Figure 11.6. A class imbalance can lead to models incorrectly predicting the majority class without penalty, as they may still achieve good predictive accuracy [349]. This effect is visible in the confusion matrices, for example with GLCM Energy. The GLCM Energy feature (1st row, 4th column, Figure 11.4) appears to have a decent level of agreement with ECS (69%, Table 11.1), but inspection of Figure 11.4 shows all predictions were for the same Krenning Score of 3. There are techniques available to artificially over-sample any under-sampled classes, but these were not investigated in this work, as they have the potential to introduce poor quality data points [81]. It was felt the sampled data accurately reflected the Krenning Score class distribution in the wider patient population. SRI is usually requested after the patient has biopsy-confirmed NETs, therefore the number of patients with low Krenning Scores (e.g. 0 or 1) is usually fairly limited. In such cases, target lesions for inclusion in this study may not have been identified in the radiologist's report, if superseded in clinical relevance by lesions with higher Krenning Scores. The use of multivariable regression may also be questioned due to the small number of cases in each class of Krenning Score versus the number of features examined (model over-fitting), with some studies in SRI choosing to omit multivariable analysis altogether [355].

There has been limited investigation of textural features in SPECT SRI: Wetz et al identified a textural analysis metric, asphericity, that on diagnostic ^{111}In -Octreoscan that showed a strong association with a patient's response to Lutathera treatment [84]. Their hypothesis is that greater asphericity (or, for the equivalent metric in the Lifex software, lower sphericity) implies a greater surface area where somatostatin receptor sites may be expressed. However, in this study sphericity showed poorer inter-rater variability than most other features (ICC = 0.78; confidence interval = 0.66 – 0.87. The Wetz study makes no reference for testing of intra- or inter-observer variability. Univariate analysis of this data also showed it to be a poor predictor of visual Krenning Score by experts (Agreement = 55.2%, κ = -0.151). A more robust parameter, according to this study, would be surface area of the VOI, which shows significantly less inter-rater variability (ICC = 0.99; confidence interval = 0.97 – 0.99) and should demonstrate the hypothesised effect in a more direct fashion. The results for predicting expert Krenning Score were poor however (Agreement = 10.3%, κ = 0), though there is not necessarily a direct correlation between a pre-therapy visual Krenning Score and treatment response (albeit Krenning Score is still the key imaging criteria for selecting patient eligibility for PRRT). The correlation of this metric with treatment response (through surrogates, such as Overall Survival) will be studied in Chapter 13.

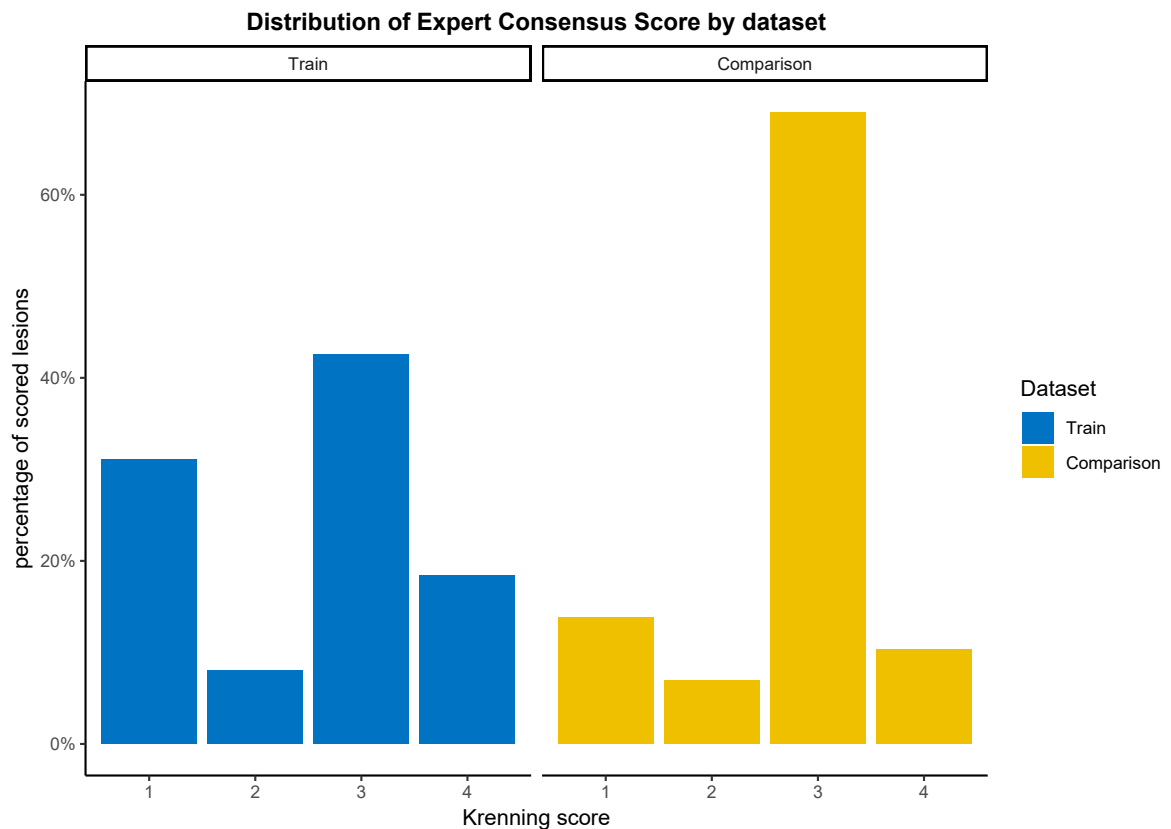


Figure 11.6: Distribution of Expert Consensus Scores by their visual assessment of Krenning Score, for each of the two datasets, "Train" used for model training, and "Comparison" used to test the models.

This study of textural analysis features suffers the same limitations as the previous work in Chapter 10. The use of an ECS as the gold standard is limiting, as it is subject to the issues inherent with any subjective visual assessment. Both studies, utilising the same subject population, may have limited power due to small numbers, with the sample further spread across the distribution of the five levels of the Krenning Score. This is amplified for this study of textural analysis features, as though a successful reduction of features was performed, there remained a limited number of results per feature. Expanding the patient numbers in further study could be of benefit, while widening the scope to include data from other centres would be necessary if useful features were to be found, to enable wider applicability to be assessed.

11.5 Conclusion and relevance for the thesis

While analysis of textural analysis features demonstrated that some show good agreement with an expert consensus reading of the Krenning Score, they demonstrated poorer agreement scores than the quantitative Krenning Score derived in Chapter 10. Given that multivariable models are in general harder to explain than the features identified in Chapter

10 and subject to issues with class imbalance and under-sampling, at present they cannot be recommended as the optimal choice as a quantitative alternative to the Krenning Score.

Chapter 12

The Tumour Sink Effect in ^{99m}Tc -EDDA/HYNIC-TOC quantitative SPECT

12.1 Introduction

For radiopharmaceuticals where there is significant concentration in sites of disease, it has been postulated that a Tumour Sink Effect (TSE) can occur. This involves cases where either extremely avid tumours or a large total tumour volume monopolises uptake of the radiopharmaceutical, resulting in reduced uptake elsewhere (either in the smaller or less-avid sites of disease, or reduced uptake in physiological processes). Examples of TSE have been reported for different radiopharmaceuticals including ^{68}Ga -PSMA for prostate cancer [356] and [^{18}F]FDG [357], while the “superscan” appearance in bone scans is perhaps the best known example of TSE. The presence of TSE in SRI has been suggested in a number of commentary and review articles [358–361]. However, these reviews seem to be based on a limited number of clinical studies using ^{68}Ga -labelled DOTA-peptides, and commonly reference a single study by Beauregard et al [362]. A study based on pharmacokinetic modelling of ^{68}Ga -DOTA-TATE suggested a significant TSE would be quantifiable with a total tumour burden above 550 ml [363]. To determine if the reported TSE extended to SPECT SRI, this study investigated TSE in ^{99m}Tc -EDDA/HYNIC-TOC quantitative SPECT-CT.

12.2 Materials and Methods

A total of 177 patients who received quantitative ^{99m}Tc -EDDA/HYNIC-TOC SPECT-CT for SRI over a 27-month span were included – 8 were excluded due to previous splenectomy. Patients were first divided based on pre-treatment with LAR-SSA. Patients were further sub-categorised based on presence or absence of ^{99m}Tc -EDDA/HYNIC-TOC-avid disease, a visual assessment of extent of disease and the location(s) of disease.

The spleen was selected to represent physiological uptake, as an organ known to have high somatostatin receptor density, that is not part of the excretory process for ^{99m}Tc -EDDA/HYNIC-TOC, and is rarely a site of disease.

For each patient with avid disease on SRI, up to 5 of the most visually prominent lesions were segmented, in line with PERCIST criteria (and as described in Chapter 9 and "Materials and Methods" section of [364]). Correlation with total tumour burden was performed for summed tumour volume (TV) and summed TLSRE using Pearson's correlation test.

Comparisons were also made between patient body size and splenic uptake to explore if the volume of distribution for the radiopharmaceutical affected the physiological uptake. Patient body habitus was measured initially by weight, with the further calculated metrics used:

- Body Mass Index (BMI)
- Body Surface Area (BSA) using the Mosteller correction.
- Lean Body Mass (LBM) using both James and Janma formulas [365].

12.3 Results

Patient's self-reported use of LAR-SSA resulted in groupings of 117 patients naïve to LAR-SSA therapy, with 52 currently receiving LAR-SSA injections (patient-reported mean time since administration 12.4 days, range 0-28 days).

For patients' naïve to LAR-SSA, there was no significant difference between splenic uptake (as measured by SUV_{mean}) for patients without avid disease ($n=49$) and those with disease ($n=68$). This applied for both solitary or multiple sites of disease (Figure 12.1 and Figure 12.2). There was also no dependence on the location of the disease (Figure 12.3).

Comparative analysis was unviable for patients on LAR-SSA, due to the limited numbers of patients without Tektrotyd-avid disease ($n=1$, 1.9%) or with solitary sites of disease ($n=4$, 7.7%). Comparison of the location of sites of diseases was also not amenable to statistical analysis, again due to low numbers in some groupings, but visual inspection suggests there are not significant differences in splenic uptake amongst these groupings (Figure 12.4).

Although on the borderline of statistical significance (with p-values slightly below and above $p=0.05$ respectively), there are weak correlations shown for TV (Figure 12.5) and TLSRE (Figure 12.6) for LAR-SSA-naïve patients. However, these correlations are highly susceptible to variation, due to the small number of data points at large tumour burdens. For patients on LAR-SSA, the weak correlations for both TV and TLSRE are not statistically significant.

For LAR-SSA naïve patients, only 2 of 68 patients had tumour burden > 550 ml (median burden = 39.4 ml; IQR = 63.0). The average SUV_{mean} of the spleen for tumour burden < 550 ml was 22.4 ± 6.5 , which when compared to the values of SUV_{mean} for the two cases of

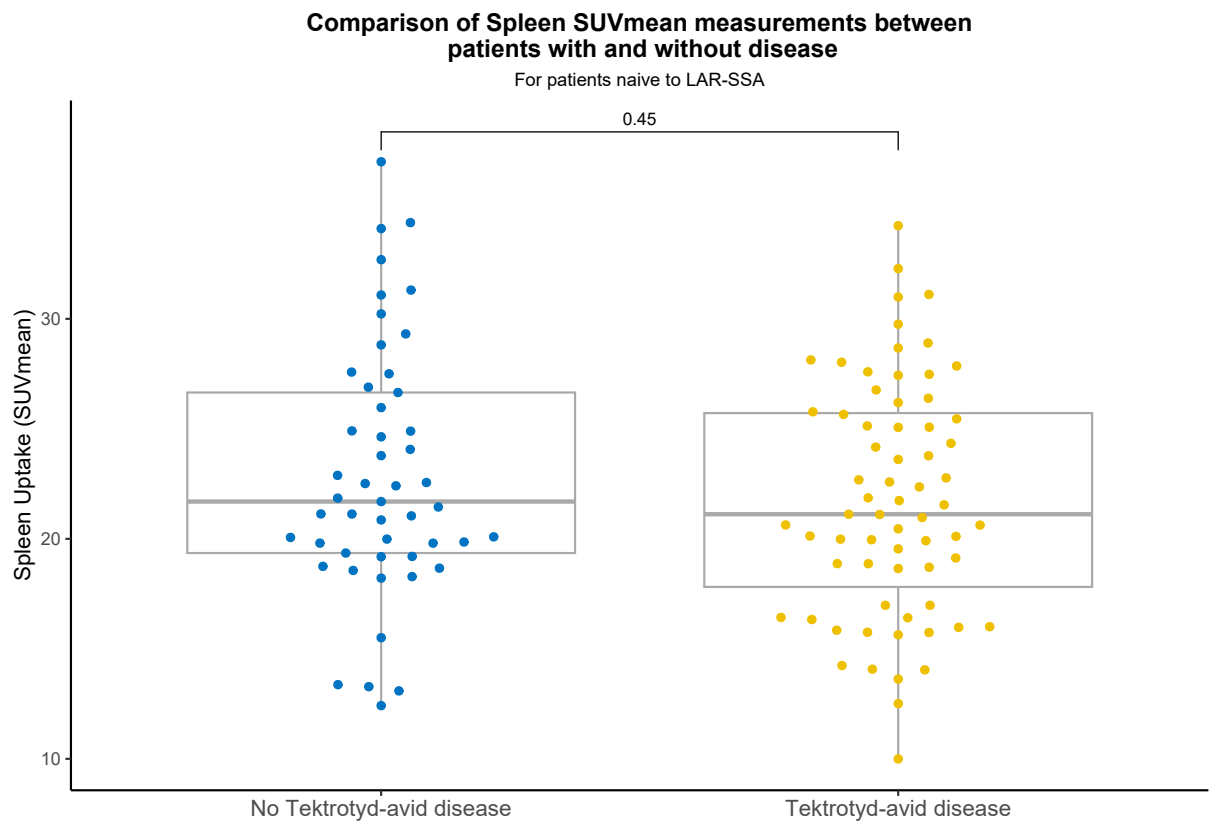


Figure 12.1: Measurement of ^{99m}Tc -EDDA/HYNIC-TOC SUV_{mean} in the spleen for LAR-SSA naive patients, comparing those avid and no disease.

higher tumour burden (18.9 and 15.8 respectively) was not statistically different, although no solid conclusions can be drawn from the small sample size.

For patients on LAR-SSA, 2 of 51 patients had tumour burden > 550 ml (median burden = 64.0 ml; IQR = 98.3). The average SUV_{mean} for burden < 550 ml was 7.9 ± 3.6 , with the two cases of higher tumour burden having SUV_{mean} of 5.2 and 13.7 respectively. This was again not statistically significant.

Analysis of splenic uptake (SUV_{mean}) by correlation with patient body habitus showed a weak, but statistically significant positive correlation for 4 out of 5 metrics for LAR-naïve patients, but only for BMI in the patients receiving LAR-SSA (Table 12.1).

12.4 Discussion

A limited number of studies have investigated TSE in SRI: the most-cited article at time of writing, by Beauregard et al, was limited to 10 patients using ^{68}Ga -DOTA-TATE and only showed significant correlation between tumour volume and uptake of the kidney [362]. The use of uptake in the kidney to define the presence of a TSE has disadvantages as the

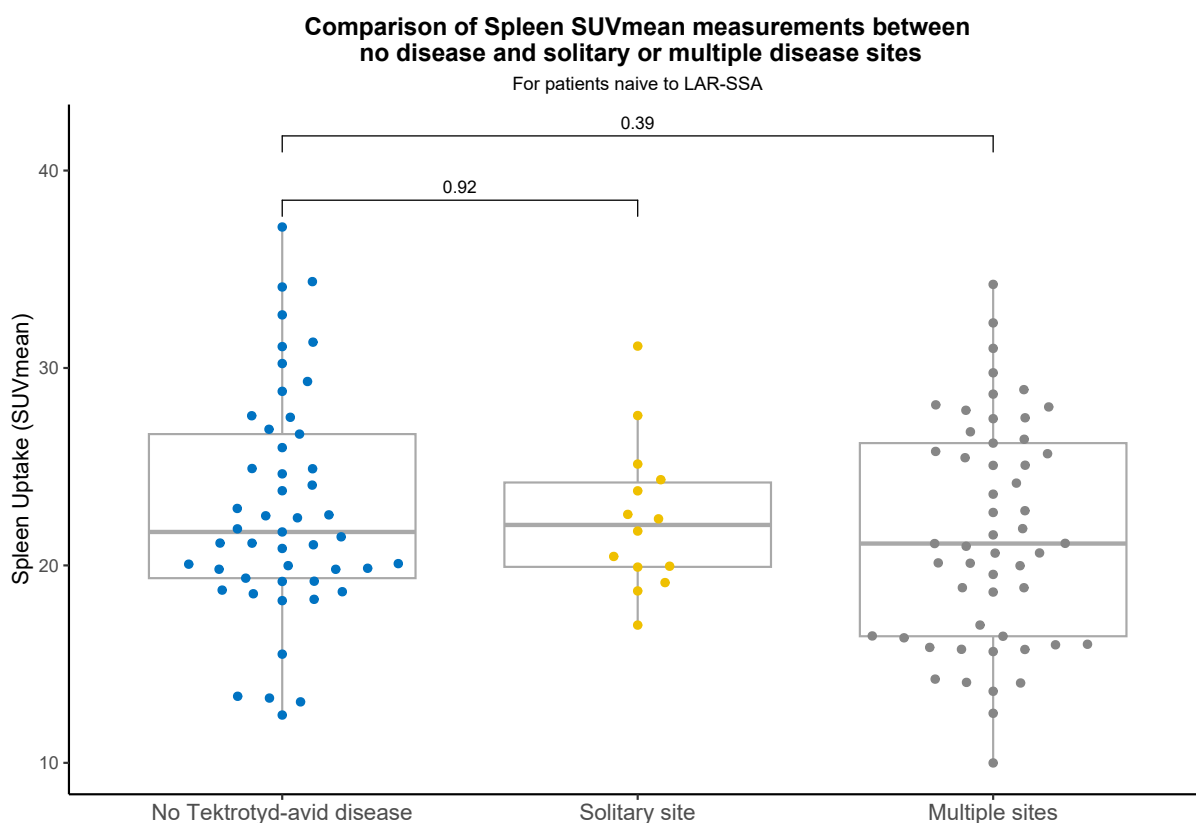


Figure 12.2: Measurement of ^{99m}Tc -EDDA/HYNIC-TOC SUV_{mean} in the spleen for LAR-SSA naïve patients, comparing those with no disease to the presence of solitary or widespread disease.

renal tract is the primary route of excretion in SRI radiopharmaceuticals [171]. This means quantitative measurements will be variable with time as the radiopharmaceutical is processed by the kidneys. There will also be inter-patient differences due to natural and pathological variations in hydration and kidney function. Therefore this present study has focussed on the spleen as a more consistent measure of physiological uptake. Unlike the kidney, there is no physiological process to suggest splenic uptake will vary over short time frames and uptake is less likely to be affected by disease (as compared to the liver, for example, where metastases are common).

Whilst Beauregard et al did describe a negative correlation between tumour volume and splenic uptake, it did not reach statistical significance [362]. There is evidence that splenic uptake is affected by the administration of LAR-SSA (see chapter 9). Therefore analysis was split into two separate groups in this study: LAR-SSA naïve and those on LAR-SSA treatment. Beauregard et al specified that their patients were asked to halt any somatostatin analogue treatment for four weeks [362]: given blood serum levels have been reported to be above clinically significant levels at 56 days in normal volunteers [366], it is unclear if the four week time period would be sufficient to see a significant decrease in blood serum levels.

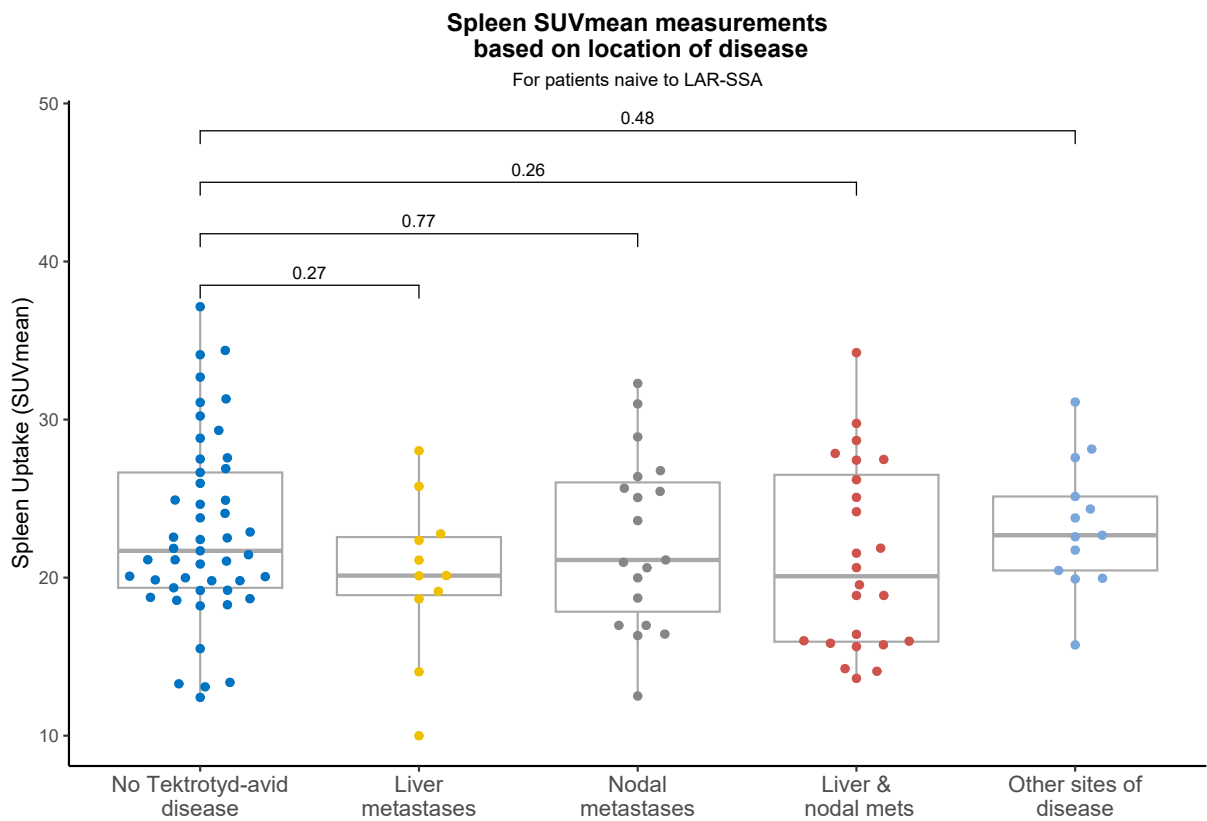


Figure 12.3: Measurement of ^{99m}Tc -EDDA/HYNIC-TOC SUV_{mean} in the spleen for LAR-SSA naive patients, comparing those with no disease to different locations of avid disease.

Another study using ^{68}Ga -DOTA-TOC to examine tumour and physiological uptake briefly mentions the presence of TSE, as defined as 24% of their scans with metastatic disease showing SUV_{max} lower than their measured mean physiological uptake in spleen [175]. It is unclear exactly how this demonstrates a TSE, as it is likely a not insignificant fraction of patients could have disease uptake lower than the physiological mean.

A study simulating uptake of ^{68}Ga -DOTA-TATE using a physiological-based pharmacokinetic model predicted that a notable TSE of >20% decrease in spleen SUV_{max} or SUV_{peak} would only be appreciable for total tumour burdens in excess of 550 ml [363]. The limited number of patients with total tumour volume >550 ml in this study (only two for each of the LAR-naive and LAR-treated groups: 2.9% and 3.9% respectively) means direct comparison is of limited value. However, given an average tumour burden per patient is reported as 69.2 ml [367], even the modelled decrease in splenic uptake of 3.7%-12.7% for tumour burdens under <200 ml would likely only be seen at the lower end in clinical practice (the mean of total TV in this study was 85.4 ml and 123.4 ml for LAR-naïve and LAR-treated respectively, albeit for a limited subset of lesions). The use of spleen SUV_{max} or SUV_{peak} may also limit applicability of this simulation to the clinical setting, as for physiological uptake measurement SUV_{mean} may be considered more appropriate.

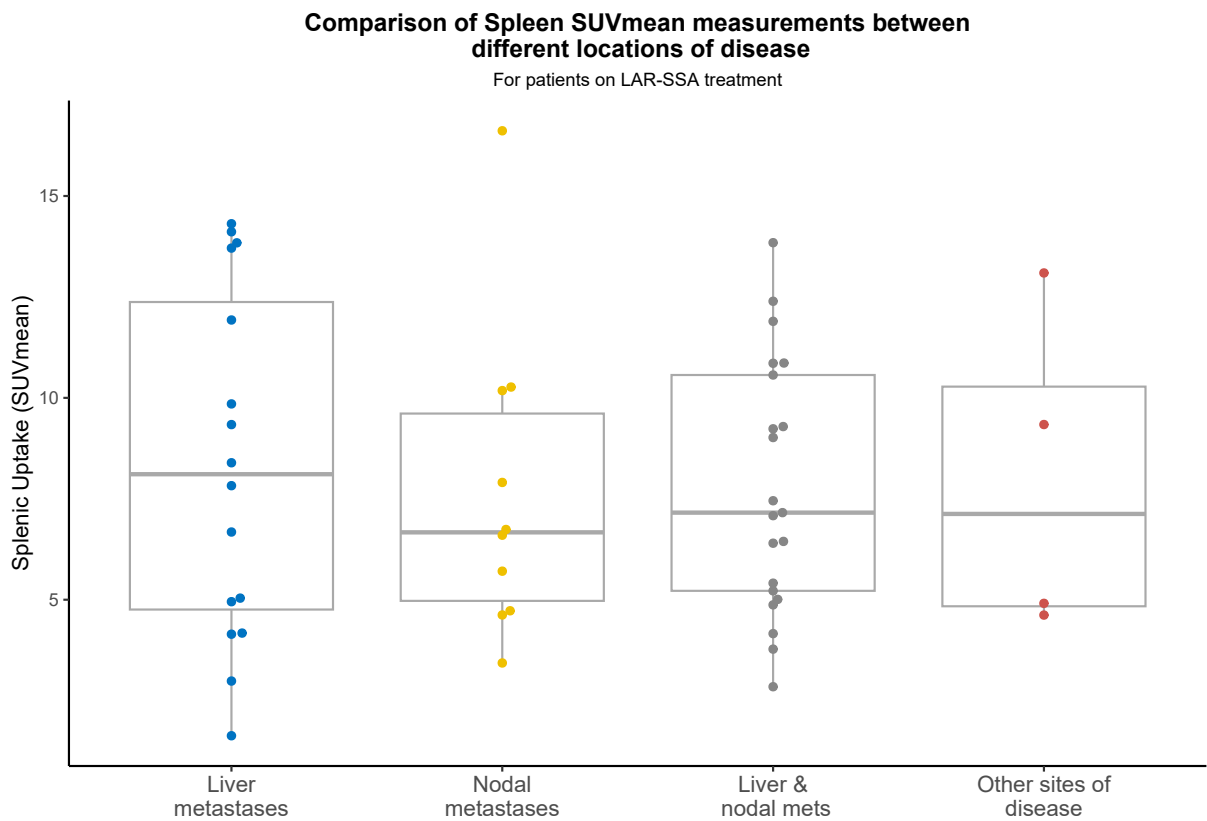


Figure 12.4: Measurement of ^{99m}Tc -EDDA/HYNIC-TOC SUV_{mean} in the spleen for patients on LAR-SSA treatment, comparing different locations of avid disease.

It is hypothesised by Werner et al that the higher SSTR affinity of ^{68}Ga -DOTA-TATE may be responsible for the instances of TSE previously reported, as their study of ^{68}Ga -DOTA-TOC showed no evidence of correlation between fractional uptake by tumours and fractional uptake in physiological organs [254]. The reported lower affinity of ^{99m}Tc -EDDA/HYNIC-TOC [176] could also explain the lack of findings for a TSE in this study. Werner et al did identify a positive correlation between an increase in body habitus and increasing SUV_{mean} in the spleen [254], which this study also demonstrated for weight and BMI (Table 12.1). This finding could be a reason for differences in physiological uptake reported elsewhere.

A limitation of this study is that it measured uptake for up to only five tumour lesions for each patient, whilst it has been proposed that the total volume of disease may affect the TSE [362, 363]. The total tumour volume was not measured as there was not suitable automatic software available to delineate all lesions in a scan. Manual delineation of every lesion would have been excessively time consuming, and a visual assessment of total tumour burden would have required arbitrary thresholds and been subject to inter-observer variation. Therefore, these methods were not considered suitable for analysis in this study. Instead, lesion selection per patient relied on the PERCIST criteria [4], which involves selection of a limited number of lesions (up to 5 per patient) and which has been validated in PET imaging.

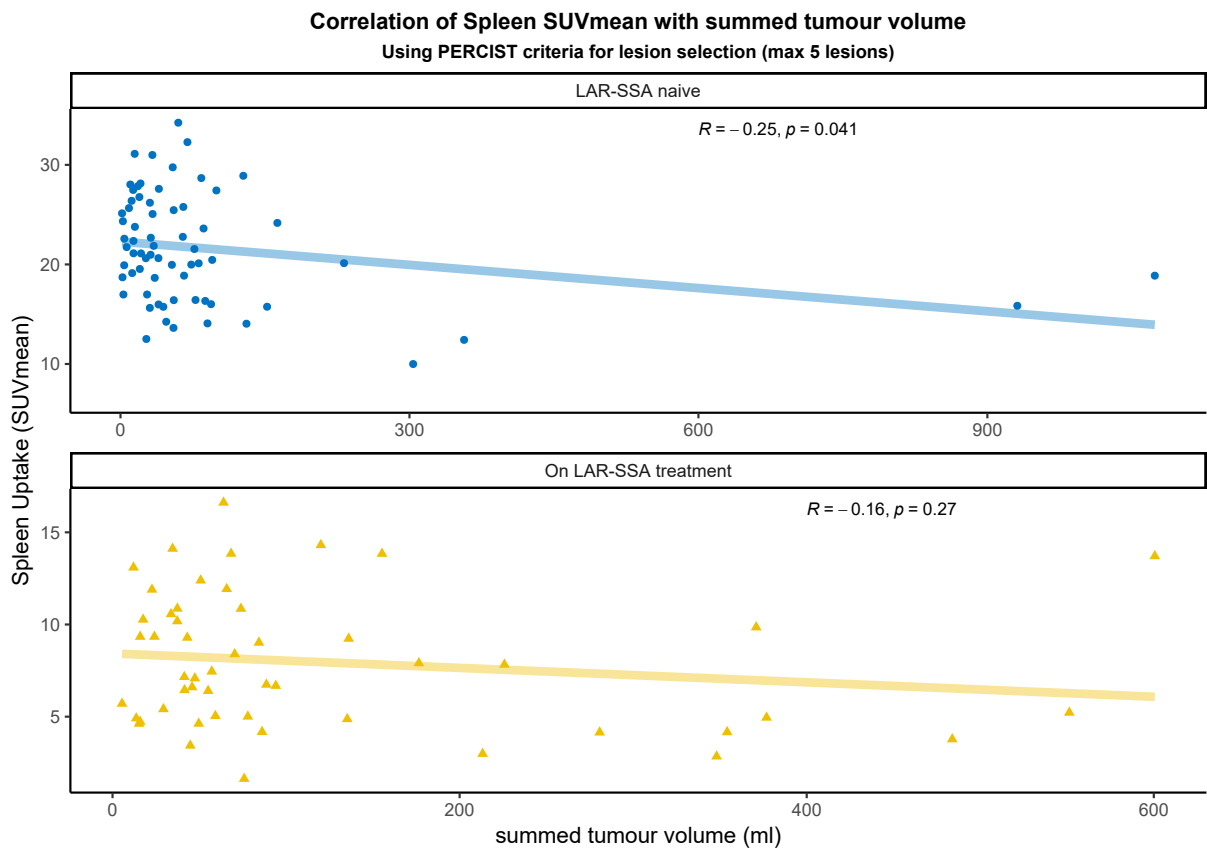


Figure 12.5: Correlation of the summed Tumour Volume (max. 5 lesions, as per PERCIST) with the SUV_{mean} in the spleen, for both LAR-SSA naive and patients on LAR-SSA treatment.

Only LAR-SSA naive patients for summed tumour volume showed a significant correlation with splenic SUV_{mean} (though this was only a weak correlation). The presence of multiple sites of disease compared with solitary sites of disease was also used as a proxy of overall tumour burden for analysis of LAR-SSA-naive patients, with no significant difference between the two. It is expected that the PERCIST method will significantly underweight the total tumour burden for patients with high numbers of metastases. Further study of this area, preferably with automated segmentation of the total tumour burden and with more patients demonstrating high total tumour burden than were included in this study, would be of value in evaluating the true likelihood of a TSE in ^{99m}Tc -EDDA/HYNIC-TOC imaging.

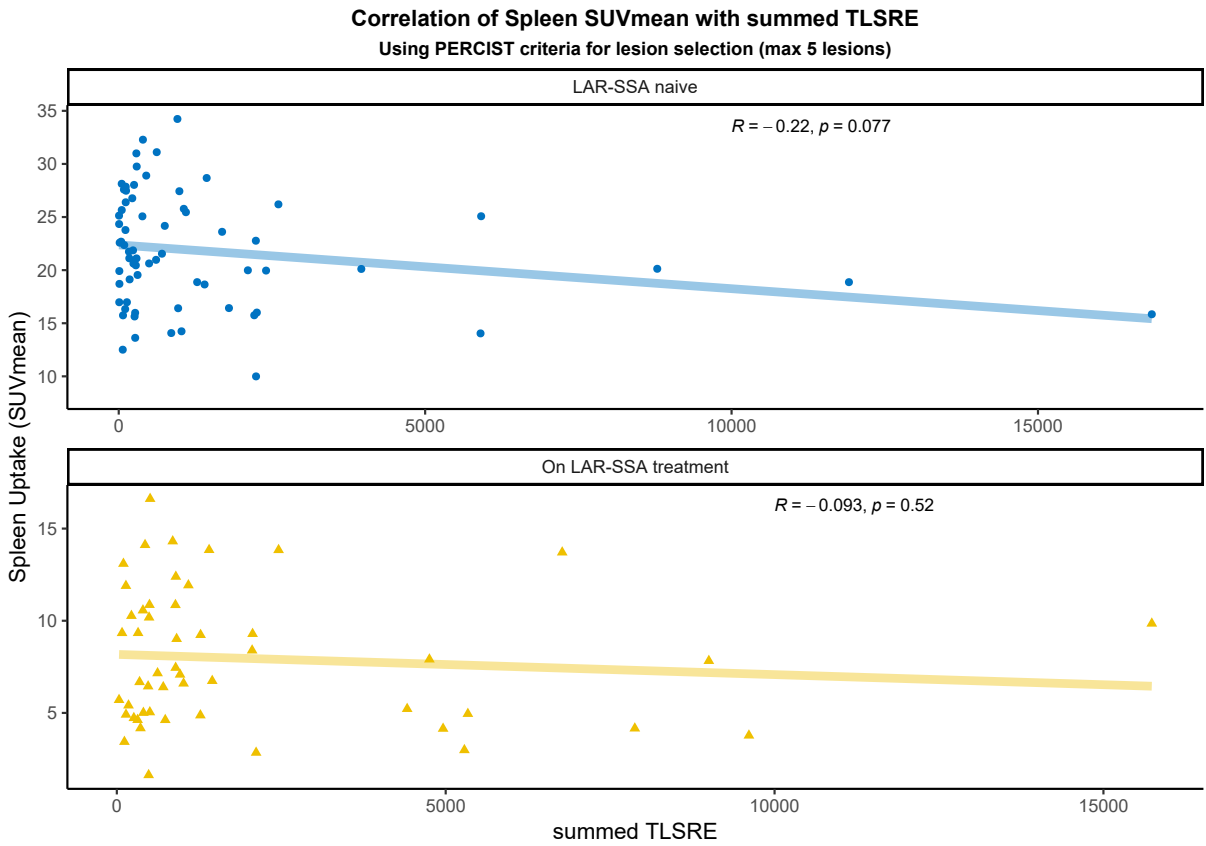


Figure 12.6: Correlation of the summed Total Lesion Somatostatin Receptor Expression (max. 5 lesions, as per PERCIST) with the SUV_{mean} in the spleen, for both LAR-SSA naive and patients on LAR-SSA treatment.

LAR-SSA status	measurement	Pearson correlation R	p-value
Not on LAR-SSA	Weight	0.250	0.006
	BMI	0.290	0.002
	BSA	0.230	0.013
	LBM (James)	0.170	0.071
	LBM (Janma)	0.220	0.017
on LAR-SSA	Weight	0.230	0.099
	BMI	0.310	0.024
	BSA	0.170	0.230
	LBM (James)	-0.009	0.950
	LBM (Janma)	0.015	0.920

Table 12.1: Correlation and significance of various body habitus measurements with splenic uptake measured by SUV_{mean} .

12.5 Conclusion and relevance for the thesis

This study did not find conclusive evidence of a TSE for ^{99m}Tc -EDDA/HYNIC-TOC, either for patients naïve to, or receiving LAR-SSA treatment. This finding may be due to lower receptor affinity than ^{68}Ga -DOTA-TATE, where a TSE has been described. Alternatively, the total tumour burden per patient may have been lower than where a significant TSE could be expected to be detected. With conflicting reports in the literature, it is apparent that further investigation is necessary to determine if TSE truly exists for SRI.

Chapter 13

Post-therapy quantitative SPECT imaging as a predictor of PRRT response

13.1 Introduction

The previous chapters on clinical implementation of quantitative SPECT in SRI have focussed on diagnostic imaging, and specifically on ^{99m}Tc -EDDA/HYNIC-TOC. This chapter will turn to post-therapy imaging of ^{177}Lu -DOTATATE, a radiopharmaceutical used to provide Peptide Receptor Radionuclide Therapy (PRRT). PRRT is an increasingly common treatment option for patients with progressive NETs, with a 180% increase in UK fractions delivered between 2018 and 2021 [368], and a number of recent clinical trials.

PRRT with ^{177}Lu -DOTATATE has proven a relatively safe and effective treatment over a prolonged period [197, 369]. However, it has been reported that 20-30% of patients will progress during the course of PRRT treatment and another 10% in the year following treatment [216]. A panel of experts classified response assessment following PRRT as "challenging" [370]. Therefore, there is merit in investigating post-therapy quantitative SPECT to determine if it may add useful information.

For this study, it is hypothesised that post-therapy imaging with quantitative SPECT provides new imaging features (either quantitative or textural analysis) which may allow separation of patients into positive and poor responders. This differentiation would allow the opportunity to personalise patient's treatments, or divert patients to a more suitable alternative treatment. This hypothesis will be investigated in a population of PRRT patients who were retrospectively reviewed at least two years after their final PRRT fraction.

The NETTER-1 trial originally showed a benefit to PFS of PRRT [225], with an updated publication describing the effect on OS [204]. Issues with PFS arise due to the nature of measurement of progression, which is dependent on the timing of assessments (either diagnostic tests or clinical assessment) [371]. A review article describes several studies that have shown PFS may not be suitable as an OS surrogate in solid tumours [372]: for example in cancers where survival post-progression is long (as applies to NETs), then PFS and OS are not well correlated [373]. For this study, PFS was discounted as a suitable endpoint due primarily to the nature of the PRRT service: a nationally-delivered service with centralised

treatment, but where patient follow-up is conducted locally. It has been noted that there are prominent local variations in both timing and nature of follow-ups (e.g. choice of CT or SRI as imaging modality), which makes collation of PFS and image-based response data difficult, and comparison of the data unreliable. Therefore OS will be the primary measure for this study.

13.2 Materials and Methods

13.2.1 Patients

Consecutive patients who underwent PRRT with ^{177}Lu -DOTATATE within our centre who also had quantitative SPECT imaging performed after (at least) the first fraction were included in this retrospective study: 77 in total, across a 38 month period (January 2020 to April 2023). Of those patients, 63 completed the four prescribed fractions. Due to camera breakdown or patient illness, only 61 of the patients had quantitative imaging for their fourth fraction. This study will focus on quantitative SPECT imaging after their first and fourth fractions.

Due to the retrospective nature of this study, the pre-treatment of these patients is not recorded (patients were referred from across Scotland, including external health boards), hence may have been subject to inter-hospital variations in preferred treatment pathways. A number of clinical factors were recorded to assess if they were factors in survival: including tumour type, tumour grade, the presence or absence of bone metastases and the presence or absence of carcinoid syndrome.

13.2.2 Measurement of survival

As discussed in Section 13.1 above, OS was the chosen measure of response in this study. This choice is due both to the retrospective nature of the study, as well as decentralised follow-up of the PRRT patients, with corresponding differences in protocols.

Data on survival was gathered from clinical information systems for patients within the tertiary referral area of our cancer centre. For other patients, the local Clinical Nurse Specialist provided information of the last contact point with NHS services. If the patient was still alive at time of response assessment then the last contact point served as date of censoring. Otherwise the date of death was used to calculate OS time (all causes) from the start of PRRT treatment. The minimum period of follow-up was 23 months, with a median of 45 months.

13.2.3 Imaging

Imaging was performed approximately 24-hours post-PRRT administration. All SPECT-CT scans were performed on a Siemens Intevo Bold, with appropriate BroadQuant calibration for ^{177}Lu (see Chapter 6). A SPECT-CT was acquired over 3 Fields of View from vertex to mid-thigh, and parameters consisted of 64 views (128 projections), with an ADS orbit, and each view being 5 seconds. The collimator used was MELP, and the energy windows included both 113 keV and 208 keV photopeaks, as evaluated in Section 7.2. Reconstruction was performed with the OSCG algorithm, using the parameters previously confirmed as optimal for quantitative analysis in Section 7.2.8: 72 iterations with 1 subset and 0 mm post-filter.

13.2.4 Image Analysis

Segmentation of tumour and physiological uptake was conducted in LifeX using the method described in Chapter 9 and "Materials and Methods" section of [364], with spherical VOI or 40% thresholding of SUV_{max} .

In a pilot evaluation of image analysis with LifeX, the maximum measured SUV_{max} in tumour lesions for ^{177}Lu -DOTATATE was higher than seen previously in $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC (compared with the measurements reported in Chapter 9). Therefore, for discretisation of data for textural analysis, an extended range of 0-40 SUV was used, with the number of bins increased to 128 to retain a fixed bin width of 0.3 SUV/bin. Although this choice meant some tumour values will still exceed the maximum, the broad distribution of SUV_{max} (range 0.6 to 121.4) means appropriate bounds are impossible to set whilst maintaining a fixed bin width in line with recommendations [88]. The alternatives result in either an excessive number of bins (~ 100 for bin width 1, >300 for bin width 0.3) which will result in sparsely populated data, or an excessive bin width (e.g. exceeding 1.5 SUV/bin). Studies using ^{68}Ga -DOTATATE have used wider bin widths of 1 [374] or 0.95 [257], but are expected to have tumour SUV_{max} more in line with $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC (as another diagnostic test) than ^{177}Lu -DOTATATE, for which there are not similar comparative studies in the literature.

For patients who had both 1st and 4th fraction imaging, the differences in feature value for each lesion were measured, using the absolute difference. These are denoted as Δ in the rest of this chapter.

13.2.5 Feature reduction and selection

The list of robust features for ^{177}Lu detailed in Table 8.5 was used as a starting point for feature reduction. These 33 features were cross-correlated for the 346 tumour lesions identified and segmented in the first fraction images. In a similar process to that outlined in Section

11.2.2, feature pairs with $|\rho| > 0.95$ on Spearman's Rank Correlation were reviewed, with redundant features selected for removal where possible. This was applied to both the 1st fraction imaging analysis, as well as to the differences between 1st and 4th fraction images.

Testing of the features for clinical utility used the Cox Proportional Hazards model with univariate and multivariable analysis. The number of features included in any multivariable model was constrained by the requirement for sufficient number of events to have occurred in the patient population (a "rule of thumb" being 10 events per feature [375]). Models were compared using the AIC (adjusted version to account for small datasets), where the lowest values represented the best fitting model [376]. To further develop the clinical utility of the significant features, partitioning into two or three subgroups was investigated (i.e. partitioning into low/high or low/intermediate/high risk categories). This partitioning utilised statistical testing of the optimum thresholds using the R package OptSurvCutR [137]. The model preference for 0, 1 or 2 cutpoints was tested, using the adjusted AIC as the metric to be optimised, before the optimal threshold was determined (using bootstrapping with multiple iterations to refine selection of the optimal cutpoint).

Once suitable features had been identified, with cutpoints and models developed, these were then tested on a per-patient basis, to better represent the clinical problem, rather than on a per-lesion basis. This testing involved partitioning based on the presence of at least one high-risk lesion per patient, with further investigation using Cox survival analysis and displayed using Kaplan-Meier curves.

13.3 Results

13.3.1 Patient demographics

Patient demographics, as displayed in Table 13.1, are similar to the NETTER-1 study for average age and sex [225], though NETTER-1 had a higher percentage of patients with tumour grade 1/low Ki-67 proliferation index, at 66% compared to 47% for this study. The percentage of patients completing the full four fractions (81.8%) was not significantly different from the completion percentage in the NETTER-1 study (75.7%, $p=0.316$) [225]. By the time of response assessment in this study, 31 of the patients had died.

This study incorporated pancreatic primaries (as SMC approval covered the GEP-NET population), whilst NETTER-1 was limited to midgut only [225]. A small number of lung carcinoid patients were also included in this study. Although not part of the NETTER-1 trial, a number of studies have investigated the use of PRRT for lung/bronchial carcinoids [369, 377] and found comparable survival rates. Although the median OS was shorter for lung carcinoid in this study (Table 13.2), it was not significantly different from the GEP cases

(Figure 13.1). Given the findings of the other studies of similar median OS which included a larger population of lung carcinoids and hence will be more representative, it was felt that lung carcinoid could be included in the population to be analysed in this study.

Three clinical factors pertinent to the patient's condition (taken from the time of their first fraction), are compared to the OS probability in Figure 13.2. Both tumour grade and the presence of carcinoid syndrome are shown to make no difference to the survival curves. For bone metastases, there is a visible difference between the survival curves, but it does not reach statistical significance. As a result, these clinical factors will not be incorporated into predictive models in this study.

Category	
Sex	
Male	43 (56%)
Female	34 (44%)
Age at 1st Fraction	66.1 $\pm$ 11.5
Tumour type	
GEP	71 (92%)
Lung	6 (8%)
Tumour grade	
1	36 (47%)
2	39 (51%)
3	1 (<1%)
Unknown	1 (<1%)
Carcinoid syndrome	
Syndromic	34 (44%)
Non-syndromic	40 (52%)
Unknown	3 (4%)

Table 13.1: Demographics for patients undergoing PRRT and quantitative SPECT imaging.

Tumour type	Study	Median OS (months)	95% CI
GEP	This study	54.8	37.2 - inf.
Lung	This study	14.8	11.1 - inf.
GE	NETTER-1 trial [225]	48.0	37.4 - 55.2
Bronchial	Brabander, et al [369]	52	49 - 55

Table 13.2: Comparison of median Overall Survival (OS) following PRRT between this retrospective study and other studies in the literature.

13.3.2 Inter-fraction differences

Figure 13.3 demonstrates changes in the SUV_{mean} between fractions 1 and 4. All three tumour groupings studied (liver metastases, nodal metastases and pancreatic lesions)

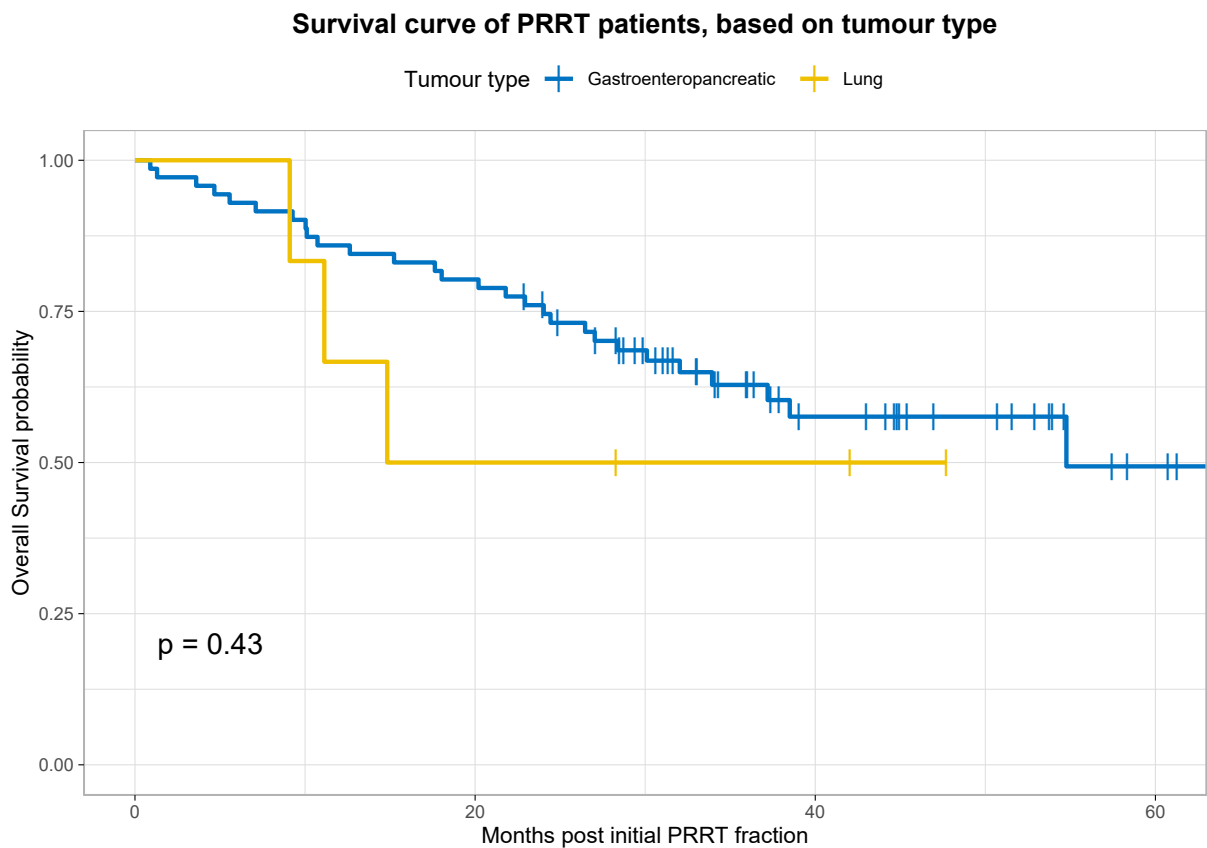


Figure 13.1: Kaplan-Meier Survival curve for PRRT patients, differentiated by tumour type.

showed significantly higher mean uptake ($p < 0.001$) at the first fraction than at the fourth fraction. Whilst measurement of SUV_{mean} in physiological Blood Pool was not significantly different between the fractions, the splenic uptake was significantly higher in Fraction 4 (the inverse of the tumour uptake).

Whilst these differences could be used to infer a drop in tumour uptake between Fraction 1 and Fraction 4, it is helpful to confirm this by looking at pair-wise data for individual lesions. The graphs in Figure 13.4 plot the SUV_{mean} in Fraction 4 against the SUV_{mean} in Fraction 1: they demonstrate that for the majority of lesions the Fraction 1 SUV_{mean} was higher. When considered with the changes in splenic uptake, where Fraction 4 was generally higher than Fraction 1 (part a of Figure 13.4), it raises the possibility of a TSE being demonstrated (as discussed in Chapter 12). The TSE hypothesis assumes that as less radiopharmaceutical is taken up in the tumours in the later fraction, it becomes more available for physiological uptake, with the most prominent location for this being the spleen. This could potentially be due to reduced receptor sites in the tumours if they become de-differentiated; however, this is unlikely to be occurring in such a large number of patients simultaneously, and is more likely attributable to cell death from the three fractions already administered.

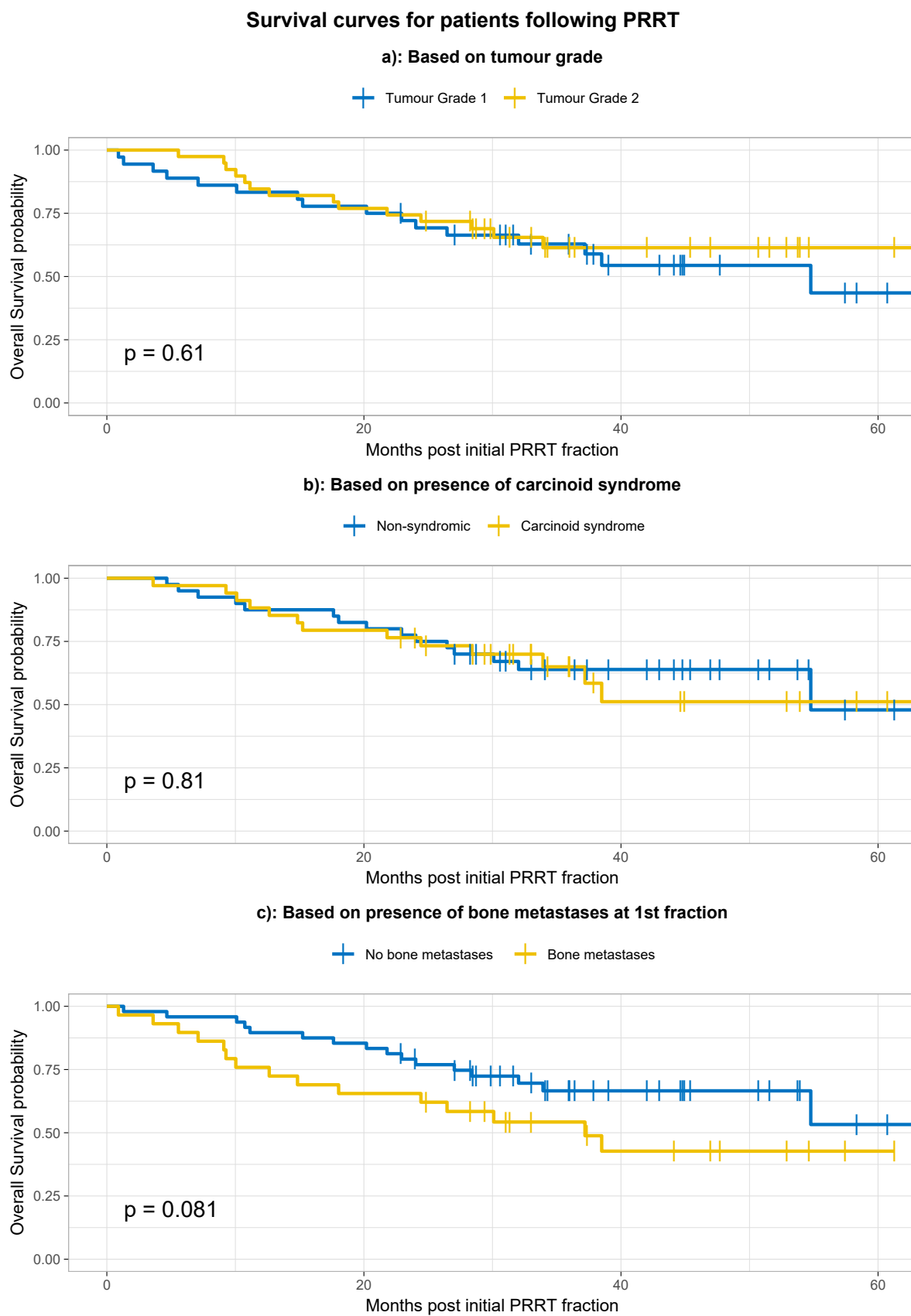


Figure 13.2: Kaplan Meier Survival curves for PRRT patients, based on a) tumour grades 1 and 2; b) presence of carcinoid syndrome; and c) presence of bone metastases on first fraction imaging.

Differences in SUVmean between PRRT fractions

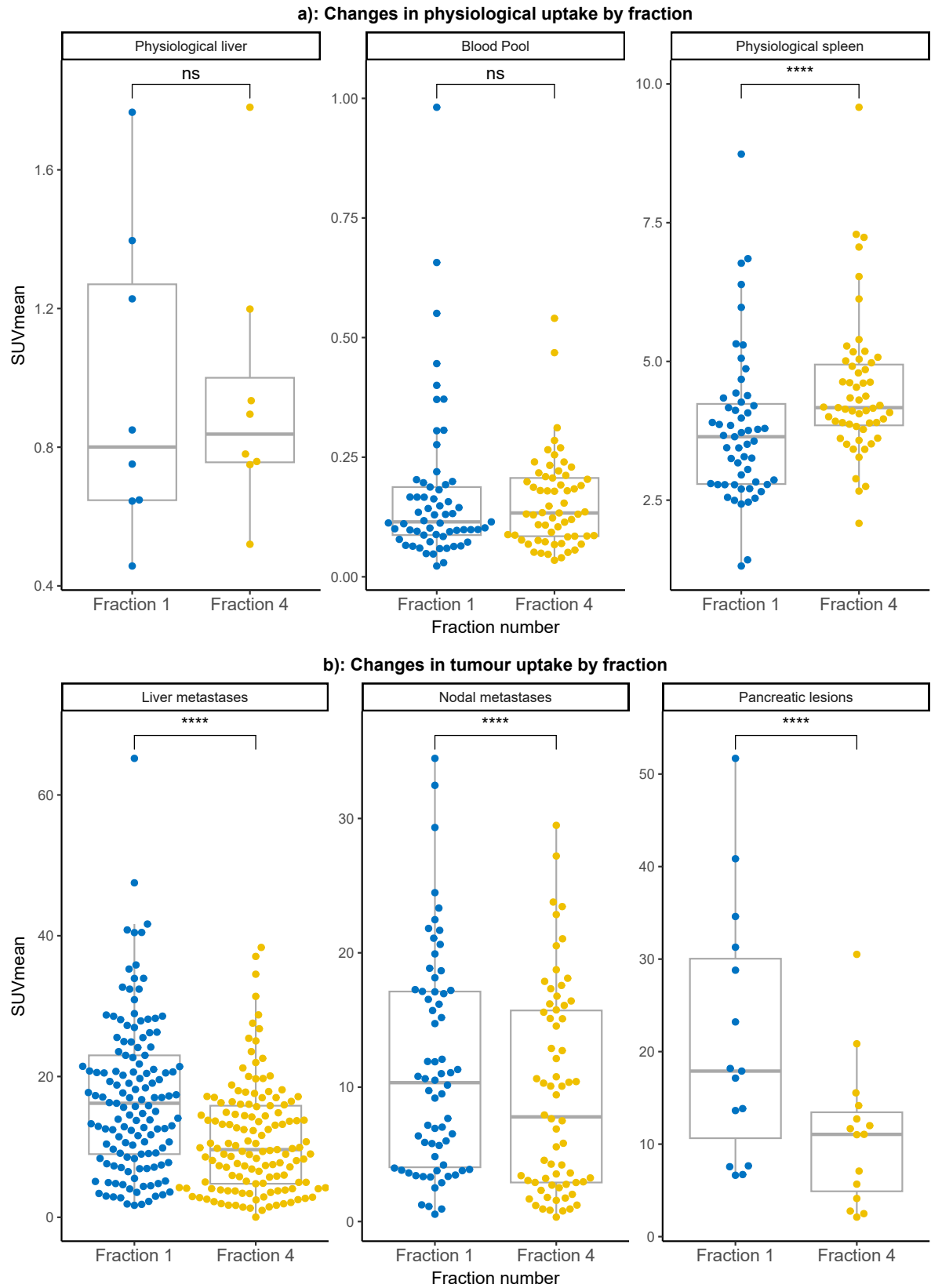


Figure 13.3: Comparison of the average values of SUV_{mean} between PRRT fractions 1 and 4, for both a) physiological uptake and b) three sites of tumour uptake. **** equates to $p < 0.0001$.

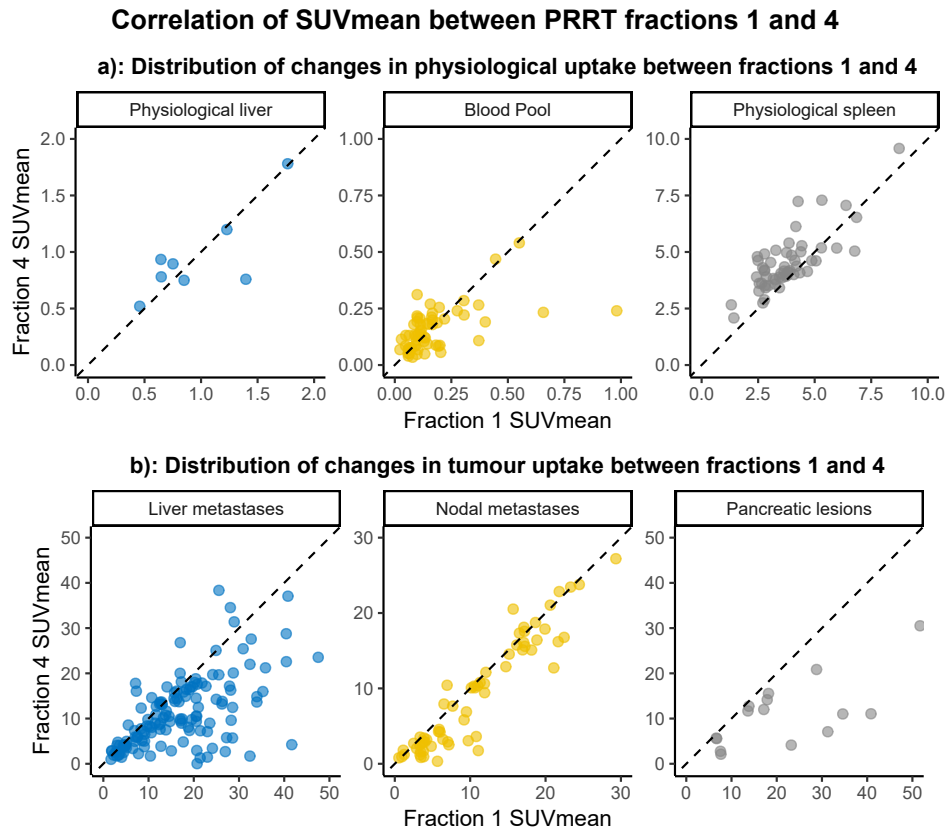


Figure 13.4: The distribution of paired SUV_{mean} between PRRT fractions 1 and 4, for both a) physiological uptake and b) three sites of tumour uptake.

Changes in volume of tumour lesions between Fractions 1 and 4 are demonstrated in Figure 13.5. The plots demonstrate the general trend for the active volume of lesions to have decreased in size between the first and fourth fractions. However, as the lesions were demarcated by means of an SUV threshold of 40% of the maximum value in the lesion, it is possible that change in the uptake between fractions could affect the volume that is segmented, rather than demonstrating shrinkage of the tumour volume.

13.3.3 Feature reduction

The initial set of quantitative and textural analysis features was taken from the work performed in Chapter 8 and listed in Table 8.5. A correlation matrix was produced, as shown in Figure 13.6, with the features showing correlations > 0.95 investigated further. This led to the following features being either removed for redundancy or retained:

- A number of the SUV features were strongly correlated (with the exception of SUV_{min}). SUV_{mean} , SUV_{max} and $SUV_{peak\ 0.5ml}$ were retained, in order to determine if a particular SUV metric was more predictive than the others.

- Entropy and Energy features were strongly correlated, as was Homogeneity. Entropy had the least number of correlations with other features so was retained (to remain consistent with Chapter 11, GLCM Entropy Log2 was retained).
- A number of features showed close correlations with each other (Homogeneity, GLCM Dissimilarity, GLRLM SRE, GLRLM RP, GLZLM ZP and NGLDM Contrast). NGLDM Contrast had no other strong correlations if the former factors were removed, so was chosen to be the sole feature to be retained.

In total, 20 features were removed for redundancy, resulting in 13 features being retained for further analysis.

13.3.4 Univariate and multivariable analysis

For the 1st fraction data, univariate analysis per lesion produced the results shown in Table 13.3. There are three features showing major variation from Hazard Ratio of 1, plus a p-score significant at the 95% level: Sphericity, GLCM Correlation and NGLDM Contrast. Whilst multivariable models consisting of the three features were also tested for completeness and did have significant p-values, their AIC did not differ significantly from those for the univariate analysis, hence the simpler models were preferred. On testing of partitioning, all features showed substantial evidence for dichotomisation (one cutpoint) over the default of no partitioning. A number of the features showed a slight preference for three groupings (two cutpoints), but the difference in AIC was minimal and therefore the simpler dichotomous model was selected in all cases.

Feature	Hazard Ratio	95% CI	p-value
SUVmin	0.97	0.95-1.00	0.034
SUVmean	0.98	0.97-1.00	0.031
SUVmax	0.99	0.98-1.00	0.053
SUVpeak0.5	0.99	0.98-1.00	0.062
TLSRE	1	1.00-1.00	0.077
Discretised Histogram Skewness	0.99	0.93-1.06	0.8
Volume	1	1.00-1.00	0.002
Sphericity	0.07	0.02-0.31	<0.001
Surface area	1	1.00-1.00	0.001
GLCM Correlation	6.33	1.20-33.5	0.03
GLCM Entropy log2	0.97	0.90-1.03	0.3
NGLDM Coarseness	0	0.00-inf.	0.2
NGLDM Contrast	0.34	0.15-0.77	0.009

Table 13.3: Results for univariate Cox survival analysis, based on 1st fraction imaging. Features selected for further analysis are in bold text.

model type	features	AIC	Optimal cutpoint
univariate	Sphericity	1498.7	0.794
univariate	GLCM Correlation	1487.8	0.747
univariate	NGLDM Contrast	1501.0	0.155
multivariable	Sphericity & NGLDM Contrast	1495.3	n/a
multivariable	Sphericity, NGLDM Contrast & GLCM Correlation	1478.8	n/a

Table 13.4: Comparison of models by the Akaike Information Criterion (AIC), based on 1st fraction imaging; and identification of the optimal cutpoint following bootstrap testing.

Analysis of the data comprising the differences between Fraction 1 and Fraction 4 did not demonstrate any features with both an important deviation from Hazard Ratio (HR)=1 and a significant p-score (Table 13.5). For example, although $\Delta \text{SUV}_{\min}$ had a HR= 0.97, the 95% confidence interval included 1, and it was not felt this was a significant deviation from the HR=1. Two features which did have HR that differed from 1, though with p-values that bordered the significant p-value threshold of $p=0.05$, were $\Delta \text{GLCM Correlation}$ and $\Delta \text{GLCM Entropy log2}$. The AIC scores for both were similar, so both features were subject of further analysis. Multivariable analysis of a combined model did not show significant p-values or Hazard Ratios. The AIC difference between one- and two-cutpoint models was minimal, hence dichotomisation was preferred.

Feature	Hazard Ratio	95% CI	p-value
$\Delta \text{SUV}_{\min}$	0.97	0.94-1.00	0.028
$\Delta \text{SUV}_{\text{mean}}$	0.98	0.96-1.00	0.094
$\Delta \text{SUV}_{\text{max}}$	0.99	0.98-1.00	0.076
$\Delta \text{SUV}_{\text{peak0.5}}$	0.99	0.97-1.00	0.066
ΔTLSRE	1.00	1.00-1.00	0.089
$\Delta \text{Discretised Histogram Skewness}$	1.04	0.95-1.14	0.4
ΔVolume	1.00	0.99-1.01	>0.9
$\Delta \text{Sphericity}$	0.45	0.02-12.4	0.6
$\Delta \text{Surface area}$	1.00	1.00-1.00	0.3
$\Delta \text{GLCM Correlation}$	0.13	0.01-1.42	0.094
$\Delta \text{GLCM Entropy log2}$	0.90	0.80-1.01	0.073
$\Delta \text{NGLDM Coarseness}$	0.00	0.00-Inf	0.5
$\Delta \text{NGLDM Contrast}$	1.04	0.46-2.35	>0.9

Table 13.5: Results for univariate Cox survival analysis, based on differences between 1st and 4th fraction imaging. Features selected for further analysis are in bold text.

model type	features	AIC	Optimal cutpoint
Univariate	Δ GLCM Correlation	821.9	-0.07
Univariate	Δ GLCM Entropy log2	822.2	-2.75

Table 13.6: Comparison of models by the Akaike Information Criterion (AIC), based on differences between 1st and 4th fraction imaging; and identification of the optimal cutpoint following bootstrap testing.

Results per patient

As there is no practical way to summarise the features values per patient (e.g. averaging of the values of many of the features is meaningless), per-patient analysis had to be conducted after the feature values were partitioned.

For the first fraction images, each lesion was classified as high or low-risk based on each of the three features (Sphericity, GLCM Correlation and NGLDM Contrast). Per-patient analysis was then conducted based on a patient having at least one high-risk lesion. Only sphericity had a significant difference in OS between the groupings ($p=0.013$) when dichotomised (for GLCM Correlation and NGLDM Contrast, the p -values were 0.96 and 0.19 respectively). As every patient had at least one lesion that was flagged as high-risk by at least one of the three features, a different approach was required for combining the features. If all three features agreed (i.e. all high-risk, or all low-risk) this became the lesion risk factor. If one feature was discordant from the other two, then the lesion risk factor was classified as intermediate. This classification was then collated on a per-patient basis, again based on at least one high-risk lesion, using this combined risk factor. A diagram detailing this process is shown in Figure 13.7. The combined risk demonstrated a significant difference between the high and low risk group on survival analysis, with $p=0.0082$ (Figure 13.8). Two-, three- and four-year survival percentages were also calculated and are displayed in Table 13.7.

1st fraction imaging: risk factors	2-year survival	3-year survival	4-year survival
Based on combined risk factors			
High-risk	64% (49%-85%)	45% (29%-69%)	37% (21%-65%)
Low-risk	80% (69%-92%)	72% (60%-86%)	68% (55%-84%)
Based on risk factor using Sphericity			
High-risk	69% (58%-82%)	52% (39%-68%)	44% (31%-63%)
Low-risk	86% (73%-100%)	86% (73%-100%)	86% (73%-100%)

Table 13.7: Survival rates for 1st fraction post-PRRT imaging risk factors. 95% confidence intervals are given in brackets.

1st vs 4th fraction: risk factors	2-year survival	3-year survival	4-year survival
Based on risk factor using Δ GLCM Correlation			
High-risk	87% (71%-100%)	53% (32%-86%)	53% (32%-86%)
Low-risk	91% (84%-100%)	82% (71%-95%)	78% (65%-94%)

Table 13.8: Survival rates for changes in post-PRRT between 1st and 4th fractions, risk factors. 95% confidence intervals are given in brackets.

Similar analysis was performed for the differences between 1st and 4th fraction imaging: only Δ GLCM Correlation showed a significant p-value once analysed on a per-patient basis (Figure 13.9 and Table 13.8).

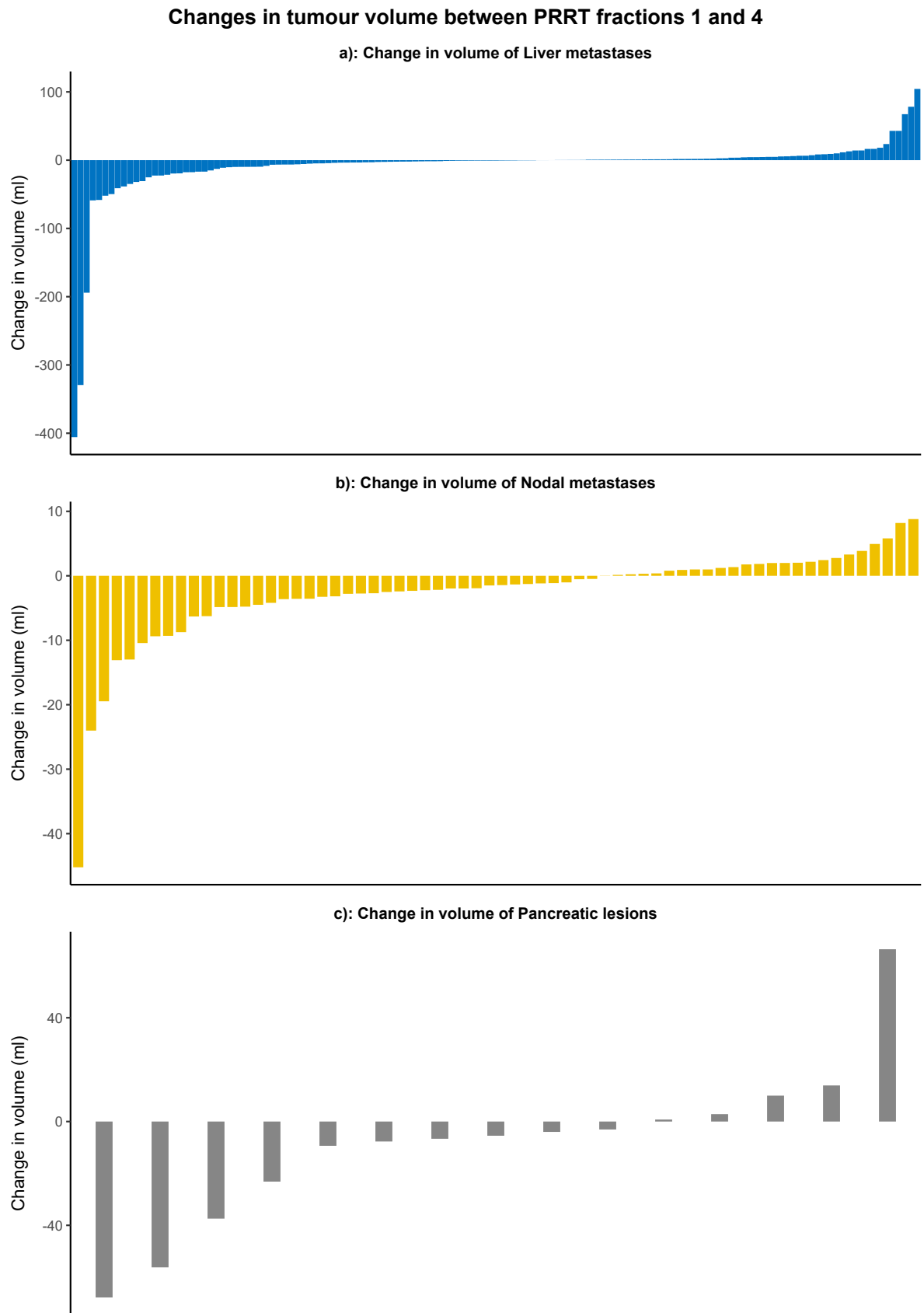


Figure 13.5: Measured changes in the active volume of tumours between fractions 1 and 4 a): for liver metastases, b): for nodal metastases and c): for pancreatic lesions.

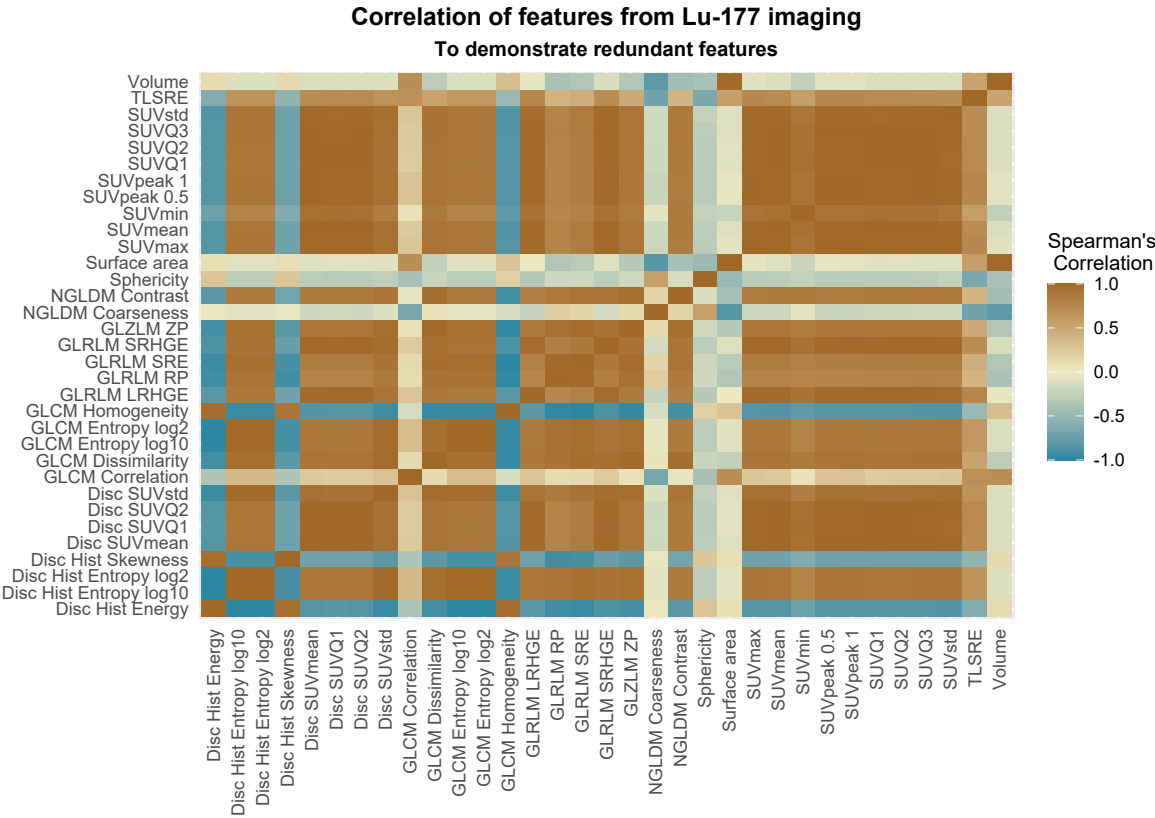


Figure 13.6: Correlation matrix for quantitative & textural analysis features from ¹⁷⁷Lu-DOTATATE images, to determine features that can be removed due to redundancy.

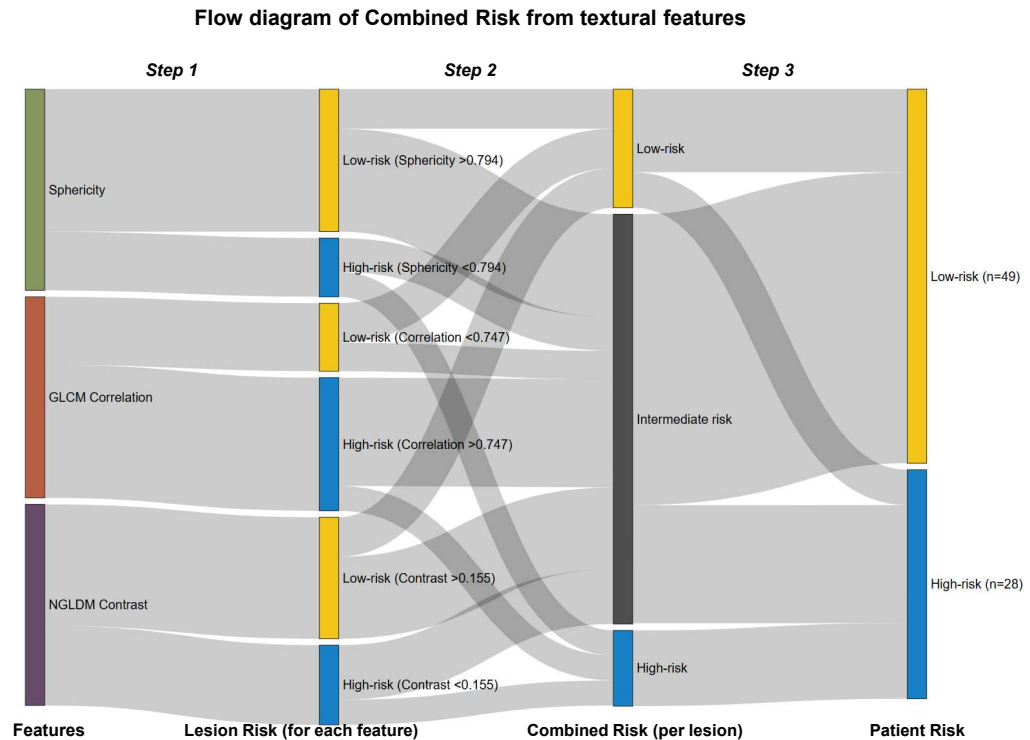


Figure 13.7: A flow diagram of the process of combining the risk factors from three textural analysis features into a dichotomised risk for patients undergoing PRRT. Step 1: Lesions classified by each of the three features, using cut-points detailed in Table 13.4. Step 2: Produce a single risk rating per lesion, by combining the individual factors from the 3 features: if 3/3 agree, then low- or high-risk respectively; mixed risks (e.g. 1/3 low- and 2/3 high-risk, or vice-versa), then lesion is classified as intermediate risk. Step 3: A single risk rating per patient is created: if patient has at least one high-risk lesion, then patient is high-risk; if no high-risk lesions, then low-risk.

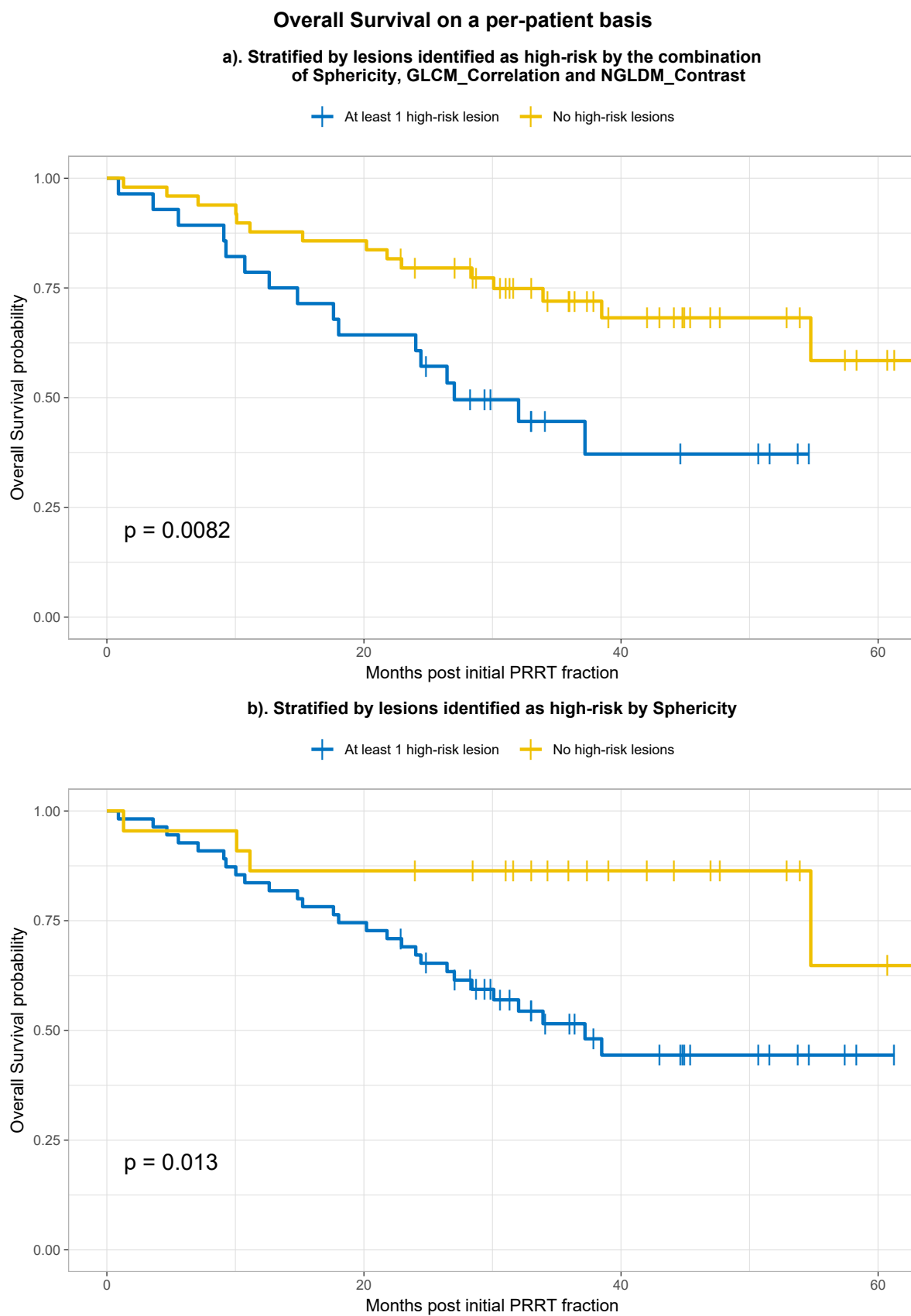


Figure 13.8: Kaplan-Meier curves for Overall Survival per patient, stratified by relevant features on first fraction imaging. a) A combined risk factor, b) risk factor based on dichotomised Sphericity values.

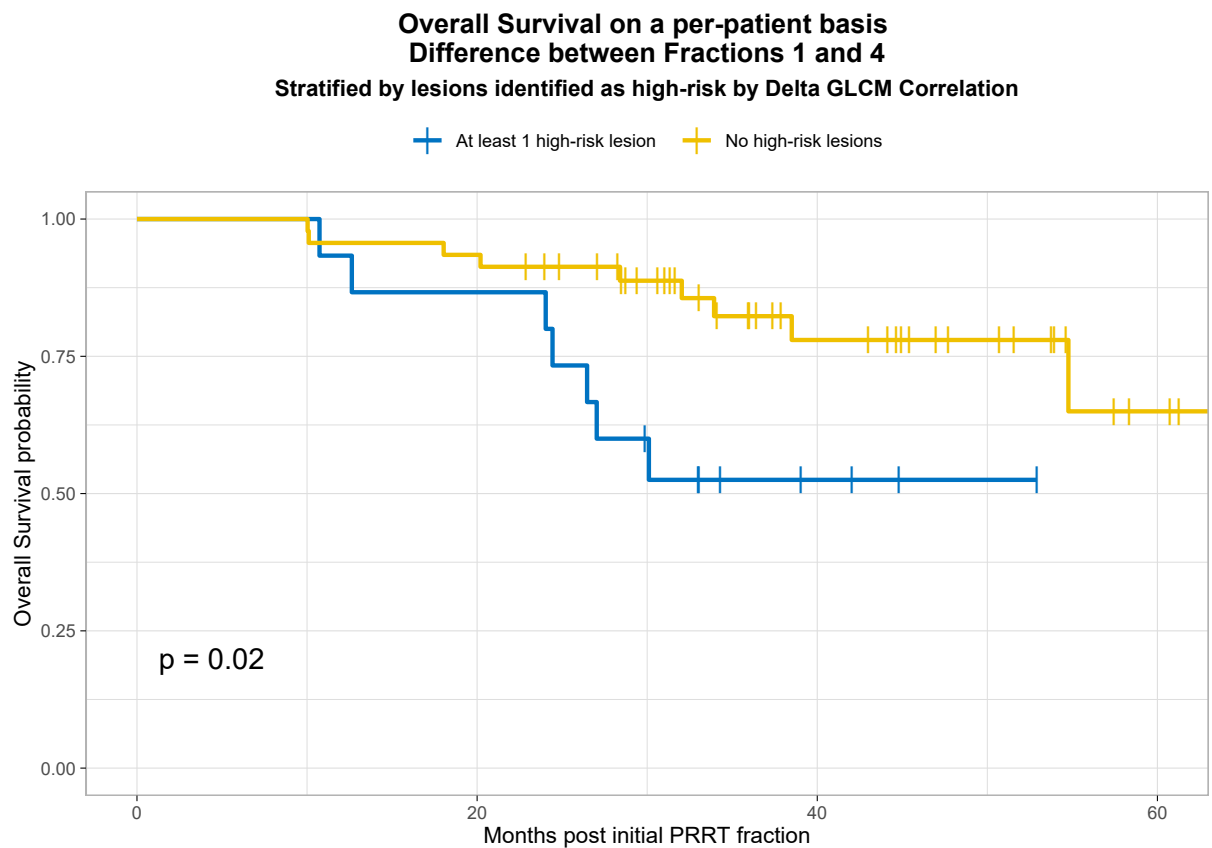


Figure 13.9: Kaplan-Meier curves for Overall Survival per patient, stratified by differences in GLCM Correlation between 1st and 4th fraction, dichotomised into high- and low-risk groups.

13.4 Discussion

It has been reported in dosimetric studies for ^{177}Lu -DOTATATE that tumour absorbed dose is higher in the first fraction than in following fractions [259, 260, 378]. This finding has led to proposals for a "first hit" strategy, where the prescribed activity in the first fraction is increased in order to have greater dosimetric benefit. The findings of this study, demonstrated in Figures 13.3 and 13.4, do suggest this holding true for SUV_{mean} for ^{177}Lu -DOTATATE, albeit our study only compares Fraction 1 with Fraction 4. However, it is in agreement with findings elsewhere; Roth et al measured decreases in activity concentration across a number of fractions (though tumour volume reduced on average, it was not statistically significant) [379].

Prediction of OS from post-therapy scans may have important clinical applications. Those patients adjudged to be high-risk may have more frequent follow-up scheduled (clinical or radiological assessment), with those in the low-risk criteria seen less frequently. Those in the high-risk category may be treated more aggressively if their disease progresses further, and are less likely to be referred for a repeat course of PRRT if it has already proven sub-optimal.

Selection of features from 1st fraction imaging highlighted three potentially useful features: Sphericity, GLCM Correlation and NGLDM Contrast. The sphericity of lesions was found to be a significant factor in predicting PRRT response in Wetz et al [261]. They found 1st fraction lesions with high asphericity (i.e. low sphericity) had a higher rate of non-response, which is mirrored by the findings here (taking into account the inversion of sphericity to asphericity). A search of the literature failed to identify any studies that evaluated the change in quantitative or textural features between PRRT fractions as a predictor of OS. Therefore, there is no external corroboration that textural changes, particularly in GLCM Correlation, can be used to predict survival. The survival rates on a per-patient basis show clear delineation between high- and low-risk groups for 1st fraction imaging with the combined risk and for sphericity; the Δ GLCM Correlation between 1st and 4th patients also demonstrates separation. The Sphericity feature on 1st fraction imaging has the advantage of being observable after the first fraction only and non-overlapping confidence intervals for 3- and 4-year survival, suggesting it may be the optimal predictor.

Outwith the Wetz paper, there have been few studies on predictive value of post-PRRT SPECT. However, there have been numerous studies that have investigated pre-therapy diagnostic imaging with ^{68}Ga -labelled DOTA SRI. These have generally been single-site studies without external validation of the identified prognostic features. Features identified as having value include Entropy, skewness (both [141]), and heterogeneity [254]. Others have determined that SUV features do not correlate well with OS [197], even up to 12 years post-therapy. The use of total tumour burden, as measured by summed SRETV or summed TLSRE has shown promise, with higher values correlating with shorter PFS and OS [355,

380] on pre-therapy imaging, and a decrease in SRETV on post 1st PRRT fraction ^{68}Ga SRI associated with longer PFS [259]. These factors evaluating summed volumes could not be fully assessed in this current study, due to limitations with LifeX software precluding total body segmentation of lesions. Whilst in other parts of this thesis concerned with studies using diagnostic SRI a limited number of lesions were summed to give a summed score, the substantially higher number of lesions seen in most post-PRRT imaging was judged to render this approach unreliable.

The use of cutpoints or thresholds to dichotomise continuous variables when evaluating prognostic or predictive value is generally not recommended by the statistical community. The choice of cutpoint is highly dependent on the sample population used to calculate it [381] and using the same sample for deriving a cutpoint and testing its effect on survival prognosis also adversely biases the p-values produced by the log-rank analysis [375]. However, for developing a method for clinical use, it may be more appropriate to dichotomise the data [382]. Whilst response to PRRT is likely to be on a continuum, the results will inform a clinical decision that is by its nature binary: to treat or not to treat. Thus results are more easily interpretable if they are dichotomised, rather than displayed using a proportional hazard model, as used in Cox regression. The method chosen in this study, firstly for determining optimal number of cutpoints and then identifying their optimal value (by optimising AIC) avoids the downsides demonstrated in the mathematical examples where cutpoints are determined from either median values [383] or arbitrary thresholds [384]. As we do not use the optimal p-score from the outcome to determine our cutpoints, our independent assessment of cutpoints (and further testing with cross-validation via bootstrapping) reduces the possible complaints against the choice to dichotomise. There is also an issue with SPECT data in that it is inherently statistically noisy (and therefore dissimilar to the data in many of the studies from the social sciences arguing against dichotomisation). Dichotomisation may reduce the effect of this noise on outcomes, and in fact it is recognised that in some situations dichotomisation may be legitimate [383].

A limitation of this study is that demographic and medical factors are not built into our multivariable models. This choice was made for two reasons: the nature of the patient population is that it is geographically distributed across Scotland (as the PRRT service is delivered on a national basis) and hence such data (e.g. CgA or the patient's previous treatment regimens) were not accessible for this retrospective study. The second reason is that the limited patient numbers included in the study (constituting the entire population that met inclusion criteria over 3 years) and the generally long survival time for NETs meant the number of events used for survival analysis was low, limiting the number of factors that would be applicable in multivariable models. Further work with larger datasets may allow such multi-factorial models to be investigated.

A further limitation is that the models developed were not tested with independent data from other centres, so cannot be determined to have wider relevance. Local validation was through the use of bootstrapping, as the limited dataset meant splitting the data into testing and validation sets was deemed impractical. Therefore, further work is planned on local testing with a separate dataset.

13.5 Conclusion and relevance for the thesis

This single centre study has demonstrated that post-PRRT quantitative SPECT imaging may provide useful information to predict a patient's likely response to PRRT. The use of textural analysis identified several features that had potential predictive value, both for 1st fraction imaging alone, but also for changes in the feature value for a given tumour between the 1st and 4th fractions. Though dichotomisation of continuous values has drawbacks, the process allowed for stratification into clinically relevant risk groups.

Quantification, therefore, raises the the possibility of altering patient treatment pathways based on imaging results. For those patients with poorer response to PRRT, it may be this still remains the "least worst" option for their treatment, and could be prolonging their survival beyond what could be achieved with other treatments. Investigating this is beyond the scope of this thesis, but would be an interesting area for future clinical trials.

Chapter 14

Patient-reported Quality of Life metrics following PRRT and the predictive value of SPECT quantification

14.1 Introduction

The preceding chapter assessed if quantitative images could be used to predict a patient's survival following PRRT. For patients with NETs, where survival with the disease may be measured in years, continuing quality of life is also an important factor in judging the effectiveness of a treatment. Factors such as the frequency and severity of symptoms, how the patient experiences pain (including sleep disruption) and if the disease prevents them from working or having a social life may all affect patients to varying degrees.

Patient-Reported Outcomes (PROs) is a qualitative method for analysing how a patient perceives the effects of their disease [385]. They are generally measured through the use of a questionnaire, which attempts to assess the state of a patient's health and mental wellbeing, summarised under the umbrella term of Health-Related Quality of Life (HR-QoL). Use of a standardised questionnaire, with a pre-treatment baseline and regular follow-ups, allows for longitudinal monitoring. In cancer care and research, the European Organisation for Research and Treatment in Cancer (EORTC) QLQ-C30 questionnaire has been widely used [386], with additional modules available for specific cancer types (for example, QLQ-GI.NET21 for NETs).

The use of PROs could be considered especially relevant for NETs, given the expected (relatively) long survival times may mean a continuing burden of symptoms that could have significant effect on a patient's quality of life. Studies have shown patients with NETs who are about to start PRRT report lower QoL scores and higher symptom scores than the general population [387, 388]. A systematic review highlighted several studies showing no difference in reported QoL between the standard treatment of (non-radioactive) somatostatin analogue and placebo [389], which contrasts with a recent sub-analysis of the NETTER-1 trial showing that PRRT (using ^{177}Lu -DOTATATE) delivered a longer period before deterioration in several PRO metrics [390].

An early study of the effect of PRRT on HR-QoL found improvement in PROs at 6 weeks after completion of PRRT compared to baseline for overall HR-QoL, as well as fatigue, nausea, pain and diarrhoea [391]. It also found symptoms improved even in patients showing stable disease on CT or MRI [391], suggesting structural imaging may have deficiencies in detecting changes in the patient's disease after PRRT. Only one study was found in the literature discussing the relationship between SRI and HR-QoL: Ohlsson et al compared pre-therapy SRETV and TLSRE with HR-QoL metrics and found no correlation between either of these tumour volume measurements and a summary score for HR-QoL [392]. However, there were weak correlations with some symptoms such as dyspnoea, diarrhoea and flushing [392]. Taking these previous studies into account, this current study hypothesises that the intensity and/or heterogeneity of tumour lesions may have a stronger relationship with severity of symptoms than active tumour volume and that intra-therapy imaging may be instructive. This chapter will test this hypothesis by correlating quantitative measures on post-PRRT ^{177}Lu -DOTATATE imaging against HR-QoL metrics.

14.2 Materials and Methods

The EORTC QLQ-C30 questionnaire (version 3.0) and EORTC QLQ-GI.NET21 (version 1.1) were used to measure PROs – permission was given by the EORTC for their use in clinical practice and for the purpose of this research study, which was also covered by local ethics approval.

Patients were given printed copies of both of the questionnaires immediately before the start of their first PRRT fraction, and before each subsequent fraction. Completion of the questionnaire was voluntary. Over the period of this investigation, 66 patients completed the questionnaires pre-therapy and also had post 1st fraction quantitative imaging; 40 of these patients also completed the questionnaires at 6 months (before the 4th fraction) with further post-therapy imaging. Completed questionnaires were gathered by the Clinical Nurse Specialist and the responses entered into a spreadsheet. Due to the disseminated nature of patient follow-up care for this national service, it was not possible to gather PROs at a later time-point post-therapy.

Imaging and image analysis was carried out as described in Section 13.2 of the previous chapter. Based on the work in Chapter 8 and applied to clinical ^{177}Lu -DOTATATE images in Section 13.3.3, a subset of 13 quantitative and textural analysis features were utilised.

14.2.1 Statistical analysis

Scoring of the QLQ-C30 questionnaire involves dividing the 30 questions into 15 scales, covering different aspects of a patient's health. Scores are standardised to a value out of

100 to allow comparison, using a methodology outlined in the official EORTC scoring manual [393]. In addition to the 15 scales, a summary score has been validated, which has shown superior discriminatory powers for clinically-relevant groupings than the standard scales [394]. This summary score includes 13 of the 15 scales from the questionnaire, excluding the global health status/QoL and the financial impact scales.

Questionnaire	Scale name	Abbreviation	Scale type
QLQ-C30	Global health status	QOL	High score → high QoL
	Physical functioning	PF	High score → good level of function
	Role functioning	RF	High score → good level of function
	Emotional functioning	EF	High score → good level of function
	Cognitive functioning	CF	High score → good level of function
	Social functioning	SF	High score → good level of function
	Fatigue	FA	High score → high degree of symptoms
	Nausea and vomiting	NV	High score → high degree of symptoms
	Pain	PA	High score → high degree of symptoms
	Dyspnoea	DY	High score → high degree of symptoms
	Insomnia	SL	High score → high degree of symptoms
	Appetite loss	AP	High score → high degree of symptoms
	Constipation	CO	High score → high degree of symptoms
	Diarrhoea	DI	High score → high degree of symptoms
	Financial difficulties	FI	High score → high degree of symptoms
	QLQ Summary Score	QLQ_Total	Summary of QLQ-C30, high score → better QoL

Table 14.1: Patient-Reported Quality of Life metrics for the EORTC QLQ-C30 questionnaire.

Conversion of the question responses into the EORTC QLQ-C30 scored scales [393] was performed using the R package PROscorer [395]. For QLQ-GI.NET21 there was no equivalent analysis program. Therefore, the conversion of scores was performed manually

Questionnaire	Scale name	Abbreviation	Scale type
QLQ-GI.NET21	Endocrine Symptoms	ED	High score → high degree of symptoms
	Gastrointestinal Symptoms	GI	High score → high degree of symptoms
	Treatment-Related Symptoms	TR	High score → high degree of symptoms
	Disease-Related Worries	DRW	High score → high degree of symptoms
	Social Functioning	SF	High score → high degree of symptoms
	Weight Loss	WL	High score → high degree of symptoms
	Weight Gain	WG	High score → high degree of symptoms
	Muscle and/or Bone Pain	MBP	High score → high degree of symptoms
	Information	INF	High score → high degree of symptoms
	Sexual Functioning	SX	High score → high degree of symptoms

Table 14.2: Patient-Reported Quality of Life metrics for the EORTC QLQ-GI.NET21 questionnaire.

using the instructions available in the scoring manual [396]. Both questionnaire's scores are standardised to 100. The scales are described in Tables 14.1 and 14.2.

After the conversion to scales, each scale was correlated against the imaging results for the 13 features using Spearman's Correlation test (the hottest lesion for each patient was used for comparison). This test would determine if any of the features had a strong relationship with a symptom.

Differences in scores between the pre-therapy and 6-month questionnaire (denoted by delta) were calculated such that a positive number indicated an improvement in the patient's symptoms and a negative number a deterioration in a patient's symptoms (regardless of the original scoring scale). Significant changes in scores have been reported in patients receiving radiotherapy [394], which may be analogous to patients undergoing PRRT. A clinically Minimal Important Difference (MID) of 10 (out of 100) was used to identify at least moderate-sized effect in change of score [397]. Delta was thus classified as an improvement in score, a deterioration in score, or no substantial change. Utilising the MID, boxplots were created to determine if any textural analysis features would identify a PRO scale that showed either improvement or deterioration in a symptom and significance testing performed using a Kruskal-Wallis test. For features demonstrating significance, cutpoints were established and,

where possible, compared with the cutpoints derived in Chapter 13. Testing of correlation and significance with quantitative imaging data and other analyses were performed in R statistical software [398].

14.3 Results

Scale name	Abbreviation	This study ¹	Other studies ²
Global health status	QOL	66 ± 21	64 - 71.2
Physical functioning	PF	82 ± 19	81 - 85.2
Role functioning	RF	75 ± 29	72 - 83.6
Emotional functioning	EF	79 ± 21	75.3 - 79.8
Cognitive functioning	CF	81 ± 24	83.1 - 87.3
Social functioning	SF	78 ± 26	75 - 80.8
Fatigue	FA	33 ± 25	28.3 - 33
Nausea and vomiting	NV	7 ± 16	4.2 - 12
Pain	PA	23 ± 25	18.5 - 28.4
Dyspnoea	DY	17 ± 23	17 - 20.2
Insomnia	SL	28 ± 29	24 - 27.7
Appetite loss	AP	18 ± 29	7.5 - 15.2
Constipation	CO	13 ± 21	5.7 - 7
Diarrhoea	DI	30 ± 31	36.6 - 43.3
Financial difficulties	FI	8 ± 18	6.6 - 23.3
QLQ Summary Score	QLQ_Total	79 ± 16	82.3 (one study)
Endocrine Symptoms	ED	18 ± 20	12.3 - 22
Gastrointestinal Symptoms	GI	22 ± 18	20.4 - 24
Treatment-Related Symptoms	TR	8 ± 14	11.6 - 17.2
Disease-Related Worries	DRW	45 ± 26	36 - 43.7
Social Functioning	SF	34 ± 24	27.3 - 69
Weight Loss	WL	16 ± 25	18.3 - 20
Weight Gain	WG	16 ± 27	3 - 3.9
Muscle and/or Bone Pain	MBP	40 ± 30	24 - 29
Information	INF	24 ± 28	5.4 - 84
Sexual Functioning	SX	20 ± 33	28.5 - 69

¹ Mean ± standard deviation.

² Range of mean values from studies in the literature [388, 390, 392].

Table 14.3: Baseline HR-QoL results before starting PRRT, with comparison to other published results [388, 390, 392]. Blue shading defines a Functional scale, where high score represents a healthy level of functioning. Yellow shading defines a Symptom scale, where high score represents a high level of problematic symptoms.

The baseline QoL scores were compared against those reported in the literature [388, 390, 392] and showed similar levels of reported QoL for many of the scales (Table 14.3). Those where a significantly different result were recorded for local patients were worse levels of appetite loss and constipation, and greater problems with weight gain. Outliers in the other studies were a higher level of financial difficulties in the QoL measurements from the

NETTER-1 study [390] (a high percentage of patients were recruited to NETTER-1 from the USA, whilst the other studies recruited solely in Sweden). A higher impact on Social Functioning on the QLQ-GI.NET21 form was noted for the study by Edfeldt et al [388]: this was not mirrored by a similar difference in the Social Functioning questions in QLQ-C30, so is likely due to effects noted by the patient on next of kin and the ability to travel, but there is no obvious explanation as to why it would deviate from that reported in other studies.

14.3.1 Correlation of quantitative data and baseline QoL

QoL responses from immediately before PRRT have been compared to the quantitative measurements taken from the hottest lesion per patient on the 1st fraction post-PRRT scan (Figure 14.1).

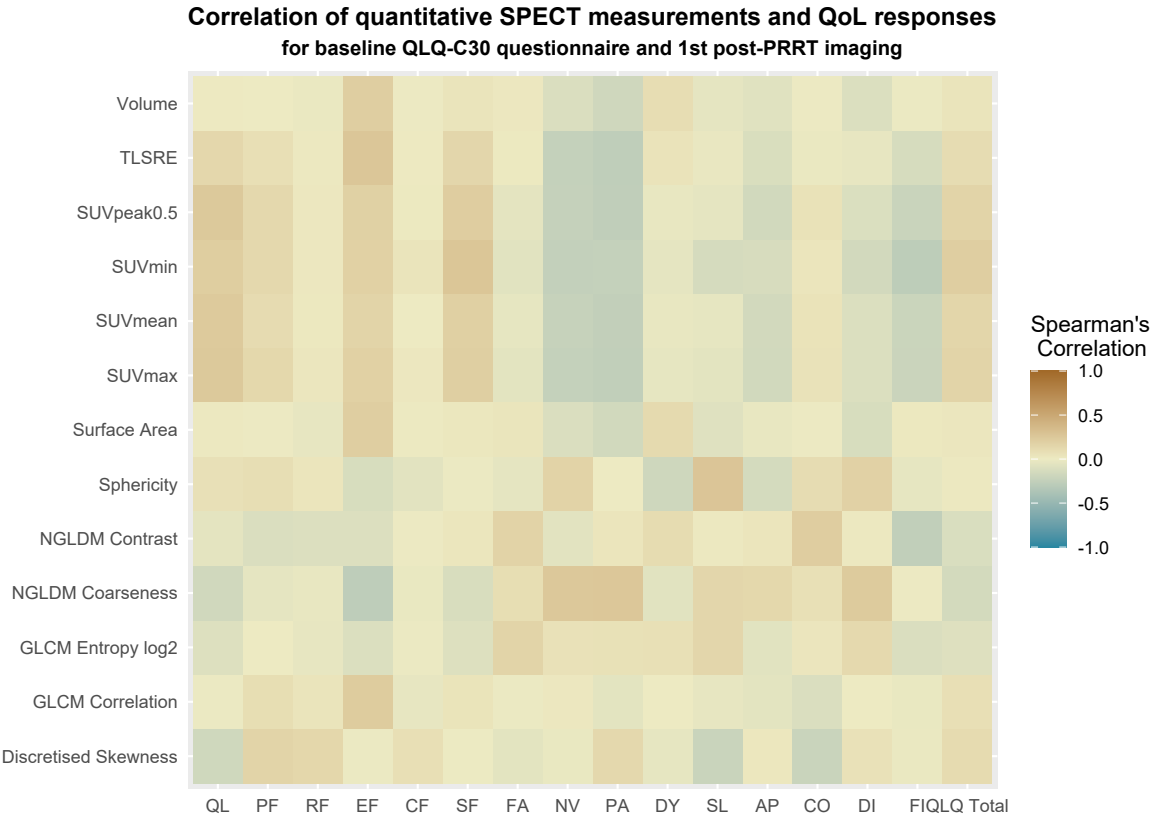


Figure 14.1: Correlation of quantitative and textural features from 1st fraction quantitative SPECT imaging with baseline QLQ-C30 responses on HR-QoL.

Only weak correlations ($|p| < 0.3$) are demonstrated, so little can be inferred other than the patient's symptoms pre-PRRT do not clearly identify who is likely to show high uptake on post-PRRT scans. This may be due to low scores for several criteria (e.g. pain). As it is a generic questionnaire for multiple types of cancer, the questions relating to symptoms may not be specific enough for NETs. However, the specialised NET questionnaire is similar in showing weak or no correlation (Figure 14.2).

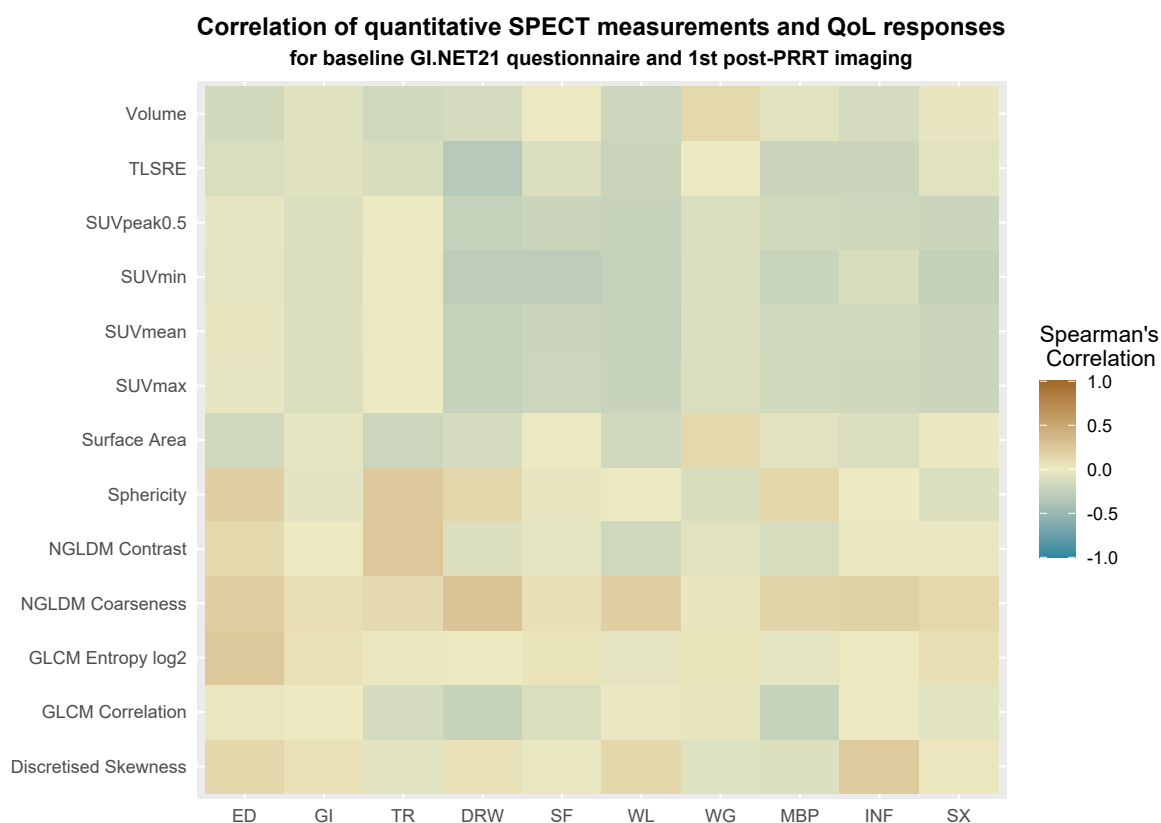


Figure 14.2: Correlation of quantitative and textural features from 1st fraction quantitative SPECT imaging with baseline QLQ-GI.NET21 responses on HR-QoL.

14.3.2 Changes in QoL after 6 months

Forty patients had both a baseline pre-PRRT questionnaire and a pre-4th fraction questionnaire: the median time interval was 168 days. The changes in score, based on a MID of 10, are shown in Figure 14.3 for the QLQ-C30 questionnaire (7 out of 16 scales show more patients with improving symptoms than deteriorating). Figure 14.4 shows the changes for the QLQ-GI.NET21 questionnaire (8 out of 10 scales show improvement in symptoms).

The greatest number of patients reporting an improvement was for Disease Related Worries (58%), followed by Social Functioning (43%), then Endocrine Symptoms (38%) and Gastrointestinal Symptoms and Fatigue (both 35%). However, the most common result is “no significant change”: given the patients had not yet finished their treatment when completing the 6-month questionnaire (with one fraction still to be performed) and that response to PRRT might be expected to be slow, this is unsurprising. A study measuring HR-QoL until first restaging of the patients showed some scales may continue to show improvement at 3 months following the final fraction [387].

The results for changes in the PRO were compared to the 1st fraction imaging, for each of the textural analysis features. Visual inspection of boxplots demonstrated that for many features

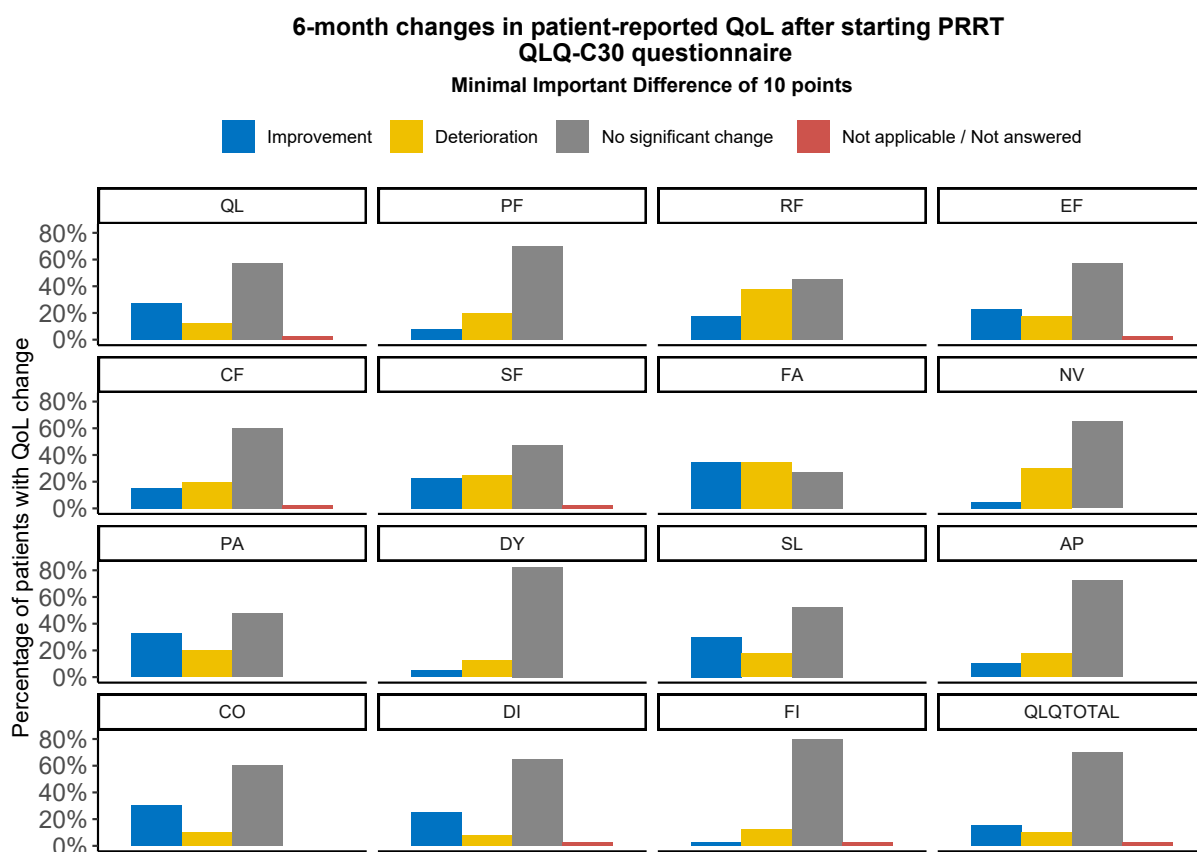


Figure 14.3: Percentage of patients with changes in HR-QoL from QLQ-C30 questionnaire, 6 months after starting PRRT. Uses a Minimal Important Difference of 10 in the score on 6-month questionnaire compared to baseline.

there was overlap in values between the groups segmented using the MID criteria alone (Improvement, Deterioration or No significant change). Therefore to develop a descriptor with potentially more clinical relevance, the data was dichotomised in two separate methods:

- Improvement versus No improvement (incorporating MID of 10).
- Deterioration versus No deterioration (incorporating MID of 10).

An example for one feature and for one PRO scale is given in Figure 14.5 to demonstrate how dichotomisation helps to confirm how the textural analysis feature relates to the patient response.

Features with significant p-scores ($p < 0.05$) on Kruskal-Wallis test were identified (Table 14.4) and a process of rationalisation was performed before cutpoints were established.

- Improvement to Appetite Loss (AP) showed significant correlation with several textural features: the lowest p-value was for TLSRE. Therefore, this was selected as the preferred feature.

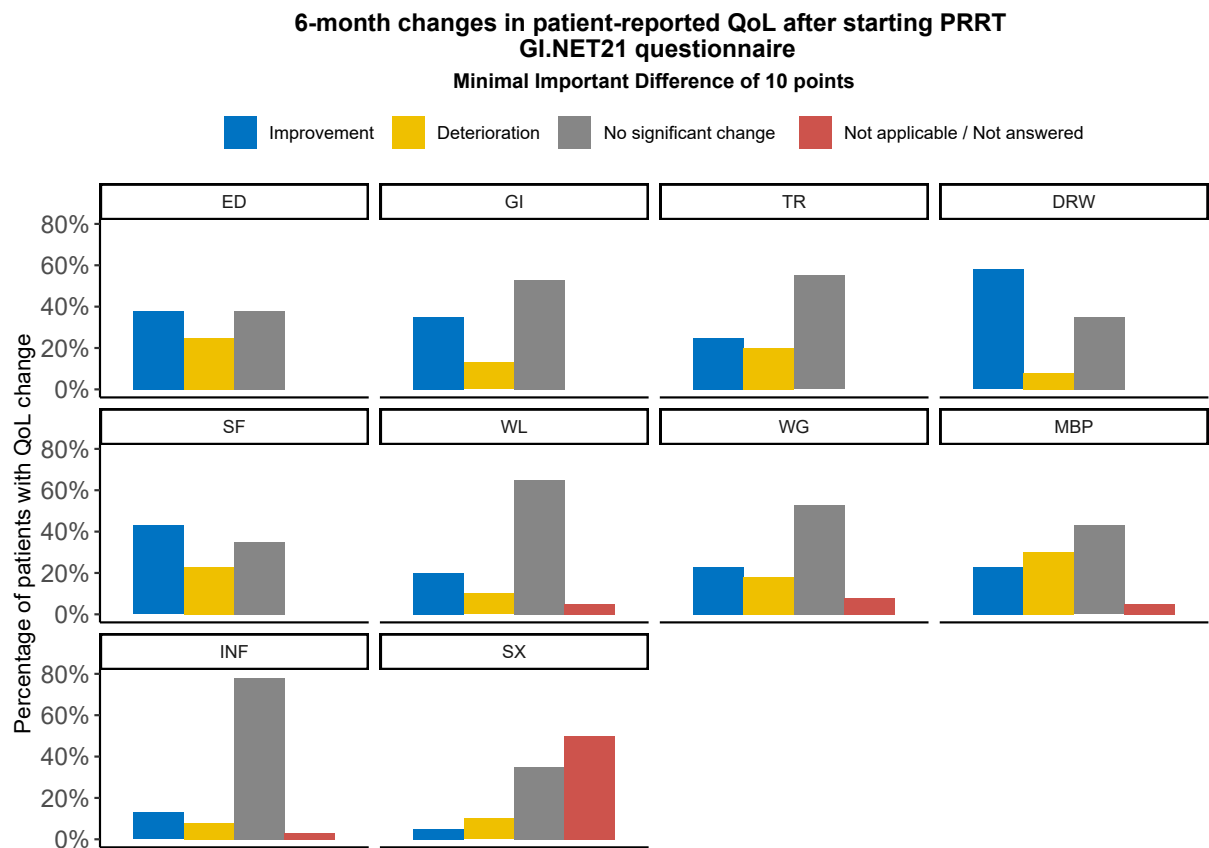


Figure 14.4: Percentage of patients with changes in HR-QoL from GI.NET21 questionnaire, 6 months after starting PRRT. Uses a Minimal Important Difference of 10 in the score on 6-month questionnaire compared to baseline.

- Multiple features were significant for deterioration of Fatigue (FA) and Insomnia (SL) symptoms: these were mainly SUV-related. Hence SUV_{max} was preferred as, on average, it had the lowest p-score.
- Skewness demonstrated a lower p-value than Sphericity for deterioration in Diarrhoea (DI).
- For Sphericity, the correlation with deterioration in Financial Difficulties (FI) was ignored, due to there being no explainable mechanism for this correlation, as was Discretised Skewness for Disease-Related Worries (DRW). Both are suspected to be false positives.
- Deterioration in Endocrine Symptoms (ED) corresponded with several features (most were related to volume (e.g. Surface Area, Volume, TLSRE), though NGLDM Coarseness had the most significant p-score.
- A relationship between deterioration in Gastrointestinal Symptoms (GI) and Weight Loss (WL) was noted, with several features showing significance for both. NGLDM Coarseness was selected for cutpoint analysis due to overall lower p-values.

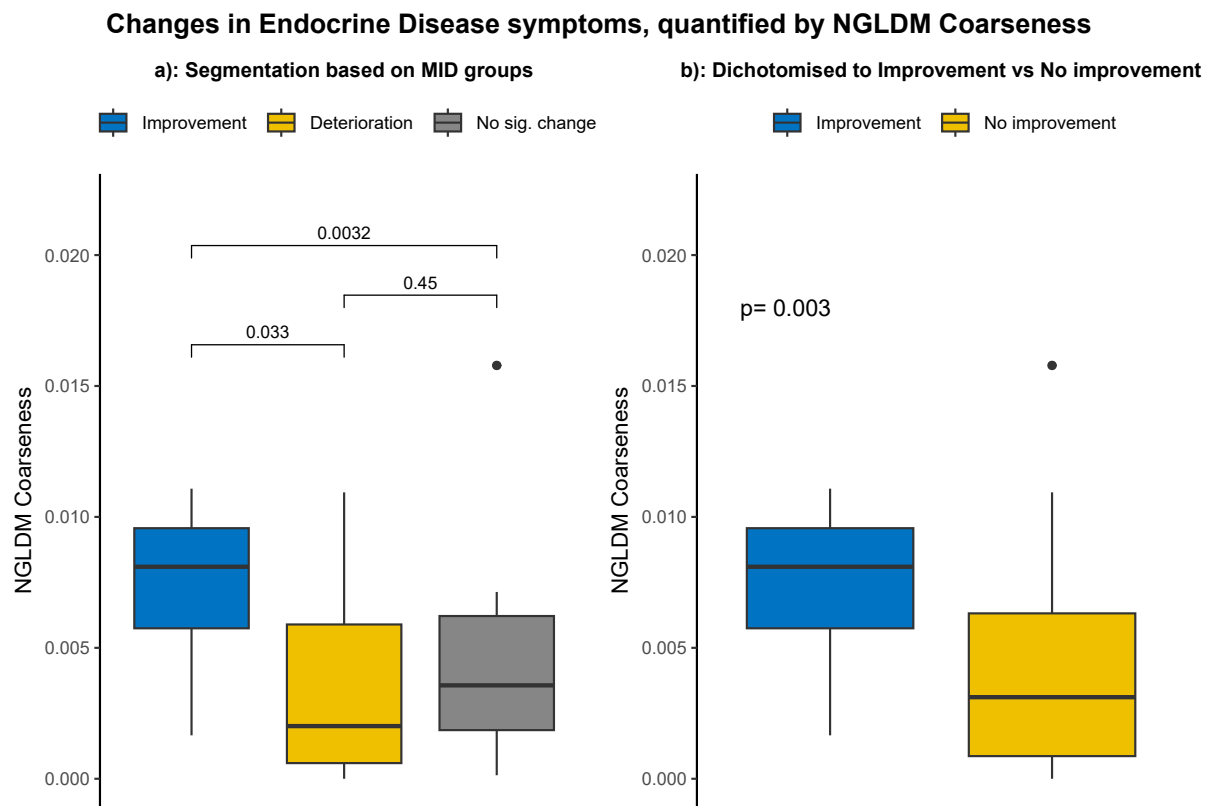


Figure 14.5: Changes in Endocrine Disease symptoms 6 months after starting PRRT, quantified by NGLDM Coarseness. A Minimal Important Difference of 10 was used to identify clinically-relevant changes for a): categorisation into Improvement, Deterioration and No significant change; b): dichotomisation into Improvement or No improvement of the symptoms. Kruskal-Wallis & Wilcoxon tests used to determine significance.

- Several volume-related factors were significant for deterioration of Social Functioning (SF). It is difficult to explain a causal factor for the measurement of volume in a single lesion (though may be more relatable if it was a surrogate to total tumour burden).

Cutpoints were established for the resultant five features which demonstrated significance for a scale. Each cutpoint had a 95% confidence interval established, using bootstrapping of 1000 runs. Their accuracy for correctly predicting the patient's change in symptoms was measured using Sensitivity and Specificity and results are shown in Table 14.5.

14.3.3 Comparison with results from Overall Survival prediction

As Chapter 13 demonstrated that Sphericity could be predictive of OS following PRRT, the previously determined optimal cutpoint from Chapter 13 of 0.794 was compared to cutpoints determined from the data in this chapter. The Sensitivity and Specificity are compared in Table 14.6. Although Table 14.4 demonstrates a significant p-value for the GI scale, further investigation of the cutpoints showed a Receiver Operating Characteristic graph with a diagonal curve, demonstrating the predictor is no better than random. Therefore, it was

Questionnaire	Metric	Feature	Scale	p-value
QLQ-C30	Improvement vs No improvement	Discretised Skewness	CF	0.017
		TLSRE	AP	0.025
		Surface Area	AP	0.032
		Volume	AP	0.042
		NGLDM Contrast	NV	0.044
		NGLDM Contrast	PF	0.040
QLQ-C30	Deterioration vs No deterioration	Discretised Skewness	DI	0.003
		Sphericity	DI	0.032
		Sphericity	FI	0.007
		GLCM Entropy log2	FA	0.041
		SUV _{max}	FA	0.014
		SUV _{max}	SL	0.022
		SUV _{mean}	FA	0.013
		SUV _{mean}	SL	0.028
		SUV _{peak}	FA	0.015
		SUV _{peak}	SL	0.025
		TLSRE	FA	0.021
GI.NET21	Improvement vs No improvement	NGLDM Coarseness	ED	0.003
		Sphericity	ED	0.022
		TLSRE	ED	0.028
		Surface Area	ED	0.013
		Volume	ED	0.021
		Discretised Skewness	DRW	0.044
GI.NET21	Deterioration vs No deterioration	NGLDM Coarseness	GI	0.012
		NGLDM Coarseness	WL	0.043
		TLSRE	GI	0.028
		TLSRE	WL	0.04
		Sphericity	GI	0.043
		Surface Area	SF	0.026
		Volume	SF	0.028
		Discretised Skewness	WL	0.044
		Discretised Skewness	TR	0.021

Table 14.4: Features from SPECT imaging with significant p-values for prediction of patient-reported response in symptoms at 6 months post-PRRT.

Feature	Scale	cutpoint (95% CI)	Sensitivity	Specificity	Interpretation
Discretised Skewness	CF	0.0484 (0.0198-0.453)	0.871	0.5	<0.0484 indicates an expected improvement in Cognitive Function.
Discretised Skewness	DI	0.588 (0.0198-0.453)	1	0.91	>0.588 indicates an expected deterioration in Diarrhoea symptoms.
Discretised Skewness	TR	0.381 (0.07-0.46)	1	0.77	>0.381 indicates an expected deterioration in Treatment-Related symptoms.
TLSRE	AP	376.5 (271-720)	1	0.55	>376.5 indicates an expected improvement in Appetite Loss.
NGLDM Contrast	NV	0.215 (0.150-0.555)	1	0.71	<0.215 indicates an expected improvement in Nausea and Vomiting.
NGLDM Contrast	PF	0.377 (0.311-0.523)	1	0.65	>0.377 indicates an expected improvement in Physical Functioning.
NGLDM Coarseness	ED	0.007 (0.006-0.008)	0.73	0.88	>0.007 indicates an expected improvement in Endocrine Disease symptoms
NGLDM Coarseness	GI	0.005 (0.004-0.007)	0.9	0.7	>0.005 indicates an expected deterioration in Gastrointestinal symptoms
NGLDM Coarseness	WL	0.006 (0.004-0.007)	0.86	0.81	>0.006 indicates an expected deterioration in Weight loss
SUV _{max}	FA	21.0 (17.1-28.2)	0.88	0.57	<21.0 indicates an expected deterioration in Fatigue.
SUV _{max}	SL	19.3 (16.9-24.3)	0.85	0.57	<19.3 indicates an expected deterioration in Insomnia.

Table 14.5: Cutpoints for quantitative SPECT-CT features established as predictive for changes in patient-reported outcomes following PRRT. Confidence intervals established using bootstrapping (1000 runs).

not useful for this comparison. For the two scales investigated, Diarrhoea and Endocrine symptoms, Table 14.6 shows that the cutpoint of 0.794 demonstrates similar Sensitivity and Specificity when applied to the HR-QoL data as the optimal cutpoints derived directly from the HR-QoL data. Therefore, it would be possible to use the same cutpoint of Sphericity for prediction both of OS response to PRRT, and also the prediction of changes to the specified patient-reported HR-QoL metrics.

Feature	Scale	Cutpoint source	cutpoint (95% CI)	Sensitivity	Specificity
Sphericity	DI	QoL data	0.769 (0.689-0.803)	0.83	0.67
Sphericity	DI	OS data	0.794 (0.654-0.825)	0.78	1.00
Sphericity	ED	QoL data	0.851 (0.777-0.885)	0.80	0.52
Sphericity	ED	OS data	0.794 (0.654-0.825)	0.93	0.40

Table 14.6: Comparison of optimal cutpoints established with QoL data and the cutpoint generated from Overall Survival data (Chapter 13).

14.4 Discussion

Our results demonstrate that many patients undergoing PRRT perceive an improvement in symptoms during the course of their treatment (and generally at least no significant worsening of symptoms). The majority of patients reported improvements in disease-related worries, while improvement was the most common result for endocrine disease-related symptoms and social functioning.

Correlation of baseline symptoms with quantitative imaging results from the first fraction does not demonstrate any strong correlations. Changes in HR-QoL at 6 months after treatment have also been compared to 1st fraction quantitative imaging, with 5 features identified as showing significant relationships with changes to symptoms. With cutpoints identified to allow for discrimination, it allows for the possibility of predicting how a patient's symptoms might change during the course of their PRRT, based on the experience of previous patients. Further, the use of a cutpoint established for Sphericity in earlier work on Overall Survival (Chapter 13) shows that it could be used interchangeably (with minimal change in accuracy) to also predict changes in HR-QoL for certain HR-QoL scales.

A limitation of our study is that PROs were obtained only during the course of treatment, with no long-term follow-up. Therefore, the final measurement of patient HR-QoL was taken before fraction 4 and hence before completion of the full course of PRRT. The patients may gain additional benefit from this final fraction that will not be recorded, though studies have suggested the highest absorbed dose (and hence potentially greatest physiological impact) occurs with the first fraction [378]. The limited number of lesions segmented per patient (as

discussed in detail in Section 12.4) means the total burden of disease (i.e. total volume, or summed TLSRE) is not tested for correlation with either initial QoL or changes to QoL.

The patient population who met the criteria of quantitative imaging and completion of PROs was small, which limits the power of the work carried out. The limited numbers also impacted on the ability to verify the results obtained in a separate sample: this is an opportunity for further work. The large number of features tested is also subject to the potential errors of multiple statistical testing, which may be alleviated by application of corrections to p-values, though this may result in more false negatives [399].

Larger samples could possibly clarify inconsistent or unexplainable (in the textural analysis sense) findings. Inconsistent findings included some patients reporting weight loss, and others reporting weight gain; some features showed a correlation with seemingly contradictory results (e.g high NGLDM Coarseness values showing both an improvement in endocrine disease symptoms, but also a deterioration in gastrointestinal symptoms). This inconsistency may be due to differences in the patient population (e.g. those suffering carcinoid syndrome versus those who do not) and every patient experiences their disease differently, both due to the unpredictable effects of the hormones released by NETs but also variance in how they report their symptoms. Some results that are difficult to explain on a purely physiological basis include significant relationships between features and, for example, disease-related worries or financial difficulties. Although a plausible physiological reason may be constructed (hormones released by the tumour causing general anxiety), perhaps the most likely explanation is that multiple statistical testing resulted in significant readings by chance. Correction to p-values (e.g. Bonferroni correction or equivalent) can reduce this issue, but can raise the number of false negatives [399]. Confidence in the selected features and their relationships to the scales is increased by several factors:

- Related factors showing significance for the same scale. This can be seen by a number of volume-related features having significance for changes in Appetite Loss.
- The same factors having multiple significant relationships (e.g. Skewness).
- Closely related scales (such as insomnia and fatigue; or weight loss and gastrointestinal symptoms) demonstrating significance with similar factors, in the same direction.

Clinical trials such as NETTER-1 have concentrated on monitoring HR-QoL in long-term follow-up after patients have completed a full treatment course of PRRT (usually four fractions). There has been less investigation of how HR-QoL changes during the course of PRRT. One study by Martini et al tracked HR-QoL differences between baseline, 2nd fraction and final fraction, as well as at 3 months post-PRRT [387]. For small intestine NETs it found patients reported significant improvements in diarrhoea and appetite loss (female patients

only) by the 2nd fraction, with physical functioning and social functioning also showing significant changes by 3 months. For pancreatic NETs, social functioning only improved in those patients undergoing ^{90}Y -DOTATOC therapy and not ^{177}Lu -DOTATATE, and emotional functioning only for male patients [387]. Our results were more conservative, with patients generally reporting no significant change in these categories, albeit for diarrhoea more patients reported improvement than deterioration.

Only one study in the literature used quantitative or textural analysis for comparison to HR-QoL [392]. Ohlsson and colleagues evaluated the summed somatostatin-avid tumour volume and summed TLSRE in ^{68}Ga -DOTA-TATE and TOC. They found no correlation with overall QoL, though some symptoms did show a weak correlation (Pearson's $r < 0.325$) with summed tumour volume (diarrhoea, flushing and dyspnoea). As only the single hottest lesion per patient was included in this present study, a direct comparison is not possible: though TLSRE for individual lesions did show similar levels of weak correlation for pain, disease-related worries and emotional functioning. TLSRE was also able to predict an improvement in appetite loss if the value measured > 376.5 in the hottest lesion.

The format of PROs leads to some inherent issues about asking questions as to how a patient feels. Naturally, all patients have different thresholds for symptoms such as pain, or how they feel their life has been affected by a disease. There can also be longitudinal differences in how patients perceive their symptoms and their health (termed a response shift effect [390]), where patients adapt to ongoing symptoms and hence score them differently. There is also evidence that some patients find the most commonly used questionnaires (EORTC QLQ-C30 and QLQ-GI.NET21) as not fully reflecting their HR-QoL [400]. QLQ-GI.NET21 may also be less appropriate for those with non-functioning NETs, as opposed to syndromic patients [400]. It is important to consider that other factors apart from the physical changes delivered by PRRT treatment can impact a patient's psychological state; it may be that the act of starting treatment relieves worries about the treatment itself, or makes the patient feel that progress is being made in treating their disease. Our results showed that Disease-Related Worries was the only area in which the majority of patients reported improvements (58%), while the most common improvement in physical symptoms (Endocrine Symptoms) only occurred in 38% of patients. Other studies have shown reduction in patient-reported disease-related worries as early as after one fraction [401]; this may not be completely attributable to the physiological impact of the PRRT, and instead reflect a psychological aspect.

14.5 Conclusion and relevance for the thesis

For the local population, PROs on HR-QoL following PRRT are in line with those reported elsewhere. Analysis has shown that at six months following the start of treatment, the

majority of patients show either improvement or no significant change to most symptoms and functional categories.

There is limited correlation between quantitative measurements of post-PRRT imaging and baseline reports of HR-QoL, suggesting that imaging has limited scope to measure how the patient feels related to disease symptoms at the start of their PRRT treatment.

Five features have been shown as having significant relationships with patient-reported changes in symptoms following PRRT, and cutpoints identified to help classify patients. An imaging metric identified as a significant predictor of Overall Survival (lesion Sphericity) was also found to predict patient-reported changes in endocrine disease symptoms and diarrhoea. While PROs have well-reported deficiencies, and further testing is needed to validate these features, it is hoped that they may provide an imaging indicator as to how patients will respond to PRRT with ¹⁷⁷Lu-DOTATATE.

Chapter 15

Harmonisation of clinical quantitative data

15.1 Introduction

The problem of combining data from multiple centres (or acquired using different equipment/parameters) has been examined in several scientific fields outwith molecular imaging. For example, in laboratory medicine, where differences between ‘batches’ of samples acquired on different equipment (or even following changes to a reagent) at the same centre are important when combining datasets [284]. Methods developed for this issue originally required large numbers of samples in each batch, reducing the utility. One of the most common methods for statistical harmonisation is the Location-Scale (L-S) method, introduced in Chapter 5, Section 5.3.5.

To improve robustness for small batch sizes, Johnson et al proposed a method using an empirical Bayes method to adjust data for batch effects [284]. This method, dubbed “ComBat”, has proven effective in laboratory medicine, but also in amalgamating data in various imaging modalities, including CT, MRI and PET [286, 343, 402]. This experience in other modalities prompts the hypothesis that ComBat could also be used to harmonise quantitative SPECT data. This study will apply harmonisation using both the L-S and Bayesian ComBat methods to clinical quantitative SPECT data acquired on two gamma cameras at different hospitals, with the aim of determining if harmonisation of quantitative features is achievable.

15.2 Materials and Methods

15.2.1 Patients

This study included two groups of patients who underwent quantitative SPECT-CT imaging with ^{99m}Tc -EDDA/HYNIC-TOC for a suspicion or confirmed case of NET at one of two hospitals within the same city, referred to as ‘Hospital A’ (122 patients) or ‘Hospital B’ (39 patients). Only patients who were naïve to LAR-SSA treatment were included. The characteristics of the patient populations are listed in Table 15.1.

All patients had provided signed informed consent for the use of their images for research, and ethics approval was granted by local committee (GSH24RA002).

	Hospital A (N=122)	Hospital B (N=39)	p-value
Sex			0.4
Male	66 (54%)	18 (46%)	
Female	56 (46%)	21 (54%)	
Age	62 (53,71)	65 (52,71)	0.8
Weight (kg)	80 (68,91)	78 (63,88)	0.4
Disease location			0.1
Liver metastases	81 (46%)	12 (43%)	
Nodal metastases	77 (44%)	9 (32%)	
Small bowel primary	18 (10%)	7 (25%)	

Table 15.1: Demographics for patients undergoing ^{99m}Tc -EDDA/HYNIC-TOC imaging, compared between the two hospital sites.

15.2.2 Imaging and acquisition protocols

	Hospital A	Hospital B
Camera	Siemens Intevo Bold	Siemens Intevo Bold
Quantification software	Siemens BroadQuant	Siemens BroadQuant
Energy window	140 keV $\pm$ 7.5%	140 keV $\pm$ 7.5%
Matrix	256 x 256	256 x 256
Orbit	Non-circular	Non-circular
Motion	Acquire During Step	Step and Shoot
Number of views	64	60
Time per view	16 seconds	20 seconds

Table 15.2: Acquisition parameters for ^{99m}Tc -EDDA/HYNIC-TOC SPECT imaging for each hospital site.

Both sites utilised a Siemens Intevo Bold SPECT-CT gamma camera (Siemens Healthineers, Erlangen, Germany). Quantitative calibration was carried out as per the manufacturer's instructions for Siemens BroadQuant; regular QC was performed to ensure the accuracy of the quantitative calibration. The respective imaging protocols for the two hospitals were those used locally for clinical studies and are summarised in Table 15.2.

	Hospital A	Hospital B
Algorithm	Ordered-Subset Conjugate Gradient	Ordered-Subset Conjugate Gradient
Iterations	24	48
Subsets	2	1
Filtering	0 mm (All-pass)	5 mm Gaussian

Table 15.3: Reconstruction parameters for ^{99m}Tc -EDDA/HYNIC-TOC SPECT imaging for each hospital site.

Image reconstruction for both sites was performed using an OSCG algorithm. The iterations and post-reconstruction filtering differed between sites and are listed in Table 15.3.

15.2.3 Image analysis

Image analysis was performed by one experienced operator, to exclude any potential inter-operator variability. Intra-operator variability has been studied previously (Chapter 10) and shown to be minimal for this imaging procedure and software. Analysis used the LifeX package [98]. VOIs were created on the quantified SPECT data using a 40% of SUV_{max} threshold for physiological uptake in the spleen, and in up to 5 sites of disease, as indicated in the clinical report. In order to avoid any sites of tumour metastases in the liver for physiological liver uptake measurement, spherical VOIs were placed manually, based on visual interpretation of the SPECT and CT data. A spherical VOI was also used for blood pool uptake, drawn in the left ventricle.

For all images, SUV_{min} , SUV_{max} , SUV_{mean} and SUV_{peak} (where peak is the average voxel value in a 1 ml sphere centred on the maximal pixel in the drawn VOI) were determined. A further 10 textural analysis features were extracted, based on the features found to be clinically robust in Chapter 10 (though TLSRE was not utilised for physiological measurements, due to its dependence on the VOI volume).

15.2.4 Harmonisation method

Harmonisation was applied using the “NeuroCombat” package [287] in R [398]. Each of the physiological and tumour sites were harmonised independently using both the L-S method and the Bayesian adaptation to L-S using priors. Additionally, for each of the L-S and Bayesian methods the options for adjusting for covariates and reference site were tested. The number of significantly different features and the average p-score were calculated for each harmonisation method, to allow for comparison between the methods.

15.2.5 Statistical analysis

Data was tested for normality using the Wilks-Shapiro test. To examine if the data from the two systems was significantly different before and after harmonisation with ComBat, the Mann-Whitney U test was used. Comparison was made for each of the sites of physiological uptake and tumour sites for each of the quantitative features. The level of significance was taken as a p value less than 0.05.

15.3 Results

Data was not normally distributed for the features, as per Wilks-Shapiro test; therefore, all further analysis was performed using non-parametric methods.

Sites of disease were selected for further analysis if they had at least 10 cases identified in data from the proposed reference site (Hospital A). This held for liver metastases, nodal metastases and small bowel primary.

Patient weight, age and sex were all found to have correlation with factor values for a substantial number of the textural analysis features, and hence were included as covariates in the model.

The p-value before harmonisation was compared to the p-value after harmonisation for each of the combination of harmonisation methods, use of covariates, and use of reference sites. Full results for each combination are shown in the tables in Appendix D. An example of the difference between pre- and post-harmonisation data is shown for one feature using a density plot (Figure 15.1): this plot shows the harmonisation process has effectively brought the distribution of values for this feature much closer together.

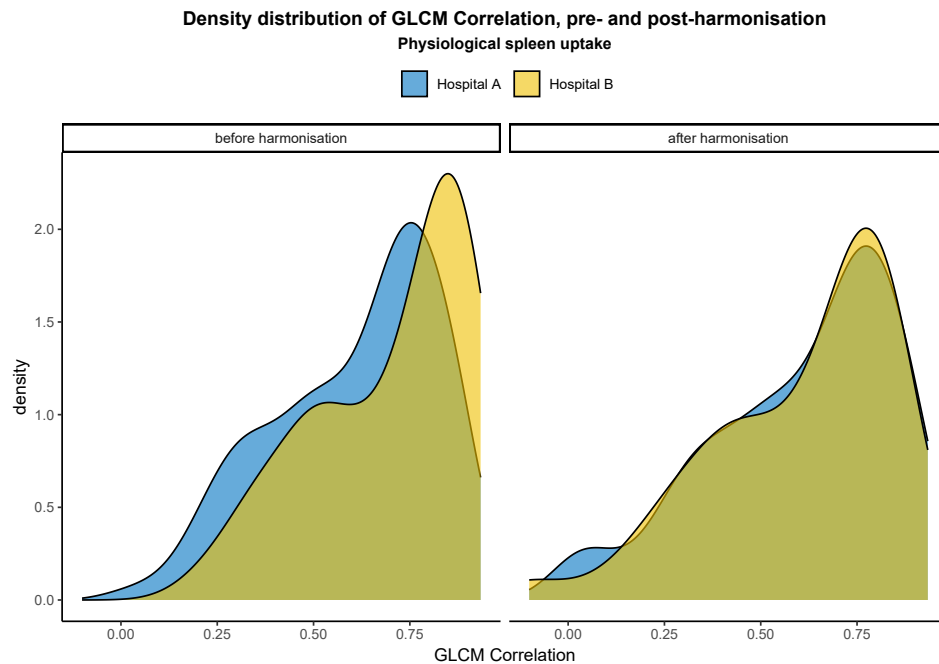


Figure 15.1: Example of the distribution of values for two different clinical sites of a textural analysis feature (GLCM Correlation), before and after harmonisation.

The results following harmonisation show a reduction in differences between the two sets of data for the majority of metrics when measuring physiological uptake. For tumour uptake, fewer metrics were significantly different before harmonisation, but there is a noted increase in p-scores (moving towards a reduced chance of significant difference) for most metrics, particularly the SUV metrics. The algorithms without covariates applied generally showed better performance. The use of a reference site was in general the equal-best or best-performing algorithm; this method is preferred as it is the most likely to be used in a clinical scenario. The relative poorer performance of the Bayesian algorithm when compared to the

L-S algorithm differs from findings reported by other groups using harmonisation. However, it is postulated that the features assessed in this study do not have the close relationships that may be expected in, for example, genes for laboratory testing, therefore utilising a prior based on the harmonisation results for other features may not be of benefit. Results for the best-performing combination are demonstrated in Figure 15.2 and summarised in Table 15.4.

	Features with $p < 0.05$ before harmonisation	Features with $p < 0.05$ after harmonisation	Increased p-value after harmonisation
Physiological liver	10	0	13
Physiological spleen	2	1	11
Physiological blood pool	11	1	12
Liver metastases	1	1	11
Nodal metastases	2	2	9
Small bowel primary	2	0	11

Table 15.4: Summary of harmonisation effects on p-values for L-S method, without covariates and Hospital A as reference site. The full tables for each organ / site of disease and each variant of harmonisation parameters are shown in Appendix D.

15.4 Discussion

Harmonisation of data gathered from two quantitative SPECT-CT systems was performed using the L-S and ComBat algorithms, based on methodology first developed for use in batch harmonisation of laboratory samples.

Differences between the two sets of acquired data were minimal in respect to the acquisition parameters: although Hospital B had a longer overall acquisition (by 14%), the use of ADS by Hospital A will result in more counts being acquired than if acquisition had used Step and Shoot (local unpublished work), hence effectively equalising the counts per image. The reconstruction parameters deviate in a more significant way; though the effective iterations are identical for both hospitals (and it can be seen from the figures in Appendix A that there is limited difference in the quantitative accuracy between the use of 1 or 2 subsets (Figures A.1, A.2 and A.3), the difference in filtering applied is more likely to be significant in terms of values of the quantitative features.

A downside of the statistical harmonisation process is that it doesn't account for the clinical or physical properties of the features being harmonised – for example, SUV features, which by definition must have a positive value, can be revised to a negative value after harmonisation. This was apparent mostly for the SUV_{min} feature in the present work but also occurred for one tumour lesion for SUV_{mean} . Applying the natural logarithm to the SUV before harmonising, then reversing the process on harmonised data, can resolve this issue. Use of a reference

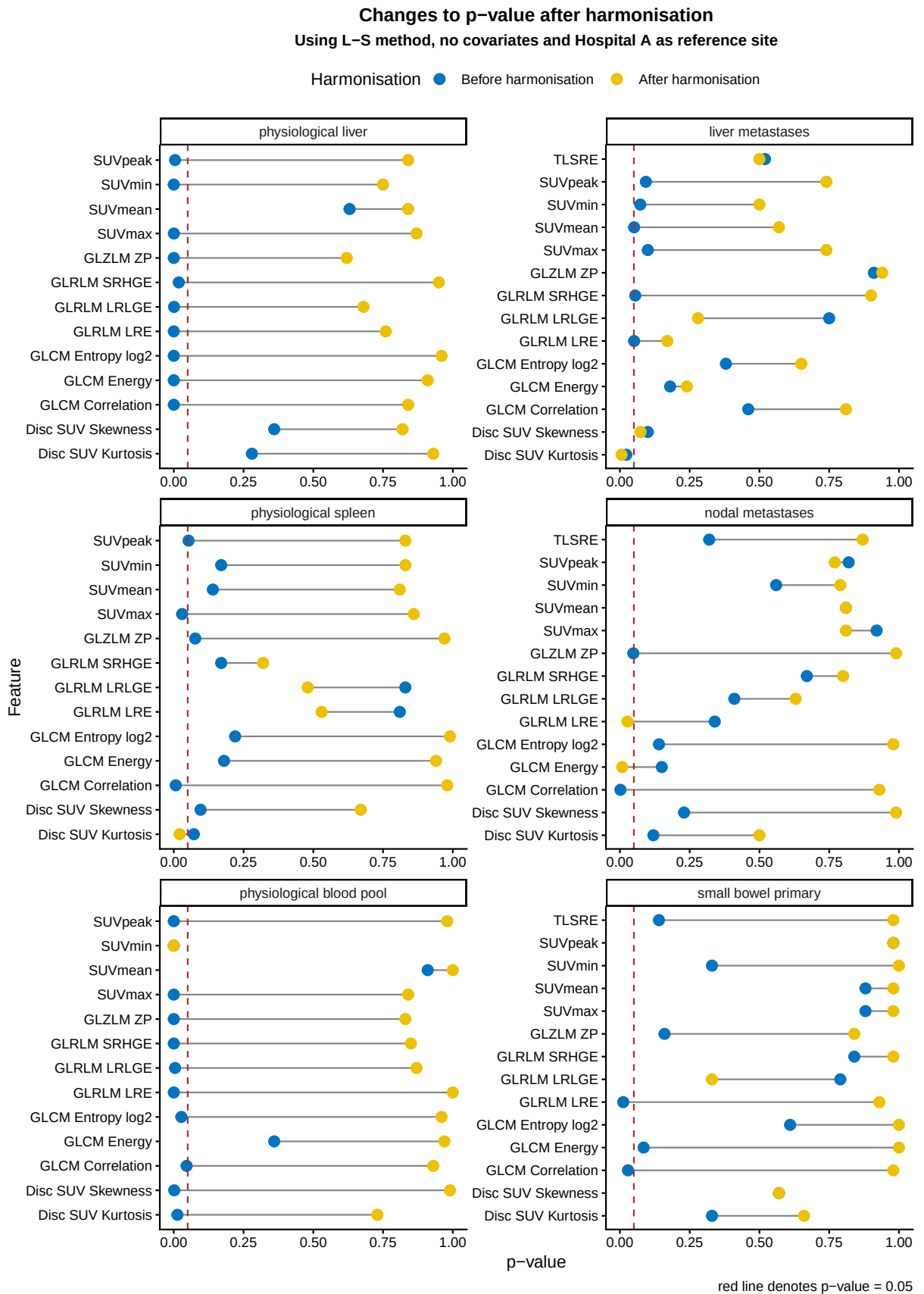


Figure 15.2: Changes in p-value following harmonisation for each feature. Using the L-S method, without covariates and Hospital A as reference site. Red dotted line is p=0.05.

site also limits the chances of negative SUVs being produced. The use of these statistical methods without consideration of the circumstances of the data being used is clearly not appropriate, a point emphasised by Yadav et al who found for harmonisation of CT data that methods varied in effectiveness when applied to differing data sets [292].

The ComBat algorithm is easy to implement and can be applied to retrospective data without any knowledge of the hardware or reconstruction protocols (albeit the user must be confident that the systems have been properly calibrated to deliver accurate quantitative values). Unlike other techniques for harmonisation of molecular imaging data, it does not require access to the original imaging data (which can cause issues with anonymisation and data transfer), nor is filtering applied to meet the lowest common denominator in camera performance (which degrades image quality for one of the systems). Whilst removing features from analysis that are sensitive to variations between centres is a possible alternative methodology, this approach does risk removing features that hold useful information about the disease state or treatment. Limitations to the ComBat method are that the corrections applied to the data cannot easily be replicated on subsequent data, without inclusion of the original dataset (though this functionality is currently under development for the NeuroCombat package [287]) and there is no opportunity to harmonise individual patient data with those acquired at a different centre. The algorithm also assumes the samples in each centre's dataset are similar: for molecular imaging studies this means factors such as medication that may interfere with radiopharmaceutical uptake, tumour type and uptake time must be similar. With some implementations of ComBat, covariates can be implemented in order to account for these factors. Otherwise, the ability to differentiate between different groups within the data can be compromised.

Other studies have used modified versions of ComBat in attempts to improve downsides of the original algorithm. For example, the base version of ComBat can only shift data based on the overall mean of the feature from the combined centres. This restriction could render the harmonised feature meaningless if, for example, it created a feature with negative values that do not relate to real-world conditions. A modification to shift data based on a chosen reference 'batch' (or centre) was proposed by Stein et al [290], but a study applying this modification to radiomic data from MRI, PET and CT found no improvement to performance over the base ComBat package [288]. The same group also tested a modification to provide bootstrapping and Monte Carlo modelling of the estimated Bayesian parameters. This adapted algorithm, termed B-ComBat, did reduce variability in the harmonised data, however the improvement was small [288].

All of the studies referenced above have used the Bayesian algorithm of ComBat [284]. However, this study looked at both the underpinning L-S model without the Bayesian prior, and with the Bayesian process. This study found on average the Bayesian model resulted in poorer

harmonisation than the basic L-S model: this result does not correlate with findings elsewhere [284], though the majority of imaging studies using ComBat have used the Bayesian method without questioning the suitability of utilising a prior based on the harmonisation of different features [286, 343]. It is hypothesised that as the majority of features tested are not strongly correlated (a deliberate decision to reduce dimensionality) then assigning priors based on the other features cannot provide a benefit and may instead hinder an accurate harmonisation. Therefore, use of prior knowledge on the harmonisation needed between batches for SUV features, for example, may not be a positive factor when harmonising unrelated textural features such as entropy or grey-level run length. The designation of a reference batch made only a small difference to the average p-score after harmonisation or the number of features significantly different, but may be of value if multiple sites are to be harmonised and one site can be designated the reference site. Application of covariates resulted in poorer harmonisation for the physiological features, though for tumour sites there was little difference or slightly improved performance. The reasoning for this is unclear – the strong correlations of weight and sex with SUV measurements of physiological uptake are likely to be a consequence of the lack of weight-based administrations, and indeed show improved harmonisation with covariates in splenic background measurements, but not in liver background.

While variation in gamma camera performance is well known, even for systems from the same manufacturer [263, 264], this has not been widely tested for quantitative data derived from patient images. However, extrapolating from the field of quantitative PET data it is not surprising that SPECT data could not be immediately pooled without harmonisation. The lack of agreed harmonisation protocols or accreditation standards for SPECT-CT imaging is partially due to the relative newness of quantitative imaging in SPECT and will hopefully be addressed by multi-centre collaborations in future. However, the wide spectrum of possible radionuclides and radiopharmaceuticals that can be used in SPECT-CT may mean guidelines will be limited to specific tests, and other methods of harmonisation may have a role if SPECT-CT is to develop into a widely used quantitative modality with a role in multi-centre trials. A method applicable for small batches of data is particularly important for molecular imaging data such as PET or SPECT, as patient cohorts are often limited in numbers (especially compared to other imaging modalities such as CT or MRI) due to the limited clinical indications of specific radiopharmaceuticals and limited availability of PET-CT scanners or quantitative-capable SPECT-CT cameras.

15.4.1 Further work

This work focussed on patients who were not currently receiving Long-Acting Release Octreotide, as Chapter 9 has shown a significant difference for measured SUV between patients receiving LAR-SSA and those naïve to treatment. The ComBat algorithm can deal

with differing patient characteristics between the groups, by defining covariates of interest. Orlhac et al have shown that implementation of the covariate technique can separate out groups for [^{18}F]FDG PET [286], and such comparison is a source of further work for LAR SSA treatment (though use of covariates such as sex, weight and age were not successful in the work described here). While the ComBat package has shown promise in harmonising imaging data, and here applied to SPECT data for the first time, other approaches are available which could also be applied to SPECT: they are discussed in detail by several review papers [294, 403].

15.5 Conclusion and relevance for the thesis

The Location-Scale method of statistical harmonisation proved beneficial in reducing statistical differences between quantitative SPECT data acquired on separate gamma cameras, using different reconstruction protocols. The Bayesian modifications to the L-S algorithm proved less effective, postulated to be due to the unrelated nature of the features not providing useful a-priori information to the harmonisation process. Despite downsides to the retrospective nature of the process, statistical harmonisation may have a role in amalgamation of datasets for quantitative SPECT, for example in the clinical trial setting.

Part 4:

Summary and Conclusions

Data-driven predictions can succeed - and they can fail. It is when we deny our role in the process that the odds of failure rise. Before we demand more of our data, we need to demand more of ourselves.

Nate Silver

The Signal and the Noise, 2012

The final part of this thesis will begin by summarising the main findings of the work described previously. Areas for further development of the work of this thesis will be identified. A final chapter will draw conclusions, focussing on how this thesis has contributed to the field of quantitative SPECT and NET imaging.

Chapter 16

Summary of the thesis

This thesis has explored several issues regarding SPECT quantification and its application to NETs imaging, specifically for ^{99m}Tc and ^{177}Lu radionuclides.

16.1 Findings on calibration and quantification

Though several studies in the literature have examined the accuracy of the OSCG algorithm for SPECT quantification (some using the Siemens BroadQuant package), there has not, to date, been an extensive analysis of the calibration process or a longitudinal study of the stability of quantitative measurements. Chapter 6 split the calibration process into the composite parts, performing repeated measurements to assess repeatability, reproducibility and linearity. This work provided an opportunity to make recommendations as to how the calibration process could be improved, over and above the manufacturer's protocols, with the suggestions being:

- Preferentially measure the syringe in the radionuclide calibrator, rather than the point source, as this mirrors the circumstance for patient administration, reducing differences between calibration and clinical processes.
- If it is chosen to measure the point source, rather than the syringe, a method to ensure replicable positioning in the calibrator should be used - such as the jig designed and manufactured for this thesis (Figure 6.4).
- Point source volume should be at least 0.4 ml to aid reproducibility. The manufacturer's recommendation to use gauze in the point source to absorb the liquid should be followed.

This thesis has described a methodology for calculating the estimated uncertainty for the calibration process. These experiments resulted in an estimated standard uncertainty of 1.53% for ^{99m}Tc and 2.26% for ^{177}Lu . The result for ^{177}Lu compares well with uncertainties estimated by the NPL as an exemplar in their guide [308]. The availability of a traceable activity standard for ^{177}Lu (through NPL) gives lower calculated uncertainties than studies conducted in countries where such traceability is not available.

Chapter 7 evaluated the quantitative accuracy of the calibration and measured stability over prolonged periods of time. The accuracy for both ^{99m}Tc and ^{177}Lu proved to be comparable with similar studies in the literature, as well as the manufacturer's white paper. Longitudinal stability, as measured by repeated SUV measurements and Shewhart chart analysis for ^{99m}Tc , was acceptable over periods up to 9 months. The infrequent availability of ^{177}Lu meant that analysis was not as detailed as for ^{99m}Tc . Nonetheless, repeated SUV measurements were within tolerance of $\pm 5\%$. Calculation of Coefficients of Variation (1.78% and 1.63% respectively for ^{99m}Tc and ^{177}Lu) agreed well with the one comparable study in the literature and suggested consistency in quantitative measurement.

In order to implement textural analysis in clinical studies, it was necessary to determine which of the multitude of features would be robust across a number of imaging studies. Reducing the number of features was also required to avoid the risk of model over-fitting. This was achieved in the work described in Chapter 8, where phantom studies recommended in EANM guidelines on radiomics were applied to quantitative SPECT for the first time (to the author's knowledge). For both ^{99m}Tc and ^{177}Lu , a uniform phantom was used to determine features that showed a dependency on volume. A "revolver" phantom to mimic inhomogeneous lesions was adapted for SPECT imaging, and used to assess feature reproducibility across multiple scans. Combining the results of both phantom experiments produced lists of reproducible features for ^{99m}Tc and ^{177}Lu which could then be utilised in the subsequent clinical studies.

16.2 Findings on clinical applications

Quantification in SRI has mainly focussed on PET and ^{68}Ga -DOTA-peptides. This thesis has instead investigated SPECT quantification, with a particular focus on ^{99m}Tc -EDDA/HYNIC-TOC. Chapter 9 detailed the impact of LAR-SSA medication on quantitative results. This work demonstrated that physiological uptake was lower in patients on LAR-SSA, while most sites of disease (with the exception of bone metastases) were unaffected. This raises issues regarding accurate reporting of tumour-to-background ratios between the two groups, especially for visual analysis. Importantly, it also raised the possibility that professional body recommendations to halt LAR-SSA before PRRT are sub-optimal, as pre-treatment with LAR-SSA may reduce dose to organs at risk whilst maintaining dose to targeted disease.

One of the key uses for ^{99m}Tc -EDDA/HYNIC-TOC imaging is to determine patient eligibility for PRRT, generally based on the Krenning Score. Assessed visually, the Krenning Score compares uptake in tumour sites to physiological uptake, to determine if a patient's disease is sufficiently avid to benefit from the therapy. In Chapter 10, LAR-SSA naive patients imaged with ^{99m}Tc -EDDA/HYNIC-TOC were assessed visually by expert reviewers to determine an Expert Consensus Score, which was compared both to a group of less-experienced reporters and to quantitative measurements. Visual assessment of the Krenning Score was found

to have significant inter-operator variability, and whilst most non-expert reporters showed good agreement with the ECS, quantitative measurement was found to produce similar levels of agreement but with much-improved inter-observer variability. A quantitative Krenning Score (utilising tumour to background ratios with SUV_{max}) was identified as the preferred measure. The comparison of quantification to Expert Consensus review was extended further in Chapter 11 to incorporate textural analysis features. The clinical imaging data was used to further reduce the set of ^{99m}Tc features reported in Chapter 8 based on redundancy. However, the features selected did not show better agreement with ECS than the best-performing quantitative features from Chapter 10.

A Tumour Sink Effect has been reported in ^{68}Ga -DOTA-TATE, where radiopharmaceutical uptake is reduced in physiological organs (and/or other tumours) due to monopolisation of uptake by extremely large or avid tumours. This was explored in Chapter 12 for SPECT imaging of ^{99m}Tc -EDDA/HYNIC-TOC. No conclusive evidence for a link between tumour burden and physiological uptake was found. This finding suggested there was no TSE in this cohort of patients imaged with ^{99m}Tc -EDDA/HYNIC-TOC.

This thesis has identified that quantitative SPECT and textural analysis could have a useful role in predicting response to PRRT with ^{177}Lu -DOTATATE. Quantitative analysis of imaging after the first PRRT fraction was found to be predictive of length of Overall Survival in Chapter 13, with Sphericity identified as the most useful feature. When stratified based on the presence of at least one high-risk lesion (where high-risk is indicated by Sphericity < 0.794), the 4-year survival rate is 44% for high-risk patients, compared with 86% for low-risk patients. Sphericity was also found to be amongst the useful features for predicting patient-reported response in Chapter 14. Using the same cut-off, Sphericity was found to predict which patients would show improvement in endocrine disease symptoms (sensitivity: 0.93, specificity: 0.4), and which patients would avoid deterioration of their diarrhoea symptoms (sensitivity: 0.78, specificity: 1.0).

Harmonisation of quantitative SPECT data could be an important tool to help deliver clinical trials examining the clinical utility of quantitative SPECT. The use of statistical methods for harmonisation was tested in Chapter 15, utilising ^{99m}Tc -EDDA/HYNIC-TOC data acquired on two different gamma cameras with varying reconstruction parameters. The Location-Scale method was found to reduce statistical differences across a range of textural analysis features, confirming the possibility of using statistical harmonisation to combine quantitative SPECT data from multiple centres.

Chapter 17

Future work

Inevitably, the work generated as part of this thesis cannot provide the definitive answer to all possible questions on SPECT quantification in NETs. All research in this field is a continuum and may create as many additional questions as it provide answers. This Section will detail some areas of further study following on from this work and how they may be addressed.

17.1 Further work on calibration and quantification

The work presented in Chapter 6 on calibration would benefit from replication using a different calibrator and/or gamma camera to provide additional confidence in the findings, in particular the jig for the point source container, designed and manufactured for this thesis, should be tested independently to confirm it reduces the variability of point source measurement.

Limited availability of ^{177}Lu at sufficient concentration to prepare a point source meant that the linearity experiment did not use the required activity to cover the manufacturer's suggested range of up to 1110 MBq. Longitudinal stability measurements for ^{177}Lu would ideally be tested with greater frequency over a single calibration period.

The accuracy of the quantification process was measured through use of standard nuclear medicine phantoms, such as a uniform cylinder and the NEMA IEC phantom. These phantoms have drawbacks in their simplified approach to radiopharmaceutical distribution, and do not accurately replicate the anatomy of patient studies. A more sophisticated approach would utilise anthropomorphic phantoms, an example of which was demonstrated by Robinson et al [128]. With the kind assistance of the original authors, a copy of this phantom has been manufactured locally, with the aim of performing accuracy checks for both $^{99\text{m}}\text{Tc}$ and ^{177}Lu in patient-replicating geometry. The phantom will also prove useful in expanding the work in Chapter 8 in identifying robust textural analysis features by more accurately replicating the clinical use case.

17.2 Further work on clinical applications

As the work evaluating a quantitative Krenning Score in Chapters 10 and 11 was performed only in patients who were naive to LAR-SSA, it is planned to expand this study to include patients receiving LAR-SSA at the time of imaging. This work will be facilitated by expanding

the recruitment to incorporate multiple gamma cameras within the Health Board, for the purposes of validating the derived thresholds in a wider population, and will incorporate the findings of the harmonisation work performed in Chapter 15. A reporting tool is also under development in conjunction with Hermes Medical Solutions, which will implement the quantitative Krenning Score proposed in Chapter 10.

Although a TSE was not identified for ^{99m}Tc -EDDA/HYNIC-TOC in the work in Chapter 12, a simulation study has suggested that a total tumour burden of >550 ml is required to demonstrate a notable effect [363]. As only four patients exceeded this threshold, it is possible that a larger cohort of patients with extensive tumour burden would be more likely to demonstrate a TSE.

Further long-term follow-up of the surviving patients who were studied in Chapter 13 could help to better define predictive factors for 3- and 4-year survival. It is also planned to identify patients who had pre-PRRT diagnostic imaging with ^{99m}Tc -EDDA/HYNIC-TOC on a SPECT gamma camera calibrated for quantification, as predictive factors for response to PRRT on diagnostic imaging would provide the opportunity to change patient management before initiating treatment.

The use of harmonisation techniques for SPECT quantification has the potential to improve gathering and comparison of quantitative data, potentially aiding clinical trials. This thesis has demonstrated the utility of statistical harmonisation using the Location-Scale method for a number of textural analysis features in ^{99m}Tc -EDDA/HYNIC-TOC imaging. However, further work is needed to analyse the impact of harmonisation on multi-centre data and inter-operator variability on segmentation, as well as for other radiopharmaceuticals and for other radionuclides. It is important to evaluate how harmonisation may affect features that have proven predictive in single centre studies (for example, sphericity in ^{177}Lu -DOTATATE post-therapy SPECT, as shown in this thesis). Harmonisation could assist efforts to independently validate such predictive features, by combining datasets and reducing inter-camera and inter-operator variability. Such validation studies are comparatively rare at present, but are necessary to initiate widespread adoption of textural analysis as a tool for predictive or prognostic imaging.

The implementation of quantification into the clinical workflow requires a holistic approach. Whilst a given textural analysis feature may predict those with better response to a particular treatment, and may help with patient selection, it does not help those who 'fail' selection. For these patients, we must aim to provide better alternatives. This could involve, for example, another treatment line. Perhaps the initial treatment would be more effective with an adjustment, such as a administration of a radiosensitiser, or an adjustment to the dosing schedule. These questions require strong multi-disciplinary working involving

medical practitioners outside the nuclear medicine community, and aligned approaches from professional bodies, pharmaceutical companies and medical licencing regulators.

The problems of integrating SPECT quantification into patient pathways cannot be addressed by any single piece of work, but it is hoped that this thesis, and other studies by researchers across the nuclear medicine community, will be a step towards the clinical integration of SPECT quantification and improved diagnosis and treatment for patients.

Chapter 18

Conclusion

This thesis has two overarching aims: to characterise the accuracy and stability of the BroadQuant package for quantitative SPECT; and to demonstrate clinical applications of quantitative SPECT, specifically for NETs. Chapter 16 describes in detail how these goals were met. There are several key areas of novel work presented in this thesis that contribute to the development of quantitative SPECT as a clinical tool. These include:

- Validation of the accuracy of BroadQuant for quantitative imaging for ^{99m}Tc and ^{177}Lu , identifying improvements to the calibration process and determining the duration that calibrations remain stable.
- Publication of evidence that LAR-SSA affects physiological uptake in ^{99m}Tc -EDDA/HYNIC-TOC SPECT imaging, whilst uptake in the majority of tumour lesions is not affected.
- Proposal of a quantitative Krenning Score that demonstrates greater robustness than visual interpretation, and greater accuracy than a litany of textural analysis features.
- Establishment of threshold values for textural analysis features that can accurately stratify a patient's likely response to PRRT.
- Demonstration that the Location-Scale method for statistical harmonisation of textural analysis features were superior to a methodology that incorporated Bayesian priors, when applied to quantitative ^{99m}Tc -EDDA/HYNIC-TOC SPECT.

This thesis has demonstrated that SPECT quantification and textural analysis can be robust and accurate, with the potential to add value to clinical interpretation of somatostatin receptor imaging. Improvements to diagnostic accuracy and reproducibility, as well as prediction of response to molecular radiotherapy treatment with PRRT, has been highlighted by several facets of the work of this thesis. Presentation of the strong foundations of SPECT quantification and the utility of clinical applications demonstrates that the work reported in this thesis can improve the diagnosis, care and treatment of patients with NETs.

Appendices

Appendix A

Optimisation of reconstruction parameters for Siemens BroadQuant

The selection of the optimal reconstruction parameters for ^{99m}Tc and ^{177}Lu involved testing with a NEMA IEC phantom. The standard set up of the phantom was used, with six spheres from 10mm to 37mm diameter and with the "lung" insert included. A sphere:background ratio of 5.2:1 for ^{99m}Tc was prepared while for ^{177}Lu (where clinical imaging is generally performed >24 hours after administration and washout of the radiopharmaceutical from background is expected to be greater) a ratio of 10.0:1 was used.

The acquisition parameters selected as the local standard (as defined in Section 7.2.8) were used, with acquisition times of 16 seconds per view. The reconstruction parameters for the OSCG algorithm in the Siemens BroadQuant package were varied, using the following combination of parameters.

Iterations and Subsets:

- 24 iterations and 2 subsets
- 48 iterations and 1 subset
- 60 iterations and 1 subset
- 72 iterations and 1 subset

Post-reconstruction filtering:

- 0 mm (i.e. all-pass filter)
- 5 mm
- 10 mm
- 14 mm (maximum allowed by Siemens OSCG algorithm)

Image analysis was performed in Hermes Hybrid Viewer (version 4), with Volumes of Interest drawn based on the nominal sphere and lung insert size and positioned by use of the co-registered CT.

Three metrics were chosen for analysis, based on the recommendation of the EANM guidelines for SPECT quantification [11] for determining optimal reconstruction parameters:

- Recovery Coefficients were used to assess convergence of activity with alteration of acquisition parameters and were calculated using Equation 7.2 and results plotted in Figure A.1.
- The Signal to Noise Ratio was assessed to ensure noise levels did not become excessive with increasing iterations. SNR was calculated by comparing the measured activity concentration in the sphere to the standard deviation in a background volume, as defined in [33].

$$\text{Signal to Noise Ratio (SNR)} = \frac{a_j - a_{\text{background}}}{SD_{\text{background}}} \quad (\text{A.1})$$

- The lung error was calculated to determine the amount of spill-out from a "hot" background into a nominally "cold" insert: helping to test the impact of scatter correction on quantitative accuracy. The percentage error was calculated using Equation A.2:

$$\text{Lung error (\%)} = \frac{\text{measured mean concentration in lung insert}}{\text{known true concentration in background}} \quad (\text{A.2})$$

Results were assessed from Figures A.1, A.2 and A.3 before a quantitative analysis combining the result using the Root Mean Square Deviation (Equation 7.3), as detailed in section 7.2.8. Where multiple sets of parameters produced similar results for RMSD, priority was given to RC measurements as the highest level of quantitative accuracy was the focus of the thesis.

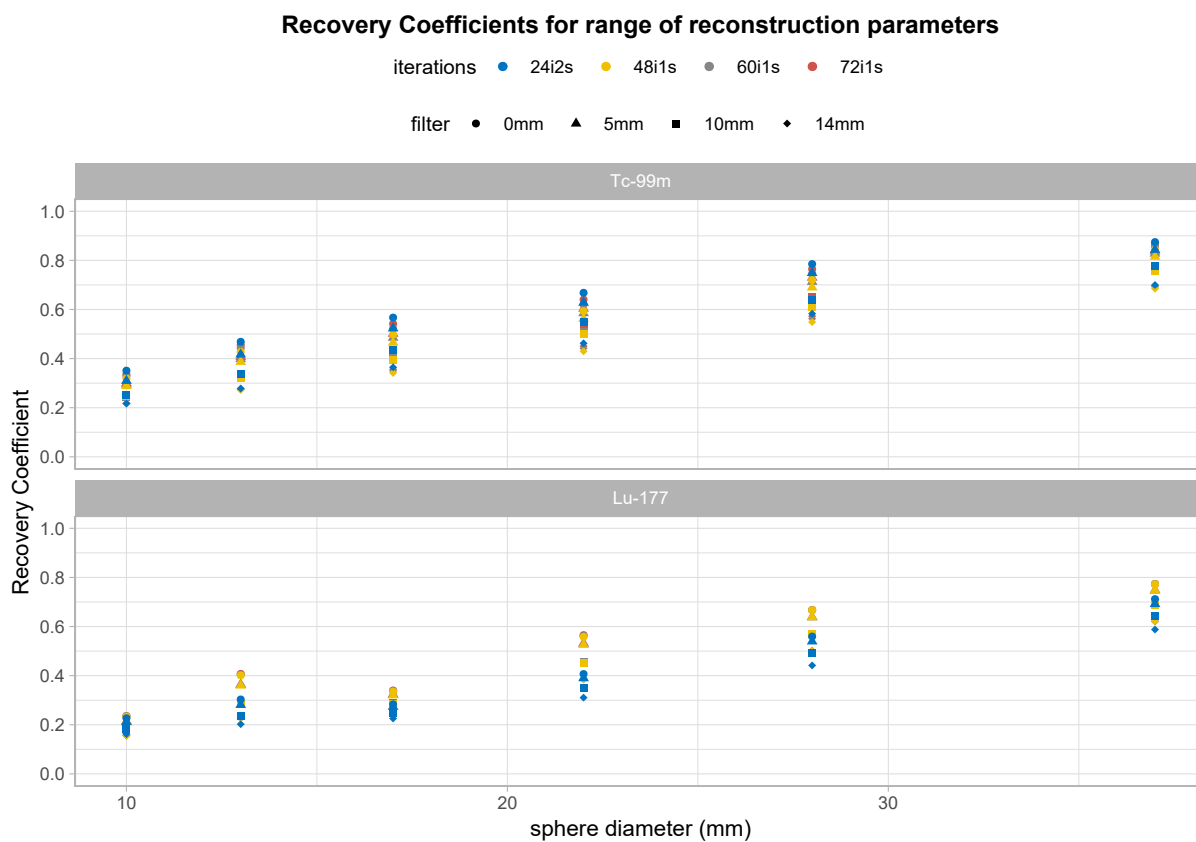


Figure A.1: Recovery Coefficients for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu .

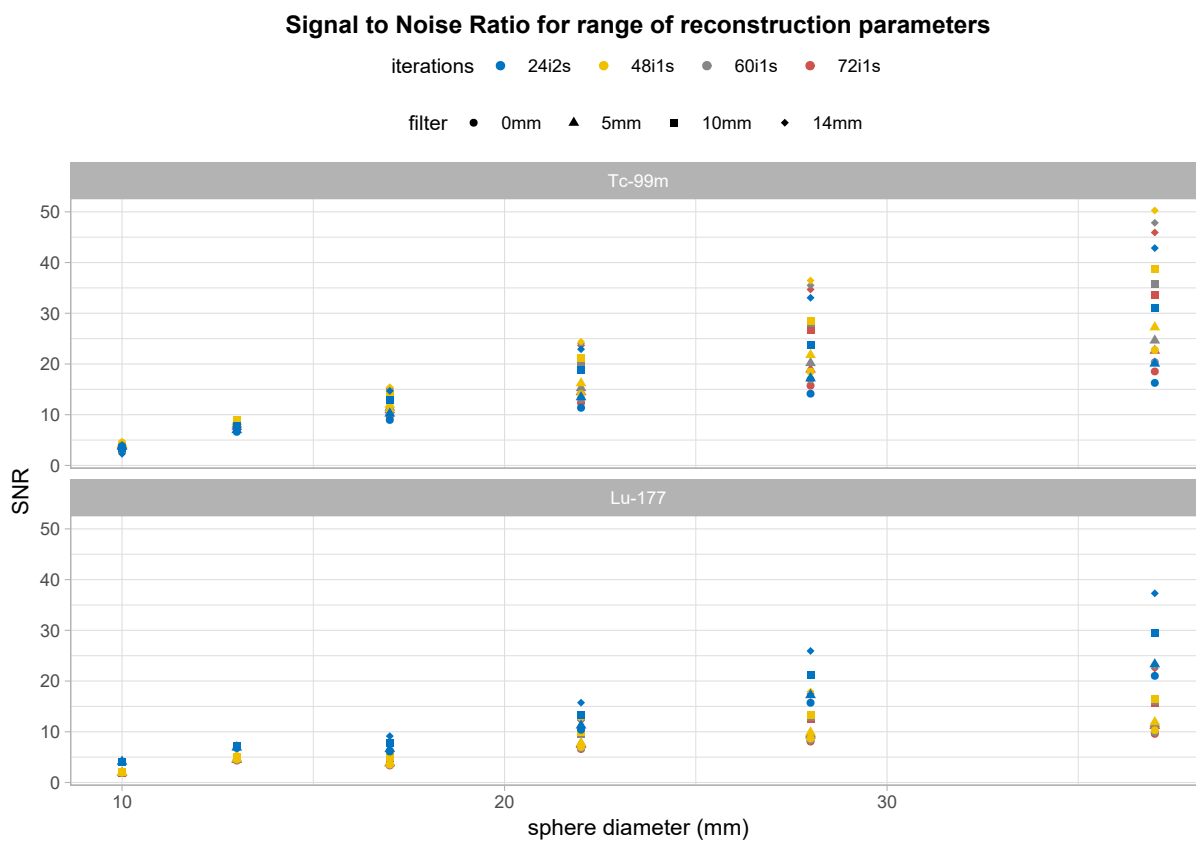


Figure A.2: Signal to Noise Ratio for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu .

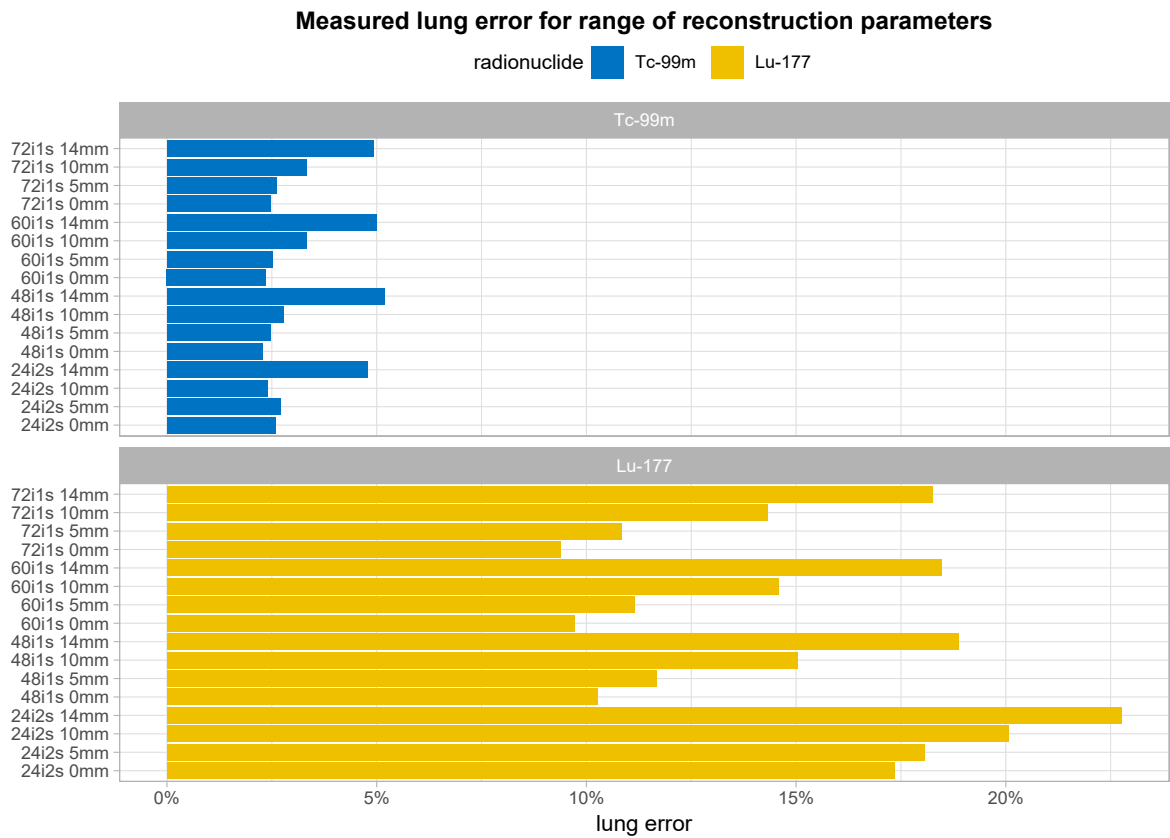


Figure A.3: Measured lung error for a range of reconstruction parameters for ^{99m}Tc and ^{177}Lu .

Appendix B

Supplementary materials, Chapter 8

This appendix reproduces the Supplementary Material published in Gemmell, et al [404] and reproduced in this thesis as Chapter 8.

Figures B.1 and B.2 (each consisting of 3 pages) are discussed in Section 8.3.1 and demonstrate how textural analysis features display different patterns of correlation with changes in the outlined volume in a uniform phantom.

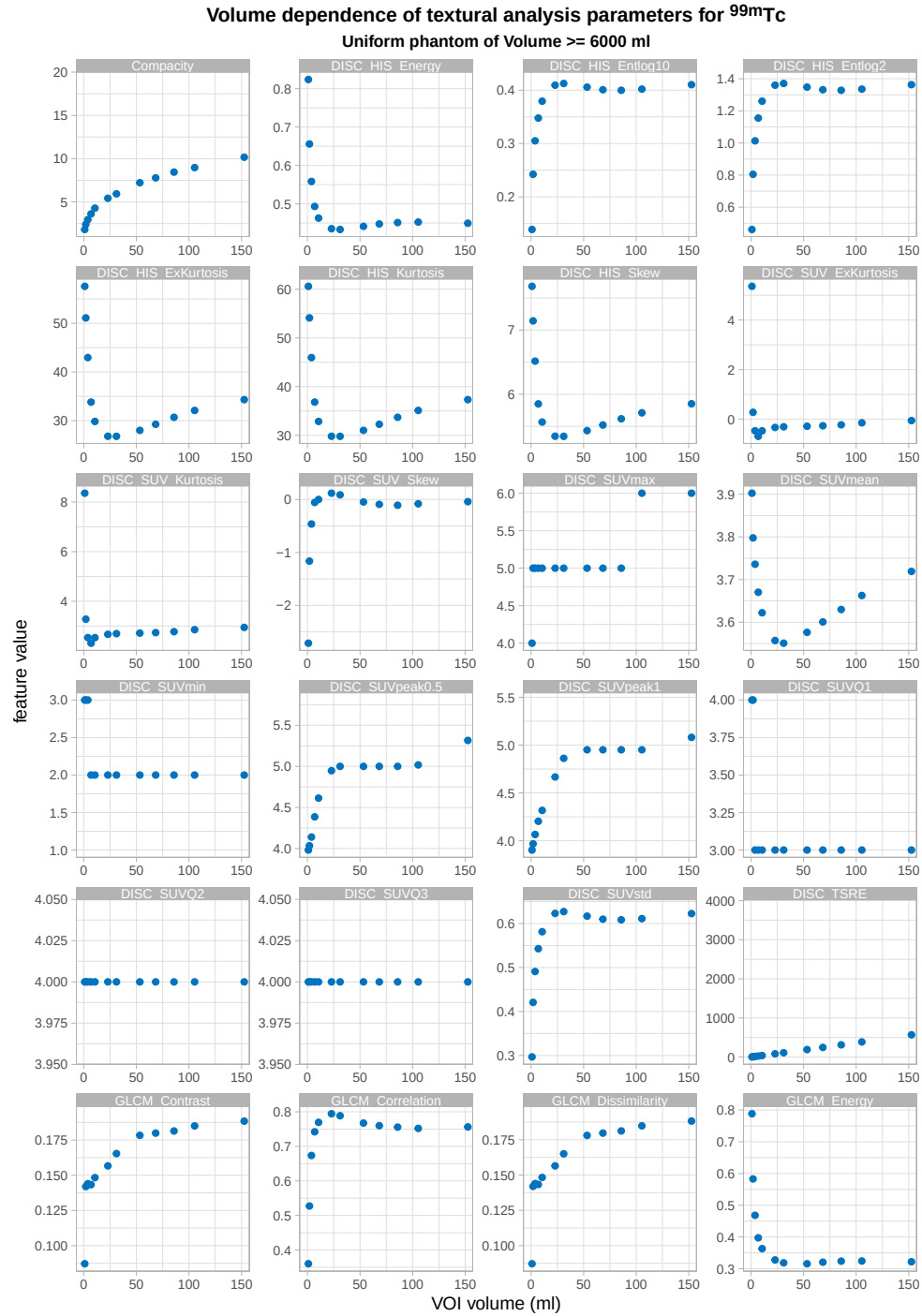


Figure B.1: Volume dependence of textural analysis parameters for ^{99m}Tc .

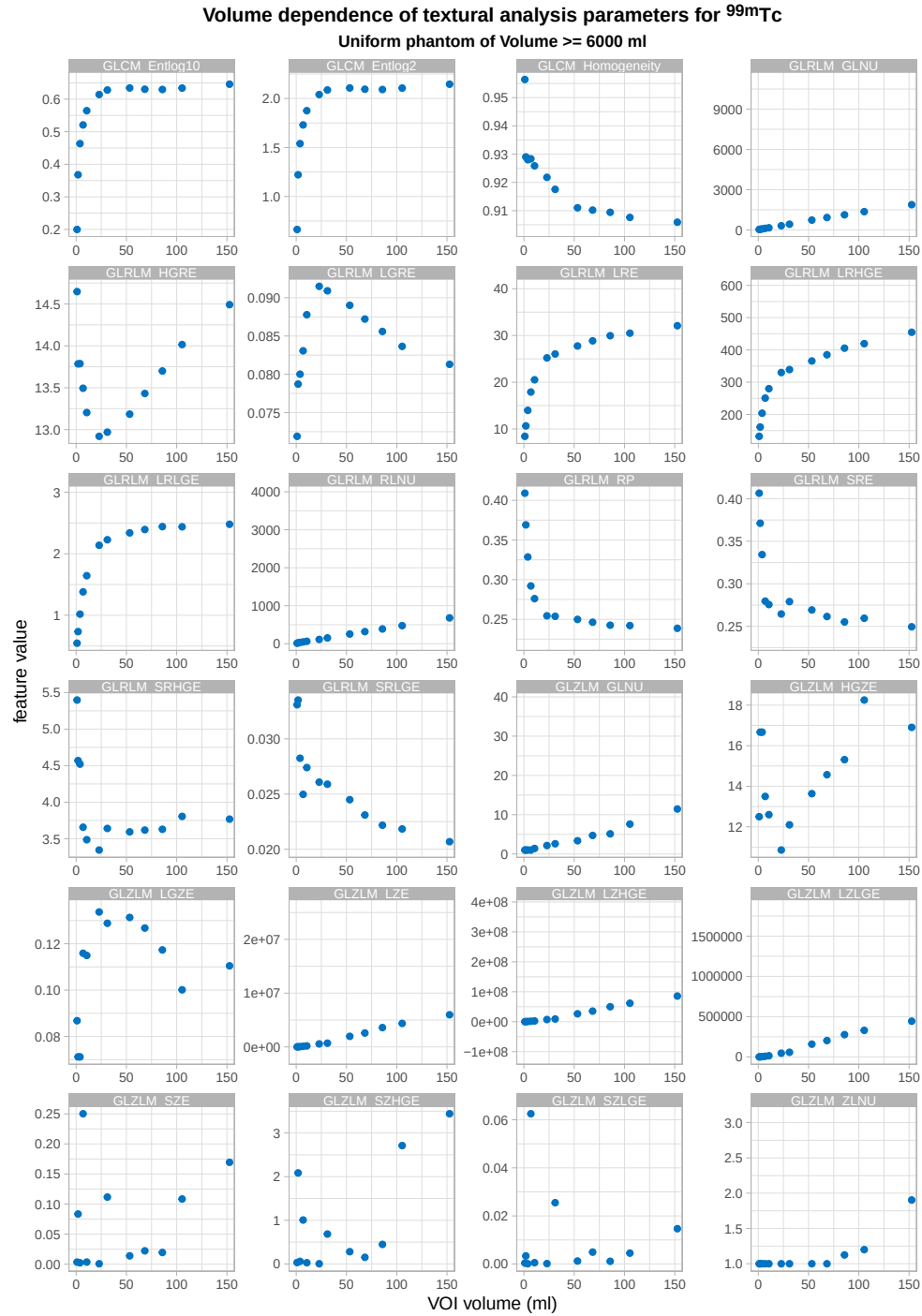


Figure B.1: Volume dependence of textural analysis parameters for ^{99m}Tc .

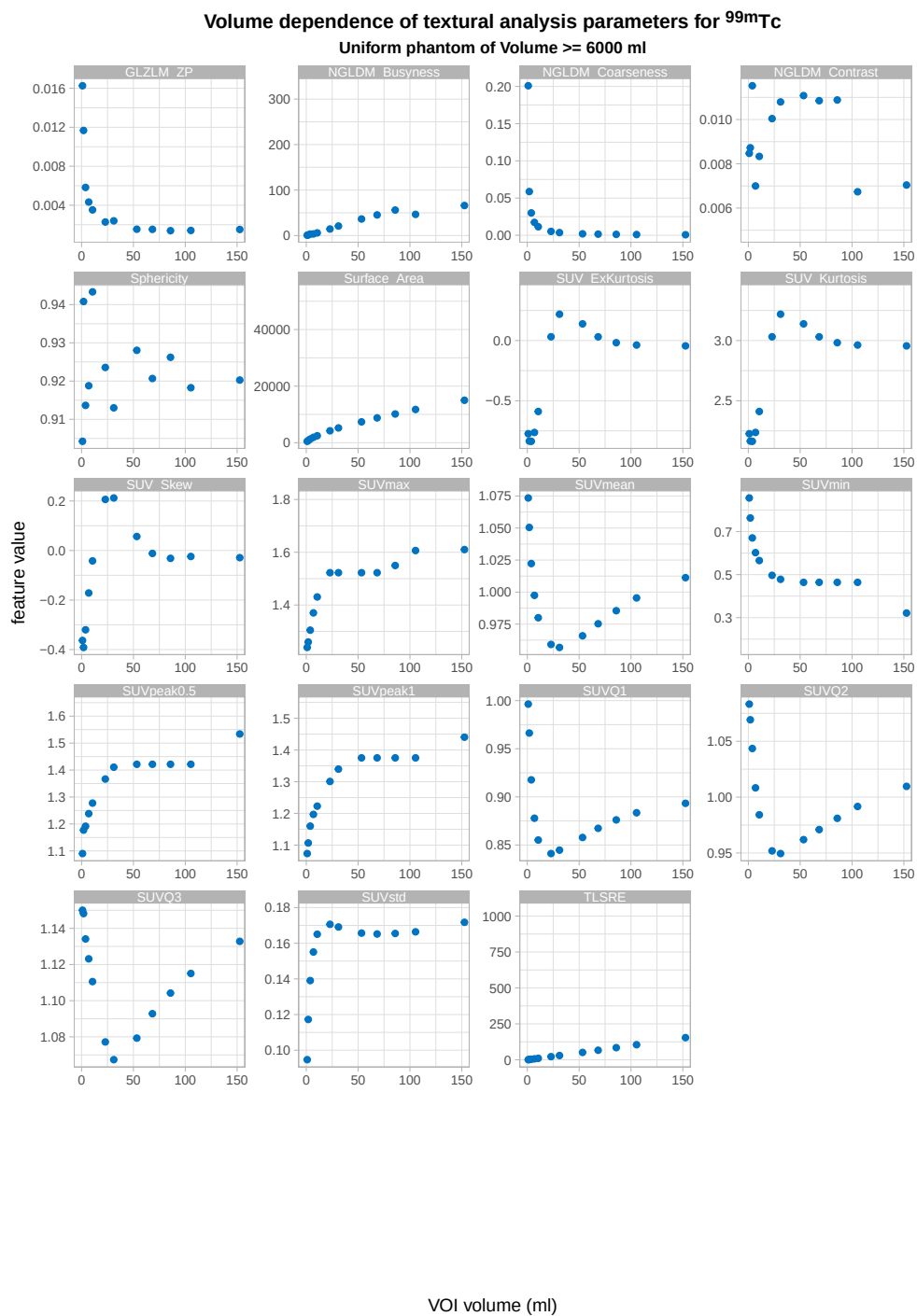


Figure B.1: Volume dependence of textural analysis parameters for ^{99m}Tc .

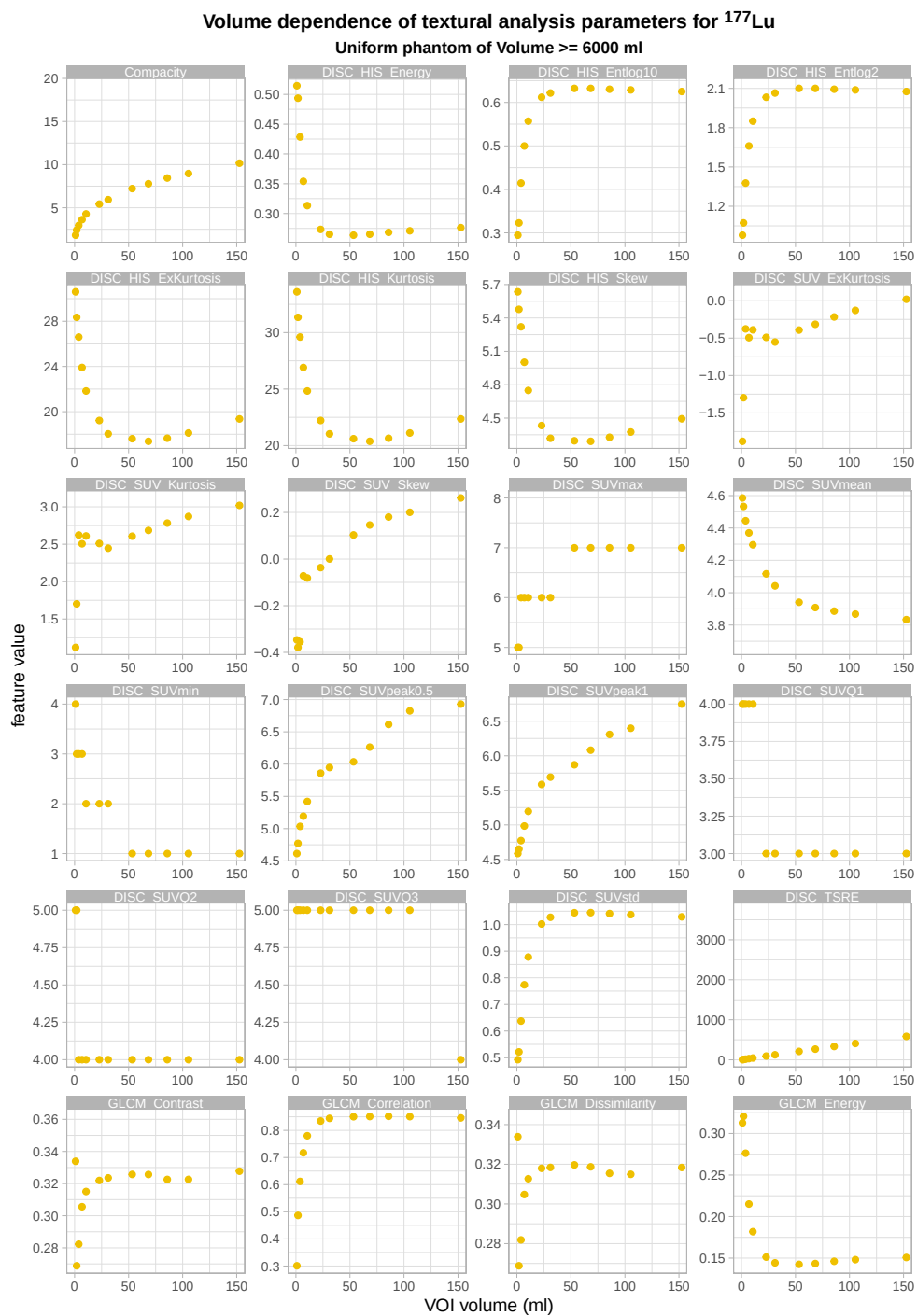


Figure B.2: Volume dependence of textural analysis parameters for ^{177}Lu .

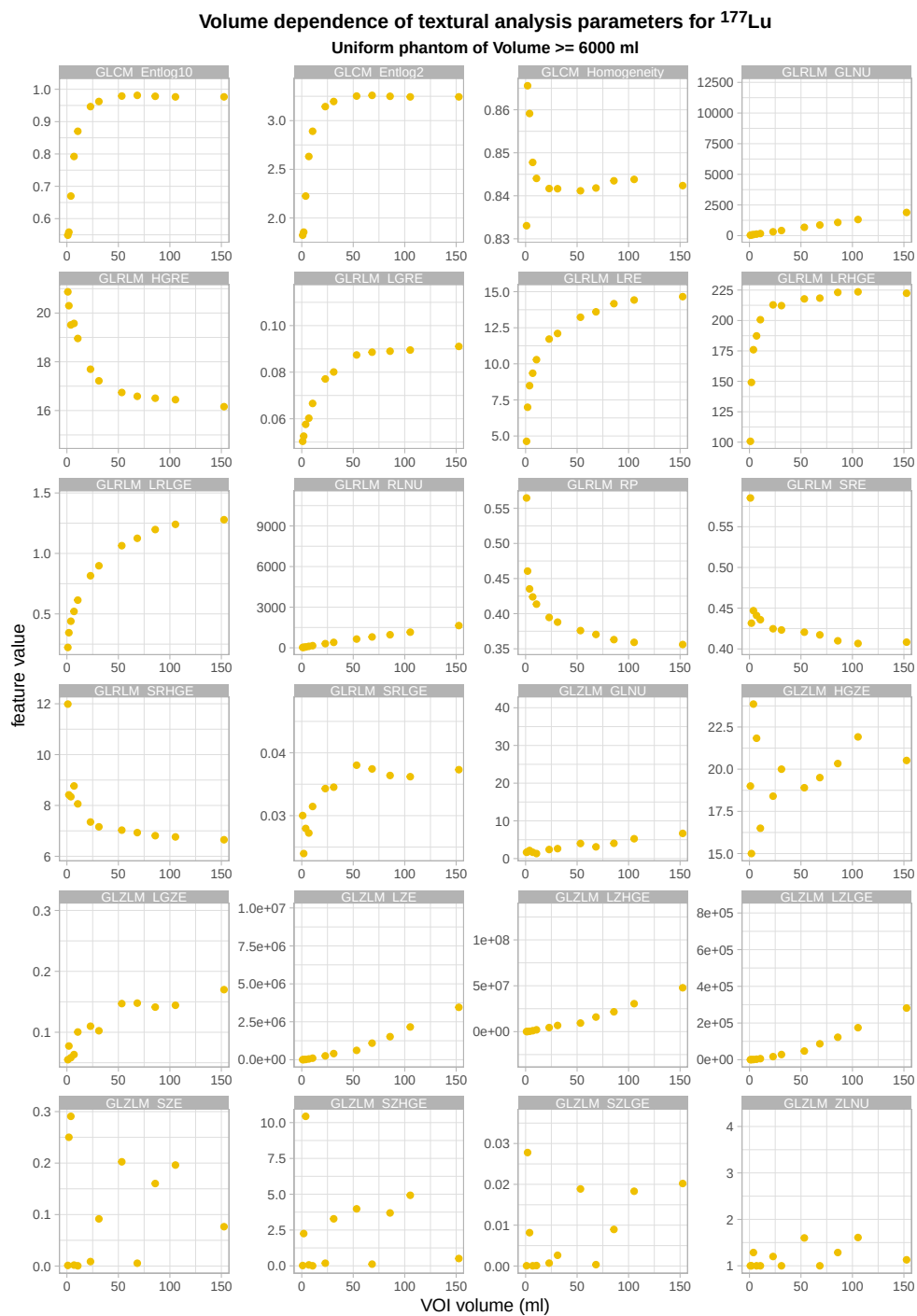


Figure B.2: Volume dependence of textural analysis parameters for ^{177}Lu .

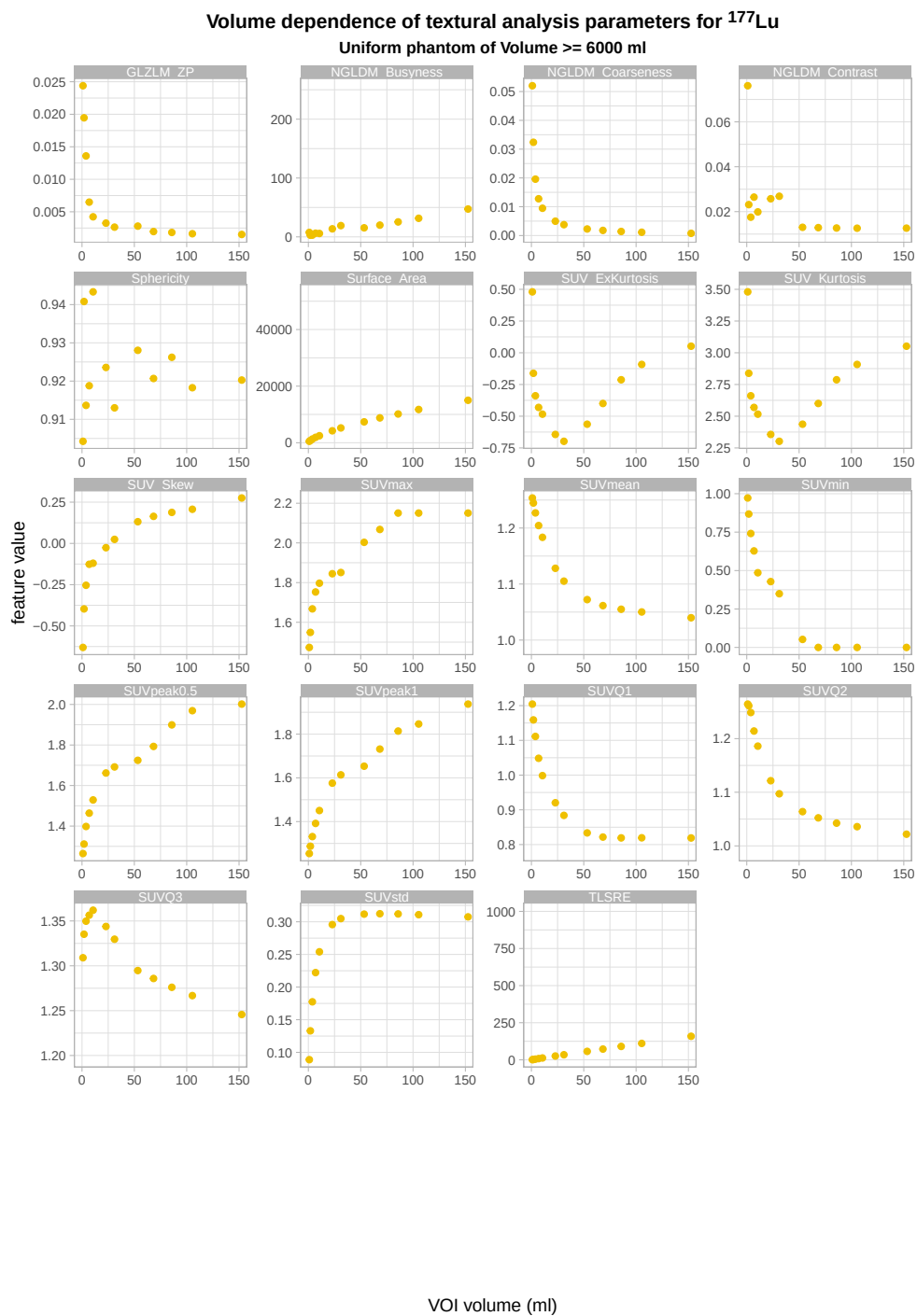


Figure B.2: Volume dependence of textural analysis parameters for ^{177}Lu .

Appendix C

List of Textural Analysis features used in LifeX v6.3

In addition to the features listed in Table C.1, there are 19 discretised variants of features in the SUV and Histogram classes. A full mathematical description for all of the features can be seen in the LifeX v6.3 manual and online at [405].

Notation	Description
SUVmin	The minimum Standardised Uptake Value in the volume
SUVmean	The mean Standardised Uptake Value in the volume
SUVstd	The standard deviation of the Standardised Uptake Value in the volume
SUVmax	The maximum Standardised Uptake Value in the volume
SUVQ1	The 1st quartile of the Standardised Uptake Values in the volume
SUVQ2	The 2nd quartile of the Standardised Uptake Values in the volume
SUVQ3	The 3rd quartile of the Standardised Uptake Values in the volume
SUV Skewness	The asymmetry of the intensity distribution
SUV Kurtosis	Measure of "peakedness" in the intensity distribution
SUV Excess Kurtosis	Measure of "peakedness" in the intensity distribution, normalised against the normal distribution
SUVpeak 0.5ml	The mean SUV of a 0.5ml sphere centred around the maximum pixel
SUVpeak 1ml	The mean SUV of a 1ml sphere centred around the maximum pixel
TLG	Total Lesion Glycolysis - when used with SRI, becomes Total Lesion Somatostatin Receptor Expression (TLSRE)
Histogram Skewness	The asymmetry of the grey level distribution in the histogram

Notation	Description
Histogram Kurtosis	Describes the shape (peaked versus flatness) of the grey level distribution to a normal distribution
Histogram Excess Kurtosis	Compares the shape of the grey level distribution to the normal distribution
Histogram Entropy log10	Randomness of the grey level distribution
Histogram Entropy log2	Randomness of the grey level distribution
Histogram Energy/Uniformity	The uniformity of the grey level distribution
Volume	The volume in ml of the VOI
Sphericity	A measure of how close to a perfect sphere (=1) the VOI is
Surface Area	Surface area of the VOI
Compacity	How compact the VOI is, based on surface and volume of the VOI
GLCM Homogeneity	Grey-Level Co-occurrence Matrix, homogeneity of voxel pairs
GLCM Energy	Grey-Level Co-occurrence Matrix, Uniformity or Second Angular Moment
GLCM Contrast	Grey-Level Co-occurrence Matrix, local variations in the matrix (also called Variance or Inertia)
GLCM Correlation	Grey-Level Co-occurrence Matrix, linear dependency of grey levels
GLCM Entropy log10	Grey-Level Co-occurrence Matrix, randomness of the voxel pairs
GLCM Entropy log2	Grey-Level Co-occurrence Matrix, randomness of the voxel pairs
GLCM Dissimilarity	Grey-Level Co-occurrence Matrix, variation of the voxel pairs
GLRLM SRE	Grey-Level Run Length Matrix, Short-Run Emphasis
GLRLM LRE	Grey-Level Run Length Matrix, Long-Run Emphasis
GLRLM LGRE	Grey-Level Run Length Matrix, Low Grey-level Run Emphasis
GLRLM HGRE	Grey-Level Run Length Matrix, High Grey-level Run Emphasis
GLRLM SRLGE	Grey-Level Run Length Matrix, Short Run Low Grey-level Run Emphasis
GLRLM SRHGE	Grey-Level Run Length Matrix, Short Run High Grey-level Run Emphasis

Notation	Description
GLRLM LRLGE	Grey-Level Run Length Matrix, Long Run Low Grey-level Run Emphasis
GLRLM LRHGE	Grey-Level Run Length Matrix, Long Run High Grey-level Run Emphasis
GLRLM GLNU	Grey-Level Run Length Matrix, Grey-Level Non-Uniformity
GLRLM RLNU	Grey-Level Run Length Matrix, Run-Level Non-Uniformity
GLRLM RP	Grey-Level Run Length Matrix, Run Percentage
NGLDM Coarseness	Neighbourhood Grey-Level Difference Matrix, rate of change of intensity
NGLDM Contrast	Neighbourhood Grey-Level Difference Matrix, intensity difference between neighbouring regions
NGLDM Busyness	Neighbourhood Grey-Level Difference Matrix, spatial frequency of change in intensity
GLZLM SZE	Grey-Level Zone Length Matrix, Short-Zone Emphasis
GLZLM LZE	Grey-Level Zone Length Matrix, Long-Zone Emphasis
GLZLM LGZE	Grey-Level Zone Length Matrix, low grey level zone emphasis
GLZLM HGZE	Grey-Level Zone Length Matrix, high grey level zone emphasis
GLZLM SZLGE	Grey-Level Zone Length Matrix, short zone low grey level emphasis
GLZLM SZHGE	Grey-Level Zone Length Matrix, short zone high grey level emphasis
GLZLM LZLGE	Grey-Level Zone Length Matrix, long zone low grey level emphasis
GLZLM LZHGE	Grey-Level Zone Length Matrix, long zone high grey level emphasis
GLZLM GLNU	Grey-Level Zone Length Matrix, grey level non-uniformity
GLZLM ZLNU	Grey-Level Zone Length Matrix, zone length non-uniformity
GLZLM ZP	Grey-Level Zone Length Matrix, zone percentage, the homogeneity of the homogeneous zones

Table C.1: Abbreviation and description of textural analysis features from LifeX version 6.3 used in the thesis.

Appendix D

Detailed results, Chapter 15

The statistical harmonisation process described in Chapter 15 using ComBat had many combinations of parameters that could be tested:

- Location-Scale (L-S) method, or L-S with Bayesian adjustment
- The use of co-variates, or not
- The use of a reference site to anchor the results, or not

In addition, the measurements made in the patient could be classified by the organ or disease site being measured, as it was demonstrated in Chapter 9 that uptake varies in different organs. Only if there were at least 10 examples per site was a test performed to harmonise the data, which resulted in the following being tested:

- Physiological liver uptake
- Physiological spleen uptake
- Physiological blood pool uptake
- Liver metastases
- Nodal metastases
- Small bowel primary

The following tables demonstrate the post-harmonisation results in terms of the Mann Whitney U test p-value for each combination of parameters, as well as compare the pre-harmonisation p-value. A significantly different p-value (<0.05) is marked by an asterisk. The average p-value across all the features, as well as the number of features that showed a significant difference is displayed, to allow comparison of the different methods to determine the most effective. The findings are discussed in Chapter 15.

<i>Physiological liver Features</i>	<i>Raw data</i>	<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>Bayes</i>		<i>Bayes</i>		<i>Bayes</i>	
		no cov	no ref	cov	no ref	No cov	Cov	no cov	Cov	no ref	Cov	No Cov	ref site	ref site	Cov
SUVmin	0*	0.86		0.95		0.75	0.96	0.17	0.17		0.17	0.47		0.6	0.6
SUVmean	0.63	0.83		0.77		0.84	0.79	0.66	0.42		0.42	0.41		0.26	0.26
SUVmax	0.00053*	0.87		0.62		0.87	0.65	0.43	0.17		0.17	0.69		0.3	0.3
SUVpeak	0.0047*	0.85		0.64		0.84	0.69	0.9	0.49		0.49	0.57		0.91	0.91
Disc SUV	0.36	0.82		0.75		0.82	0.76	0.43	0.45		0.45	0.5		0.53	0.53
Skewness															
Disc SUV	0.28	0.93		0.84		0.93	0.83	0.6	0.73		0.73	0.91		0.99	0.99
Kurtosis															
GLCM Energy	0.0001*	0.89		0.87		0.91	0.38	0.32	0.92		0.92	0.0001*		0.059	0.059
GLCM	0.00033*	0.85		0.88		0.84	0.87	0.39	0.57		0.57	0.96		1	1
Correlation															
GLCM Entropy	0.0001*	0.95		0.45		0.96	0.4	0.00013*	0.0001*		0.0001*	0.0001*		0.0001*	0.0001*
log2															
GLRLM LRE	0.0001*	0.75		0.71		0.76	0.57	0.00013*	0.00016*		0.00016*	0.0001*		0.0001*	0.0001*
GLRLM SRHGE	0.018*	0.94		0.56		0.95	0.62	0.66	1		1	0.1		0.3	0.3
GLRLM LRLGE	0.0009*	0.73		0.85		0.68	0.87	0.021*	0.63		0.63	0.025*		0.51	0.51
GLZLM ZP	0.0001*	0.61		0.5		0.62	0.46	0.0001*	0.0001*		0.0001*	0.0001*		0.0001*	0.0001*
TLSRE	n/a	n/a		n/a		n/a	n/a	n/a	n/a		n/a	n/a		n/a	n/a
Average p-value	0.100	0.837		0.722		0.828	0.681	0.352	0.427		0.427	0.357		0.420	0.420
number with	10	0		0		0	0	4	3		3	5		3	3
significant diff															

Table D.1: p-values for harmonised data from Hospitals A and B, for physiological liver uptake. An asterisk denotes $p < 0.05$.

<i>Physiological spleen Features</i>	<i>Raw data</i>	<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>Bayes</i>		<i>Bayes</i>		<i>Bayes</i>	
		no cov	cov	No cov	Cov	no cov	Cov	no ref	Cov	No Cov	Cov	no ref	Cov
		no ref	no ref	ref site	ref site	no ref	ref site	no ref	ref site	ref site	ref site	no ref	ref site
SUVmin	0.17	0.83	0.95	0.83	0.97	0.41	0.64	0.64	0.91	0.64	0.91	0.64	0.91
SUVmean	0.14	0.81	0.95	0.81	0.93	0.88	0.90	0.90	0.53	0.68	0.53	0.68	0.53
SUVmax	0.03*	0.87	0.95	0.86	0.97	0.45	0.47	0.47	0.60	0.61	0.60	0.61	0.60
SUVpeak	0.05	0.83	0.99	0.83	0.99	0.63	0.64	0.64	0.91	0.88	0.91	0.88	0.91
Disc SUV	0.10	0.67	0.14	0.67	0.17	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*
Skewness													
Disc SUV	0.07	0.021*	0.35	0.021*	0.48	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*
Kurtosis													
GLCM Energy	0.18	0.94	0.75	0.94	0.77	0.27	0.17	0.17	0.27	0.49	0.27	0.49	0.27
GLCM	0.0071*	0.99	0.65	0.98	0.68	0.08	0.06	0.06	0.16	0.21	0.16	0.21	0.16
Correlation													
GLCM Entropy	0.22	0.98	0.78	0.99	0.77	0.58	0.51	0.51	0.61	0.77	0.61	0.77	0.61
log2													
GLRLM LRE	0.81	0.52	0.24	0.53	0.25	0.93	0.46	0.46	0.52	0.76	0.52	0.76	0.52
GLRLM SRHGE	0.17	0.33	0.24	0.32	0.24	0.53	1.00	1.00	0.10	0.32	0.10	0.32	0.10
GLRLM LRLGE	0.83	0.46	0.21	0.48	0.22	0.87	0.44	0.44	0.52	0.77	0.52	0.77	0.52
GLZLM ZP	0.08	0.96	0.74	0.97	0.77	0.97	0.90	0.90	0.89	0.74	0.89	0.74	0.89
TLSRE	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Average	0.220	0.709	0.611	0.710	0.632	0.508	0.476	0.476	0.463	0.528	0.463	0.528	0.463
number with	2	1	0	1	0	2	2	2	2	2	2	2	2
significant diff													

Table D.2: p-values for harmonised data from Hospitals A and B, for physiological spleen uptake. An asterisk denotes $p < 0.05$.

<i>Physiological Blood Pool Features</i>	<i>Raw data</i>	<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>Bayes</i>		<i>Bayes</i>		<i>Bayes</i>	
		no cov	cov	no cov	cov	no cov	cov	no cov	cov	no cov	cov	no cov	cov
		no ref	no ref	ref site	ref site	no ref	ref site	no ref	ref site	no ref	ref site	no ref	ref site
SUVmin	0.0001*	0.0001*	0.012*	0.0001*	0.0032*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*	0.0001*
SUVmean	0.91	1.00	0.70	1.00	0.71	0.14	0.28	0.27	0.33	0.27	0.33	0.27	0.33
SUVmax	0.0001*	0.85	0.80	0.84	0.81	0.48	0.24	0.60	0.30	0.60	0.30	0.60	0.30
SUVpeak	0.0001*	0.98	0.76	0.98	0.75	0.92	0.84	0.65	0.88	0.65	0.88	0.65	0.88
Disc SUV	0.0015*	0.99	0.76	0.99	0.76	0.13	0.06	0.12	0.042*	0.12	0.042*	0.12	0.042*
Skewness													
Disc SUV	0.013*	0.740	0.94	0.730	0.96	0.81	0.61	0.90	0.45	0.90	0.45	0.90	0.45
Kurtosis													
GLCM Energy	0.36	0.96	0.77	0.97	0.77	0.40	0.80	0.57	0.92	0.57	0.92	0.57	0.92
GLCM	0.046*	0.93	0.74	0.93	0.73	0.0001*	0.0001*	0.0001*	0.03	0.0001*	0.03	0.0001*	0.03
Correlation													
GLCM Entropy	0.027*	0.96	0.92	0.96	0.93	0.63	0.52	0.54	0.35	0.54	0.35	0.54	0.35
log2													
GLRLM LRE	0.0001*	0.99	0.37	1.00	0.34	0.00	0.00	0.01	0.08	0.01	0.08	0.01	0.08
GLRLM SRHGE	0.0001*	0.83	0.41	0.85	0.41	0.46	0.67	0.13	0.27	0.13	0.27	0.13	0.27
GLRLM LRLGE	0.0047*	0.88	0.32	0.87	0.31	0.0001*	0.0001*	0.00	0.00	0.00	0.00	0.00	0.00
GLZLM ZP	0.0001*	0.83	0.70	0.83	0.69	0.03	0.00	0.09	0.10	0.09	0.10	0.09	0.10
TLSRE	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Average	0.105	0.842	0.631	0.842	0.629	0.307	0.309	0.298	0.289	0.298	0.289	0.298	0.289
number with	11	1	1	1	1	3	3	2	2	2	2	2	2
significant diff													

Table D.3: p-values for harmonised data from Hospitals A and B, for physiological blood pool. An asterisk denotes $p < 0.05$.

<i>Liver metastases Features</i>	<i>Raw data</i>	<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>Bayes</i>		<i>Bayes</i>		<i>Bayes</i>	
		no cov	no ref	cov	no ref	No cov	ref site	Cov	ref site	no cov	no ref	Cov	no ref	No Cov	ref site
SUVmin	0.073	0.51	0.07	0.07	0.50	0.50	0.07	0.07	0.24	0.048*	0.25	0.047*	0.25	0.047*	0.047*
SUVmean	0.051	0.58	0.40	0.40	0.57	0.57	0.42	0.42	0.99	0.25	0.97	0.22	0.97	0.22	0.22
SUVmax	0.1	0.74	0.42	0.42	0.74	0.74	0.45	0.45	0.74	0.52	0.68	0.56	0.52	0.68	0.56
SUVpeak	0.093	0.74	0.47	0.47	0.74	0.74	0.51	0.51	0.77	0.49	0.73	0.51	0.49	0.73	0.51
Disc SUV	0.1	0.07	0.0058*	0.0058*	0.07	0.07	0.0058*	0.0058*	0.11	0.46	0.18	0.29	0.46	0.18	0.29
Skewness															
Disc SUV	0.023*	0.0061*	0.00066*	0.00066*	0.0061*	0.0061*	0.00018*	0.00018*	0.0058*	0.0082*	0.0058*	0.0058*	0.0082*	0.0058*	0.0058*
Kurtosis															
GLCM Energy	0.18	0.23	0.91	0.91	0.240	0.240	0.89	0.89	0.270	0.65	0.220	0.67	0.65	0.220	0.67
GLCM	0.460	0.81	0.97	0.97	0.81	0.81	0.97	0.97	0.730	0.900	0.760	0.950	0.900	0.760	0.950
Correlation															
GLCM Entropy	0.38	0.61	0.62	0.62	0.65	0.65	0.62	0.62	0.40	0.59	0.95	0.52	0.59	0.95	0.52
log2															
GLRLM LRE	0.051	0.17	0.23	0.23	0.170	0.170	0.23	0.23	0.210	0.24	0.180	0.23	0.24	0.180	0.23
GLRLM SRHGE	0.055	0.89	0.81	0.81	0.90	0.90	0.81	0.81	0.28	0.23	0.40	0.33	0.23	0.40	0.33
GLRLM LRLGE	0.75	0.28	0.25	0.25	0.28	0.28	0.25	0.25	0.24	0.007*	0.21	0.00062*	0.007*	0.21	0.00062*
GLZLM ZP	0.910	0.92	0.96	0.96	0.94	0.94	0.98	0.98	0.680	0.970	0.610	0.930	0.970	0.610	0.930
TLSRE	0.52	0.50	0.52	0.52	0.5	0.5	0.55	0.55	0.26	0.24	0.43	0.24	0.24	0.43	0.24
Average	0.268	0.504	0.474	0.474	0.509	0.509	0.483	0.483	0.423	0.400	0.470	0.393	0.400	0.470	0.393
number with	1	1	2	2	1	1	2	2	1	3	1	3	3	1	3
significant diff															

Table D.4: p-values for harmonised data from Hospitals A and B, for liver metastases. An asterisk denotes $p < 0.05$.

<i>Nodal metastases Features</i>	<i>Raw data</i>	<i>L-S</i>		<i>L-S</i>		<i>L-S</i>		<i>Bayes no cov</i>		<i>Bayes Cov</i>		<i>Bayes No Cov</i>		<i>Bayes Cov</i>	
		no cov	no ref	cov	no ref	No cov	Cov	no ref	no ref	no ref	ref site	no ref	ref site	no ref	ref site
SUVmin	0.56	0.79	0.90	0.90	0.79	0.79	0.91	0.60	0.72	0.56	0.66	0.72	0.56	0.72	0.66
SUVmean	0.81	0.84	1.00	1.00	0.81	0.81	0.99	0.88	0.80	0.83	0.83	0.80	0.83	0.80	0.83
SUVmax	0.92	0.80	1.00	1.00	0.81	0.81	0.99	0.98	0.62	1.00	0.62	0.62	1.00	0.62	0.62
SUVpeak	0.82	0.79	0.97	0.97	0.77	0.77	0.98	0.83	0.87	0.80	0.88	0.87	0.80	0.87	0.88
Disc SUV	0.23	0.97	0.66	0.66	0.99	0.99	0.63	0.59	0.16	0.08	0.05	0.16	0.08	0.16	0.05
Skewness															
Disc SUV	0.12	0.500	0.76	0.76	0.500	0.500	0.76	0.0001*	0.75	0.0001*	0.75	0.75	0.0001*	0.75	0.75
Kurtosis															
GLCM Energy	0.15	0.0083*	0.77	0.77	0.0083*	0.0083*	0.78	0.048*	0.90	0.0083*	0.78	0.90	0.0083*	0.78	0.78
GLCM	0.0019*	0.94	0.93	0.93	0.93	0.93	0.94	0.012*	0.0024*	0.014*	0.0018*	0.0024*	0.014*	0.0024*	0.0018*
Correlation															
GLCM Entropy	0.14	1.00	0.95	0.95	0.98	0.98	0.98	0.09	0.08	0.08	0.07	0.08	0.08	0.08	0.07
log2															
GLRLM LRE	0.34	0.027*	0.44	0.44	0.027*	0.027*	0.43	0.0091*	0.56	0.0065*	0.57	0.56	0.0065*	0.56	0.57
GLRLM SRHGE	0.67	0.80	0.94	0.94	0.80	0.80	0.98	0.99	0.87	0.98	0.94	0.87	0.98	0.87	0.94
GLRLM LRLGE	0.41	0.62	0.43	0.43	0.63	0.63	0.41	0.18	0.11	0.28	0.13	0.11	0.28	0.11	0.13
GLZLM ZP	0.048*	0.97	0.93	0.93	0.99	0.99	0.91	0.03*	0.015*	0.027*	0.014*	0.015*	0.027*	0.015*	0.014*
TLSRE	0.32	0.88	0.67	0.67	0.87	0.87	0.68	0.34	0.61	0.32	0.6	0.61	0.32	0.61	0.6
Average	0.396	0.710	0.811	0.811	0.708	0.708	0.812	0.398	0.505	0.356	0.492	0.505	0.356	0.492	0.492
number with significant diff	2	2	0	0	2	2	0	5	2	5	2	2	5	2	2

Table D.5: p-values for harmonised data from Hospitals A and B, for nodal metastases. An asterisk denotes $p < 0.05$.

Bibliography

- [1] S. Carlson, "A glance at the history of nuclear medicine," *Acta Oncologica*, vol. 34, no. 8, pp. 1095–1102, 1995, ISSN: 0284186X. DOI: 10.3109/02841869509127236
- [2] B. Cassen, L. Curtis, and C. W. Reed, "A sensitive directional gamma-ray detector," *Nucleonics*, vol. 2, pp. 78–81, 1950.
- [3] H. Anger, "A new instrument for mapping gamma-ray emitters," *Biology and Medicine Quarterly Report*, vol. 3653, no. January, p. 38, 1957.
- [4] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors," *Journal of Nuclear Medicine*, vol. 50, no. SUPPL. 1, pp. 122–150, 2009, ISSN: 01615505. DOI: 10.2967/jnumed.108.057307
- [5] N. Avril et al., "Prediction of response to neoadjuvant chemotherapy by sequential F-18-fluorodeoxyglucose positron emission tomography in patients with advanced-stage ovarian cancer," *Journal of Clinical Oncology*, vol. 23, no. 30, pp. 7445–7453, 2005, ISSN: 0732183X. DOI: 10.1200/JCO.2005.06.965
- [6] U. Tateishi, C. Gamez, S. Dawood, H. W. Yeung, M. Cristofanilli, and H. A. Macapinlac, "Bone metastases in patients with metastatic breast cancer: Morphologic and metabolic monitoring of response to systemic therapy with integrated PET/CT," *Radiology*, vol. 247, no. 1, pp. 189–196, 2008, ISSN: 00338419. DOI: 10.1148/radiol.2471070567
- [7] R. S. Lawson, "Gamma Camera SPECT," in *Practical SPECT/CT in Nuclear Medicine*, D. W. Jones, P. Hogg, and E. Seeram, Eds., 1st Editio, London: Springer London, 2013, ch. 4, pp. 47–76, ISBN: 978-1-4471-4702-2. DOI: 10.1007/978-1-4471-4703-9
- [8] A. Yordanova et al., "Theranostics in nuclear medicine practice," *OncoTargets and Therapy*, pp. 10–4821, 2017.
- [9] S. Frangos and J. R. Buscombe, "Why should we be concerned about a "g"?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, no. 2, p. 519, 2019, ISSN: 16197089. DOI: 10.1007/s00259-018-4204-z
- [10] U. Eberlein, M. Cremonesi, and M. Lassmann, "Individualized dosimetry for theranostics: Necessary, nice to have, or counterproductive?" *Journal of Nuclear*

- Medicine*, vol. 58, 97S–103S, 2017, ISSN: 2159662X. DOI: 10.2967/jnumed.116.186841
- [11] J. C. Dickson et al., “EANM practice guideline for quantitative SPECT-CT,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 50, no. 4, pp. 980–995, Mar. 2023, ISSN: 1619-7070. DOI: 10.1007/s00259-022-06028-9
- [12] D. L. Bailey and K. P. Willowson, “Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, no. SUPPL. 1, 2014, ISSN: 16197089. DOI: 10.1007/s00259-013-2542-4
- [13] J. Dickson, J. Ross, and S. Vöö, “Quantitative SPECT: the time is now,” *EJNMMI Physics*, vol. 6, no. 1, pp. 1–7, 2019, ISSN: 21977364. DOI: 10.1186/s40658-019-0241-3
- [14] J. C. Villforth, “Problems in Radium Control,” *Public health reports*, vol. 79, no. 4, pp. 337–342, 1964, ISSN: 00946214. DOI: 10.2307/4592124
- [15] B. Cassen, L. Curtis, C. Reed, and R. Libby, “Instrumentation for ¹³¹I use in medical studies,” *Nucleonics*, vol. 9, pp. 46–50, 1951.
- [16] P. Richards, W. D. Tucker, and S. C. Srivastava, “Technetium-99m: An historical perspective,” *The International Journal Of Applied Radiation And Isotopes*, vol. 33, no. 10, pp. 793–799, 1982, ISSN: 0020708X. DOI: 10.1016/0020-708X(82)90120-X
- [17] W. C. Eckelman, “Unparalleled Contribution of Technetium-99m to Medicine Over 5 Decades,” *JACC: Cardiovascular Imaging*, vol. 2, no. 3, pp. 364–368, 2009, ISSN: 1936878X. DOI: 10.1016/j.jcmg.2008.12.013
- [18] J. G. Graham L. Stephen. Kereiakes, C. Harris, and M. B. Cohen, “From Nuclear Becquerel Medicine to the Present,” *RadioGraphics*, vol. 9, no. 6, p. 14, 1989.
- [19] J. Jaszczak, E. Floyd, and C. Craig, “Improved SPECT Quantification Using Compensation for Scattered Photons,” *Journal of Nuclear Medicine*, vol. 25, no. 8, pp. 893–900, 1984.
- [20] A. Kojima, M. Matsumoto, M. Takahashi, Y. Hirota, and H. Yoshida, “Effect of spatial resolution on SPECT quantification values,” *Journal of Nuclear Medicine*, vol. 30, no. 4, pp. 508–514, 1989, ISSN: 01615505.
- [21] K. Lange and R. Carson, “EM reconstruction algorithms for emission and transmission tomography,” *Journal of Computer Assisted Tomography*, vol. 8, no. 2, pp. 306–316, 1984, ISSN: 03638715.
- [22] H. N. Wagner, “Clinical PET: Its time has come?” *Journal of Nuclear Medicine*, vol. 32, pp. 561–564, 1991, ISSN: 01615505.
- [23] W. A. Weber, S. I. Ziegler, R. Thödtmann, A. R. Hanauske, and M. Schwaiger, “Reproducibility of metabolic measurements in malignant tumors using FDG PET,” *Journal of Nuclear Medicine*, vol. 40, no. 11, pp. 1771–1777, 1999, ISSN: 01615505.

- [24] G. Lucignani, G. Paganelli, and E. Bombardieri, "The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective," *Nuclear Medicine Communications*, vol. 25, no. 7, 2004, ISSN: 0143-3636.
- [25] M. Bocher et al., "Gamma camera-mounted anatomical X-ray tomography: Technology, system characteristics and first images," *European Journal of Nuclear Medicine*, vol. 27, no. 6, pp. 619–627, 2000, ISSN: 03406997. DOI: 10.1007/s002590050555
- [26] M. T. Madsen, "Recent advances in SPECT imaging," *Journal of Nuclear Medicine*, vol. 48, no. 4, pp. 661–673, 2007, ISSN: 01615505. DOI: 10.2967/jnumed.106.032680
- [27] S. Shcherbinin, A. Celler, T. Belhocine, R. Vanderwerf, and A. Driedger, "Accuracy of quantitative reconstructions in SPECT/CT imaging," *Physics in Medicine and Biology*, vol. 53, no. 17, pp. 4595–4604, 2008, ISSN: 00319155. DOI: 10.1088/0031-9155/53/17/009
- [28] K. P. Willowson, D. L. Bailey, and C. Baldock, "Quantitative SPECT reconstruction using CT-derived corrections," *Physics in Medicine and Biology*, vol. 53, no. 12, pp. 3099–3112, 2008, ISSN: 00319155. DOI: 10.1088/0031-9155/53/12/002
- [29] J. Zeintl, A. H. Vija, A. Yahil, J. Horneegger, and T. Kuwert, "Quantitative accuracy of clinical 99mTc SPECT/CT using ordered-subset expectation maximization with 3-dimensional resolution recovery, attenuation, and scatter correction," *Journal of Nuclear Medicine*, vol. 51, no. 6, pp. 921–928, 2010, ISSN: 01615505. DOI: 10.2967/jnumed.109.071571
- [30] D. L. Bailey and K. P. Willowson, "An evidence-based review of quantitative SPECT imaging and potential clinical applications," *Journal of Nuclear Medicine*, vol. 54, no. 1, pp. 83–89, 2013, ISSN: 01615505. DOI: 10.2967/jnumed.112.111476
- [31] J. C. Dickson, "Quantitative SPECT: a survey of current practice in the UK Nuclear Medicine Community," *Nuclear medicine communications*, vol. 40, no. 10, pp. 986–994, 2019, ISSN: 14735628. DOI: 10.1097/MNM.0000000000001059
- [32] N. Bebbington, P. Anderson, and L. Jenkins, "Quantitative accuracy of SPECT reconstructions using Hermes SUV SPECT software for commonly used radionuclides (poster)," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, no. Suppl 2, S328, 2014, ISSN: 1619-7089. DOI: 10.1007/s00259-014-2901-9
- [33] I. S. Armstrong and S. A. Hoffmann, "Activity concentration measurements using a conjugate gradient (Siemens xSPECT) reconstruction algorithm in SPECT/CT," *Nuclear Medicine Communications*, vol. 37, no. 11, pp. 1212–1217, 2016, ISSN: 14735628. DOI: 10.1097/MNM.0000000000000586
- [34] S. Gnesin, P. Leite Ferreira, J. Malterre, P. Laub, J. O. Prior, and F. R. Verdun, "Phantom Validation of Tc-99m Absolute Quantification in a SPECT/CT Commercial

- Device,” *Computational and Mathematical Methods in Medicine*, vol. 2016, 2016, ISSN: 17486718. DOI: 10.1155/2016/4360371
- [35] N. Bebbington, P. Holdgaard, and H. Larsen, “Accuracy, Repeatability and Reproducibility of xSPECT Quant Sensitivity Calibrations using a NIST-traceable Point Source (poster),” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. S2, S388, Oct. 2017, ISSN: 1619-7070. DOI: 10.1007/s00259-017-3822-1
- [36] Siemens, “Siemens White Paper: SPECT/CT - Overview of Quantification,” Siemens Medical Solutions, Inc, Tech. Rep., 2018.
- [37] P. F. Sharp and K. A. Goatman, “Nuclear Medicine Imaging,” in *Practical Nuclear Medicine*, P. F. Sharp, H. G. Gemmell, and A. D. Murray, Eds., 3rd Edition, London: Springer London, 2005, ch. 1, pp. 1–20, ISBN: 978-1-85233-875-6.
- [38] Y. Yamamoto et al., “Improvement in the Quantification of Striatal Tracer Uptake in Single-photon Emission Computed Tomography With 123I-ioflupane Using a Cadmium-zinc-telluride Semiconductor Camera,” *Shimane J. Med. Sci*, vol. 34, pp. 35–40, 2017.
- [39] D. Bellevre et al., “Quantitation of myocardial 99mTc-HMDP uptake with new SPECT/CT cadmium zinc telluride (CZT) camera in patients with transthyretin-related cardiac amyloidosis: Ready for clinical use?” *Journal of Nuclear Cardiology*, vol. 29, no. 2, pp. 506–514, 2022, ISSN: 15326551. DOI: 10.1007/s12350-020-02274-2
- [40] S. L. Heller and P. N. Goodwin, “SPECT instrumentation: Performance, lesion detection, and recent innovations,” *Seminars in Nuclear Medicine*, vol. 17, no. 3, pp. 184–199, 1987, ISSN: 00012998. DOI: 10.1016/S0001-2998(87)80033-8
- [41] A. Khorshidi, “Assessment of SPECT images using UHRFB and other low-energy collimators in brain study by Hoffman phantom and manufactured defects,” *European Physical Journal Plus*, vol. 135, no. 2, pp. 1–19, 2020, ISSN: 21905444. DOI: 10.1140/epjp/s13360-020-00238-6
- [42] Quantitative Imaging Biomarkers Alliance, “SPECT-CT: Estimating the Concentration of 99mTc Based Imaging Biomarkers in Large and Small Volumes of Interest Profile 2019-10-21,” Quantitative Imaging Biomarkers Alliance, Tech. Rep., 2019, pp. 1–65.
- [43] Y. K. Dewaraja et al., “MIRD Pamphlet No. 24: Guidelines for Quantitative 131 I SPECT in Dosimetry Applications,” *Journal of Nuclear Medicine*, vol. 54, no. 12, pp. 2182–2188, Dec. 2013, ISSN: 0161-5505. DOI: 10.2967/jnumed.113.122390
- [44] D. M. Huizing, M. Sinaasappel, M. C. Dekker, M. P. Stokkel, and B. J. de Wit – van der Veen, “177Lutetium SPECT/CT: Evaluation of collimator, photopeak and scatter correction,” *Journal of Applied Clinical Medical Physics*, vol. 21, no. 9, pp. 272–277, 2020, ISSN: 15269914. DOI: 10.1002/acm2.12991

- [45] J. A. Biesz and E. G. Hawman, "Evaluation of SPECT angular sampling effects: Continuous versus step-and-shoot acquisition," *Journal of Nuclear Medicine*, vol. 28, no. 8, pp. 1308–1314, 1987, ISSN: 01615505.
- [46] V. Picone, N. Makris, F. Boutevin, S. Roy, M. Playe, and M. Soussan, "Clinical validation of time reduction strategy in continuous step-and-shoot mode during SPECT acquisition," *EJNMMI Physics*, vol. 8, no. 1, p. 10, Dec. 2021, ISSN: 2197-7364. DOI: 10.1186/s40658-021-00354-x
- [47] Z. J. Cao, C. Maunoury, C. C. Chen, and L. E. Holder, "Comparison of continuous step-and-shoot versus step-and-shoot acquisition SPECT," *Journal of Nuclear Medicine*, vol. 37, no. 12, pp. 2037–2040, 1996, ISSN: 01615505.
- [48] H. G. Gemmell and R. T. Staff, "Single Photon Emission Computed Tomography (SPECT)," in *Practical Nuclear Medicine*, P. F. Sharp, H. G. Gemmell, and A. D. Murray, Eds., Third Edit, London: Springer, 2005, ch. 2, pp. 21–33, ISBN: 978-1-85233-875-6.
- [49] M. W. Groch and W. D. Erwin, "SPECT in the year 2000: Basic principles," *Journal of Nuclear Medicine Technology*, vol. 28, no. 4, pp. 233–244, 2000, ISSN: 00914916.
- [50] P. Knoll et al., "Comparison of advanced iterative reconstruction methods for SPECT/CT," *Zeitschrift für Medizinische Physik*, vol. 22, no. 1, pp. 58–69, 2012, ISSN: 09393889. DOI: 10.1016/j.zemedi.2011.04.007
- [51] H. Vija, "Introduction to xSPECT Technology: Evolving Multi-modal SPECT to Become Context-based and Quantitative," Siemens Medical Solutions, Inc., Tech. Rep., 2013, MI-1138.TM.JV.
- [52] J. Zeintl, A. H. Vija, A. Yahil, X. Ding, J. Hornegger, and T. Kuwert, "Towards quantitative SPECT: Error estimation of SPECT OSEM with 3D resolution recovery, attenuation correction and scatter correction," *IEEE Nuclear Science Symposium Conference Record*, pp. 4106–4111, 2008, ISSN: 10957863. DOI: 10.1109/NSSMIC.2008.4774185
- [53] M. Tsujimoto et al., "Fluctuation of quantitative values on acquisition time and the reconstruction conditions in ^{99m}Tc-SPECT," *Nuclear Medicine Communications*, vol. 39, no. 7, pp. 601–609, 2018, ISSN: 14735628. DOI: 10.1097/MNM.0000000000000854
- [54] T. S. Kangasmaa, C. Constable, E. Hippeläinen, and A. O. Sohlberg, "Multicenter evaluation of single-photon emission computed tomography quantification with third-party reconstruction software," *Nuclear Medicine Communications*, vol. 37, no. 9, pp. 983–987, 2016, ISSN: 14735628. DOI: 10.1097/MNM.0000000000000538
- [55] A. H. Vija, "Characteristics of the xSPECT reconstruction method," Siemens Healthineers, Tech. Rep., 2017.
- [56] P. P. Bruyant, "Analytic and iterative reconstruction algorithms in SPECT," *Journal of Nuclear Medicine*, vol. 43, no. 10, pp. 1343–1358, 2002, ISSN: 01615505.

- [57] J. Tran-Gia and M. Lassmann, "Characterization of noise and resolution for quantitative ^{177}Lu SPECT/CT with XSPECT quant," *Journal of Nuclear Medicine*, vol. 60, no. 1, pp. 50–59, 2019, ISSN: 2159662X. DOI: 10.2967/jnumed.118.211094
- [58] H. Zaidi and B. Hasegawa, "Determination of the Attenuation Map in Emission Tomography," *Journal of Nuclear Medicine*, vol. 44, no. 2, 291 LP –315, Feb. 2003, ISSN: 0161-5505.
- [59] J. S. Fleming, "A technique for using CT images in attenuation correction and quantification in SPECT," *Nuclear Medicine Communications*, vol. 10, no. 2, 1989, ISSN: 0143-3636.
- [60] *Siemens Syngo MI Applications online help*, 2019.
- [61] M. Salas-Ramirez, J. Leube, M. Lassmann, and J. Tran-Gia, "Effect of kilovoltage and quality reference mAs on CT-based attenuation correction in ^{177}Lu SPECT/CT imaging: a phantom study," *EJNMMI Physics*, vol. 11, no. 1, 2024, ISSN: 21977364. DOI: 10.1186/s40658-024-00622-6
- [62] T. Ichihara, K. Ogawa, N. Motomura, A. Kubo, and S. Hashimoto, "Compton scatter compensation using the triple-energy window method for single- and dual-isotope SPECT," *Journal of Nuclear Medicine*, vol. 34, no. 12, pp. 2216–2221, 1993, ISSN: 01615505.
- [63] D. J. Kadrmas, E. C. Frey, S. S. Karimi, and B. M. W. Tsui, "Fast implementations of reconstruction-based scatter compensation in fully 3D SPECT image reconstruction," *Physics in Medicine and Biology*, vol. 43, no. 4, pp. 857–873, Apr. 1998, ISSN: 0031-9155. DOI: 10.1088/0031-9155/43/4/014
- [64] T. Bexelius and A. Sohlberg, "Implementation of GPU accelerated SPECT reconstruction with Monte Carlo-based scatter correction," *Annals of Nuclear Medicine*, vol. 32, no. 5, pp. 337–347, 2018, ISSN: 18646433. DOI: 10.1007/s12149-018-1252-1
- [65] H. Ryu, S. R. Meikle, K. P. Willowson, E. M. Eslick, and D. L. Bailey, "Performance evaluation of quantitative SPECT/CT using NEMA NU 2 PET methodology," *Physics in Medicine and Biology*, vol. 64, no. 14, 2019, ISSN: 13616560. DOI: 10.1088/1361-6560/ab2a22
- [66] J. Gustafsson, G. Brolin, M. Cox, M. Ljungberg, L. Johansson, and K. S. Gleisner, "Uncertainty propagation for SPECT/CT-based renal dosimetry in ^{177}Lu peptide receptor radionuclide therapy," *Physics in Medicine and Biology*, vol. 60, no. 21, pp. 8329–8346, 2015, ISSN: 13616560. DOI: 10.1088/0031-9155/60/21/8329
- [67] D. Finocchiaro et al., "Partial volume effect of SPECT images in PRRT with ^{177}Lu labelled somatostatin analogues: A practical solution," *Physica Medica*, vol. 57, no. June 2018, pp. 153–159, 2019, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2018.12.029

- [68] M. Ljungberg, A. Celler, M. W. Konijnenberg, K. F. Eckerman, Y. K. Dewaraja, and K. Sjögreen-Gleisner, "MIRD pamphlet no. 26: Joint EANM/MIRD guidelines for quantitative ^{177}Lu SPECT applied for dosimetry of radiopharmaceutical therapy," *Journal of Nuclear Medicine*, vol. 57, no. 1, pp. 151–162, 2016, ISSN: 01615505. DOI: 10.2967/jnumed.115.159012
- [69] C. A. Van Gils, C. Beijst, R. Van Rooij, and H. W. De Jong, "Impact of reconstruction parameters on quantitative I-131 SPECT," *Physics in Medicine and Biology*, vol. 61, no. 14, pp. 5166–5182, 2016, ISSN: 13616560. DOI: 10.1088/0031-9155/61/14/5166
- [70] B. Moore, J. Galt, D. Mirando, A. Nelson, and J. Nye, "Phantom evaluation of a vendor neutral quantitative SPECT/CT reconstruction package," *Journal of Nuclear Medicine*, vol. 60, no. supplement 1, p. 1359, 2019, ISSN: 0161-5505.
- [71] A. H. Vija, C. Von Gall, and P. Ghosh, "White Paper: Accurate, reproducible, and standardized quantification," Siemens Healthineers, Tech. Rep., 2019, https://static.healthcare.siemens.com/siemens_hwem.
- [72] D. Vriens, E. P. Visser, L. F. De Geus-Oei, and W. J. Oyen, "Methodological considerations in quantification of oncological FDG PET studies," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 7, pp. 1408–1425, 2010, ISSN: 16197070. DOI: 10.1007/s00259-009-1306-7
- [73] I. Kuji, T. Yamane, A. Seto, Y. Yasumizu, S. Shirotake, and M. Oyama, "Skeletal standardized uptake values obtained by quantitative SPECT/CT as an osteoblastic biomarker for the discrimination of active bone metastasis in prostate cancer," *European Journal of Hybrid Imaging*, vol. 1, no. 1, pp. 1–16, 2017, ISSN: 2510-3636. DOI: 10.1186/s41824-017-0006-y
- [74] A. Collarino et al., "Experimental validation of absolute SPECT/CT quantification for response monitoring in breast cancer," *Medical Physics*, vol. 45, no. 5, pp. 2143–2153, 2018, ISSN: 00942405. DOI: 10.1002/mp.12880
- [75] P. R. Scully et al., "DPD Quantification in Cardiac Amyloidosis: A Novel Imaging Biomarker," *JACC: Cardiovascular Imaging*, vol. 13, no. 6, pp. 1353–1363, 2020, ISSN: 18767591. DOI: 10.1016/j.jcmg.2020.03.020
- [76] Y. Sugawara, K. R. Zasadny, A. W. Neuhoﬀ, and R. L. Wahl, "Reevaluation of the standardized uptake value for FDG: Variations with body weight and methods for correction," *Radiology*, vol. 213, no. 2, pp. 521–525, 1999, ISSN: 00338419. DOI: 10.1148/radiology.213.2.r99nv37521
- [77] C. Brendle, J. Kupferschläger, K. Nikolaou, C. La Fougère, S. Gatidis, and C. Pfannenberger, "Is the standard uptake value (SUV) appropriate for quantification in clinical PET imaging? - Variability induced by different SUV measurements and

- varying reconstruction methods,” *European Journal of Radiology*, vol. 84, no. 1, pp. 158–162, 2015, ISSN: 18727727. DOI: 10.1016/j.ejrad.2014.10.018
- [78] M. A. Lodge, M. A. Chaudhry, and R. L. Wahl, “Noise considerations for PET quantification using maximum and peak standardized uptake value,” *Journal of Nuclear Medicine*, vol. 53, no. 7, pp. 1041–1047, 2012, ISSN: 01615505. DOI: 10.2967/jnumed.111.101733
- [79] C. Lapa et al., “Prognostic value of positron emission tomography-assessed tumor heterogeneity in patients with thyroid cancer undergoing treatment with radiopeptide therapy,” *Nuclear Medicine and Biology*, vol. 42, no. 4, pp. 349–354, 2015, ISSN: 18729614. DOI: 10.1016/j.nucmedbio.2014.12.006
- [80] R. T. Leijenaar et al., “The effect of SUV discretization in quantitative FDG-PET Radiomics: The need for standardized methodology in tumor texture analysis,” *Scientific Reports*, vol. 5, no. November 2014, pp. 1–10, 2015, ISSN: 20452322. DOI: 10.1038/srep11075
- [81] A. Zwanenburg, “Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, no. 13, pp. 2638–2655, 2019, ISSN: 16197089. DOI: 10.1007/s00259-019-04391-8
- [82] V. Rajkumar et al., “Texture analysis of 125I-A5B7 anti-CEA antibody SPECT differentiates metastatic colorectal cancer model phenotypes and anti-vascular therapy response,” *British Journal of Cancer*, vol. 112, no. 12, pp. 1882–1887, 2015, ISSN: 15321827. DOI: 10.1038/bjc.2015.166
- [83] A. Rahmim et al., “Application of texture analysis to DAT SPECT imaging: Relationship to clinical assessments,” *NeuroImage: Clinical*, vol. 12, e1–e9, 2016, ISSN: 22131582. DOI: 10.1016/j.nicl.2016.02.012
- [84] C. Wetz et al., “Predictive Value of Asphericity in Pretherapeutic [111In]DTPA-Octreotide SPECT/CT for Response to Peptide Receptor Radionuclide Therapy with [177Lu]DOTATATE,” *Molecular Imaging and Biology*, vol. 19, no. 3, pp. 437–445, 2017, ISSN: 18602002. DOI: 10.1007/s11307-016-1018-x
- [85] J. C. Sanders, T. Kuwert, J. Hornegger, and P. Ritt, “Quantitative SPECT/CT Imaging of 177Lu with In Vivo Validation in Patients Undergoing Peptide Receptor Radionuclide Therapy,” *Molecular Imaging and Biology*, vol. 17, no. 4, pp. 585–593, 2015, ISSN: 18602002. DOI: 10.1007/s11307-014-0806-4
- [86] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973. DOI: 10.1109/TSMC.1973.4309314

- [87] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016, ISSN: 15271315. DOI: 10.1148/radiol.2015151169
- [88] F. Orlhac, C. Nioche, I. Klyuzhin, A. Rahmim, and I. Buvat, "Radiomics in PET Imaging," *PET Clinics*, vol. 16, no. 4, pp. 597–612, Oct. 2021, ISSN: 15568598. DOI: 10.1016/j.cpet.2021.06.007
- [89] A. Zwanenburg et al., "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020, ISSN: 15271315. DOI: 10.1148/radiol.2020191145
- [90] M. Hatt et al., "Joint EANM/SNMMI guideline on radiomics in nuclear medicine," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 50, pp. 352–375, 2022, ISSN: 1619-7070. DOI: 10.1007/s00259-022-06001-6
- [91] S. Ha, H. Choi, J. C. Paeng, and G. J. Cheon, "Radiomics in Oncological PET/CT: a Methodological Overview," *Nuclear Medicine and Molecular Imaging*, vol. 53, no. 1, pp. 14–29, 2019, ISSN: 18693482. DOI: 10.1007/s13139-019-00571-4
- [92] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present. . . any future?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 1, pp. 151–165, 2017, ISSN: 16197089. DOI: 10.1007/s00259-016-3427-0
- [93] G. M. Kalisvaart et al., "Design and evaluation of a modular multimodality imaging phantom to simulate heterogeneous uptake and enhancement patterns for radiomic quantification in hybrid imaging: A feasibility study," *Medical Physics*, vol. 49, no. 5, pp. 3093–3106, 2022, ISSN: 24734209. DOI: 10.1002/mp.15537
- [94] I. Shiri, A. Rahmim, P. Ghaffarian, P. Geramifar, H. Abdollahi, and A. Bitarafan-Rajabi, "The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies," *European Radiology*, vol. 27, no. 11, pp. 4498–4509, 2017, ISSN: 14321084. DOI: 10.1007/s00330-017-4859-z
- [95] M. C. Adams, T. G. Turkington, J. M. Wilson, and T. Z. Wong, "A systematic review of the factors affecting accuracy of SUV measurements," *American Journal of Roentgenology*, vol. 195, no. 2, pp. 310–320, 2010, ISSN: 15463141. DOI: 10.2214/AJR.10.4923
- [96] R. Boellaard, N. C. Krak, O. S. Hoekstra, and A. A. Lammertsma, "Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: A simulation study," *Journal of Nuclear Medicine*, vol. 45, no. 9, pp. 1519–1527, 2004, ISSN: 01615505.

- [97] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, "Radiomics in medical imaging—"how-to" guide and critical reflection," *Insights into Imaging*, vol. 11, no. 1, 2020, ISSN: 18694101. DOI: 10.1186/s13244-020-00887-2
- [98] C. Nioche et al., "Lifex: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity," *Cancer Research*, vol. 78, no. 16, pp. 4786–4789, 2018, ISSN: 15387445. DOI: 10.1158/0008-5472.CAN-18-0125
- [99] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014, ISSN: 18790534. DOI: 10.1016/j.compbiomed.2014.04.014
- [100] H. Arabi, A. AkhavanAllaf, A. Sanaat, I. Shiri, and H. Zaidi, "The promise of artificial intelligence and deep learning in PET and SPECT imaging," *Physica Medica*, vol. 83, no. February, pp. 122–137, 2021, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2021.03.008
- [101] E. A. Carlsen et al., "A convolutional neural network for total tumor segmentation in [64Cu]Cu-DOTATATE PET/CT of patients with neuroendocrine neoplasms," *EJNMMI Research*, vol. 12, no. 1, 2022, ISSN: 2191-219X. DOI: 10.1186/s13550-022-00901-2
- [102] N. M. deSouza et al., "Standardised lesion segmentation for imaging biomarker quantitation: a consensus recommendation from ESR and EORTC," *Insights into Imaging*, vol. 13, no. 1, 2022, ISSN: 18694101. DOI: 10.1186/s13244-022-01287-4
- [103] L. Papp, I. Rausch, M. Grahovac, M. Hacker, and T. Beyer, "Optimized feature extraction for radiomics analysis of 18F-FDG PET imaging," *Journal of Nuclear Medicine*, vol. 60, no. 6, pp. 864–872, 2019, ISSN: 2159662X. DOI: 10.2967/jnumed.118.217612
- [104] F. Orlhac, M. Soussan, K. Chouahnia, E. Martinod, and I. Buvat, "18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer," *PLoS ONE*, vol. 10, no. 12, pp. 1–16, 2015, ISSN: 19326203. DOI: 10.1371/journal.pone.0145063
- [105] F. Orlhac, C. Nioche, M. Soussan, and I. Buvat, "Understanding changes in tumor texture indices in PET: A comparison between visual assessment and index values in simulated and patient data," *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 387–392, 2017, ISSN: 2159662X. DOI: 10.2967/jnumed.116.181859
- [106] F. J. Brooks and P. W. Grigsby, "The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake," *Journal of Nuclear Medicine*, vol. 55, no. 1, pp. 37–42, 2014, ISSN: 01615505. DOI: 10.2967/jnumed.112.116715
- [107] V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1,

- no. 2, pp. 207–226, Mar. 2016, ISSN: 2380-8993. DOI: 10.1080/23808993.2016.1164013
- [108] O. Morin et al., “A Deep Look Into the Future of Quantitative Imaging in Oncology: A Statement of Working Principles and Proposal for Change,” *International Journal of Radiation Oncology Biology Physics*, vol. 102, no. 4, pp. 1074–1082, 2018, ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2018.08.032
- [109] I. Shahzadi et al., “Analysis of MRI and CT-based radiomics features for personalized treatment in locally advanced rectal cancer and external validation of published radiomics models,” *Scientific Reports*, vol. 12, no. 1, pp. 1–15, 2022, ISSN: 20452322. DOI: 10.1038/s41598-022-13967-8
- [110] A. Traverso, L. Wee, A. Dekker, and R. Gillies, “Repeatability and Reproducibility of Radiomic Features: A Systematic Review,” *International Journal of Radiation Oncology Biology Physics*, vol. 102, no. 4, pp. 1143–1158, 2018, ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2018.05.053
- [111] M. L. Welch et al., “Vulnerabilities of radiomic signature development: The need for safeguards,” *Radiotherapy and Oncology*, vol. 130, pp. 2–9, 2019, ISSN: 18790887. DOI: 10.1016/j.radonc.2018.10.027
- [112] M. J. Pencina, R. B. D’Agostino, and R. S. Vasan, “Statistical methods for assessment of added usefulness of new biomarkers,” *Clinical Chemistry and Laboratory Medicine*, vol. 48, no. 12, pp. 1–7, Jan. 2010, ISSN: 1437-4331. DOI: 10.1515/CCLM.2010.340
- [113] E. Vittinghoff and C. E. McCulloch, “Relaxing the rule of ten events per variable in logistic and cox regression,” *American Journal of Epidemiology*, vol. 165, no. 6, pp. 710–718, 2007, ISSN: 00029262. DOI: 10.1093/aje/kwk052
- [114] R. A. Werner et al., “Impact of Tumor Burden on Quantitative [68Ga] DOTATOC Biodistribution,” *Molecular Imaging and Biology*, vol. 21, no. 4, pp. 790–798, 2019, ISSN: 18602002. DOI: 10.1007/s11307-018-1293-9
- [115] M. Vallières et al., “Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer,” *Scientific Reports*, vol. 7, no. 1, pp. 1–14, 2017, ISSN: 20452322. DOI: 10.1038/s41598-017-10371-5
- [116] S. S. Yip et al., “Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non–Small Cell Lung Cancer,” *Journal of Nuclear Medicine*, vol. 58, no. 4, pp. 569–576, Apr. 2017, ISSN: 0161-5505. DOI: 10.2967/jnumed.116.181826
- [117] C. Bailly et al., “Revisiting the robustness of PET-based textural features in the context of multi-centric trials,” *PLoS ONE*, vol. 11, no. 7, pp. 1–16, 2016, ISSN: 19326203. DOI: 10.1371/journal.pone.0159984
- [118] M. J. Nyflot, F. Yang, D. Byrd, S. R. Bowen, G. A. Sandison, and P. E. Kinahan, “Quantitative radiomics: impact of stochastic effects on textural feature analysis implies

- the need for standards,” *Journal of Medical Imaging*, vol. 2, no. 4, p. 041 002, 2015, ISSN: 2329-4302. DOI: 10.1117/1.jmi.2.4.041002
- [119] X. Fave et al., “Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer,” *Translational Cancer Research*, vol. 5, no. 4, pp. 349–363, 2016, ISSN: 22196803. DOI: 10.21037/tcr.2016.07.11
- [120] M. McNitt-Gray et al., “Standardization in quantitative imaging: A multicenter comparison of radiomic features from different software packages on digital reference objects and patient data sets,” *Tomography*, vol. 6, no. 2, pp. 118–128, 2020, ISSN: 2379139X. DOI: 10.18383/j.tom.2019.00031
- [121] A. Kaalep et al., “Feasibility of state of the art PET/CT systems performance harmonisation,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 8, pp. 1344–1361, 2018, ISSN: 16197089. DOI: 10.1007/s00259-018-3977-4
- [122] E. Pfaehler et al., “Experimental Multicenter and Multivendor Evaluation of the Performance of PET Radiomic Features Using 3-Dimensionally Printed Phantom Inserts,” *Journal of Nuclear Medicine*, vol. 61, no. 3, pp. 469–476, 2020, ISSN: 2159662X. DOI: 10.2967/JNUMED.119.229724
- [123] E. J. Limkin et al., “Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology,” *Annals of Oncology*, vol. 28, no. 6, pp. 1191–1206, 2017, ISSN: 15698041. DOI: 10.1093/annonc/mdx034
- [124] G. Kayal et al., “Generation of clinical ¹⁷⁷Lu SPECT/CT images based on Monte Carlo simulation with GATE,” *Physica Medica*, vol. 85, no. March, pp. 24–31, 2021, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2021.04.002
- [125] E. Wikberg, M. van Essen, T. Rydén, J. Svensson, P. Gjerdtsson, and P. Bernhardt, “Evaluation of reconstruction methods and image noise levels concerning visual assessment of simulated liver lesions in ¹¹¹In-octreotide SPECT imaging,” *EJNMMI Physics*, vol. 10, no. 1, 2023, ISSN: 21977364. DOI: 10.1186/s40658-023-00557-4
- [126] A. Skretting, O. Glomset, and T. V. Bogsrud, “A phantom for investigation of tumour signal and noise in pet reconstruction with various smoothing filters: Experiments and comparisons with simulated intensity diffusion,” *Radiation Protection Dosimetry*, vol. 139, no. 1-3, pp. 191–194, 2010, ISSN: 01448420. DOI: 10.1093/rpd/ncq051
- [127] F. Gallivanone, M. Interlenghi, D. D’Ambrosio, G. Trifirò, and I. Castiglioni, “Parameters influencing PET imaging features: A phantom study with irregular and heterogeneous synthetic lesions,” *Contrast Media and Molecular Imaging*, vol. 2018, 2018, ISSN: 15554317. DOI: 10.1155/2018/5324517
- [128] A. P. Robinson et al., “Development of a validation imaging dataset for Molecular Radiotherapy dosimetry multicenter intercomparison exercises based on

- anthropomorphic phantoms,” *Physica Medica*, vol. 109, p. 102583, May 2023, ISSN: 11201797. DOI: 10.1016/j.ejmp.2023.102583
- [129] D. Didierlaurent et al., “A new respiratory gating device to improve 4D PET/CT,” *Medical Physics*, vol. 40, no. 3, 2013, ISSN: 00942405. DOI: 10.1118/1.4789487
- [130] D. C. Vines, H. Keller, J. D. Hoisak, and S. L. Breen, “Quantitative PET comparing gated with nongated acquisitions using a NEMA phantom with respiratory-simulated motion,” *Journal of Nuclear Medicine Technology*, vol. 35, no. 4, pp. 246–251, 2007, ISSN: 00914916. DOI: 10.2967/jnmt.107.040782
- [131] S. Williams, K. Jones, C. Paterson, J. Robinson, and N. Goodfield, “Development of a Bespoke Cardiac Prototype Phantom for Dynamic Imaging,” in *British Nuclear Medicine Society Autumn Meeting 2021. Nuclear Medicine Communications*, 2021, pp. 1162–1185. DOI: 10.1097/MNM.0000000000001479
- [132] S. Chicklore, V. Goh, M. Siddique, A. Roy, P. K. Marsden, and G. J. Cook, “Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 1, pp. 133–140, 2013, ISSN: 16197070. DOI: 10.1007/s00259-012-2247-0
- [133] F. J. Brooks, “On some misconceptions about tumor heterogeneity quantification,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 8, pp. 1292–1294, 2013, ISSN: 16197070. DOI: 10.1007/s00259-013-2430-y
- [134] “Advanced Issues and Deeper Insights,” in *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, K. P. Burnham and D. R. Anderson, Eds., New York, NY: Springer New York, 2002, pp. 267–351, ISBN: 978-0-387-22456-5. DOI: 10.1007/978-0-387-22456-5{_}6
- [135] S. Ashrafinia, P. Dalaie, T. H. Schindler, M. G. Pomper, and A. Rahmim, “Standardized Radiomics Analysis of Clinical Myocardial Perfusion Stress SPECT Images to Identify Coronary Artery Calcification,” *Cureus*, vol. 15, no. 8, 2023. DOI: 10.7759/cureus.43343
- [136] U. Bashir et al., “The effects of segmentation algorithms on the measurement of 18F-FDG PET texture parameters in non-small cell lung cancer,” *EJNMMI Research*, vol. 7, pp. 1–9, 2017, ISSN: 2191219X. DOI: 10.1186/s13550-017-0310-3
- [137] P. T. Yau, “OptSurvCutR: Validated Cut-point Selection for Survival Analysis,” Oct. 2025. DOI: 10.1101/2025.10.08.681246
- [138] M. Hatt et al., “18F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort,” *Journal of Nuclear Medicine*, vol. 56, no. 1, pp. 38–44, 2015, ISSN: 01615505. DOI: 10.2967/jnumed.114.144055
- [139] F. Orlhac, M. Soussan, J. A. Maisonobe, C. A. Garcia, B. Vanderlinden, and I. Buvat, “Tumor texture analysis in 18F-FDG PET: Relationships between texture parameters,

- histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis,” *Journal of Nuclear Medicine*, vol. 55, no. 3, pp. 414–422, 2014, ISSN: 01615505. DOI: 10.2967/jnumed.113.129858
- [140] A. Akhavanallaf et al., “The predictive value of pretherapy [68Ga]Ga-DOTA-TATE PET and biomarkers in [177Lu]Lu-PRRT tumor dosimetry,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 50, no. 10, pp. 2984–2996, 2023, ISSN: 16197089. DOI: 10.1007/s00259-023-06252-x
- [141] C. Atkinson et al., “Radiomics-Based Texture Analysis of 68Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With 177Lu-DOTATATE,” *Frontiers in Oncology*, vol. 11, no. August, pp. 1–9, 2021, ISSN: 2234943X. DOI: 10.3389/fonc.2021.686235
- [142] F. C. Staal et al., “GEP-NET radiomics: a systematic review and radiomics quality score assessment,” *European Radiology*, vol. 32, no. 10, pp. 7278–7294, 2022, ISSN: 14321084. DOI: 10.1007/s00330-022-08996-w
- [143] R. Gutsche et al., “Multicentric 68Ga-PSMA PET radiomics for treatment response assessment of 177Lu-PSMA-617 radioligand therapy in patients with metastatic castration-resistant prostate cancer,” *Frontiers in Nuclear Medicine*, vol. 3, no. September, pp. 1–9, 2023, ISSN: 26738880. DOI: 10.3389/fnume.2023.1234853
- [144] K. H. Leung et al., “Deep learning and radiomics framework for PSMA-RADS classification of prostate cancer on PSMA PET,” *EJNMMI Research*, vol. 12, no. 1, pp. 1–15, 2022, ISSN: 2191219X. DOI: 10.1186/s13550-022-00948-1
- [145] A. Gafita et al., “Nomograms to predict outcomes after 177Lu-PSMA therapy in men with metastatic castration-resistant prostate cancer: an international, multicentre, retrospective study,” *The Lancet Oncology*, vol. 22, no. 8, pp. 1115–1125, 2021, ISSN: 14745488. DOI: 10.1016/S1470-2045(21)00274-6
- [146] E. Yazdani, P. Geramifar, N. Karamzade-Ziarati, M. Sadeghi, P. Amini, and A. Rahmim, “Radiomics and Artificial Intelligence in Radiotheranostics: A Review of Applications for Radioligands Targeting Somatostatin Receptors and Prostate-Specific Membrane Antigens,” *Diagnostics*, vol. 14, no. 2, 2024, ISSN: 20754418. DOI: 10.3390/diagnostics14020181
- [147] M. E. Klontzas et al., “Machine Learning Integrating 99mTc Sestamibi SPECT/CT and Radiomics Data Achieves Optimal Characterization of Renal Oncocytic Tumors,” *Cancers*, vol. 15, no. 14, 2023, ISSN: 20726694. DOI: 10.3390/cancers15143553
- [148] L. Feng, X. Yang, C. Wang, H. Zhang, W. Wang, and J. Yang, “Predicting event-free survival after induction of remission in high-risk pediatric neuroblastoma: combining

- 123I-MIBG SPECT-CT radiomics and clinical factors,” *Pediatric Radiology*, vol. 54, no. 5, pp. 805–819, 2024, ISSN: 14321998. DOI: 10.1007/s00247-024-05901-z
- [149] H. Jiang et al., “Radiomics incorporating deep features for predicting Parkinson’s disease in 123I-iodoflupane SPECT,” *EJNMMI Physics*, vol. 11, no. 1, pp. 1–14, 2024, ISSN: 21977364. DOI: 10.1186/s40658-024-00651-1
- [150] M. Edalat-Javid et al., “Cardiac SPECT radiomic features repeatability and reproducibility: A multi-scanner phantom study,” *Journal of Nuclear Cardiology*, vol. 28, no. 6, pp. 2730–2744, 2021, ISSN: 15326551. DOI: 10.1007/s12350-020-02109-0
- [151] O. Gharibi et al., “Myocardial perfusion SPECT radiomic features reproducibility assessment: Impact of image reconstruction and harmonization,” *Medical Physics*, no. October 2023, pp. 965–977, 2024, ISSN: 24734209. DOI: 10.1002/mp.17490
- [152] M. Amini et al., “Machine learning-based diagnosis and risk classification of coronary artery disease using myocardial perfusion imaging SPECT: A radiomics study,” *Scientific Reports*, vol. 13, no. 1, pp. 1–12, 2023, ISSN: 20452322. DOI: 10.1038/s41598-023-42142-w
- [153] C. Zhou et al., “Radiomics Nomogram Derived from Gated Myocardial Perfusion SPECT for Identifying Ischemic Cardiomyopathy,” *Journal of Imaging Informatics in Medicine*, vol. 37, no. 6, pp. 2784–2793, 2024, ISSN: 2948-2933. DOI: 10.1007/s10278-024-01145-3
- [154] S. Ashrafinia, “Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics,” Ph.D. dissertation, Johns Hopkins University, 2019.
- [155] Y. Liu et al., “Diagnostic precision in thyroid-associated ophthalmopathy using multi-center radiomics with 99mTc-DTPA SPECT/CT,” *Scientific Reports*, vol. 14, no. 1, p. 25810, Oct. 2024, ISSN: 2045-2322. DOI: 10.1038/s41598-024-76018-4
- [156] G. Hajianfar et al., “Pulmonary perfusion deficiency detection in lung subsegments of SPECT/CT images using radiomics and machine learning algorithms,” *Journal of Nuclear Medicine*, vol. 65, no. supplement 2, p. 241915, 2024, ISSN: 0161-5505.
- [157] L. M. Montuenga et al., “The diffuse endocrine system: From embryogenesis to carcinogenesis,” *Progress in Histochemistry and Cytochemistry*, vol. 38, no. 2, pp. 153–272, 2003, ISSN: 00796336. DOI: 10.1016/s0079-6336(03)80004-9
- [158] T. S. Genus et al., “Impact of neuroendocrine morphology on cancer outcomes and stage at diagnosis: a UK nationwide cohort study 2013–2015,” *British Journal of Cancer*, vol. 121, no. 11, pp. 966–972, 2019, ISSN: 15321827. DOI: 10.1038/s41416-019-0606-3
- [159] S. Das and A. Dasari, “Epidemiology, Incidence, and Prevalence of Neuroendocrine Neoplasms: Are There Global Differences?” *Current Oncology Reports*, vol. 23, no. 4, 2021, ISSN: 15346269. DOI: 10.1007/s11912-021-01029-7

- [160] S. T. Lee, H. R. Kulkarni, A. Singh, and R. P. Baum, "Theranostics of Neuroendocrine Tumors," *Visceral Medicine*, vol. 33, no. 5, pp. 358–366, 2017, ISSN: 2297475X. DOI: 10.1159/000480383
- [161] B. G. Taal and O. Visser, "Epidemiology of neuroendocrine tumours," *Neuroendocrinology*, vol. 80, no. SUPPL. 1, pp. 3–7, 2004, ISSN: 00283835. DOI: 10.1159/000080731
- [162] I. Uri and S. Grozinsky-Glasberg, "Current treatment strategies for patients with advanced gastroenteropancreatic neuroendocrine tumors (GEP-NETs)," *Clinical Diabetes and Endocrinology*, vol. 4, no. 1, pp. 1–10, 2018, ISSN: 2055-8260. DOI: 10.1186/s40842-018-0066-3
- [163] W. A. Breeman et al., "Somatostatin receptor-mediated imaging and therapy: Basic science, current knowledge, limitations and future perspectives," *European Journal of Nuclear Medicine*, vol. 28, no. 9, pp. 1421–1429, 2001, ISSN: 03406997. DOI: 10.1007/s002590100502
- [164] Y. C. Patel, "Somatostatin and Its Receptor Family," *Frontiers in Neuroendocrinology*, vol. 20, no. 3, pp. 157–198, Jul. 1999, ISSN: 00913022. DOI: 10.1006/frne.1999.0183
- [165] A. Sundin et al., "ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: Radiological, Nuclear Medicine and Hybrid Imaging," *Neuroendocrinology*, vol. 105, no. 3, pp. 212–244, 2017, ISSN: 14230194. DOI: 10.1159/000471879
- [166] O. N. Y. Lee, K. V. Tan, V. Tripathi, H. Yuan, W. W. L. Chan, and K. W. H. Chiu, "The Role of 68Ga-DOTA-SSA PET/CT in the Management and Prediction of Peptide Receptor Radionuclide Therapy Response for Patients With Neuroendocrine Tumors: A Systematic Review and Meta-analysis," *Clinical Nuclear Medicine*, vol. 47, no. 9, pp. 781–793, 2022, ISSN: 15360229. DOI: 10.1097/RLU.0000000000004235
- [167] J. J. Zaknun et al., "The joint IAEA, EANM, and SNMMI practical guidance on peptide receptor radionuclide therapy (PRRNT) in neuroendocrine tumours," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 5, pp. 800–816, 2013, ISSN: 16197089. DOI: 10.1007/s00259-012-2330-6
- [168] R. P. Baum, H. R. Kulkarni, and C. Carreras, "Peptides and receptors in image-guided therapy: Theranostics for neuroendocrine neoplasms," *Seminars in Nuclear Medicine*, vol. 42, no. 3, pp. 190–207, 2012, ISSN: 15584623. DOI: 10.1053/j.semnuclmed.2012.01.002
- [169] E. C. S. De Camargo Etchebehere et al., "68Ga-DOTATATE PET/CT, 99mTc-HYNIC-octreotide SPECT/CT, and whole-body MR imaging in detection of neuroendocrine tumors: A prospective trial," *Journal of Nuclear Medicine*, vol. 55, no. 10, pp. 1598–1604, 2014, ISSN: 01615505. DOI: 10.2967/jnumed.114.144543

- [170] E. Bombardieri et al., "111In-pentetreotide scintigraphy: Procedure guidelines for tumour imaging," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 7, pp. 1441–1448, 2010, ISSN: 16197070. DOI: 10.1007/s00259-010-1473-6
- [171] Polatom, "TEKTROTYD 16 micrograms kit summary of product characteristics," MHRA, <https://mhraproducts4853.blob.core.windows.net/docs/a48b6388d64755814c678be4addf93cc4877f79d>, Tech. Rep., 2019.
- [172] C. Decristoforo, S. J. Mather, W. Cholewinski, E. Donnemiller, G. Riccabona, and R. Moncayo, "(99m)Tc-EDDA/HYNIC-TOC: A new (99m)Tc-labelled radiopharmaceutical for imaging somatostatin receptor-positive tumours: First clinical results and intra-patient comparison with 111In-labelled octreotide derivatives," *European Journal of Nuclear Medicine*, vol. 27, no. 9, pp. 1318–1325, 2000, ISSN: 03406997. DOI: 10.1007/s002590000289
- [173] M. Gabriel et al., "An inpatient comparison of 99mTc-EDDA/HYNIC-TOC with 111In-DTPA-octreotide for diagnosis of somatostatin receptor-expressing tumors," *Journal of Nuclear Medicine*, vol. 44, no. 5, pp. 708–716, 2003, ISSN: 01615505.
- [174] V. Prasad and R. P. Baum, "Biodistribution of the Ga-68 labeled somatostatin analogue DOTA-NOC in patients with neuroendocrine tumors," *Q J Nucl Med Mol Imaging*, vol. 54, no. 1, pp. 61–67, 2010.
- [175] A. Kroiss et al., "68Ga-DOTA-TOC uptake in neuroendocrine tumour and healthy tissue: Differentiation of physiological uptake and pathological processes in PET/CT," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 4, pp. 514–523, 2013, ISSN: 16197070. DOI: 10.1007/s00259-012-2309-3
- [176] J. Kunikowska, V. Lewington, and L. Krolicki, "Optimizing somatostatin receptor imaging in patients with neuroendocrine tumors: The impact of 99mTc-HYNICTOC SPECT/SPECT/CT versus 68Ga-DOTATATE PET/CT upon clinical management," *Clinical Nuclear Medicine*, vol. 42, no. 12, pp. 905–911, 2017, ISSN: 15360229. DOI: 10.1097/RLU.0000000000001877
- [177] E. Fortunati, N. Bonazzi, L. Zanoni, S. Fanti, and V. Ambrosini, "Molecular imaging Theranostics of Neuroendocrine Tumors," *Seminars in Nuclear Medicine*, vol. 53, no. 4, pp. 539–554, 2023, ISSN: 15584623. DOI: 10.1053/j.semnuclmed.2022.12.007
- [178] M. Liu et al., "A Prospective Evaluation of [18F]AIF-NOTA-LM3 in Patients with Well-differentiated Neuroendocrine Tumors: Head-to-head Comparison with 68GaDOTATATE," in *EANM'23 Abstract Book Congress Sep 9-13, 2023. Eur J Nucl Med Mol Imaging*, 2023, S132.
- [179] J. Zhang et al., "First-in-human Study of an Optimized, potential Kit-type, SSTR Antagonist 68Ga-DATA5m-LM4 in Neuroendocrine Tumors," in *EANM'23 Abstract Book Congress Sep 9-13, 2023. Eur J Nucl Med Mol Imaging*, 2023, S131.

- [180] W. Zhu, M. Liu, Y. Cheng, R. Jia, C. Bai, and L. Huo, "Retrospective Analysis of 666 Gallium-68 Labeled Somatostatin Receptor Antagonist PET/CT Imaging in Over 500 Patients: Experience from a Single Center in China," in *EANM'23 Abstract Book Congress Sep 9-13, 2023. Eur J Nucl Med Mol Imaging*, 2023, S28.
- [181] D. A. Pattison and M. S. Hofman, "Role of fluorodeoxyglucose PET/computed tomography in targeted radionuclide therapy for endocrine malignancies," *PET Clinics*, vol. 10, no. 4, pp. 461–476, 2015, ISSN: 18799809. DOI: 10.1016/j.cpet.2015.05.005
- [182] M. F. Bozkurt et al., "Guideline for PET/CT imaging of neuroendocrine neoplasms with 68Ga-DOTA-conjugated somatostatin receptor targeting peptides and 18F-DOPA," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 9, pp. 1588–1601, 2017, ISSN: 16197089. DOI: 10.1007/s00259-017-3728-y
- [183] M. Pavel et al., "ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site," *Neuroendocrinology*, vol. 103, no. 2, pp. 172–185, 2016, ISSN: 14230194. DOI: 10.1159/000443167
- [184] R. Arnold et al., "Somatostatin analogue octreotide and inhibition of tumour growth in metastatic endocrine gastroenteropancreatic tumours," *Gut*, vol. 38, no. 3, pp. 430–438, 1996, ISSN: 00175749. DOI: 10.1136/gut.38.3.430
- [185] J. Guillermet-Guibert, H. Lahlou, P. Cordelier, C. Bousquet, S. Pyronnet, and C. Susini, "Physiology of somatostatin receptors," eng, *Journal of endocrinological investigation*, vol. 28, no. 11 Suppl International, pp. 5–9, 2005, ISSN: 0391-4097.
- [186] L. Anthony and P. U. Freda, "From somatostatin to octreotide LAR: evolution of a somatostatin analogue," *Current Medical Research and Opinion*, vol. 25, no. 12, pp. 2989–2999, Dec. 2009, ISSN: 0300-7995. DOI: 10.1185/03007990903328959
- [187] E. S. Delpassand et al., "Safety and efficacy of radionuclide therapy with high-activity in-111 pentetreotide in patients with progressive neuroendocrine tumors," *Cancer Biotherapy and Radiopharmaceuticals*, vol. 23, no. 3, pp. 292–300, 2008, ISSN: 10849785. DOI: 10.1089/cbr.2007.0448
- [188] D. J. Kwekkeboom et al., "Peptide Receptor Radionuclide Therapy in Patients With Gastroenteropancreatic Neuroendocrine Tumors," *Seminars in Nuclear Medicine*, vol. 40, no. 2, pp. 78–88, 2010, ISSN: 00012998. DOI: 10.1053/j.semnuclmed.2009.10.004
- [189] D. J. Kwekkeboom et al., "Treatment with the radiolabeled somatostatin analog [177Lu-DOTA0,Tyr3]octreotate: Toxicity, efficacy, and survival," *Journal of Clinical Oncology*, vol. 26, no. 13, pp. 2124–2130, 2008, ISSN: 0732183X. DOI: 10.1200/JCO.2007.15.2553

- [190] D. Delacroix, J. P. Guerre, P. Leblanc, and C. Hickman, *Radionuclide and radiation protection data handbook 2002*. Nuclear Technology Publishing, 2002, vol. 98, p. 7, ISBN: 1 870965 87 6.
- [191] M. Ljungberg and P. H. Pretorius, "SPECT/CT: an update on technological developments and clinical applications," *The British Journal of Radiology*, vol. 91, no. 1081, p. 20160402, Jan. 2018, ISSN: 0007-1285. DOI: 10.1259/bjr.20160402
- [192] R. P. Baum and H. R. Kulkarni, "Duo-PRRT of Neuroendocrine Tumors Using Concurrent and Sequential Administration of Y-90-and Lu-177-Labeled Somatostatin Analogues," in *Somatostatin Analogues: From Research to Clinical Practice*, A. Hubalewska-Dydejczyk, A. Signore, M. de Jong, R. A. Diercks, J. Buscombe, and C. Van de Wiele, Eds., 1st, John Wiley & Sons, Inc, 2015, ch. 6.4, pp. 264–272, ISBN: 978-1-118-52153-3.
- [193] D. M. V. Huizing, B. J. de Wit-van der Veen, M. Verheij, and M. P. M. Stokkel, "Dosimetry methods and clinical applications in peptide receptor radionuclide therapy for neuroendocrine tumours: a literature review," *EJNMMI Research*, vol. 8, 2018, ISSN: 2191219X. DOI: 10.1186/s13550-018-0443-z
- [194] K. Kacperski et al., "Individual dosimetry and monitoring response to treatment in tandem PRRT with [177Lu/90Y]-DOTA-TATE," *Polish Journal of Medical Physics and Engineering*, vol. 30, no. 4, pp. 294–299, 2024, ISSN: 18980309. DOI: 10.2478/pjmpe-2024-0037
- [195] B. L. Kam et al., "Lutetium-labelled peptides for therapy of neuroendocrine tumours," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 39, no. SUPPL.1, 2012, ISSN: 16197070. DOI: 10.1007/s00259-011-2039-y
- [196] M. Cremonesi, M. Ferrari, L. Bodei, G. Tosi, and G. Paganelli, "Dosimetry in peptide radionuclide receptor therapy: A review," *Journal of Nuclear Medicine*, vol. 47, no. 9, pp. 1467–1475, 2006, ISSN: 01615505.
- [197] M. Gabriel, B. Nilica, B. Kaiser, and I. J. Virgolini, "Twelve-year follow-up after peptide receptor radionuclide therapy," *Journal of Nuclear Medicine*, vol. 60, no. 4, pp. 524–529, 2019, ISSN: 2159662X. DOI: 10.2967/jnumed.118.215376
- [198] J. Strosberg et al., "Phase 3 trial of 177lu-dotatate for midgut neuroendocrine tumors," *New England Journal of Medicine*, vol. 376, no. 2, pp. 125–135, 2017, ISSN: 15334406. DOI: 10.1056/NEJMoa1607427
- [199] Advanced Accelerator Applications, "Lutathera 370 MBq/mL solution for infusion: Summary of Product Characteristics," MHRA, <https://www.medicines.org.uk/emc/product/12723/smpc#gref>, Tech. Rep., 2021.
- [200] P. Courault et al., "Amino Acid Solutions for 177Lu-Oxodotreotide Premedication: A Tolerance Study," *Cancers*, vol. 14, no. 21, p. 5212, Oct. 2022, ISSN: 2072-6694. DOI: 10.3390/cancers14215212

- [201] T. A. Hope et al., "NANETS/SNMMI Consensus Statement on Patient Selection and Appropriate Use of ^{177}Lu -DOTATATE Peptide Receptor Radionuclide Therapy," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 61, no. 2, pp. 222–227, 2020, ISSN: 15355667. DOI: 10.2967/jnumed.119.240911
- [202] A. Pearce, "NPL REPORT IR 6: Recommended Nuclear Decay Data," no. January, 2008.
- [203] K. Zwirtz et al., "Comparison of Choi, RECIST and Somatostatin Receptor PET/CT Based Criteria for the Evaluation of Response and Response Prediction to PRRT," *Pharmaceutics*, vol. 14, no. 6, p. 1278, 2022. DOI: 10.3390/pharmaceutics14061278
- [204] J. R. Strosberg et al., " ^{177}Lu -Dotatate plus long-acting octreotide versus high-dose long-acting octreotide in patients with midgut neuroendocrine tumours (NETTER-1): final overall survival and long-term safety results from an open-label, randomised, controlled, phase 3 trial," *The Lancet Oncology*, vol. 22, no. 12, pp. 1752–1763, 2021, ISSN: 14745488. DOI: 10.1016/S1470-2045(21)00572-6
- [205] Scottish Medicines Consortium, "Approval for use of Lutathera in NHS Scotland, SMC No 1337/18," <https://www.scottishmedicines.org.uk/media/3557/lutetium-177lu-oxodotreotide-lutathera-final-june-2018-for-website.pdf>, Tech. Rep. 1337, 2018, pp. 1–14.
- [206] T. Nayak, J. Norenberg, T. Anderson, and R. Atcher, "A comparison of high- versus low-linear energy transfer somatostatin receptor targeted radionuclide therapy in vitro," *Cancer Biotherapy and Radiopharmaceuticals*, vol. 20, no. 1, pp. 52–57, 2005, ISSN: 10849785. DOI: 10.1089/cbr.2005.20.52
- [207] A. Sabet et al., "Can peptide receptor radionuclide therapy be safely applied in florid bone metastases? TT - Kann die Peptid-Radiorezeptorthherapie bei diffuser Knochenmetastasierung noch sicher angewendet werden?" DE, *Nuklearmedizin*, vol. 53, no. 02, pp. 54–59, 2014, ISSN: 0029-5566 DOI - 10.3413/Nukmed-0614-13-08.
- [208] H. Bergsma et al., "Persistent hematologic dysfunction after peptide receptor radionuclide therapy with ^{177}Lu -DOTATATE: Incidence, course, and predicting factors in patients with gastroenteropancreatic neuroendocrine tumors," *Journal of Nuclear Medicine*, vol. 59, no. 3, pp. 452–458, 2018, ISSN: 2159662X. DOI: 10.2967/jnumed.117.189712
- [209] S. Ballal, M. P. Yadav, C. Bal, R. K. Sahoo, and M. Tripathi, "Broadening horizons with ^{225}Ac -DOTATATE targeted alpha therapy for gastroenteropancreatic neuroendocrine tumour patients stable or refractory to ^{177}Lu -DOTATATE PRRT: first clinical experience on the efficacy and safety," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 4, pp. 934–946, 2020, ISSN: 16197089. DOI: 10.1007/s00259-019-04567-2

- [210] E. Delpassand, I. Tworowska, R. Esfandiari, J. Torgue, J. Hurt, and R. Nuñez, “Targeted Alpha-Emitter Therapy with ^{212}Pb -DOTAMTATE in Neuroendocrine Tumor Subjects who Progressed Following Prior PRRT,” *Journal of Nuclear Medicine*, vol. 63, no. supplement 2, p. 2216, 2022, ISSN: 0161-5505.
- [211] J. Kunikowska and L. Królicki, “Targeted α -Emitter Therapy of Neuroendocrine Tumors,” *Seminars in Nuclear Medicine*, vol. 50, no. 2, pp. 171–176, 2020, ISSN: 15584623. DOI: 10.1053/j.semnuclmed.2019.11.003
- [212] E. Merola and C. M. Grana, “Peptide Receptor Radionuclide Therapy (PRRT): Innovations and Improvements,” *Cancers*, vol. 15, no. 11, 2023, ISSN: 20726694. DOI: 10.3390/cancers15112975
- [213] D. J. Kwekkeboom et al., “ENETS consensus guidelines for the standards of care in neuroendocrine tumors: Peptide receptor radionuclide therapy with radiolabeled somatostatin analogs,” *Neuroendocrinology*, vol. 90, no. 2, pp. 220–226, 2009, ISSN: 00283835. DOI: 10.1159/000225951
- [214] R. J. Hicks et al., “ENETS standardized (synoptic) reporting for molecular imaging studies in neuroendocrine tumours,” *Journal of Neuroendocrinology*, vol. 34, no. 3, pp. 1–10, 2022, ISSN: 13652826. DOI: 10.1111/jne.13040
- [215] E. A. Eisenhauer et al., “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1),” *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, 2009, ISSN: 09598049. DOI: 10.1016/j.ejca.2008.10.026
- [216] V. Liberini et al., “The challenge of evaluating response to peptide receptor radionuclide therapy in gastroenteropancreatic neuroendocrine tumors: The present and the future,” *Diagnostics*, vol. 10, no. 12, pp. 1–31, 2020, ISSN: 20754418. DOI: 10.3390/diagnostics10121083
- [217] C. K. Kuhl et al., “Validity of RECIST version 1.1 for response assessment in metastatic cancer: A prospective, multireader study,” *Radiology*, vol. 290, no. 3, pp. 349–356, 2019, ISSN: 15271315. DOI: 10.1148/radiol.2018180648
- [218] T. Brabander et al., “Pitfalls in the response evaluation after peptide receptor radionuclide therapy with [^{177}Lu -DOTA0,Tyr3] octreotate,” *Endocrine-Related Cancer*, vol. 24, no. 5, pp. 243–251, 2017, ISSN: 14796821. DOI: 10.1530/ERC-16-0524
- [219] S. J. Galgano, A. Iravani, L. Bodei, G. El-Haddad, M. S. Hofman, and G. Kong, “Imaging of Neuroendocrine Neoplasms: Monitoring Treatment Response—AJR Expert Panel Narrative Review,” *American Journal of Roentgenology*, vol. 218, no. 5, pp. 767–780, 2022, ISSN: 15463141. DOI: 10.2214/AJR.21.27159
- [220] R. A. Werner et al., “SSTR-RADS version 1.0 as a reporting system for SSTR PET imaging and selection of potential PRRT candidates: A proposed standardization framework,” *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1085–1091, 2018, ISSN: 2159662X. DOI: 10.2967/jnumed.117.206631

- [221] D. L. Chan et al., "Dual somatostatin receptor/FDG PET/CT imaging in metastatic neuroendocrine tumours: Proposal for a novel grading scheme with prognostic significance," *Theranostics*, vol. 7, no. 5, pp. 1149–1158, 2017, ISSN: 18387640. DOI: 10.7150/thno.18068
- [222] L. Bodei et al., "PRRT neuroendocrine tumor response monitored using circulating transcript analysis: the NETest," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 4, pp. 895–906, 2020, ISSN: 16197089. DOI: 10.1007/s00259-019-04601-3
- [223] L. Bodei et al., "Interim Analysis of a Prospective Validation of 2 Blood-Based Genomic Assessments (PPQ and NETest) to Determine the Clinical Efficacy of 177Lu-DOTATATE in Neuroendocrine Tumors," *Journal of Nuclear Medicine*, vol. 64, no. 4, pp. 567–573, 2023, ISSN: 15355667. DOI: 10.2967/jnumed.122.264363
- [224] J. R. Strosberg et al., "Risk of bowel obstruction in patients with mesenteric or peritoneal disease receiving peptide receptor radionuclide therapy," *Journal of Nuclear Medicine*, vol. 62, no. 1, pp. 69–72, 2021, ISSN: 2159662X. DOI: 10.2967/jnumed.120.242875
- [225] J. Strosberg et al., "Phase 3 Trial of 177 Lu-Dotatate for Midgut Neuroendocrine Tumors," *New England Journal of Medicine*, vol. 376, no. 2, pp. 125–135, 2017, ISSN: 0028-4793. DOI: 10.1056/nejmoa1607427
- [226] J. Barriuso et al., "Prognostic and predictive biomarkers for somatostatin analogs, peptide receptor radionuclide therapy and serotonin pathway targets in neuroendocrine tumours," *Cancer Treatment Reviews*, vol. 70, no. July, pp. 209–222, 2018, ISSN: 15321967. DOI: 10.1016/j.ctrv.2018.09.008
- [227] K. V. Ballman, "Biomarker: Predictive or prognostic?" *Journal of Clinical Oncology*, vol. 33, no. 33, pp. 3968–3971, 2015, ISSN: 15277755. DOI: 10.1200/JCO.2015.63.3651
- [228] A. Dasari et al., "Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States," *JAMA Oncology*, vol. 3, no. 10, pp. 1335–1342, 2017, ISSN: 23742445. DOI: 10.1001/jamaoncol.2017.0589
- [229] S. Ezziddin et al., "Predictors of long-term outcome in patients with well-differentiated gastroenteropancreatic neuroendocrine tumors after peptide receptor radionuclide therapy with 177Lu-octreotate," *Journal of Nuclear Medicine*, vol. 55, no. 2, pp. 183–190, 2014, ISSN: 01615505. DOI: 10.2967/jnumed.113.125336
- [230] D. Mandair et al., "Predictors of outcome in patients treated with peptide radio-labelled receptor target therapy (PRRT).," *Journal of Clinical Oncology*, vol. 35, no. 15_suppl, p. 4090, 2017. DOI: 10.1200/JCO.2017.35.15
- [231] J. Strosberg et al., "Impact of liver tumour burden, alkaline phosphatase elevation, and target lesion size on treatment outcomes with 177Lu-Dotatate: an analysis of the

- NETTER-1 study," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 10, pp. 2372–2382, 2020, ISSN: 16197089. DOI: 10.1007/s00259-020-04709-x
- [232] E. I. van vliet, B. L. Kam, J. J. Teunissen, M. De Jong, E. P. Krenning, and D. J. Kwekkeboom, "Somatostatin Analogues and Radionuclides Used in Therapy," in *Somatostatin Analogues: From Research to Clinical Practice*, A. Hubalewska-Dydejczyk, A. Signore, M. de Jong, R. A. Diercks, J. Buscombe, and C. Van de Wiele, Eds., 1st editio, 2015, pp. 214–229, ISBN: 978-1-118-52153-3.
- [233] T. Yamaguchi et al., "Clinical validation of the gastrointestinal NET grading system: Ki67 index criteria of the WHO 2010 classification is appropriate to predict metastasis or recurrence," *Diagnostic Pathology*, vol. 8, no. 1, p. 65, Dec. 2013, ISSN: 1746-1596. DOI: 10.1186/1746-1596-8-65
- [234] G. Rindi et al., "Overview of the 2022 WHO Classification of Neuroendocrine Neoplasms," *Endocrine Pathology*, vol. 33, no. 1, pp. 115–154, Mar. 2022, ISSN: 1046-3976. DOI: 10.1007/s12022-022-09708-2
- [235] M. Vljaković et al., "The influence of Ki-67 proliferation index on Tc-99m-EDDA-HYNIC-TOC somatostatin receptor scintigraphy in patients with carcinoids," *Open Chemistry*, vol. 13, no. 1, pp. 339–345, 2015, ISSN: 23915420. DOI: 10.1515/chem-2015-0042
- [236] A. Sabet et al., "Specific efficacy of peptide receptor radionuclide therapy with ¹⁷⁷Lu-octreotate in advanced neuroendocrine tumours of the small intestine," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 42, no. 8, pp. 1238–1246, 2015, ISSN: 16197089. DOI: 10.1007/s00259-015-3041-6
- [237] J. R. Black, S. R. Atkinson, A. Singh, J. Evans, and R. Sharma, "The Inflammation-Based Index Can Predict Response and Improve Patient Selection in NETs Treated with PRRT: A Pilot Study," *Journal of Clinical Endocrinology and Metabolism*, vol. 104, no. 2, pp. 285–292, 2018, ISSN: 19457197. DOI: 10.1210/jc.2018-01214
- [238] E. Pauwels et al., "[⁶⁸Ga]Ga-DOTATATE-avid tumor volume, uptake and inflammation-based index correlate with survival in neuroendocrine tumor patients treated with [¹⁷⁷Lu]Lu-DOTATATE PRRT," *Am J Nucl Med Mol Imaging*, vol. 12, no. 5, pp. 152–162, 2022.
- [239] J. Hofland, T. Brabander, and W. W. De Herder, "Letter to the Editor: "the Inflammation-Based Index Can Predict Response and Improve Patient Selection in NETs Treated with PRRT: A Pilot Study"," *Journal of Clinical Endocrinology and Metabolism*, vol. 104, no. 11, pp. 5104–5105, 2019, ISSN: 19457197. DOI: 10.1210/jc.2019-00681
- [240] L. Bodei et al., "Measurement of circulating transcripts and gene cluster analysis predicts and defines therapeutic efficacy of peptide receptor radionuclide therapy (PRRT) in neuroendocrine tumors," *European Journal of Nuclear Medicine and*

- Molecular Imaging*, vol. 43, no. 5, pp. 839–851, 2016, ISSN: 16197089. DOI: 10.1007/s00259-015-3250-z
- [241] T. A. Hope, J. Calais, L. Zhang, W. Dieckmann, and C. Millo, “¹¹¹In-Pentetreotide Scintigraphy Versus ⁶⁸Ga-DOTATATE PET: Impact on Krenning Scores and Effect of Tumor Burden,” *Journal of Nuclear Medicine*, vol. 60, no. 9, pp. 1266–1269, Sep. 2019, ISSN: 0161-5505. DOI: 10.2967/jnumed.118.223016
- [242] R. Sharma et al., “⁶⁸Ga-DOTATATE PET/CT parameters predict response to peptide receptor radionuclide therapy in neuroendocrine tumours,” *Radiotherapy and Oncology*, no. 141, pp. 108–115, 2019, ISSN: 18790887. DOI: 10.1016/j.radonc.2019.09.003
- [243] C. Kratochwil et al., “SUV of [⁶⁸Ga]DOTATOC-PET/CT Predicts Response Probability of PRRT in Neuroendocrine Tumors,” *Molecular Imaging and Biology*, vol. 17, no. 3, pp. 313–318, 2015, ISSN: 18602002. DOI: 10.1007/s11307-014-0795-3
- [244] A. R. Haug et al., “⁶⁸Ga-DOTATATE PET/CT for the early prediction of response to somatostatin receptor-mediated radionuclide therapy in patients with well-differentiated neuroendocrine tumors,” *Journal of Nuclear Medicine*, vol. 51, no. 9, pp. 1349–1356, 2010, ISSN: 01615505. DOI: 10.2967/jnumed.110.075002
- [245] C. Ortega et al., “Quantitative ⁶⁸Ga-Dotatate PET/CT parameters for the prediction of therapy response in patients with progressive metastatic neuroendocrine tumors treated with ¹⁷⁷Lu-Dotatate,” *Journal of Nuclear Medicine*, vol. 62, no. 10, 2021, ISSN: 2159662X. DOI: 10.2967/jnumed.120.256727
- [246] A. Torihara et al., “Prognostic value of somatostatin receptor expressing tumor volume calculated from ⁶⁸Ga-DOTATATE PET/CT in patients with well-differentiated neuroendocrine tumors,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, no. 11, pp. 2244–2251, 2019, ISSN: 16197089. DOI: 10.1007/s00259-019-04455-9
- [247] M. Weber et al., “Treatment-related changes in neuroendocrine tumors as assessed by textural features derived from ⁶⁸Ga-DOTATOC PET/MRI with simultaneous acquisition of apparent diffusion coefficient,” *BMC Cancer*, vol. 20, no. 1, pp. 1–12, 2020, ISSN: 14712407. DOI: 10.1186/s12885-020-06836-y
- [248] P. Thuillier et al., “Prognostic value of whole-body PET volumetric parameters extracted from ⁶⁸Ga-DOTATOC-PET/CT in well-differentiated neuroendocrine tumors,” *Journal of Nuclear Medicine*, vol. 63, no. 7, 2022, ISSN: 2159662X. DOI: 10.2967/jnumed.121.262652
- [249] L. Urso et al., “Relevance of Volumetric Parameters Applied to [⁶⁸Ga]Ga-DOTATOC PET/CT in NET Patients Treated with PRRT,” *Diagnostics*, vol. 13, no. 4, pp. 1–11, 2023, ISSN: 20754418. DOI: 10.3390/diagnostics13040606

- [250] R. P. Reddy et al., "The quest for an accurate functional tumor volume with 68Ga-DOTATATE PET/CT," *Journal of Nuclear Medicine*, vol. 63, no. 7, pp. 1027–1032, 2022, ISSN: 2159662X. DOI: 10.2967/jnumed.121.262782
- [251] A. Tirosh et al., "Prognostic Utility of Total 68Ga-DOTATATE-Avid Tumor Volume in Patients With Neuroendocrine Tumors," *Gastroenterology*, vol. 154, no. 4, pp. 998–1008, 2018, ISSN: 15280012. DOI: 10.1053/j.gastro.2017.11.008
- [252] A. Santilli et al., "Automated full body tumor segmentation in DOTATATE PET/CT for neuroendocrine cancer patients," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, pp. 2083–2090, 2023, ISSN: 18616429. DOI: 10.1007/s11548-023-02968-1
- [253] J. Graf et al., "Prognostic Significance of Somatostatin Receptor Heterogeneity in Progressive Neuroendocrine Tumor Treated with Lu-177 DOTATOC or Lu-177 DOTATATE," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 4, pp. 881–894, 2020, ISSN: 16197089. DOI: 10.1007/s00259-019-04439-9
- [254] R. A. Werner et al., "Pre-therapy Somatostatin Receptor-Based Heterogeneity Predicts Overall Survival in Pancreatic Neuroendocrine Tumor Patients Undergoing Peptide Receptor Radionuclide Therapy," *Molecular Imaging and Biology*, vol. 21, no. 3, pp. 582–590, 2019, ISSN: 18602002. DOI: 10.1007/s11307-018-1252-5
- [255] H. Öner, Ü. Abdülrezzak, and A. Tutuş, "Could the skewness and kurtosis texture parameters of lesions obtained from pretreatment Ga-68 DOTA-TATE PET/CT images predict receptor radionuclide therapy response in patients with gastroenteropancreatic neuroendocrine tumors?" *Nuclear Medicine Communications*, pp. 1034–1039, 2020, ISSN: 14735628. DOI: 10.1097/MNM.0000000000001231
- [256] R. Laudicella et al., "[68Ga]DOTATOC PET/CT Radiomics to Predict the Response in GEP-NETs Undergoing [177Lu]DOTATOC PRRT: The "Theragnostics" Concept," *Cancers*, vol. 14, no. 4, p. 984, Feb. 2022, ISSN: 2072-6694. DOI: 10.3390/cancers14040984
- [257] V. Liberini et al., "Impact of segmentation and discretization on radiomic features in 68Ga-DOTA-TOC PET/CT images of neuroendocrine tumor," *EJNMMI Physics*, vol. 8, no. 1, 2021, ISSN: 21977364. DOI: 10.1186/s40658-021-00367-6
- [258] D. Morland et al., "Predicting subacute hematological toxicity of 177Lu-DOTATATE therapy using healthy organs' uptake on post-treatment quantitative SPECT: A pilot study," *Medicine (United States)*, vol. 101, no. 49, E32212, 2022, ISSN: 15365964. DOI: 10.1097/MD.00000000000032212
- [259] M. Mileva et al., "Prediction of 177Lu-DOTATATE PRRT Outcome Using Multimodality Imaging in Patients with Gastroenteropancreatic Neuroendocrine Tumors: Results from a Prospective Phase II LUMEN Study," *Journal of Nuclear Medicine*, vol. 65, no. 2, pp. 236–244, 2024, ISSN: 2159662X. DOI: 10.2967/jnumed.123.265987

- [260] R. Alipour et al., "The relationship between tumour dosimetry, response, and overall survival in patients with unresectable Neuroendocrine Neoplasms (NEN) treated with ^{177}Lu DOTATATE (LuTate)," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 50, no. 10, pp. 2997–3010, 2023, ISSN: 16197089. DOI: 10.1007/s00259-023-06257-6
- [261] C. Wetz et al., "The association of intra-therapeutic heterogeneity of somatostatin receptor expression with morphological treatment response in patients undergoing PRRT with [^{177}Lu]-DOTATATE," *PLoS ONE*, vol. 14, no. 5, pp. 1–14, 2019, ISSN: 19326203. DOI: 10.1371/journal.pone.0216781
- [262] S. Leyser et al., "Prognostic value of the post-treatment ^{177}Lu -DOTATOC scintigraphy in NET patients undergoing PRRT," in *EANM'23 Abstract Book Congress Sep 9-13, 2023. Eur J Nucl Med Mol Imaging*, 2023, S133.
- [263] S. C. Kappadath, W. D. Erwin, and R. E. Wendt, "SU-FF-I-76: Observed Inter-Camera Variability of Clinically Relevant Performance Characteristics for SIEMENS Symbia® Gamma Cameras," *Medical Physics*, vol. 33, no. 6, p. 2014, 2006, ISSN: 00942405. DOI: 10.1118/1.2240756
- [264] R. Buchert et al., "Reduction in camera-specific variability in [^{123}I]FP-CIT SPECT outcome measures by image reconstruction optimized for multisite settings: impact on age-dependence of the specific binding ratio in the ENC-DAT database of healthy controls," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 43, no. 7, pp. 1323–1336, 2016, ISSN: 16197089. DOI: 10.1007/s00259-016-3309-5
- [265] F. H. Fahey, P. E. Kinahan, R. K. Doot, M. Kocak, H. Thurston, and T. Y. Poussaint, "Variability in PET quantitation within a multicenter consortium," *Medical Physics*, vol. 37, no. 7, pp. 3660–3666, 2010, ISSN: 00942405. DOI: 10.1118/1.3455705
- [266] N. E. Makris, M. C. Huisman, P. E. Kinahan, A. A. Lammertsma, and R. Boellaard, "Evaluation of strategies towards harmonization of FDG PET/CT studies in multicentre trials: Comparison of scanner validation phantoms and data analysis procedures," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 10, pp. 1507–1515, 2013, ISSN: 16197070. DOI: 10.1007/s00259-013-2465-0
- [267] A. Kaalep et al., "EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 3, pp. 412–422, 2018, ISSN: 16197089. DOI: 10.1007/s00259-017-3853-7
- [268] A. Kaalep et al., "Quantitative implications of the updated EARL 2019 PET-CT performance standards," *EJNMMI Physics*, vol. 6, no. 1, pp. 1–16, 2019, ISSN: 21977364. DOI: 10.1186/s40658-019-0257-8

- [269] G. Akamatsu, Y. Tsutsui, H. Daisaki, K. Mitsumoto, S. Baba, and M. Sasaki, "A review of harmonization strategies for quantitative PET," *Annals of Nuclear Medicine*, vol. 37, no. 2, pp. 71–88, 2023, ISSN: 18646433. DOI: 10.1007/s12149-022-01820-x
- [270] J. Tran-Gia, "Harmonisation in quantitative SPECT and its challenges," in *European Journal of Nuclear Medicine and Molecular Imaging*, Aug. 2023, S123. DOI: 10.1007/s00259-023-06333-x
- [271] J. Wevrett et al., "Inter-comparison of quantitative imaging of lutetium-177 (^{177}Lu) in European hospitals," *EJNMMI Physics*, vol. 5, no. 1, 2018, ISSN: 21977364. DOI: 10.1186/s40658-018-0213-z
- [272] S. M. Peters et al., "Towards standardization of absolute SPECT/CT quantification: a multi-center and multi-vendor phantom study," *EJNMMI Physics*, vol. 6, no. 1, 2019, ISSN: 21977364. DOI: 10.1186/s40658-019-0268-5
- [273] S. M. Peters et al., "Variability in lutetium-177 SPECT quantification between different state-of-the-art SPECT/CT systems," *EJNMMI Physics*, vol. 7, no. 1, 2020, ISSN: 21977364. DOI: 10.1186/s40658-020-0278-3
- [274] J. Dickson et al., "EANM EARL Accreditation for Quantitative ^{177}Lu SPECT," in *British Nuclear Medicine Society Spring Meeting 13th - 15th May 2024*, 2024, pp. 428–429.
- [275] M. Vanderhoek, S. B. Perlman, and R. Jeraj, "Impact of the definition of peak standardized uptake value on quantification of treatment response," *Journal of Nuclear Medicine*, vol. 53, no. 1, pp. 4–11, 2012, ISSN: 01615505. DOI: 10.2967/jnumed.111.093443
- [276] A. J. Weisman, T. J. Bradshaw, M. Namias, and R. Jeraj, "Impact of scanner harmonization on PET-based treatment response assessment in metastatic melanoma," *Physics in Medicine and Biology*, vol. 65, no. 22, 2020, ISSN: 13616560. DOI: 10.1088/1361-6560/abb6bb
- [277] S. Shearer, A. J. Gemmell, R. Gillen, C. Reilly, and H. J. Wallace, "Multi-site comparison of imaging protocols for quantitative $^{99\text{m}}\text{Tc}$ SPECT-CT," in *EANM'24 Abstract Book Congress Oct 19-23, 2024. Eur J Nucl Med Mol Imaging*, 2024, S719–S720.
- [278] C. Kratochwil et al., "SUV of ^{68}Ga]DOTATOC-PET/CT Predicts Response Probability of PRRT in Neuroendocrine Tumors," *Molecular Imaging and Biology*, vol. 17, no. 3, pp. 313–318, 2015, ISSN: 18602002. DOI: 10.1007/s11307-014-0795-3
- [279] K. K. Wong et al., "Differences in tumor-to-normal organ SUV ratios measured with ^{68}Ga -DOTATATE PET compared with ^{177}Lu -DOTATATE SPECT in patients with neuroendocrine tumors," *Nuclear Medicine Communications*, vol. 43, no. 8, pp. 892–900, 2022, ISSN: 0143-3636. DOI: 10.1097/mnm.0000000000001592
- [280] M. Weber et al., "Changes in tumor-to-blood ratio as a prognostic marker for progression-free survival and overall survival in neuroendocrine tumor patients

- undergoing PRRT,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 51, no. 3, pp. 841–851, 2024, ISSN: 16197089. DOI: 10.1007/s00259-023-06502-y
- [281] C. Lasnon et al., “Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 7, pp. 985–996, Jul. 2013, ISSN: 1619-7070. DOI: 10.1007/s00259-013-2391-1
- [282] B. Spottiswoode and D. Thomas, “White Paper: Harmonizing SUV with EQ PET,” Siemens Healthineers, Tech. Rep., 2020.
- [283] E. Quak et al., “Harmonizing FDG PET quantification while maintaining optimal lesion detection: prospective multicentre validation in 517 oncology patients,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 42, no. 13, pp. 2072–2082, 2015, ISSN: 16197089. DOI: 10.1007/s00259-015-3128-0
- [284] W. E. Johnson, C. Li, and A. Rabinovic, “Adjusting batch effects in microarray expression data using empirical Bayes methods,” *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007, ISSN: 14654644. DOI: 10.1093/biostatistics/kxj037
- [285] J. Karvanen, “The Statistical Basis of Laboratory Data Normalization,” *Therapeutic Innovation & Regulatory Science*, vol. 37, no. 1, pp. 101–107, 2003, ISSN: 21684790. DOI: 10.1177/009286150303700112
- [286] F. Orlhac et al., “A postreconstruction harmonization method for multicenter radiomic studies in PET,” *Journal of Nuclear Medicine*, vol. 59, no. 8, pp. 1321–1328, 2018, ISSN: 2159662X. DOI: 10.2967/jnumed.117.199935
- [287] J.-P. Fortin, *neuroCombat: Harmonization of multi-site imaging data with ComBat*, 2021.
- [288] R. Da-ano et al., “Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies,” *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, ISSN: 20452322. DOI: 10.1038/s41598-020-66110-w
- [289] M. Ferreira et al., “[18F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation,” *European Journal of Nuclear Medicine and Molecular Imaging*, 2021, ISSN: 16197089. DOI: 10.1007/s00259-021-05303-5
- [290] C. K. Stein et al., “Removing batch effects from purified plasma cell gene expression microarrays with modified ComBat,” *BMC Bioinformatics*, vol. 16, no. 1, pp. 1–9, 2015, ISSN: 14712105. DOI: 10.1186/s12859-015-0478-3
- [291] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, vol. 58, 2019, ISSN: 13618423. DOI: 10.1016/j.media.2019.101552

- [292] A. Yadav et al., "A comparative analysis of image harmonization techniques in mitigating differences in CT acquisition and reconstruction," *Physics in Medicine & Biology*, pp. 1–10, Jan. 2025, ISSN: 0031-9155. DOI: 10.1088/1361-6560/adabad
- [293] R. Panda, S. V. Kalmady, and R. Greiner, "Multi-Source Domain Adaptation Techniques for Mitigating Batch Effects: A Comparative Study," *Frontiers in Neuroinformatics*, vol. 16, no. DI, Apr. 2022, ISSN: 1662-5196. DOI: 10.3389/fninf.2022.805117
- [294] Y. Nan et al., "Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions," *Information Fusion*, vol. 82, pp. 99–122, 2022. DOI: 10.1016/j.inffus.2022.01.001
- [295] E. Busemann Sokole, A. Płachcińska, A. Britten, M. Lyra Georgosopoulou, W. Tindale, and R. Klett, "Routine quality control recommendations for nuclear medicine instrumentation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 3, pp. 662–671, 2010, ISSN: 16197070. DOI: 10.1007/s00259-009-1347-y
- [296] J. I. Gear et al., "EANM practical guidance on uncertainty analysis for molecular radiotherapy absorbed dose calculations," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 13, pp. 2456–2474, 2018, ISSN: 16197089. DOI: 10.1007/s00259-018-4136-7
- [297] R. Gadd et al., "National Measurement Good Practice Guide No.93: Protocol for Establishing and Maintaining the Calibration of Medical Radionuclide Calibrators and their Quality Control," National Physical Laboratory, Tech. Rep. 93, 2006.
- [298] A. Fenwick, M. Baker, K. Ferreira, and J. Keightley, "Comparison of ^{90}Y and ^{177}Lu measurement capability in UK and European hospitals," *Applied Radiation and Isotopes*, vol. 87, pp. 10–13, 2014, ISSN: 18729800. DOI: 10.1016/j.apradiso.2013.11.050
- [299] A. Sarnelli et al., "An operational approach for accurate ^{177}Lu and $^{177\text{m}}\text{Lu}$ activity quantifications to comply with the environmental release criteria: the role of GEANT4 for efficiency curve and True Coincidence Summing effect estimation," *European Physical Journal Plus*, vol. 139, no. 3, 2024, ISSN: 21905444. DOI: 10.1140/epjp/s13360-024-04875-z
- [300] W. A. McDougald, R. S. Miyaoka, A. M. Alessio, R. L. Harrison, and T. K. Lewellen, "A study of SPECT/CT camera stability for quantitative imaging," *EJNMMI Physics*, vol. 3, no. 1, pp. 1–13, 2016, ISSN: 21977364. DOI: 10.1186/s40658-016-0150-7
- [301] J. I. Gear et al., "Abdo-Man: a 3D-printed anthropomorphic phantom for validating quantitative SIRT," *EJNMMI Physics*, vol. 3, no. 1, 2016, ISSN: 21977364. DOI: 10.1186/s40658-016-0151-6

- [302] A. A. Bolster and W. Waddington, "Gamma Camera Performance: technical assessment protocol," Medical Devices Agency, Tech. Rep., 1996, p. 33.
- [303] P. Jarritt, W. Waddington, and R. S. Lawson, "Assessment of performance for single photon emission computed tomography," in *Report 111: Quality Control of Gamma Cameras and Nuclear Medicine Computer Systems*, E. Eadie, Ed., Institute of Physics in Engineering and Medicine, 2015, ch. 5, pp. 70–98, ISBN: 978-1903613603.
- [304] Siemens Healthineers, "Broad Quantification Calibration Procedures SPECT/CT," Siemens Healthineers, Tech. Rep., 2018, pp. 1–8.
- [305] D. Scott et al., "Measurement uncertainty in three sample glomerular filtration rate tests," *Nuclear Medicine Communications*, 2026, ISSN: 14735628. DOI: 10.1097/MNM.0000000000002161
- [306] M. D'ariento and M. Cox, "Uncertainty Analysis in the Calibration of an Emission Tomography System for Quantitative Imaging," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017, ISSN: 17486718. DOI: 10.1155/2017/9830386
- [307] Joint Committee For Guides In Metrology, "Evaluation of measurement data — Guide to the expression of uncertainty in measurement," Tech. Rep. September, 2008, p. 134.
- [308] K. M. Ferreira et al., "Preparation of radioactive phantoms with an activity traceable to International standards," National Physical Laboratory, Teddington, UK, Tech. Rep., Aug. 2025. DOI: 10.47120/npl.mgpg154
- [309] M. D'Arienzo et al., "Gamma camera calibration and validation for quantitative SPECT imaging with ^{177}Lu ," *Applied Radiation and Isotopes*, vol. 112, pp. 156–164, 2016, ISSN: 18729800. DOI: 10.1016/j.apradiso.2016.03.007
- [310] J. Tran-Gia et al., "A multicentre and multi-national evaluation of the accuracy of quantitative Lu-177 SPECT/CT imaging performed within the MRTDosimetry project," *EJNMMI Physics*, vol. 8, no. 1, 2021, ISSN: 21977364. DOI: 10.1186/s40658-021-00397-0
- [311] C. Saldarriaga Vargas et al., "An international multi-center investigation on the accuracy of radionuclide calibrators in nuclear medicine theragnostics," *EJNMMI Physics*, vol. 7, no. 1, 2020, ISSN: 21977364. DOI: 10.1186/s40658-020-00338-3
- [312] C. F. Uribe, P. L. Esquinas, M. Gonzalez, and A. Celler, "Characteristics of Bremsstrahlung emissions of ^{177}Lu , ^{188}Re , and ^{90}Y for SPECT/CT quantification in radionuclide therapy," *Physica Medica*, vol. 32, no. 5, pp. 691–700, 2016, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2016.04.014
- [313] J. v. Dalen, B. Hendrickx, F. Corstens, and W. Oyen, "Optimization scan protocol of Lu-177 in planar imaging using dual energy window scatter correction," *Journal of Nuclear Medicine*, vol. 48, no. supplement 2, 425P LP –425P, May 2007.

- [314] T. Melhuish, A. Nautiyal, F. Sundram, Z. Saad, and S. Michopoulou, “30-minute whole-body [177Lu]Lu-PSMA SPECT-CT: evaluating quantitative accuracy of reduced time xSPECT scans,” in *Nuclear Medicine Communications: British Nuclear Medicine Society Spring Meeting 22nd - 24th May 2023*, 2023.
- [315] Siemens Healthineers, “Using the xRecon Activity for xSPECT Reconstruction,” Siemens Healthineers, Tech. Rep., 2018.
- [316] Data Spectrum Corporation and K. Greer, “NEMA IEC Body Phantom set: User’s Manual,” Data Spectrum Corporation, Tech. Rep., 2008.
- [317] National Electrical Manufacturers Association, “NEMA NU 2-2018: Performance Measurements of Positron Emission Tomographs (PET),” Tech. Rep., 2018.
- [318] Hermes Medical Solutions, “Hermes Hybrid Viewer Handbook v4.0,” Hermes Medical Solutions, Tech. Rep., 2019.
- [319] R. Gillen, K. Erlandsson, A. M. Denis-Bacelar, K. Thielemans, B. F. Hutton, and S. J. McQuaid, “Towards accurate partial volume correction in 99mTc oncology SPECT: perturbation for case-specific resolution estimation,” *EJNMMI Physics*, vol. 9, no. 1, 2022, ISSN: 21977364. DOI: 10.1186/s40658-022-00489-5
- [320] Y. Liu, Z. Lu, G. Chen, K. Shi, and G. S. P. Mok, “Partial volume correction for Lu-177-PSMA SPECT,” *EJNMMI Physics*, vol. 11, no. 1, p. 93, Nov. 2024, ISSN: 2197-7364. DOI: 10.1186/s40658-024-00697-1
- [321] O. L. Munk, L. P. Tolbod, S. B. Hansen, and T. V. Bogsrud, “Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative,” *EJNMMI Physics*, vol. 4, no. 1, pp. 1–12, 2017, ISSN: 21977364. DOI: 10.1186/s40658-016-0169-9
- [322] A. H. Vija, E. G. Hawman, and J. C. Engdahl, “Analysis of a SPECT OSEM reconstruction method with 3D beam modeling and optional attenuation correction: Phantom studies,” *IEEE Nuclear Science Symposium Conference Record*, vol. 4, pp. 2662–2666, 2003, ISSN: 10957863. DOI: 10.1109/nssmic.2003.1352436
- [323] C. Reilly et al., “Characterisation of 99mTc EDDA/HYNIC-TOC (Tektrotyd) physiological and neuroendocrine tumour uptake using SPECT/CT standardised uptake values: initial experience,” *Nuclear medicine communications*, vol. 42, no. 8, pp. 935–939, Aug. 2021, ISSN: 1473-5628. DOI: 10.1097/MNM.0000000000001416
- [324] M. Koole et al., “EANM guidelines for PET-CT and PET-MR routine quality control,” *Zeitschrift für Medizinische Physik*, 2022, ISSN: 18764436. DOI: 10.1016/j.zemedi.2022.08.003
- [325] H. Piwowska-Bilska, S. Kurkowska, and B. Birkenfeld, “Optimized method for normal range estimation of standardized uptake values (SUVmax, SUVmean) in liver SPECT/CT images with somatostatin analog [99mTc]-HYNIC-TOC (Tektrotyd),”

- Nuclear Medicine Review*, vol. 25, no. 1, pp. 37–46, 2022, ISSN: 16444345. DOI: 10.5603/NMR.a2022.0008
- [326] M. A. Park, A. Mahmood, R. E. Zimmerman, N. Limpa-Amara, G. M. Makrigiorgos, and S. C. Moore, “Adsorption of metallic radionuclides on plastic phantom walls,” *Medical Physics*, vol. 35, no. 4, pp. 1606–1610, 2008, ISSN: 00942405. DOI: 10.1118/1.2871191
- [327] A. Seret, D. Nguyen, and C. Bernard, “Quantitative capabilities of four state-of-the-art SPECT-CT cameras,” *EJNMMI Research*, vol. 2, no. 1, p. 45, 2012, ISSN: 2191-219X. DOI: 10.1186/2191-219X-2-45
- [328] A. N. Spiess and N. Neumeyer, “An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach,” *BMC Pharmacology*, vol. 10, pp. 1–11, 2010, ISSN: 14712210. DOI: 10.1186/1471-2210-10-6
- [329] H. Marquis, K. P. Willowson, and D. L. Bailey, “Partial volume effect in SPECT & PET imaging and impact on radionuclide dosimetry estimates,” *Asia Oceania Journal of Nuclear Medicine and Biology*, vol. 11, no. 1, pp. 44–54, 2023, ISSN: 23225726. DOI: 10.22038/AOJNMB.2022.63827.1448
- [330] M. Acheson and C. Campfield, “Use of 3D-printed hollow spheres and a Jaszczak Phantom in enabling quantification of [¹⁷⁷Lu]Lu-DOTATATE activity using SPECT-CT images,” in *Nuclear Medicine Communications: British Nuclear Medicine Society Spring Meeting 13th - 15th May 2024*, 2024, pp. 420–461.
- [331] T. Ryden, M. Van Essen, I. Marin, J. Svensson, and P. Bernhardt, “Deep-learning generation of synthetic intermediate projections improves ¹⁷⁷Lu SPECT images reconstructed with sparsely acquired projections,” *Journal of Nuclear Medicine*, vol. 62, no. 4, pp. 528–535, 2021, ISSN: 2159662X. DOI: 10.2967/jnumed.120.245548
- [332] A. H. Vija, “Characteristics of the xSPECT reconstruction method: Siemens White Paper,” Tech. Rep.
- [333] C. M. Brown, “Application of Novel Corrections for Quantification of ¹²³I SPECT,” Ph.D. dissertation, University of Glasgow, 2018, p. 338. DOI: <https://theses.gla.ac.uk/8923/>
- [334] W. A. Shewhart, “Quality Control Charts,” *Bell System Technical Journal*, vol. 5, no. 4, pp. 593–603, Oct. 1926, ISSN: 00058580. DOI: 10.1002/j.1538-7305.1926.tb00125.x
- [335] A. C. Knight and E. D. Williams, “An evaluation of cusum analysis and control charts applied to quantitative gamma-camera uniformity parameters for automated quality control,” *European Journal of Nuclear Medicine*, vol. 19, no. 2, pp. 125–130, 1992, ISSN: 03406997. DOI: 10.1007/BF00184128

- [336] A. Fenwick, "Quality control charts," in *Radioactive Sample Counting: Principles and Practice (Second Edition)*, S. Michopoulou, Ed., 2nd editio, Institute of Physics, 2022, ch. Chapter 3, pp. 3.1–3.5, ISBN: 978-0-7503-3895-0.
- [337] D. Pearson and S. A. Cawte, "Long-term quality control of DXA: A comparison of Shewhart rules and cusum charts," *Osteoporosis International*, vol. 7, no. 4, pp. 338–343, 1997, ISSN: 0937941X. DOI: 10.1007/BF01623774
- [338] G. Marin et al., "Accuracy and precision assessment for activity quantification in individualized dosimetry of ¹⁷⁷Lu-DOTATATE therapy," *EJNMMI Physics*, vol. 4, no. 1, 2017, ISSN: 21977364. DOI: 10.1186/s40658-017-0174-7
- [339] R. S. Lawson, "Assessment of planar performance," in *Report 111: Quality Control of Gamma Cameras and Nuclear Medicine Computer Systems*, Institute of Physics in Engineering and Medicine, 2015, ch. 3, pp. 8–58, ISBN: 978-1903613603.
- [340] W. Rogers et al., "Radiomics: from qualitative to quantitative imaging," *The British journal of radiology*, vol. 93, no. 1108, p. 20190948, 2020, ISSN: 1748880X. DOI: 10.1259/bjr.20190948
- [341] E. Pfaehler et al., "Plausibility and redundancy analysis to select FDG-PET textural features in non-small cell lung cancer," *Medical Physics*, vol. 48, no. 3, pp. 1226–1238, 2021, ISSN: 00942405. DOI: 10.1002/mp.14684
- [342] M. Weber et al., "Textural analysis of hybrid DOTATOC-PET/MRI and its association with histological grading in patients with liver metastases from neuroendocrine tumors," *Nuclear Medicine Communications*, pp. 363–369, 2020, ISSN: 14735628. DOI: 10.1097/MNM.0000000000001150
- [343] H. Xu et al., "Radiomics prognostic analysis of PET/CT images in a multicenter head and neck cancer cohort: investigating ComBat strategies, sub-volume characterization, and automatic segmentation," *European Journal of Nuclear Medicine and Molecular Imaging*, no. 0123456789, 2023, ISSN: 16197089. DOI: 10.1007/s00259-023-06118-2
- [344] S. De Schepper, G. Gnanasegaran, J. C. Dickson, and T. Van den Wyngaert, "Absolute quantification in diagnostic spect/ct: The phantom premise," *Diagnostics*, vol. 11, no. 12, pp. 1–15, 2021, ISSN: 20754418. DOI: 10.3390/diagnostics11122333
- [345] A. Forgacs et al., "A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images," *PLoS ONE*, vol. 11, no. 10, pp. 1–14, 2016, ISSN: 19326203. DOI: 10.1371/journal.pone.0164113
- [346] M. Ljungberg, "Absolute Quantitation of SPECT Studies," *Seminars in Nuclear Medicine*, vol. 48, no. 4, pp. 348–358, 2018, ISSN: 15584623. DOI: 10.1053/j.semnuclmed.2018.02.009

- [347] S. Pathmanandavel et al., "Evaluation of ^{177}Lu -PSMA-617 SPECT/CT Quantitation as a Response Biomarker Within a Prospective ^{177}Lu -PSMA-617 and NOX66 Combination Trial (LuPIN)," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 64, no. 2, pp. 221–226, 2023, ISSN: 15355667. DOI: 10.2967/jnumed.122.264398
- [348] J. Gustafsson, A. Sundlöv, and K. Sjögreen Gleisner, "SPECT image segmentation for estimation of tumour volume and activity concentration in ^{177}Lu -DOTATATE radionuclide therapy," *EJNMMI Research*, vol. 7, no. 1, 2017, ISSN: 2191219X. DOI: 10.1186/s13550-017-0262-7
- [349] M. Hatt, C. C. Le Rest, F. Tixier, B. Badic, U. Schick, and D. Visvikis, "Radiomics: Data are also images," *Journal of Nuclear Medicine*, vol. 60, no. 9, 38S–44S, 2019, ISSN: 2159662X. DOI: 10.2967/jnumed.118.220582
- [350] R. A. Werner et al., "Survival prediction in patients undergoing radionuclide therapy based on intratumoral somatostatin-receptor heterogeneity," *Oncotarget*, vol. 8, no. 4, pp. 7039–7049, 2017, ISSN: 19492553. DOI: 10.18632/oncotarget.12402
- [351] A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, "Comparison of quantitative Krenning Scores with visual assessment in $^{99\text{mTc}}$ -EDDA/HYNIC-TOC SPECT-CT," *Nuclear Medicine Communications*, vol. 46, no. 6, pp. 515–522, Jun. 2025, ISSN: 0143-3636. DOI: 10.1097/MNM.0000000000001967
- [352] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.
- [353] K. J. Archer, "glmnetcr: An R Package for Ordinal Response Prediction in High-dimensional Data Settings," <https://api.semanticscholar.org/CorpusID:9791012>, 2010.
- [354] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977, ISSN: 0006341X. DOI: 10.2307/2529310
- [355] Z. Cheng et al., "Prognostic Value of Somatostatin Receptor-Derived Volumetric Parameters from a Hybrid Standardized Uptake Value Thresholding Method in Patients with ^{68}Ga -DOTATATE-Avid Stage IV Neuroendocrine Neoplasms: A Preliminary Study," *Neuroendocrinology*, pp. 1–11, Apr. 2023, ISSN: 0028-3835. DOI: 10.1159/000530771
- [356] A. Gafita et al., "Tumor Sink Effect in ^{68}Ga -PSMA-11 PET: Myth or Reality?" *Journal of Nuclear Medicine*, vol. 63, no. 2, pp. 226–232, 2022, ISSN: 2159662X. DOI: 10.2967/JNUMED.121.261906
- [357] B. L. Viglianti et al., "Effects of Tumor Burden on Reference Tissue Standardized Uptake for PET Imaging: Modification of PERCIST Criteria," *Radiology*, vol. 287, no. 3, pp. 993–1002, 2018, ISSN: 15271315. DOI: 10.1148/radiol.2018171356

- [358] M. S. Hofman, W. F. Eddie Lau, and R. J. Hicks, "Somatostatin receptor imaging with ^{68}Ga DOTATATE PET/CT: Clinical utility, normal patterns, pearls, and pitfalls in interpretation1," *Radiographics*, vol. 35, no. 2, pp. 500–516, 2015, ISSN: 15271323. DOI: 10.1148/rg.352140164
- [359] T. A. Hope, M. Pavel, and E. K. Bergsland, "Neuroendocrine Tumors and Peptide Receptor Radionuclide Therapy: When Is the Right Time?" *Journal of Clinical Oncology*, vol. 31, no. 24, pp. 2818–2830, 2022, ISSN: 15277755. DOI: 10.1200/JCO.22.00176
- [360] G. Kong and R. J. Hicks, "Peptide Receptor Radiotherapy: Current Approaches and Future Directions," *Current Treatment Options in Oncology*, vol. 20, no. 10, 2019, ISSN: 15346277. DOI: 10.1007/s11864-019-0677-7
- [361] L. Mansi, "Radiolabeled somatostatin analogues for the diagnosis and therapy of NETs: upcoming new strategies," *Clinical and Translational Imaging*, vol. 8, no. 1, pp. 21–27, 2020, ISSN: 22817565. DOI: 10.1007/s40336-020-00355-3
- [362] J. M. Beauregard, M. S. Hofman, G. Kong, and R. J. Hicks, "The tumour sink effect on the biodistribution of ^{68}Ga -DOTA- octreotate: Implications for peptide receptor radionuclide therapy," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 39, no. 1, pp. 50–56, 2012, ISSN: 16197070. DOI: 10.1007/s00259-011-1937-3
- [363] H. Siebinga, B. J. de Wit-van der Veen, J. H. Beijnen, T. P. Dorlo, A. D. Huitema, and J. J. Hendriks, "A physiologically based pharmacokinetic model for [^{68}Ga]Ga-(HA-)DOTATATE to predict whole-body distribution and tumor sink effects in GEP-NET patients," *EJNMMI Research*, vol. 13, no. 1, 2023, ISSN: 2191219X. DOI: 10.1186/s13550-023-00958-7
- [364] A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, "Quantitative uptake in $^{99\text{m}}\text{Tc}$ -EDDA/HYNIC-TOC somatostatin receptor imaging – the effect of long-acting release somatostatin analogue therapy," *Nuclear Medicine Communications*, vol. 44, no. 11, pp. 944–952, Nov. 2023, ISSN: 0143-3636. DOI: 10.1097/MNM.0000000000001746
- [365] A. K. Tahari, D. Chien, J. R. Azadi, and R. L. Wahl, "Optimum lean body formulation for correction of standardized uptake value in PET imaging," *Journal of Nuclear Medicine*, vol. 55, no. 9, pp. 1481–1484, 2014, ISSN: 2159662X. DOI: 10.2967/jnumed.113.136986
- [366] R. M. Antonijoan et al., "Pharmacokinetics of a new Autogel formulation of the somatostatin analogue lanreotide after a single subcutaneous dose in healthy volunteers," *Journal of Pharmacy and Pharmacology*, vol. 56, no. 4, pp. 471–476, 2010, ISSN: 0022-3573. DOI: 10.1211/0022357023123
- [367] A. Tirosh et al., "Association between neuroendocrine tumors biomarkers and primary tumor site and disease type based on total ^{68}Ga -DOTATATE-Avid tumor volume

- measurements,” *European Journal of Endocrinology*, vol. 176, no. 5, pp. 575–582, May 2017, ISSN: 0804-4643. DOI: 10.1530/EJE-16-1079
- [368] B. Rojas et al., “Nearly double the patients and dramatic changes over 14 years of UK MRT: Internal Dosimetry Users Group survey results from 2007 to 2021,” *Nuclear Medicine Communications*, vol. 45, no. 1, pp. 16–23, 2023, ISSN: 14735628. DOI: 10.1097/MNM.0000000000001780
- [369] T. Brabander et al., “Long-term efficacy, survival, and safety of [177Lu-DOTA0,Tyr3]octreotate in patients with gastroenteropancreatic and bronchial neuroendocrine tumors,” *Clinical Cancer Research*, vol. 23, no. 16, pp. 4617–4624, 2017, ISSN: 15573265. DOI: 10.1158/1078-0432.CCR-16-2743
- [370] V. Prasad et al., “Challenges in developing response evaluation criteria for peptide receptor radionuclide therapy: A consensus report from the European Neuroendocrine Tumor Society Advisory Board Meeting 2022 and the ENETS Theranostics Task Force,” *Journal of Neuroendocrinology*, no. November, pp. 1–13, 2024, ISSN: 13652826. DOI: 10.1111/jne.13479
- [371] R. L. Korn and J. J. Crowley, “Overview: Progression-free survival as an endpoint in clinical trials with solid tumors,” *Clinical Cancer Research*, vol. 19, no. 10, pp. 2607–2612, 2013, ISSN: 10780432. DOI: 10.1158/1078-0432.CCR-12-2934
- [372] M. Buyse, D. J. Sargent, A. Grothey, A. Matheson, and A. De Gramont, “Biomarkers and surrogate end points-The challenge of statistical validation,” *Nature Reviews Clinical Oncology*, vol. 7, no. 6, pp. 309–317, 2010, ISSN: 17594774. DOI: 10.1038/nrclinonc.2010.43
- [373] Y. Tang, P. Bycott, Ö. Åkerborg, L. Jönsson, S. Negrier, and C. Chen, “Interpreting overall survival results when progression-free survival benefits exist in today’s oncology landscape: A metastatic renal cell carcinoma case study,” *Cancer Management and Research*, vol. 6, pp. 365–371, 2014, ISSN: 11791322. DOI: 10.2147/CMAR.S67249
- [374] M. Weber, T. Telli, D. Kersting, and R. Seifert, “Prognostic Implications of PET-Derived Tumor Volume and Uptake in Patients with Neuroendocrine Tumors,” *Cancers*, vol. 15, no. 14, 2023, ISSN: 20726694. DOI: 10.3390/cancers15143581
- [375] R. Simon and D. G. Altman, “Statistical aspects of prognostic factor studies in oncology,” *Br. J. Cancer*, vol. 69, pp. 979–985, 1994.
- [376] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, “Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit,” *British Journal of Cancer*, vol. 89, no. 4, pp. 605–611, 2003, ISSN: 00070920. DOI: 10.1038/sj.bjc.6601120
- [377] B. G. Naraev, R. A. Ramirez, A. T. Kendi, and T. R. Halfdanarson, “Peptide Receptor Radionuclide Therapy for Patients With Advanced Lung Carcinoids,” *Clinical Lung*

- Cancer*, vol. 20, no. 3, e376–e392, 2019, ISSN: 19380690. DOI: 10.1016/j.cl1c.2019.02.007
- [378] S. Ha, Y. I. Kim, J. S. Oh, C. Yoo, B. Y. Ryoo, and J. S. Ryu, “Prediction of [177Lu]Lu-DOTA-TATE therapy response using the absorbed dose estimated from [177Lu]Lu-DOTA-TATE SPECT/CT in patients with metastatic neuroendocrine tumour,” *EJNMMI Physics*, vol. 11, no. 1, 2024, ISSN: 21977364. DOI: 10.1186/s40658-024-00620-8
- [379] D. Roth et al., “Dosimetric Quantities in Neuroendocrine Tumors over Treatment Cycles with 177Lu-DOTATATE,” *Journal of Nuclear Medicine*, vol. 63, no. 3, pp. 399–405, 2022, ISSN: 2159662X. DOI: 10.2967/JNUMED.121.262069
- [380] J. Hou, Y. Yang, N. Chen, D. Chen, and S. Hu, “Prognostic Value of Volume-Based Parameters Measured by SSTR PET/CT in Neuroendocrine Tumors: A Systematic Review and Meta-Analysis,” *Frontiers in Medicine*, vol. 8, no. November, pp. 1–9, 2021, ISSN: 2296858X. DOI: 10.3389/fmed.2021.771912
- [381] D. G. Altman, B. Lausen, W. Sauerbrei, and M. Schumacher, “Dangers of Using “Optimal” Cutpoints in the Evaluation of Prognostic Factors,” *JNCI Journal of the National Cancer Institute*, vol. 86, no. 11, pp. 829–835, Jun. 1994, ISSN: 0027-8874. DOI: 10.1093/jnci/86.11.829
- [382] C. Contal and J. O’Quigley, “An application of changepoint methods in studying the effect of age on survival in breast cancer,” *Computational Statistics and Data Analysis*, vol. 30, no. 3, pp. 253–270, 1999, ISSN: 01679473. DOI: 10.1016/S0167-9473(98)00096-6
- [383] R. C. MacCallum, S. Zhang, K. J. Preacher, and D. D. Rucker, “On the practice of dichotomization of quantitative variables,” *Psychological Methods*, vol. 7, no. 1, pp. 19–40, 2002, ISSN: 1082989X. DOI: 10.1037/1082-989X.7.1.19
- [384] R. L. Nuzzo, “Making Continuous Measurements into Dichotomous Variables,” *PM and R*, vol. 11, no. 10, pp. 1132–1134, 2019, ISSN: 19341482. DOI: 10.1002/pmrj.12228
- [385] J. Greenhalgh, “The applications of PROs in clinical practice: What are they, do they work, and why?” *Quality of Life Research*, vol. 18, no. 1, pp. 115–123, 2009, ISSN: 09629343. DOI: 10.1007/s11136-008-9430-6
- [386] N. K. Aaronson et al., “The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology,” *JNCI Journal of the National Cancer Institute*, vol. 85, no. 5, pp. 365–376, Mar. 1993, ISSN: 0027-8874. DOI: 10.1093/jnci/85.5.365
- [387] C. Martini et al., “Quality of life in patients with metastatic gastroenteropancreatic neuroendocrine tumors receiving peptide receptor radionuclide therapy: Information from a monitoring program in clinical routine,” *Journal of Nuclear Medicine*, vol. 59, no. 10, pp. 1566–1573, 2018, ISSN: 2159662X. DOI: 10.2967/jnumed.117.204834

- [388] K. Edfeldt et al., "Improved health-related quality of life during peptide receptor radionuclide therapy in patients with neuroendocrine tumours," *Journal of Neuroendocrinology*, vol. 35, no. 10, pp. 1–7, 2023, ISSN: 13652826. DOI: 10.1111/jne.13342
- [389] R. Gosain, M. Gupta, A. M. Roy, J. Strosberg, K. M. Glaser, and R. Iyer, "Health-Related Quality of Life (HRQoL) in Neuroendocrine Tumors: A Systematic Review," *Cancers*, vol. 14, no. 6, pp. 1–14, 2022, ISSN: 20726694. DOI: 10.3390/cancers14061428
- [390] J. Strosberg et al., "Health-related quality of life in patients with progressive midgut neuroendocrine tumors treated with 177 lu-dotatate in the phase III netter-1 trial," *Journal of Clinical Oncology*, vol. 36, no. 25, pp. 2578–2584, 2018, ISSN: 15277755. DOI: 10.1200/JCO.2018.78.5865
- [391] S. Khan, E. P. Krenning, M. Van Essen, B. L. Kam, J. J. Teunissen, and D. J. Kwekkeboom, "Quality of life in 265 patients with gastroenteropancreatic or bronchial neuroendocrine tumors treated with [177Lu-DOTA 0,Tyr 3]octreotate," *Journal of Nuclear Medicine*, vol. 52, no. 9, pp. 1361–1368, 2011, ISSN: 01615505. DOI: 10.2967/jnumed.111.087932
- [392] H. Ohlsson, A. Gålne, E. Trägårdh, M. Malmström, A. Sundlöv, and M. Almquist, "Relationship between somatostatin receptor expressing tumour volume and health-related quality of life in patients with metastatic GEP-NET," *Journal of Neuroendocrinology*, vol. 34, no. 6, pp. 1–9, 2022, ISSN: 13652826. DOI: 10.1111/jne.13139
- [393] P. Fayers, N. Aaronson, B. K. M. Groenvold, D. Curran, and A. Bottomley, *The EORTC QLQ-C30 Scoring Manual (3rd Edition)*. <http://www.eortc.be/qol/files/scmanualqlq-c30.pdf>: European Organisation for Research and Treatment of Cancer, 2001, vol. 30, pp. 1–67, ISBN: 2-9300 64-22-6.
- [394] J. M. Giesinger et al., "Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust," *Journal of Clinical Epidemiology*, vol. 69, pp. 79–88, 2016, ISSN: 18785921. DOI: 10.1016/j.jclinepi.2015.08.007
- [395] R. Baser, *PROscorer: Functions to score commonly-used Patient-Reported Outcome (PRO) measures and other psychometric instruments*, 2023. DOI: 10.32614/CRAN.package.PROscorer
- [396] EORTC Quality of Life Group, "EORTC QLQ-GI.NET21 Scoring Manual," EORTC, Brussels, Tech. Rep., 2013.
- [397] K. Cocks, M. T. King, G. Velikova, M. M. St-James, P. M. Fayers, and J. M. Brown, "Evidence-based guidelines for determination of sample size and interpretation of the European organisation for the research and treatment of cancer quality of life

- questionnaire core 30,” *Journal of Clinical Oncology*, vol. 29, no. 1, pp. 89–96, 2011, ISSN: 15277755. DOI: 10.1200/JCO.2010.28.0107
- [398] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2021.
- [399] T. J. Van der Weele and M. B. Mathur, “Some Desirable Properties of the Bonferroni Correction: Is the Bonferroni Correction Really So Bad?” *American Journal of Epidemiology*, vol. 188, no. 3, pp. 617–618, 2019, ISSN: 14766256. DOI: 10.1093/aje/kwy250
- [400] R. S. van Leeuwen et al., “A Multinational Pilot Study on Patients’ Perceptions of Advanced Neuroendocrine Neoplasms on the EORTC QLQ-C30 and EORTC QLQ-GINET21 Questionnaires,” *Journal of Clinical Medicine*, vol. 11, no. 5, pp. 1–10, 2022, ISSN: 20770383. DOI: 10.3390/jcm11051271
- [401] I. Mateo Vito, N. Santillan, N. Davis, and S. Navalkissoor, “Revisited: Quality of Life Outcome after each cycle in NET Patients Receiving Peptide Receptor Radionuclide Therapy (PRRT),” in *British Nuclear Medicine Society Spring Meeting 13th - 15th May 2024*, British Nuclear Medicine Society, 2024, p. 437.
- [402] J. P. Fortin et al., “Harmonization of multi-site diffusion tensor imaging data,” *NeuroImage*, vol. 161, no. March, pp. 149–170, 2017, ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.08.047
- [403] R. Da-Ano, D. Visvikis, and M. Hatt, “Harmonization strategies for multicenter radiomics investigations,” *Physics in Medicine and Biology*, vol. 65, no. 24, 2020, ISSN: 13616560. DOI: 10.1088/1361-6560/aba798
- [404] A. J. Gemmell, C. M. Brown, S. Ray, and A. Small, “Robustness of textural analysis features in quantitative 99 mTc and 177Lu SPECT-CT phantom acquisitions,” *EJNMMI Physics*, vol. 12, no. 1, p. 40, Apr. 2025, ISSN: 2197-7364. DOI: 10.1186/s40658-025-00749-0
- [405] “LifeX - Radiomic features,” <https://www.lifexsoft.org/index.php/resources/texture/radiomic-features>, Accessed on 27th June, 2024.