Quantification of Image Quality

1 Introduction

2 Methodology

3 Results

3.1 Collection of scans per clinical imaging task

3.1.1. PET

3.1.1.1 Whole body [^{18}F]FDG

3.1.1.2 Whole body [^{18}F]PSMA

3.1.1.3 Brain [^{18}F]FDG

3.1.1.4 PET Reconstruction and post-processing

3.1.2 Gamma camera

3.1.2.1 SPECT

3.1.2.1.1 Bone SPECT

3.1.2.1.2 Myocardial perfusion SPECT

3.1.2.1.3 DAT-scan SPECT

3.1.2.2 Planar

3.1.2.2.1 Planar bone scintigraphy

3.1.2.2.2 DMSA

3.1.2.2.3 Gastric Emptying

3.2 Systematic literature review for selection of Model Observer

3.3 Evaluation of applicability of standard phantoms

4 References

The current rapport was made possible by funding from SKMS, the Foundation for quality funds for medical specialists (Stichting Kwaliteitsgelden Medisch Specialisten)

1. Introduction

1.1 Problem definition

Various acquisition and reconstruction parameters have an effect on how images appear on a (diagnostic) monitor. In case of gamma cameras one can think of administered dose, time between administration and start of scan, immobilization or ambulant status during that time, and distance to the camera. For planar images, the angle to the imaged part of the body, use of random noise correction, selective background noise reduction, geometric mean calculations, scan duration, total counts acquired, camera matrix used, and postprocessing optimization may play a role. Parameters that have an effect on dynamic scanning are time resolution, for SPECT and PET time per cm FOV, reconstruction settings, mode and degree of filtering, camera matrix used, and CT reconstruction settings for attenuation correction. Finally, for PET scanners, additional parameters are number of rings, use of Time-of-Flight (TOF), use of point spread function (PSF), and potentially dietary conditions (e.g. plasma glucose levels during an FDG scan). Even in more recent PET and SPECT literature, often not all of these parameters are accounted for in the methodology section.

It is known that major variations in these parameters may exist, resulting in differences in image appearance between hospitals despite using the same cameras. These differences can have significant clinical implications as shown by Mattoli et al [1]. Little is known, however, about the degree of variation and about the desired image quality between departments of nuclear medicine in the Netherlands. This leads to very subjective preferences that differ between departments. These preferences usually are based on the acquired mental reference of image quality as developed during training as a nuclear medicine physician. However, these mental references may become obsolete in time and they not necessarily have the optimal quality for detecting all clinically relevant information that is present in the acquired data. This is especially true for PET/CT with its rapid rate of development over the last 10-20 years with subsequent evolution of image quality.

Therefore, a more objective means to establish sufficient quality from acquired diagnostic images is required. Such an objective means to establish sufficient image quality would also enable optimization of the balance between administered dose and retaining sufficient diagnostic accuracy. Therefore, these standards of image quality will indirectly function as Diagnostic Reference Levels (DRL's) for dose optimization.

1.2 Project overview

This document must be seen as part one out of three stages in establishing a guideline for quantification of image quality. Several endpoints and necessary products will be established within this first part of the guideline, but subsequent projects will be needed to complete the guideline. Ultimately, this will provide a standardized means for:

1. <u>Evaluation</u>

   A way to measure routine clinical image quality for different reconstruction settings by evaluating corresponding phantom scans using a Model Observer per specific clinical imaging task. Model Observers mimic the human eye and brain in detecting lesions within diagnostic images. In this way, they provide an objective evaluation of image quality that relates directly to the clinical task of human observers, i.e. nuclear medicine physicians.

2. <u>Optimization</u>

A reference standard of mean image quality per specific clinical imaging task based on clinical images as produced by departments of nuclear medicine in the Netherlands, i.e. an Image Quality Diagnostic Reference Level (IQ-DRL).

Part I

In this first part the following is produced:

- A defined set of specific clinical imaging tasks that encompass all types of nuclear medicine scans. Every task is then linked to 1 exemplary nuclear medicine scan.
- A collection of exemplary nuclear medicine scans from a large contingent of all departments of nuclear medicine in the Netherlands with specific information on acquisition, (post-) processing and reconstruction resulting in an established 'mean' setting.
- A systematic literature review of different Model Observer techniques to select a type of Model Observer that is best suited for evaluating image quality of nuclear medicine scans.
- Evaluation of the applicability of standard commercially available phantoms to sufficiently discern the differences in image quality from different scans.

Part II

The a priori prediction was that commercially available standard phantoms would not suffice. Because the relevant sizes of detectable lesions have decreased over the last years beyond the smallest spheres within these standard phantoms (e.g. NEMA IQ phantom: 1 cm smallest diameter). An example is the detection of millimetre sized lymph nodes with tumour cells from prostate cancer that can be detected using [^{18}F]PSMA PET/CT. Also more subtle contrast differences have become clinically relevant, such as with [^{18}F]FDG PET/CT of the brain depicting subtle differences within grey matter as an expression of (pre-stage) dementia. This has been recognised in the literature and other groups have developed dedicated phantoms for specific tasks, such as Wollenweber [2] and Kadrmas [3].

Therefore, the development of a specific phantom was considered to be part two of the present project. However, to investigate whether this would really be necessary, tests were performed in part one to assess whether recovery efficiency of the more relevant smaller spheres could be derived by extrapolation of contrast-detail curves acquired using a standard phantom with relatively large spheres. Therefore a hypothesis was tested stating that extrapolation from measurements of larger spheres form a standard NEMA Image Quality phantom yields valid data for smaller spheres as well. The rationale for including these phantom measurements in part one was a strong desire to prevent a future guideline that would be dependent on a newly developed non-commercial phantom, which would come with the additional burden of central control and deployment. If successful, the second part of this guideline development may not be needed.

Part III

Part III of the task at hand would be to build different Model Observers for all different specific clinical imaging tasks, based on the type of Model Observer selected by the structured literature review described in this document. After verification of the performance of each Model Observer for each clinical imaging task, everything needed for deployment will be established.

2. Methodology

2.1 Set of specific clinical imaging tasks

Criteria for establishing different clinical imaging tasks were:

- Type of camera and scan type (PET, SPECT, planar; static, dynamic)
- Resolution based (form over signal) vs detection based (signal over form) clinical task
- Detection type (photopenic recess / high contrast lesion detection / low contrast lesion detection

This resulted in 18 different tasks.

Next, for each task a representative examination was established based on coherence with the defined clinical task and availability, i.e. high incidence in clinical practice in order to establish a sufficient large number of clinical institutions that could contribute. A relevant example could not be identified for all defined clinical tasks due to the absence of examples or the only the presence of examples with low incidence. Since there is no use for theoretical tasks without relevant clinical practice, these were ignored. The final list is shown in table 2.1.

			detectiontype	exam
Gammacamera	*Static*	**primarly resolution**	*fotopenic recess*	Renal scintigraphy (DMSA)
		primarly detection	*high contrast*	Sentinel node scintigraphy
			low contrast	3rd phase bonescintigraphy
	Dynamic	**primarly resolution**	*fotopenic recess*	1st and 2nd phase bonescintigraphy
		primarly detection	*high contrast*	Gastric emptying
			low contrast	MUGA
	SPECT static	**primarly resolution**	*fotopenic recess*	Renal scintigraphy (DMSA)
		primarly detection	*high contrast*	DATSCAN
			low contrast	3rd phase bonescintigraphy
	SPECT dynamic	**primarly resolution**	*fotopenic recess*	
		primarly detection	*high contrast*	ECG gated myocardial scintigraphy
			low contrast	
PET-camera	*Static*	**primarly resolution**	*fotopenic recess*	FDG brain
		primarly detection	*high contrast*	PSMA
			low contrast	Oncological FDG
	Dynamic	**primarly resolution**	*fotopenic recess*	
		primarly detection	*high contrast*	Rb82/N13H myocardial perfusion PET
			low contrast	

Table 2.1: Overview of defined clinical imaging tasks and representative examinations.

2.2 Collection of scans from hospitals

2.2.1 Data collection

In September 2022 we sent a survey to medical physicists of all hospitals with a nuclear medicine department asking either a DICOM header or an image (in DICOM format) for all examinations listed in table 2.1. The requested information consisted of details concerning acquisition, reconstruction and post-processing settings. If not available in the DICOM headers, a form was provided to complement the remaining parameters. After 2 months a reminder was sent and subsequently these clinical physicists were personally contacted to ask for the requested information. First, the focus was on obtaining PET data so that, in a first run, the mean value for an established PET parameter of image quality could be calculated. With the experience of this first run, a similar procedure was implemented for handling the larger range of gamma camera examinations. Eight months later the same clinical physicists were contacted again if they had not send data for these gamma camera examinations. If needed, subsequently they were also contacted personally 2 months later. The survey was open until September 2023.

Because the aim was to derive a mean value for image quality in clinical practice in the Netherlands, we established criteria for the minimum number of departments that would be needed to produce meaningful estimates. The total number of departments of nuclear medicine in the Netherlands is 65, of which 8 are academic centres, 23 large general hospitals and 34 small general hospitals. The distinction between large and small hospitals was based on the number of first ambulant visits within the hospital, being more or less than 120,000 visits per year, respectively. The total number of departments that needed to be included had to be at least 20, with a minimum of 4 university hospitals, 5 large general hospitals and 5 small hospitals to ensure that no bias was introduced towards either high complexity, low volume or low complexity, high volume departments.

The survey parameters that were requested for PET are shown in table 2.2.

Acquisition	Reconstruction	Processing
Isotope	Iterative reconstruction algorithm used (iterations and subsets)	Filter type and parameters
Administered dose as a function of weight		Slice thickness
Scan duration as a function of weight		
Interval between administration and start scan time		
Acquisition matrix size		
Acquisition mode		

Table 2.2: Overview of survey parameters for PET (static and dynamic).

The survey parameters that were requested for Gamma Camera and SPECT investigations are shown in table 2.3.

Acquisition	Processing

Isotope	Filter type and parameters
Administered dose (and variants if not fixed)	Slice thickness
Scan duration	
Interval between administration and start scan time	
Acquisition matrix size	
Acquisition mode	

Table 2.3: Overview of survey parameters for gamma camera (planar static, planar dynamic) and SPECT (static and dynamic).

If a department had installed a new scanner within 12 months prior to the survey, they were asked to also share protocols for the previous scanner. As a result, a hospital could contribute more datasets than the number of scanners they had in use at the time. Recently introduced PET scanners with long axial field-of-view were excluded, because of the very limited number of departments with such a scanner.

Further general specifications of different scanner types were obtained as described in table 2.4, including the underlying sources.

PET/CT model	NEMA sensitivity (kcps/MBq)	Overlap (%)	Bed position length (cm)	NEMA Spatial resolution (mm) Transverse/Axial	Reference
Siemens Biograph Vision 600	16.4	49 (SS mode)	26.1	3.6 / 3.5	Van Sluis et al. [4]
Siemens Biograph Vision 450	9.1	49 (SS mode)	19.7	3.6 / 3.5	Carlier et al. [5]
Siemens Biograph mCT 4R	9.6	46	19	4.1 / 4.4	Rausch et al. 2015 [6]; Jakoby et al 2021 [7]
Siemens Biograph mCT 3R	4.6		16.2	4.1 / 4.4	Eberl et al. [8]
Philips Gemini TOF	6.6	53	18	4.8 / 4.8	Surti et al. [9]
Philips Ingenuity	7.3	50	30	4.8 / 4.7	Kolthammer et al. [10]
Philips Vereos	5.2	39	16.4	4.2 / 4.2	Rausch et al. 2019 [11]

GE Discovery MI-4	12.6		20	4.2 / 4.5	Chicheportiche et al. [12]
GE Discovery MI-5	20.8	1-44	24.9	4.2 / 5.0	Pan et al. [13]
GE Discovery 710	7.5		15.7	4.7 / 4.7	Wagatsuma et al. [14]; Vandendriessche et al. [15]

Table 2.4: Overview of NEMA sensitivity and overlap for PET/CT scanners in the present survey.

2.2.2 Analysis for PET camera's

To analyse results of the survey, data for each model PET/CT scanner were grouped and weight dependent tracer administration as well as weight dependent scan duration were determined. Considering the various scanning modes that are available to the user, such as step and shoot (SS) and continuous motion acquisitions, speed of acquisition can vary as function of anatomical region. We choose to focus the analysis on the thoracoabdominal area, which in practice takes the longest.

To compare different scanners and SS versus continuous acquisition modes, a scan rate in the thoracoabdominal region was derived from each of the reported acquisition protocols, which subsequently was normalized to the length of a single bed position in the SS mode. Variations in detector sensitivities were taken into account by using reported values of the NEMA sensitivities of the PET/CT scanners. This resulted in the following equation for the acquisition count statistics n at the time of acquisition:

$$n(w) = A(w) \times \frac{T(w)}{l\,(1-o)} \times \sigma \ \text{[c/cm]} \qquad (1)$$

where w is the weight of the patient [kg], $A(w)$ the amount of radioactivity [MBq] at the start of the acquisition, which takes into account scan specific uptake times and radioactive decay corrections for the isotope used from the time of administration. $T(w)$ is the weight dependent scan duration of a single bed position [s], l is the length of a single bed position [cm], o is the overlap (%) between subsequent bed positions (in the case of step and shoot mode acquisition) and σ is the NEMA sensitivity of the scanner [cps/MBq]. For PEC/CT systems with continuous flow acquisition, there is no overlap, and the above equation reduces to $\frac{A(w)}{v} \times \sigma$, where v is the configured acquisition speed [cm/s].

2.2.3 Analysis for gamma camera and SPECT/CT imaging

To facilitate cross-scanner protocol comparison across all clinical applications outlined in Table 2.1, we computed several parameters for each hospital. These included the (i) effective activity in the patient during the scan, (ii) total duration of the scan in seconds, (iii) angular sampling of SPECT/CT in degrees, (iv) speed of longitudinal scanning in mm/s for total body gamma camera scans, (v) an estimate of total counts in the scan (in million counts) and an (vi) estimate of million counts acquired per SPECT projection.

Determining the effective patient activity in the middle of the scan entailed taking a reference dose for a patient of 70 kg (hospitals with weight specific dose scheme) or a fixed dose value (hospitals with a fixed dose scheme) in MBq. Subsequently, the value was decay corrected for

physical isotope decay, based on the average delay between the injection and the middle of the acquisition time of the scan. This is different form PET because of a significant longer period between injection of activity and scan acquisition for most gamma camera examinations compared with PET.

$$A_{eff} = A_0 e^{-\alpha t} (2)$$

α is the physical decay constant of the isotope, A_0 the initially injected activity in MBq, and t the time interval between injection and start of the scan, as provided in the survey, plus half the duration of the scan.

For SPECT/CT acquisitions, calculating the total scan duration (T) involved multiplying the total number of angular views (angular arch θ divided by the number of angular steps Δ) by the time per view in seconds t_v. This data was derived from specific DICOM fields: (0018,1143) for the scan arc in degrees, (0018,1144) for the angle step in degrees, and (0018,1242) for the actual frame duration in seconds. In case no DICOM data were provided, the same information was requested from the hospital on a one-by-one basis.

$$T = \frac{\theta}{\Delta} t_v \qquad (3)$$

where θ is an angular arch, Δ the angular step, and t_v the time per view in seconds.

To estimate the number of counts each scan would have resulted in, it was assumed that all activity present in the patient at the start of the scan would be in a point source placed in the centre of the FOV (e.g. centre-of-rotation for SPECT). This is an oversimplification of the real situation, but still provides the opportunity to compare different imaging protocols under similar conditions.

As a result the total number of counts n within SPECT/CT acquisitions was calculated as:

$$n = \mathrm{T} \times A_{eff} \times \sigma \qquad (4)$$

where σ is the sensitivity of a single detector in counts/s per MBq at 10 cm from a specific scanner-collimator combination for the isotope being imaged. This information was provided by the scanner manufacturers at the request of the project members and is listed in Table 2.5.

For planar gamma camera acquisitions, the same procedure to calculate the total number of counts within the scan was followed by replacing T with the total duration of the 2D planar scan t_{scan}:

$$n = t_{scan} \times A_{eff} \times \sigma \qquad (5)$$

In the case of whole-body gamma camera scans, with continuous bed motion over the length of the patient, the number of uptake counts was calculated similar to the acquisition count statistic described for PET/CT. However, because of differences in the length of cameraheads between cameras a correction for the limited view of the camerahead compared with the length of the distribution of activity in the patient was required:

$$A_{FOV} = \frac{A_{eff}}{L} \times l \qquad (6)$$

where L is the length of the patient (standard length of 170 cm assumed) and l is the length of the camerahead. In addition, length of acquisition was calculated as:

$$T = \frac{L}{v} \qquad (7)$$

where v is the speed of the bed transfer between camera heads. The total counts n of a whole body scan was then calculated as:

$$n = A_{FOV} \times T \times \sigma = \frac{A_{eff}}{L} \times l \times \frac{L}{v} \times \sigma \qquad (8)$$

This can be simplified to:

$$n = \frac{A_{eff}}{v} \times l \times \sigma \qquad (9)$$

To have a similar parameter as for PET/CT this needs to be divided by the scan length (patient length) L to end with cnts/cm:

$$\frac{A_{eff}}{v} \times \frac{l}{L} \times \sigma \qquad (10)$$

Further general specifications of different scanner types were obtained as described in table 2.5. Since for gamma camera specifications, the actual isotope need to be addressed as well, it was, in contrast to PET/CT, not possible to obtain all relevant sensitivity data from public available literature data. Therefore, these data were obtained from the camera manufacturers.

Manufacturer	Scanner model	Collimator	Isotope	Sensitivity [cps/MBq] at 10 cm
Siemens	Intevo family Intevo Bold family	LEHR	Tc-99m	91
		LPHR	I-123	113
		LEAP	Tc-99m	148
		LEUHR	Tc-99m	45
		LEFB	Tc-99m	126
		MELP	Ga-67	124
		HE	I-131	61
GE	NM 670 Pro series	LEHR	Tc-99m	72
		ELEGP	Tc-99m	144
		ELEGP	I-123	101
		MEGP	Ga-67	65
		HEGP	I-131	43
	NM 870 CZT Series	WEHR	Tc-99m I-123	85 117
		MEHRS	In-111	113

Table 2.5 Isotope and collimator specific gamma camera sensitivity in cps/MBq at 10 cm from the detector.

2.3 Systematic literature review for Selection of a Model Observer

The description of the literature review, including the applied methods, is given in section 3.2.

2.4 Evaluation of applicability of standard phantoms

The hypothesis we postulated was that using a standard NEMA phantom, with spheres of 1 cm diameter and larger, it will be possible to predict detectability of more clinically relevant smaller spheres down to 4 mm (conform the resolution limit of state-of-the-art PET cameras). To this end contrast-detail curves were generated using scans of such a standard NEMA IQ phantom containing spheres with inner diameters of 10, 13, 17, 22, 28 and 37 mm. These spheres and the background region were filled with activity such that the sphere-to-background ratio was 16:1. The measured contrast-detail curves of these spheres were plotted and extrapolated down to smaller spheres of 4 mm.

Next scans were acquired using a non-standard NEMA Micro-PET IQ Phantom consisting of a NEMA IQ phantom with small diameter sphere inserts to provide actual measurements of small size spheres. These smaller spheres had inner diameters of 4, 5, 6.2 and 7.9 mm. Again these were filled with activity using a sphere-to-background ratio of 16:1. For these small diameter spheres the contrast-detail curves were generated as reference for the extrapolated contrast-detail curves of the standard NEMA IQ phantom.

In comparing results, the assumption was that an extrapolated contrast-detail value within 20% of the reference value was acceptable for developing working Model Observers to evaluate image quality. If the difference was more than 20% the use of a standard NEMA IQ phantom is likely to be insufficient to measure image quality for modern PET cameras.

3.1 Collection of scans per clinical imaging task

In total, 42 independent hospitals have responded to the PET part of the survey and 31 to the SPECT part. The information provided was a filled-in Excel questionnaire containing information about the average administered dose and the time between injection and scan and, as well as one DICOM file per study type, in principle for a "typical" patient (in the questionnaire instructions, typical was described as a patient with a weight of around 70 kg and a length of 170 cm). The level of questionnaire completeness and availability varied per hospital.

3.1.1 PET

Data for the following standard clinical tasks and corresponding representative examinations were solicited:

PET-camera	*Static*	**primarly resolution**	*fotopenic recess*	FDG brain
		primarly detection	*high contrast*	PSMA
			low contrast	Oncological FDG
	Dynamic	**primarly resolution**	*fotopenic recess*	
		primarly detection	*high contrast*	Rb82/N13H myocardial perfusion PET
			low contrast	

The distribution of hospitals for which we received data is shown in table 3.1 below.

PET study	Hospital size			Total	DRL threshold met
	small	medium	academic		
FDG brain	10	11	3	24	No/Yes
PSMA	17	17	5	39	Yes
FDG whole body	18	19	5	42	Yes
$^{82}Rb/^{13}NH$	1	3	1	5	No

Table 3.1: Overview of PET/CT camera questionnaires, collected within the project.

Of these only for $^{82}Rb/^{13}NH_3$ myocardial perfusion PET, insufficient data were obtained (total of 5 hospitals). For FDG whole body PET and PSMA PET sufficient data were obtained. For FDG brain PET, strictly speaking insufficient data were obtained, because data for an insufficient number of academic centres were obtained. However, there was a 100% response rate for all academic centres, so it is impossible to obtain data for more than 3 academic centres. Therefore, analyses for FDG whole body PET, PSMA PET and FDG brain PET are described below.

3.1.1.1 FDG Whole body [^{18}F]FDG

PET study	Hospital size			Total
	small	medium	academic	

FDG whole body	18	19	5	42

Of the 64 hospitals that were invited to participate, 7 did not have a PET/CT system. From the resulting 57 hospitals with a PET/CT system, 42 responded and (partially) shared data (survey response rate 74%). Among the participating hospitals were all University Medical Centres in the Netherlands. Data were received for a total number of 58 different PET/CT scanners from 4 different manufactures with a total of 10 different scanner models. Data from large axial field of view (total body) and PET/MR scanners were excluded from further analysis, because of the limited availability of these scanners in the Netherlands. A summary of the systems included is given in the table 3.2.

Model PET/CT	FDG whole body		
	Dose	Duration	Recon
Siemens Biograph mCT 3R	4	4	4
Siemens Biograph mCT 4R	16	16	14
Siemens Biograph Horizon	1	1	1
Siemens Biograph Vision450	4	4	4
Siemens Biograph vision600	13	13	11
Philips Vereos	7	7	7
Philips Gemini TOF	4	4	1
Philips Ingenuity	3	3	1
GE DiscoveryMI4	2	2	0
GE DiscoveryMI5	3	3	3
GE Discovery710	1	1	0
Total	58	58	46

Table 3.2: Overview of survey response for FDG whole body specified by examination type and model PET/CT scanner (in 2022).

Administered dose

The dose administered to the patient varied significantly between hospitals and examination types. Figure 3.1 gives an overview of the weight dependent administered dose for FDG whole body PET examinations.

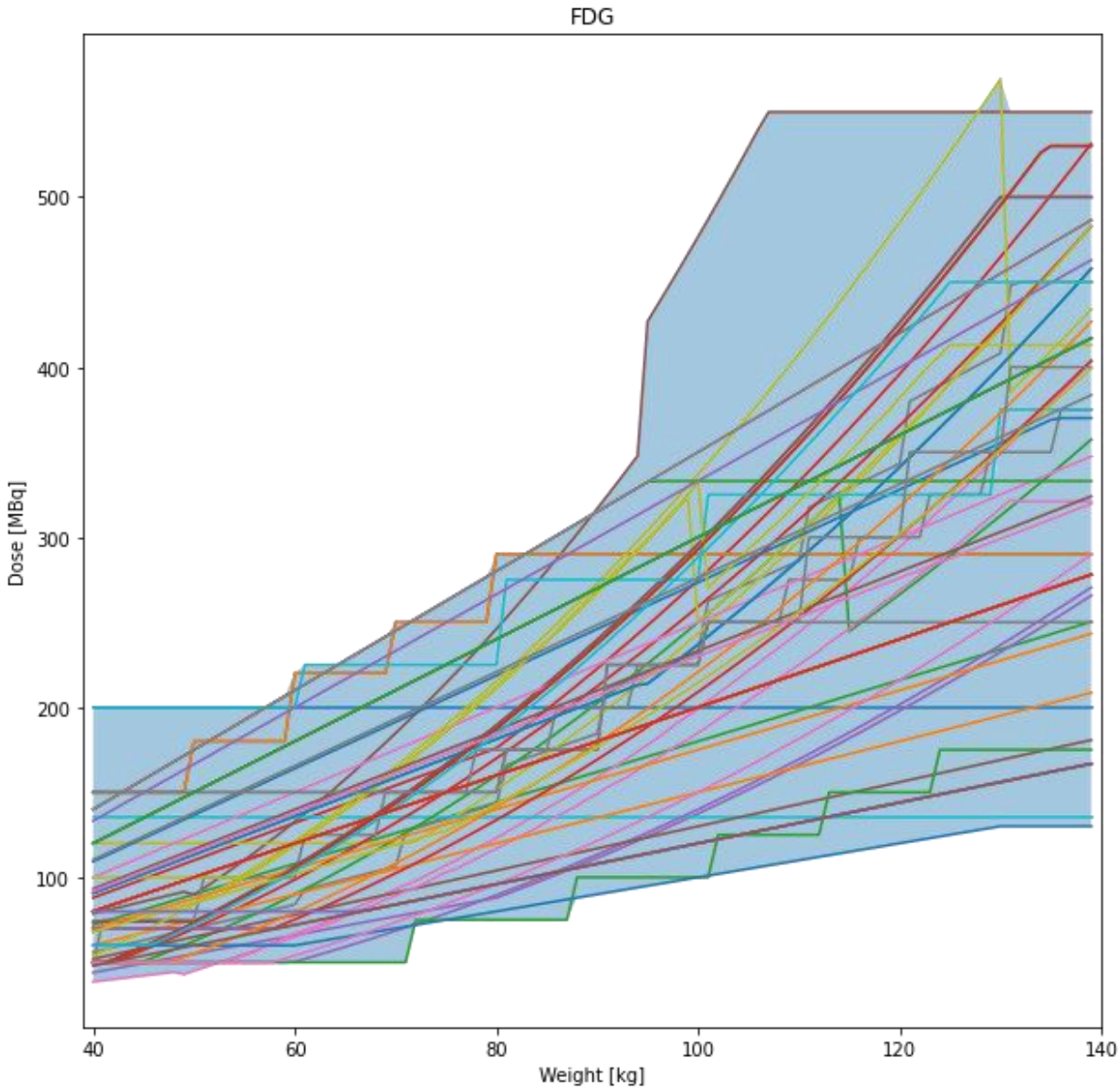

Figure 3.1: Administered dose to the patient as a function of bodyweight for FDG whole body PET examinations in the Netherlands. Each individual line corresponds to a specific PET/CT system.

In general, as expected, patients with lower body weights received lower doses than patients with higher body weights. For FDG whole body examinations there was quite some variation in type of administration protocol used, including linear, quadratic/power-law based, BMI based and hybrid combinations. The average administered dose for patients of 50, 75, and 120 kg are listed in table 3.3.

Body weight (kg)	Dose (MBq) Mean (min-max)
50	102.7 (45-200)
75	162.1 (75-263)
120	308.2 (144-550)

Table 3.3: Mean dose administered to patients for FDG whole body PET scans.

Uptake times

For FDG whole body, all departments reported a one hour physiological uptake period, in accordance with international guidelines.

Radiation dose

The effective dose coefficient for [^{18}F]FDG is equal to 0.019 mSv/MBq, resulting in an estimated mean effective FDG whole body dose of 3.1 (1.4-5.0) mSv for patients with a body weight of 75 kg.

Acquisition counts

There are various potential reasons for the differences between administered patient doses. For example, image quality might be considered as insufficient or total scan times as too long in relation to high demand and waiting lists. In addition, there are differences in effective sensitivity between different PET/CT models and, as a result, different doses are needed to achieve similar image quality. Finally, as pixel/voxel size decreases with technologically more advanced PET/CT systems, it might be necessary to increase the dose or scan time to have the same SNR per voxel compared with older systems with larger pixel/voxel size.

Figure 2 shows acquisition counts at the start of FDG whole body scans for different models. These acquisition count statistics were based on the activity present in the patient, not on actual counts detected by the scanner. A higher weight is related to more attenuation within the patient, which is the reason why higher count statistics are needed in such a patient to achieve similar actual detected count statistics by the camera.

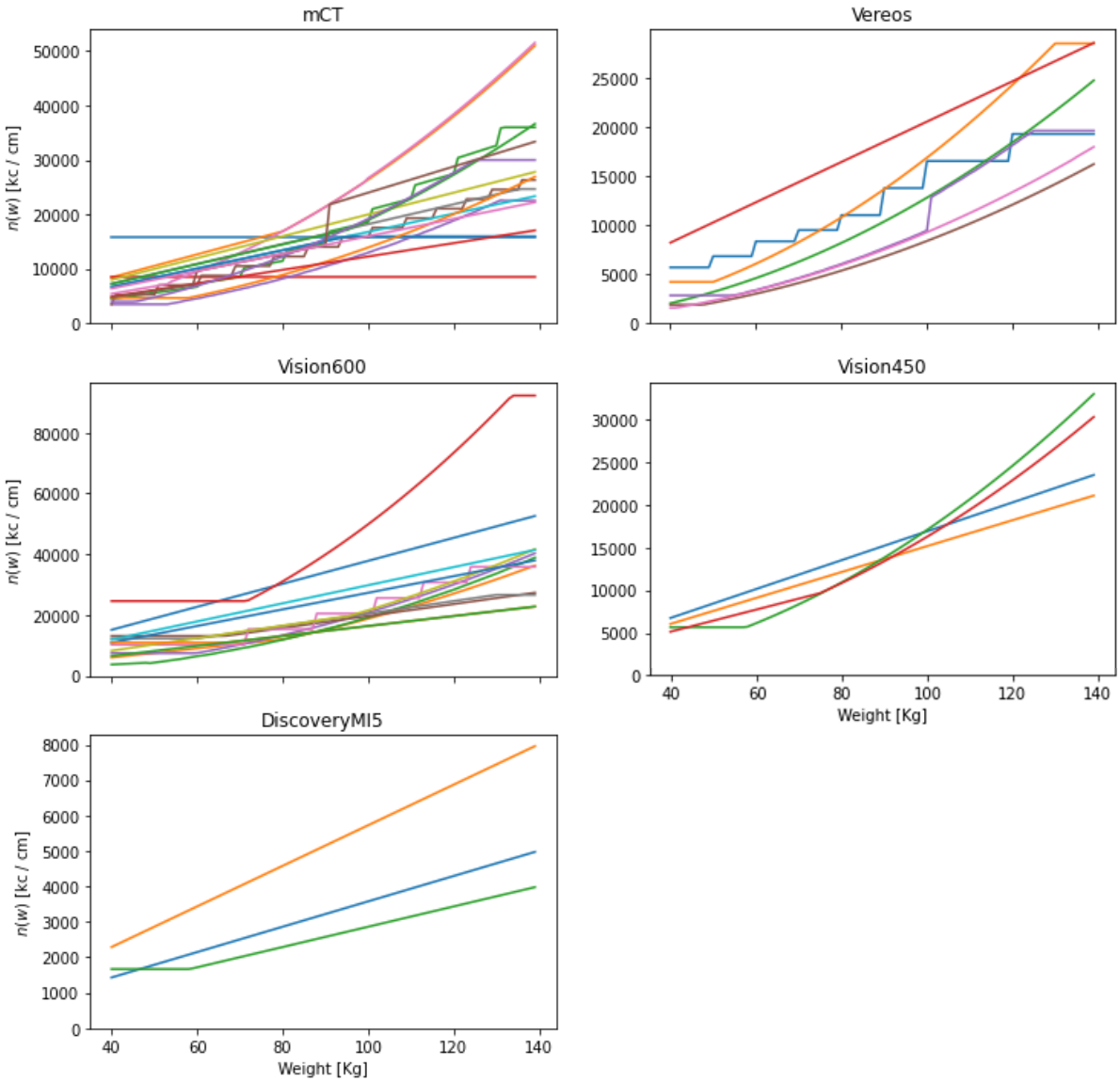

Figure 3.2: FDG whole body counts in a patient as function of bodyweight for different PET/CT models at the start of an acquisition.

As can be seen from Figures 3.1 and 3.2, the majority of FDG whole body scans follow a nonlinear dose regimen. Interestingly, there is also a large variation between the lowest and highest administered doses as a function of bodyweight.

3.1.1.2 Whole Body [^{18}F18]PSMA

PET study	Hospital size			Total
	small	medium	academic	
PSMA	17	17	5	39

Of the 64 hospitals that were invited to participate, 7 did not have a PET/CT system. From the resulting 57 hospitals with a PET/CT system, 39 responded and (partially) shared data (survey response rate 68%). Among the participating hospitals were all University Medical Centres in

the Netherlands. Data were received for a total number of 58 different PET/CT scanners from 4 different manufactures with a total of 10 different scanner models. Data from large axial field of view (total body) and PET/MR scanners were excluded from further analysis, because of the limited availability of these scanners in the Netherlands. A summary of the systems included is given in the table 3.4.

Model PET/CT	PSMA		
	Dose	Duration	Recon
Siemens Biograph mCT 3R	2	2	0
Siemens Biograph mCT 4R	13	12	11
Siemens Biograph Horizon	1	1	1
Siemens Biograph Vision450	3	3	1
Siemens Biograph vision600	12	12	5
Philips Vereos	5	5	4
Philips Gemini TOF	2	0	0
Philips Ingenuity	1	1	0
GE DiscoveryMI4	2	2	0
GE DiscoveryMI5	3	3	0
GE Discovery710	1	1	0
Total	45	42	22

Table 3.4: Overview of survey response for PSMA specified by model PET/CT scanner.

Administered dose

The PSMA dose administered to the patient varied significantly between hospitals. Figure 3.3 gives an overview of the weight dependent administered dose for PSMA examinations. It should be noted that for the calculation of the count statistics n for the [^{68}Ga]PSMA examinations we also used the NEMA sensitivity for F-18 as listed in table 2.4 as an approximation. We are aware that there is a slight difference [16,17], but for the purpose of this comparison nonsignificant.

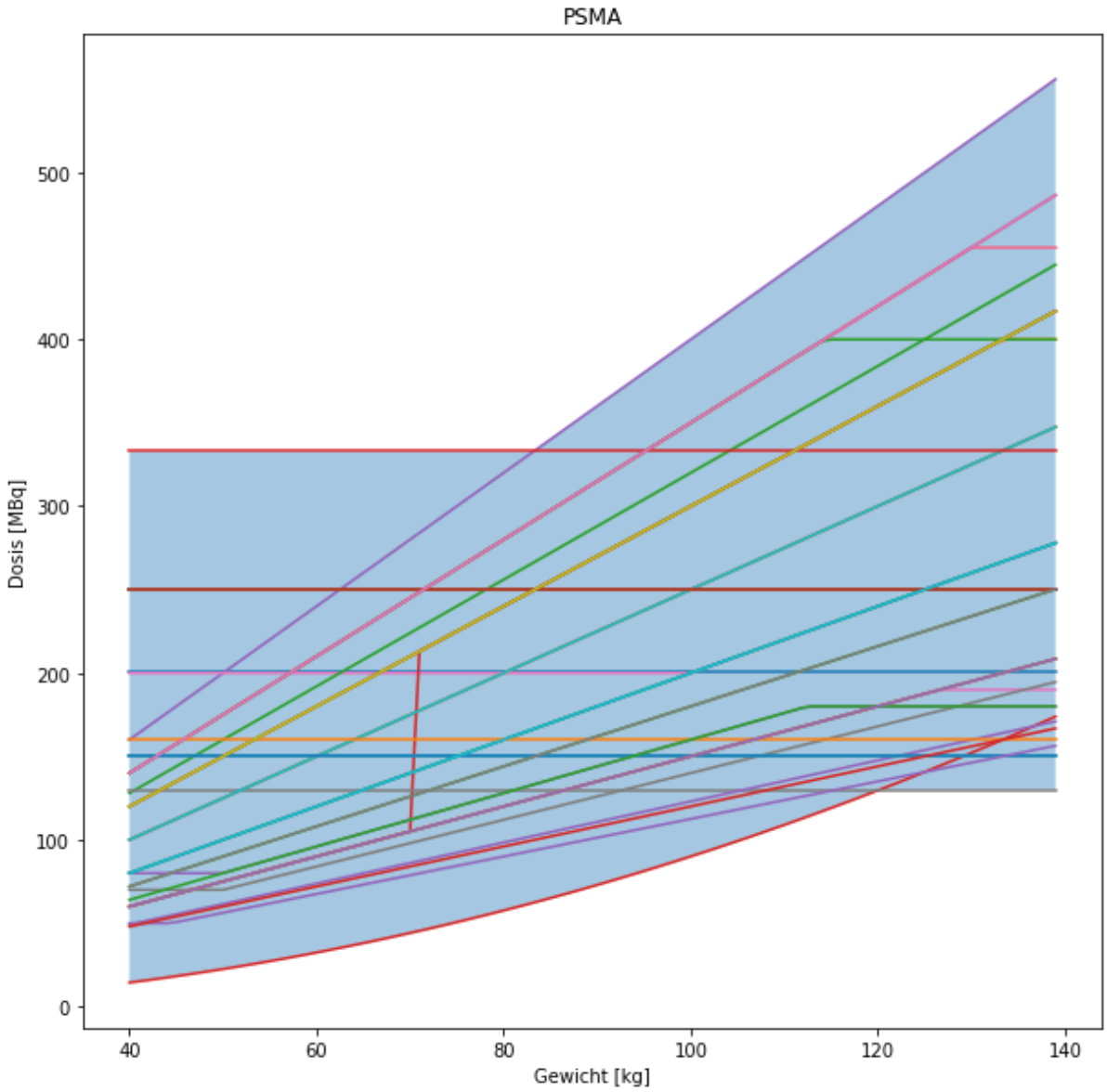

Figure 3.3: Administered dose to the patient as a function of bodyweight for PSMA PET examinations in the Netherlands. Each individual line corresponds to a specific PET/CT system.

The average administered doses for patients of 50, 75 and 120 kg are listed in table 3.5.

Body weight (kg)	F18-PSMA (MBq) Mean (min-max)	Ga68-PSMA (MBq) Mean (min-max)
50	151 .7 (80-250)	91.4 (23-201)
75	200.6 (120-300)	129.1 (51-225)
120	286.9 (130-480)	191.0 (130-360)

Table 3.5: Mean dose administered to the patient for the different PSMA tracers.

Uptake times

For PSMA PET/CT scans, significant differences in physiological uptake times between different departments were seen (Figure 3.4).

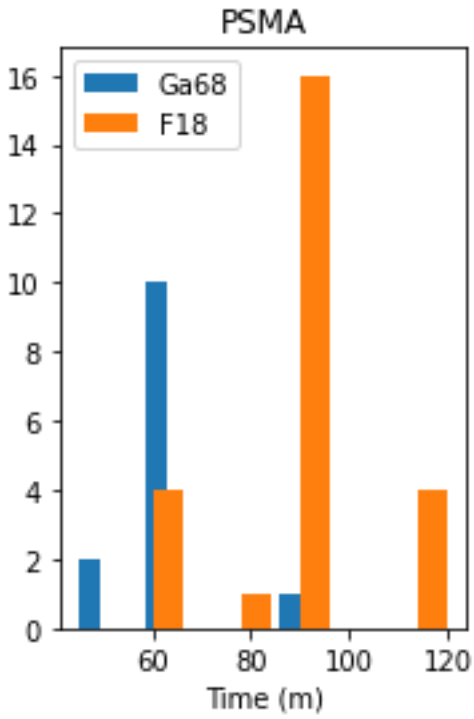

Figure 3.4: Uptake times for PSMA tracers.

There is a clear difference between Ga-68 and F-18 based tracers. Due to the shorter half-life of ^{68}Ga (68 minutes), hospitals using this tracer choose for a shorter interval between administration and scan resulting in a median physiological uptake time of 60 minutes, as compared with 90 minutes for ^{18}F.

Radiation dose

There was a wide variety in PSMA tracers used, including, [^{68}Ga]/[^{18}F]PSMA-11, [^{68}Ga]/[^{18}F]JDK-PSMA-7, [^{68}Ga]/[^{18}F]1007, and [^{18}F]DCFPYL. The dose administered to the patient is summarized in Table 3.5 for the two radionuclides used. The effective dose coefficient for the PSMA tracers varies between 0.01-0.02 mSv/MBq [18-20], resulting in a mean effective dose of 4.0 (2.4-6.0) and 2.6 (1.0-4.5) mSv for [^{18}F]PSMA and [^{68}Ga]PSMA, respectively.

Acquisition counts

There are various potential reasons for the differences between administered patient doses. For example, image quality might be considered as insufficient or total scan times as too long in relation to high demand and waiting lists. In addition, there are differences in effective sensitivity between different PET/CT models and, as a result, different doses are needed to achieve similar image quality. Finally, as pixel/voxel size decreases with technologically more advanced PET/CT systems, it might be necessary to increase the dose or scan time to have the same SNR per voxel compared with older systems with larger pixel/voxel size.

Figure 3.5 shows acquisition counts at the start of PSMA scans for different models. These acquisition count statistics were based on the activity present in the patient, not on actual counts detected by the scanner. A higher weight is related to more attenuation within the patient, which is reason why higher count statistics are needed in the patient to achieve similar actual detected count statistics by the camera.

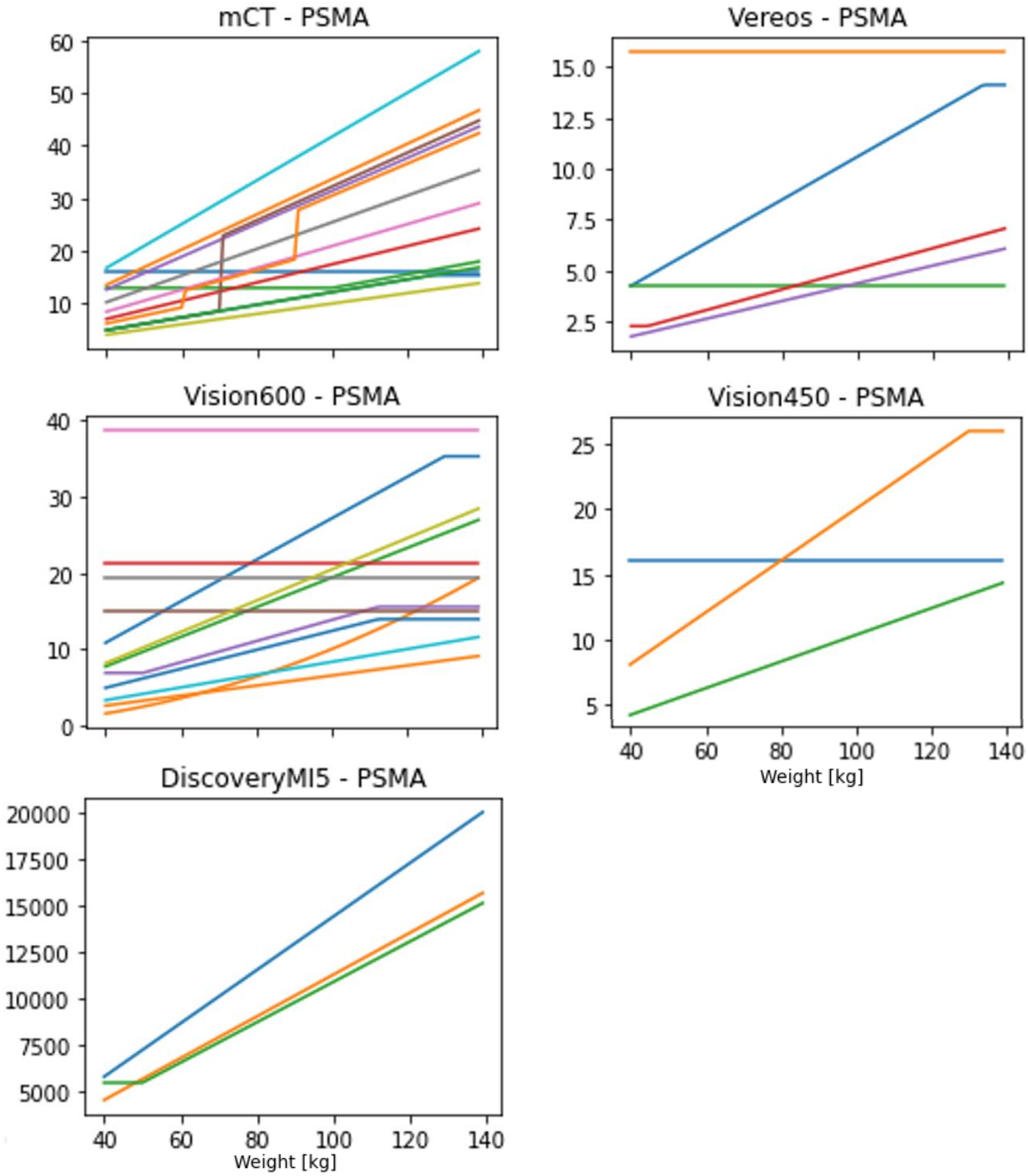

Figure 3.5: PSMA acquisition available counts in a patient as a function of body weight for different PET/CT models at the start of acquisition.

As can be seen from Figures 3.3 and 3.5, PSMA scans primarily are performed using linear dose protocols. In general, it is clear that PSMA examinations are performed with fewer acquisition counts than FDG whole body examinations for all scanner types, which can be explained by the significantly longer uptake period required for PSMA and possibly different physiological clearance and distribution compared to FDG whole body.
No correction for the differences in positron yield between F-18 and Ga-68 has been applied in the graphs, as this difference is negligible relative to the overall variability of the data.

3.1.1.3 Brain [^{18}F]FDG

PET study	Hospital size			Total
	small	medium	academic	
FDG brain	10	11	3	24

Of the 64 hospitals that were invited to participate, 7 did not have a PET/CT system. From the resulting 57 hospitals with a PET/CT system, 24 responded and (partially) shared data (survey response rate 42%). Among the participating hospitals were all University Medical Centres in the Netherlands. Data were received for a total number of 58 different PET/CT scanners from 4 different manufactures with a total of 10 different scanner models. Data from large axial field of view (total body) and PET/MR scanners were excluded from further analysis, because of the limited availability of these scanners in the Netherlands. A summary of the systems included is given in the table 3.6.

Model PET/CT	FDG brain		
	Dose	Duration	Recon
Siemens Biograph mCT 3R	2	2	0
Siemens Biograph mCT 4R	11	8	4
Siemens Biograph vision600	6	5	2
Philips Vereos	1	1	0
Philips Ingenuity	1	0	0
GE DiscoveryMI5	1	1	0
GE Discovery710	1	1	0
Total	23	16	6

Table 3.6: Overview of survey response for FDG brain specified by examination type and model PET/CT scanner.

Administered dose

The dose administered to the patient varied significantly between hospitals and examination types. Figure 3.6 gives an overview of the administered dose for FDG brain examinations. What can be seen is that both fixed doses and weight dependent doses are used in practice. The use of fixed doses, next to weight dependent doses as for FDG whole body, is explained by a Dutch Guideline where a fixed dose is advised.

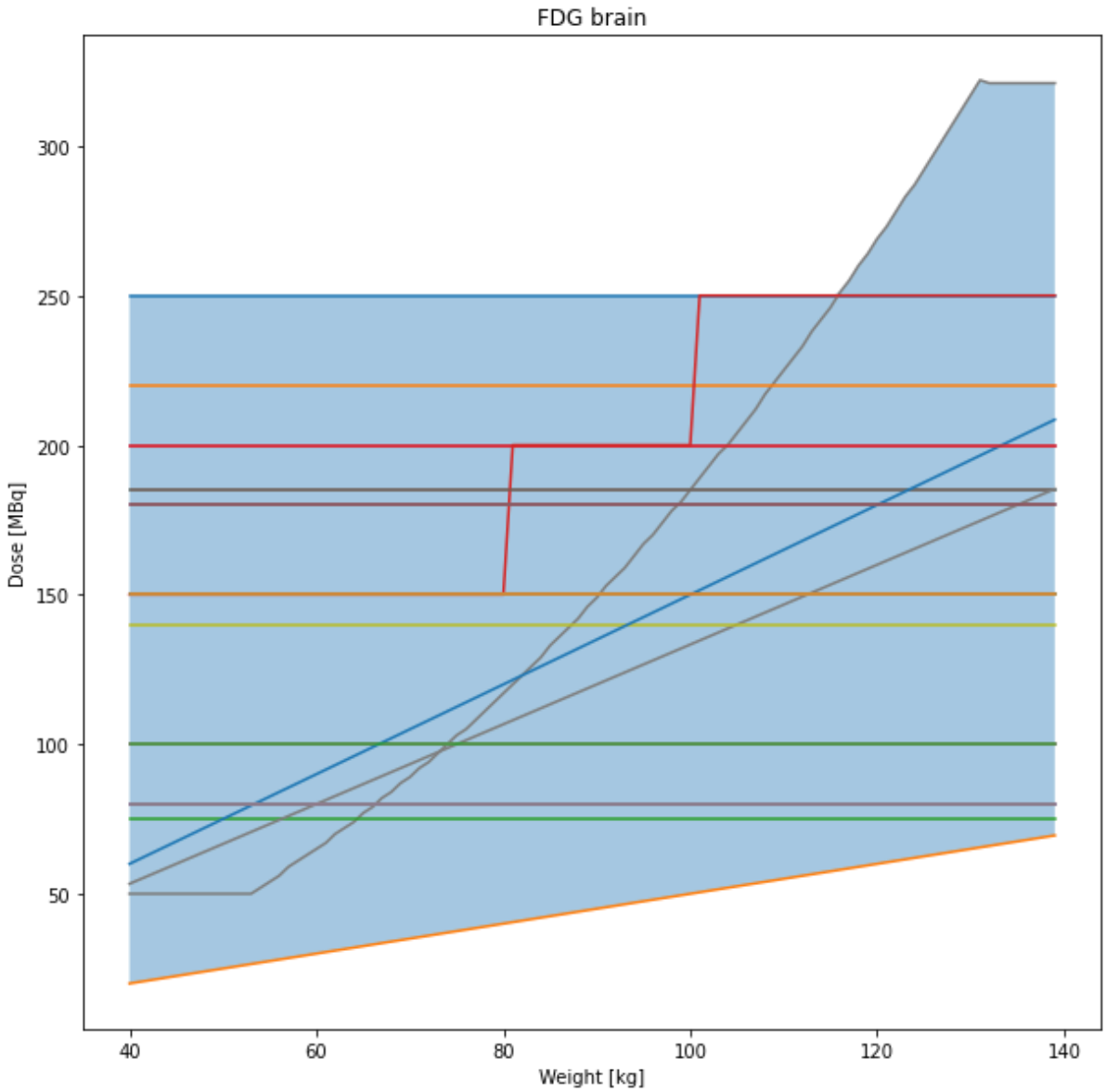

Figure 3.6: Administered dose to the patient as a function of bodyweight for FDG Brain examinations in the Netherlands. Each individual line corresponds to a specific PET/CT system.

In general, as expected, patients with lower bodyweights received lower doses than patients with higher body weights, for the sites with a weight dependent dose regimen. The average administered dose for patients of 50, 75, and 120 kg are listed in table 3.7.

	F18-Brain (MBq) Mean (min-max)
50 kg	139.8 (25-250)
75 kg	145.6 (37.5-250)
120 kg	160.2 (60-269)

Table 3.7: Mean dose administered to the patient for FDG brain scans.

Radiation dose

The effective dose coefficient for ^{18}F-FDG brain scans was 2.8 (0.7-4.8) mSv. Remarkable here is the 7 times difference between the lowest and highest dose.

Uptake times

For FDG brain scans, significant differences in physiological uptake times between different departments were seen (Figure 3.7).

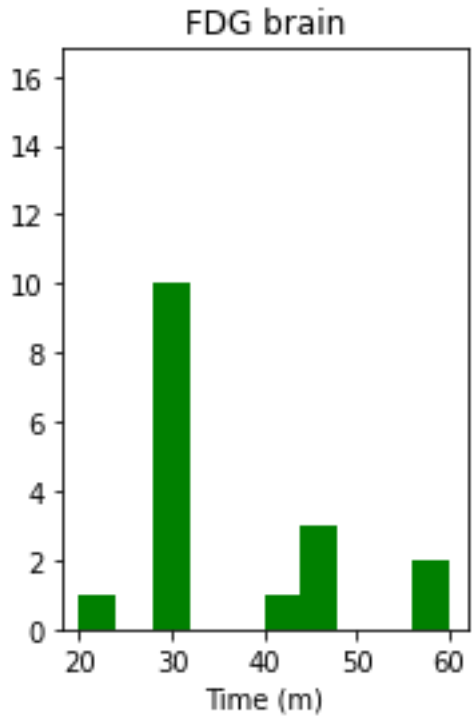

Figure 3.7: Differences in uptake times for FDG Brain.

Acquisition counts

There are various potential reasons for the differences between administered patient doses. For example, image quality might be considered as insufficient or total scan times as too long in relation to high demand and waiting lists. In addition, there are differences in effective sensitivity between different PET/CT models and, as a result, different doses are needed to achieve similar image quality. Finally, as pixel/voxel size decreases with technologically more advanced PET/CT systems, it might be necessary to increase the dose or scan time to have the same SNR per voxel compared with older systems with larger pixel/voxel size. Nevertheless, differences as large as a 7-fold increase between minimum and maximum administered dose cannot be explained by these mentioned factors.

Also, contrary to FDG whole body, with FDG brain a large variety is seen for physiological uptake time up to a 3-fold increase between minimum and maximum period.

Figure 3.8 shows acquisition counts at the start of FDG brain scans for different models. These acquisition count statistics were based on the activity present in the patient, not on actual counts detected by the scanner.

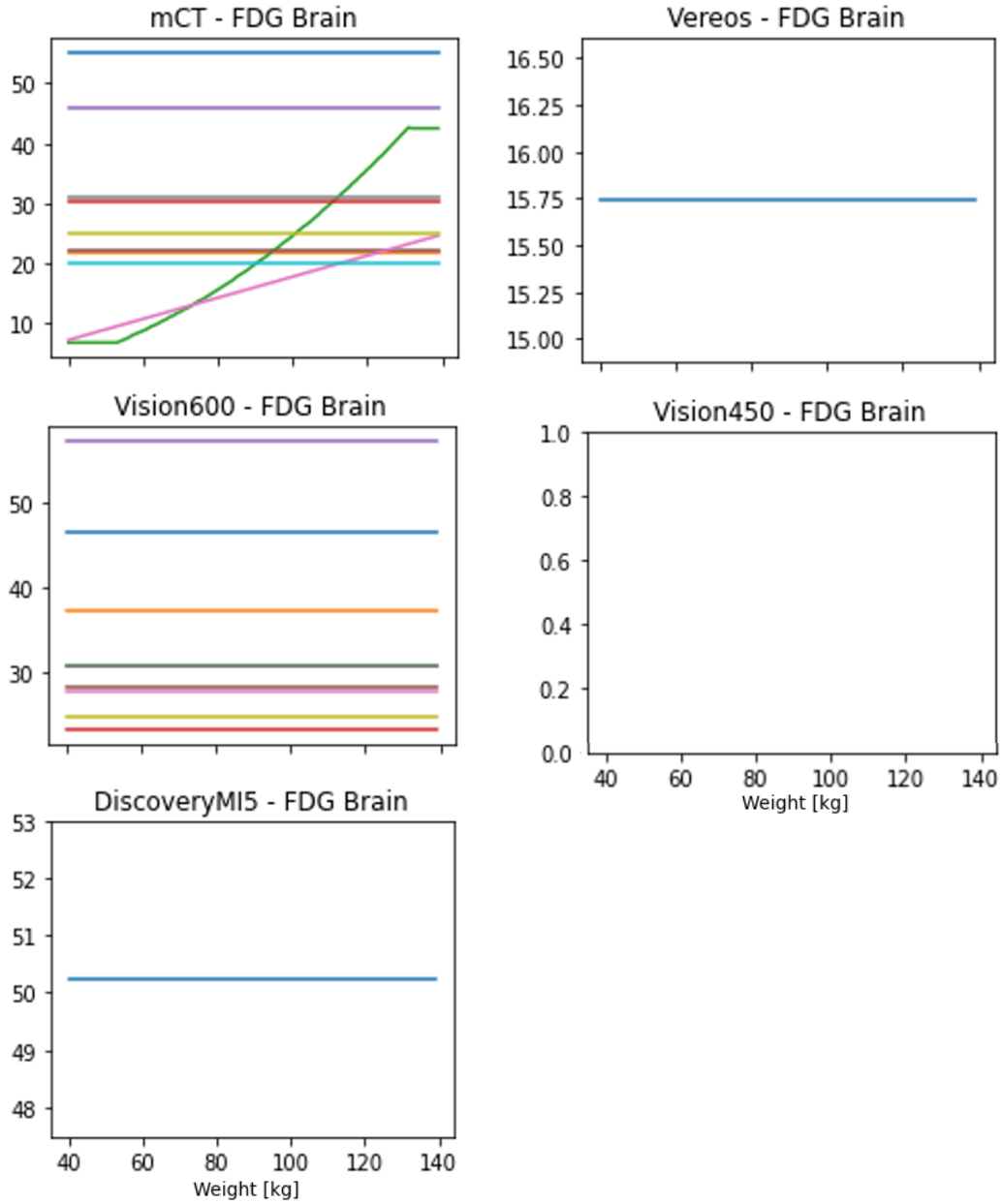

Figure 3.8: FDG-brain acquisition available counts in a patient as function of bodyweight for different PET/CT models at the start of the acquisition.

As can be seen from Figures 3.6 and 3.8 FDG brain scans often use weight independent doses.

3.1.1.4 PET Reconstruction and post-processing

For the whole body FDG protocol, the survey also included processing and reconstruction parameters used for each PET/CT scanner. Table 3.8 summarizes the unique combinations of reconstruction and processing settings for those PET/CT models for which at least 5 responses were received.

Biograph mCT					Biograph Vision600					Vereos				
Recon		Gaussian (mm)	Slice Thickness (mm)	Pixel size (mm)	Recon		Gaussian (mm)	Slice Thickness (mm)	Pixel size (mm)	Recon		Gaussian (mm)	Slice Thickness (mm)	Pixel size (mm)
PSF+TOF	2i21s	5.0	4.0	3.1	PSF+TOF	3i5s	3.00	2.0	1.6	PSF+TOF	i3s15	4.0	2.0	2.0
PSF+TOF	3i21s	5.0	3.0	2.0	PSF+TOF	3i5s	4.00	3.0	1.6	PSF+TOF	i3s6	2.0	2.0	2.0
PSF+TOF	3i21s	3.0	2.0	2.0	PSF+TOF	3i5s	2.50	2.0	1.6	OSEM+TOF	i3s11	3.0	4.0	4.0
PSF+TOF	2i21s	2.5	5.0	4.0	PSF+TOF	4i5s	2.00	3.0	1.6	OSEM+TOF	i3s13	0.0	2.0	2.0
PSF+TOF	2i21s	4.5	3.0	2.0	PSF+TOF	4i5s	4.50	3.0	1.6	OSEM+TOF	i3s5	2.5	2.0	2.0
PSF+TOF	3i21s	2.0	3.0	4.0	PSF+TOF	3i5s	0.00	2.0	1.6	OSEM+TOF	i3s9	3.0	2.0	2.0
PSF+TOF	3i21s	5.0	2.0	1.4	PSF+TOF	4i5s	4.00	3.0	1.6					
PSF+TOF	3i21s	5.0	5.0	4.0	PSF+TOF	3i5s	2.00	3.0	1.6					
PSF+TOF	3i21s	5.0	2.0	2.0	PSF+TOF	4i5s	3.50	3.0	1.6					
PSF+TOF	4i21s	5.5	3.0	4.0	PSF+TOF	4i5s	4.00	2.0	1.6					
PSF+TOF	2i21s	5.0	5.0	4.0										
PSF+TOF	2i21s	5.0	3.0	2.0										

Table 3.8: Unique combinations of reconstruction and processing settings for those PET/CT models with at least 5 responses.

Most departments make use of point spread function (PSF) and time-of-flight (TOF) based reconstruction algorithms. In addition, some degree of Gaussian filtering is applied to the reconstructed images.

Slice thickness

The slice thickness of the reconstructed images varies between different vendors, but also for a given model as shown in figure 3.9.

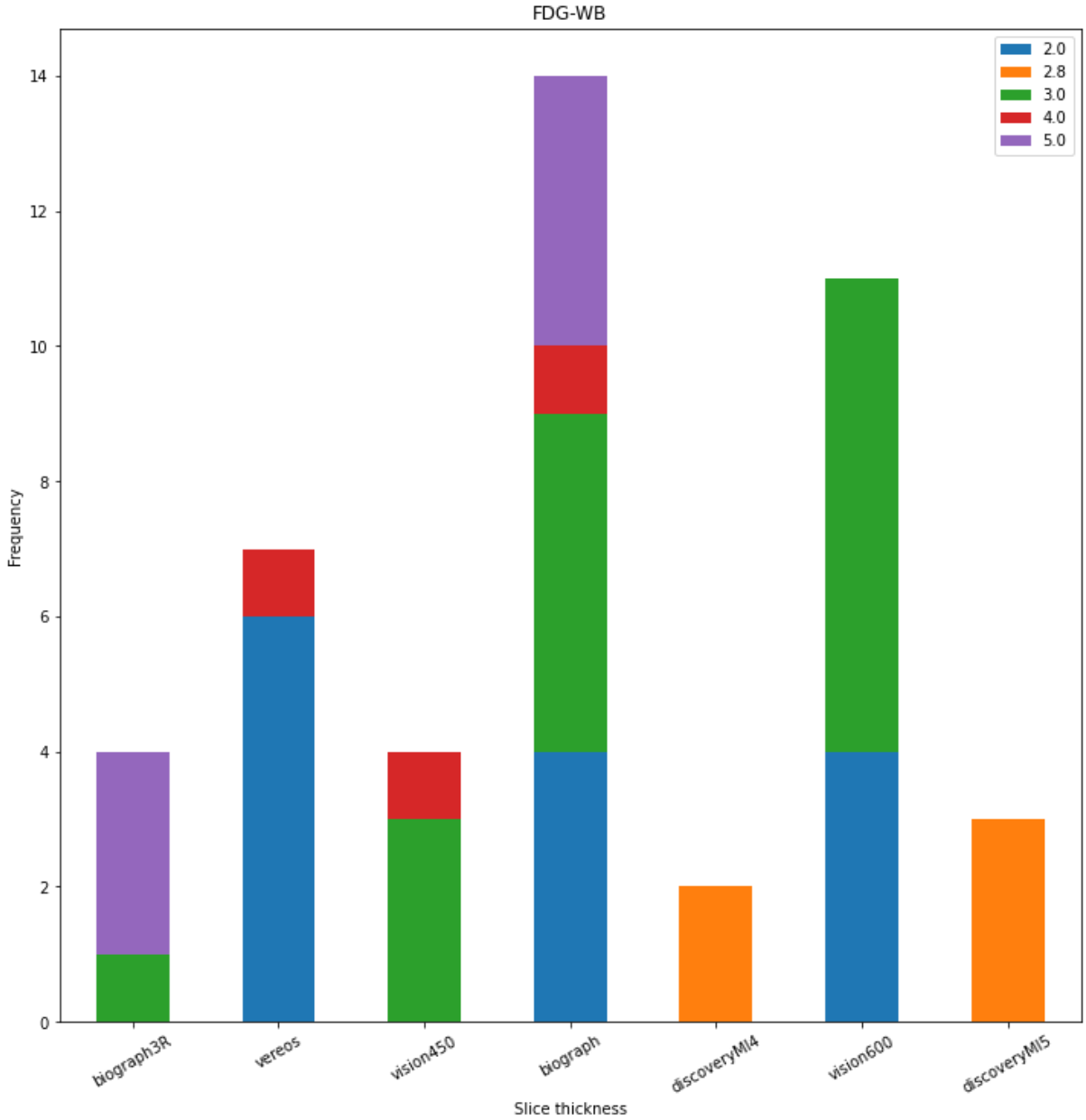

Figure 3.9: Slice thickness for different PET/CT systems sorted by NEMA sensitivity from left to right in ascending order.

From figure 9 we see there is no clear correlation between the slice thickness used and the NEMA sensitivity of the PET/CT system.

Effective resolution

In principle we can use the information from table 3.8 to determine the effective resolution used by the hospitals. Most data on the intrinsic resolution of PET/CT systems has been determined using the NEMA NU-2 2012 standard (as shown in table 2.4). Although this establishes a means for comparing systems, these measured spatial resolutions may not be the same as those used for routine clinical images, as (very) different iterative reconstruction, time of flight (TOF), point spread function (PSF) and filtering settings may be used.
The effective resolution of the produced images can be calculated by

$$FWHM_{effective}^2 = FWHM_{intrinsic}^2 + FWHM_{filter}^2 \qquad (11)$$

where $FWHM_{effective}$ is the effective full width at half maximum (FWHM) spatial resolution of a point source, $FWHM_{intrinsic}$ the FWHM in the absence of filtering and $FWHM_{filter}$ the FWHM of the (Gaussian) filter used. Equation (11) describes how filtering (blurring) will lead to degradation of the effective resolution of the reconstructed images. The positive effect of filtering is, however, to decrease noise in the images.
For Biograph mCT systems the variation in iterations and subsets settings was limited to 2 and 3 iterations with 21 subsets. For these 2 settings literature data can be used to compare spatial resolution of the resulting clinical images based on variations in applied filtering, choice of pixel size and slice thickness. Rogash et al. established the spatial resolution for reconstructions with 2 iterations and 21 subsets, TOF, PSF and a Gaussian filter of 2.0 mm [21]. From their results and after correcting for filtering, an intrinsic resolution of 3.5 mm for both axial and transverse radial directions can be derived. Similarly, Marti-Climent et al. established the spatial resolution for reconstructions with 3 iterations and 21 subsets, TOF, PSF and a Gaussian filter of 2.0 mm [22], finding an intrinsic resolution, after correcting for filtering, of 2.9 mm in the axial direction and 2.3 mm in the transverse radial direction.
Using these values as a reference for intrinsic resolution in axial (z) and transverse radial (x,y) direction, an effective resolution, $FWHM_{effective}$, for each reconstruction setting in clinical use can be determined. In addition, the degree of degrading the resolution by filtering can be calculated, as shown in table 3.9.

Biograph mCT **Recon – filter size - slice thickness - pixel size** (filter size (mm) in bold)	FWHMeff (z)	FWHMeff (x,y)	Δ (%) z	Δ (%) x,y
PSF+TOF 3i21s **5.0** 3.0 2.0	5.8	5.5	-99.3	-139.3
PSF+TOF 2i21s **4.5** 3.0 2.0	5.7	5.7	-62.9	-62.9
PSF+TOF 2i21s **5.0** 5.0 4.0	6.1	6.1	-74.4	-74.4
PSF+TOF 2i21s **5.0** 4.0 3.1	6.1	6.1	-74.4	-74.4
PSF+TOF 3i21s **5.0** 5.0 4.0	5.8	5.5	-99.3	-139.3
PSF+TOF 3i21s **3.0** 2.0 2.0	4.2	3.8	-43.9	-64.4
PSF+TOF 2i21s **5.0** 3.0 2.0	6.1	6.1	-74.4	-74.4
PSF+TOF 3i21s **2.0** 3.0 4.0	3.5	3.0	-21.5	-32.5

Table 3.9: Effective resolution for the biograph mCT systems as used in clinical studies and the deviation Δ (%) from the intrinsic resolution.

For almost every clinically used reconstruction protocol in this survey, the applied Gaussian filter leads to a more than 50% degradation in spatial resolution compared with the intrinsic resolution of the camera. As mentioned above, this is not necessarily a bad thing, given the reduction of noise in the final images. The large spread in effective resolution (and thus in blurring) between hospitals is, however, remarkable. This analysis was not possible for the other PET/CT scanners, but given the higher intrinsic resolution of many of those systems and the fact that similar filtering choices were used as for the Biograph mCT, it is likely that the spread in effective resolution will be similar for those other PET/CT models.
To identify the impact of the chosen pixel size and slice thickness on degradation of x, y and z resolution, the Nyquist-Shannon Sampling theorem can be used to identify potential undersampling issues. In contrast to filtering, there is no real benefit in undersampling. In fact, evidence suggests that in clinical practice, lesion detectability increases with smaller pixel sizes [23]. From the sampling theorem it follows that to represent the entire resolution space, for resolution r, a pixel size p is required that is smaller than r/2, where r is the effective resolution of the clinically used reconstructions. Table 3.10 shows the required minimal pixel size to

satisfy the Sampling theorem for each of the protocols used. In table 3.10 a negative percentage represents undersampling.

Biograph mCT (resp. slice thickness and pixel size (mm) in bold)	Nyquist - Shannon z	Nyquist- Shannon x,y	Δ to required sampling (%) z	Δ to required sampling (%) *x,y*
PSF+TOF 3i21s 5.0 **3.0 2.0**	2.8	2.7	-7.1	25.9
PSF+TOF 2i21s 4.5 **3.0 2.0**	2.8	2.8	-7.1	28.6
PSF+TOF 2i21s 5.0 **5.0 4.0**	3.0	3.0	-66.7	-33.3
PSF+TOF 2i21s 5.0 **4.0 3.1**	3.0	3.0	-33.3	-3.3
PSF+TOF 3i21s 5.0 **5.0 4.0**	2.8	2.7	-78.6	-48.1
PSF+TOF 3i21s 3.0 **2.0 2.0**	2.0	1.8	0.0	-11.1
PSF+TOF 2i21s 5.0 **3.0 2.0**	3.0	3.0	0.0	33.3
PSF+TOF 3i21s 2.0 **3.0 4.0**	1.7	1.5	-76.5	-166.7

Table 3.10: Nyquist-Shannon sampling theorem applied to the various protocols used on Biograph mCT systems.

Because the lowest effective resolution found from table 3.9 is smaller than 7 mm, a slice thickness or pixel size ≥ 4.0 mm will always lead to sampling issues. Even for the case where $FWHM_{effective}(z)$ is equal to 6.1 mm (PSF+TOF 2i21s 5.0 4.0 3.1), a pixel size of 3.1 mm does not satisfy the Sampling theorem.
Of course, most PET studies focus on lesion detection, which prefers high Signal to Noise Ratio (SNR), Contrast to Noise Ratio (CNR), Recovery Coefficient (RC) or Noise Equivalent Count-Rate (NECR). Nevertheless, increasing the pixel size has a degrading effect on CNR, as shown by Øen et al. [24]. In addition, the variation in number of iterations (2 and 3) had no significant effect on CNR. For small degrees of filtering, there is a negative impact on CNR but, as shown by Øen et al., this effect is limited for iteration settings of 2 and 3. Therefore, in many cases the values used in this study for slice thickness and pixel size lead to unnecessary degradation of image quality.

3.1.2 SPECT and gamma camera

Data for the following standard clinical tasks and corresponding representative examinations were solicited:

			detectiontype	exam
Gammacamera	*Static*	**primarly resolution**	*fotopenic recess*	Renal scintigraphy (DMSA)
		primarly detection	*high contrast*	Sentinel node scintigraphy
			low contrast	3rd phase bonescintigraphy
	Dynamic	**primarly resolution**	*fotopenic recess*	1st and 2nd phase bonescintigraphy
		primarly detection	*high contrast*	Gastric emptying
			low contrast	MUGA
	SPECT static	**primarly resolution**	*fotopenic recess*	Renal scintigraphy (DMSA)
		primarly detection	*high contrast*	DATSCAN
			low contrast	3rd phase bonescintigraphy
	SPECT dynamic	**primarly resolution**	*fotopenic recess*	
		primarly detection	*high contrast*	ECG gated myocardial scintigraphy
			low contrast	

Not all SPECT/CT and gamma camera studies of interest met the DRL criteria on minimal number of responses.

For SPECT/CT and gamma camera data were received as filled-in excel questionnaires, DICOM-header files and/or DICOM images. In addition, data were anonymized, which introduced differences in available parameters based on different underlying PACS-systems. Therefore, we added in table 3.11 a column “Complete” to show for how many hospitals the required parameter set was indeed complete.
The administered activity for SPECT and gamma camera examinations generally adheres to a fixed dosing regimen based on predetermined weight categories, rather than using a weight-based dosing relationship as seen with FDG-PET imaging.

a

Gamma camera study	Hospital size			Total	Complete	DRN threshold met
	small	medium	academic			
DMSA	12	8	7	28	23	yes
Sentinel node	8	7	7	23	21	yes
Bone	12	8	6	27	23	yes
Stomach	11	8	4	24	19	no
MUGA	10	5	7	23	21	yes

b

SPECT study type	Hospital size			Total	Complete	DRN threshold met
	small	medium	academic			
DMSA	5	4	6	16	16	no
DATSCAN	12	8	7	28	26	yes
Bone	12	8	7	28	22	yes
Cardio	12	8	7	28	22	yes

Table 3.11. Overview of gamma camera (a) and SPECT/CT (b) questionnaires received.

3.1.2.1 SPECT

Table 3.11 illustrates that among the four SPECT/CT examination types surveyed, only three garnered a sufficient response. Consequently, detailed analysis focused on bone scan, DAT-scan, and cardio SPECT/CT. Table 3.12 provides a comprehensive overview of the quantitative aspects for these three examination types.

Parameter	Bone SPECT/CT	DAT-Scan	Cardio SPECT
Scan duration [s]	1331±514 [945, 1575]*	3765±389 [3600, 3840]	1357±515 [873, 1600]
Angular sampling [degree]	5.6±0.7 [5.6, 6]	2.9±0.1 [2.8, 3.0]	3.8±1.4 [2.8,5.6]
Effective activity [MBq]*	257±44 [228, 284]	118±9 [109,126]	534±100 [443, 644]
Time per projection [s]	20±8 [15,21]	31±2.9 [30, 30]	26±9.3 [20, 25]
Million counts/projection	0.45±0.17 [0.33,0.54]	0.41±0.04 [0.38, 0.43]	1.2±0.6 [0.89, 1.28]
Total million counts/scan	29.5±10.5 [20.6, 36.7]	51±4.6 [46.6, 54.3]	63±25.2 [43.9, 80]

*mean ± standard deviation [Q1, Q3]

Table 3.12. Overview of quantitative SPECT/CT acquisition characteristics.

3.1.2.1 .1 Bone SPECT

An overview of quantitative acquisition parameters for SPECT/CT bone scans is shown in Figure 3.10.

Analysis of the data revealed that the effective activity in patients at the middle of the scan time remained consistent across different hospitals, typically averaging around 257±44 MBq. As for angular sampling, a crucial parameter affecting the clarity of fine details within reconstructed images, a considerable variation in settings was observed, ranging from 3.75 to 6 degrees or 30 to 48 projections per detector. A previous study has reported on the effects of varying angular sampling for various levels of counts per projection [25]. It was shown that, between 60 and 120 projections, streaking artefacts could be introduced with fewer projections. However, since 1987, reconstruction algorithms have been developed further even up to for instance deep-learning techniques to recover full-view projections from sparsely sampled projections [26]. The absolute impact of variations in angular sampling on image quality therefore is unclear, provided that overall scan statistics remain comparable, although relative differences in quality will favour higher angular sampling.

Furthermore, the analysis revealed a wide range in time per projection used by hospitals, spanning from 12 to 40 seconds. Given the consistent effective activity levels observed in patients across different hospitals, one might anticipate an inverse relationship between the number of projections and the time per projection. Surprisingly, however, no correlation between these parameters (R=-0.08) was found, suggesting that users do not adjust these settings in tandem. In addition, a very weak correlation was observed between number of projections used and total scan duration (R=0.26). Surprisingly, the scan duration was determined by the time per projection setting selected by users (R=0.97).

When looking at the last acquisition parameter described for bone SPECT/CT, namely total scan statistics, it can be noted that a wide range of count statistics are used in the Netherlands. As anticipated, based on the relationships outlined above, total scan count statistics exhibited

a correlation solely with time per projection (R=0.86). Neither effective activity in the patient nor number of projections used appeared to affect total scan count statistics.

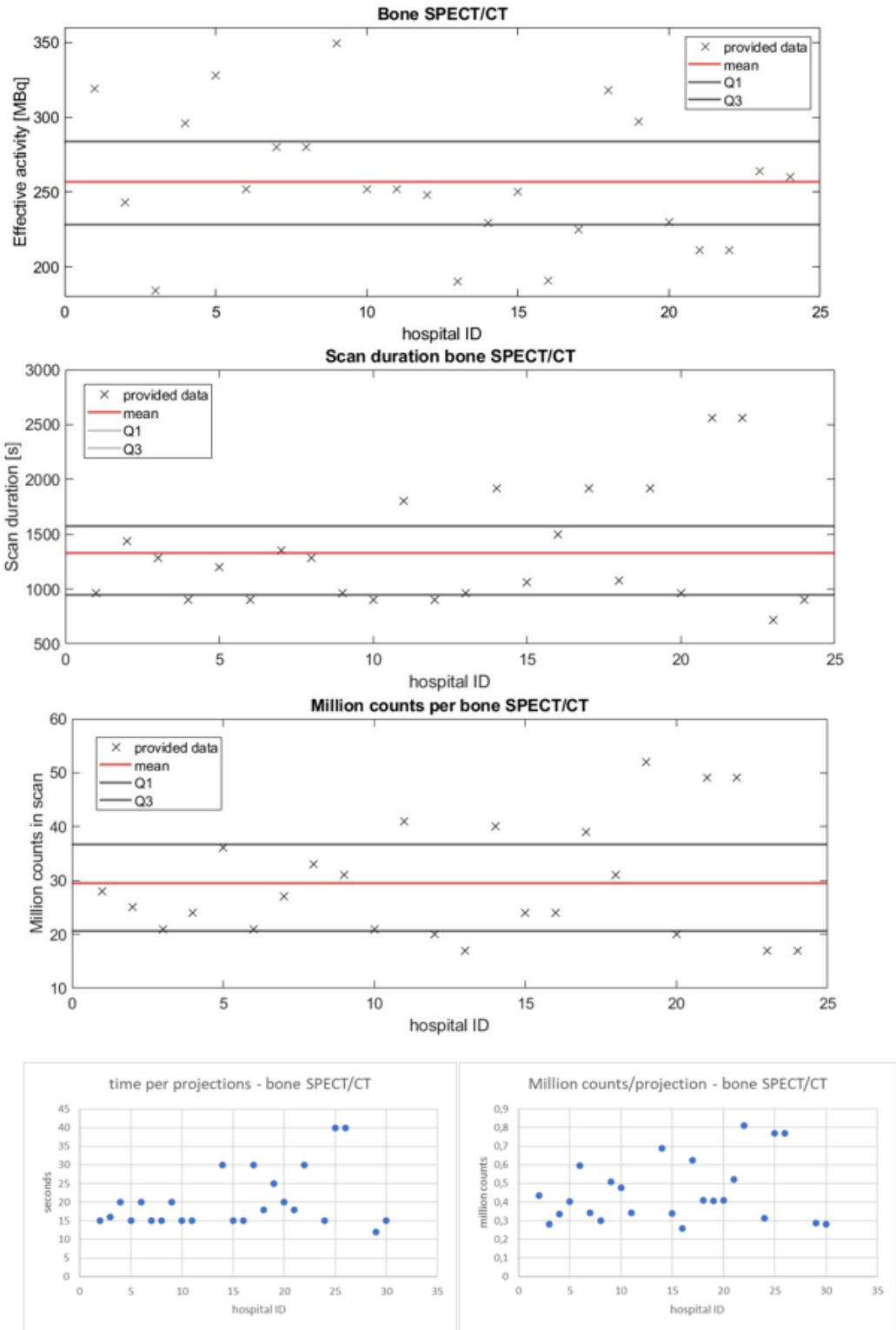

Figure 3.10. Overview of acquisition parameters for bone SPECT/CT.

3.1.2.1.2 Myocardial perfusion scintigraphy SPECT

Figure 3.11 provides an overview of quantitative acquisition parameters for SPECT/CT cardiac scans.

Upon analysing the data, significant variability across hospitals was found in the effective activity within patients, midway through the scan, averaging approximately 534 ± 100 MBq. Regarding angular sampling, similar to bone SPECT/CT scans, settings varied, spanning from 2.8 to 6.1 degrees or 16 to 34 projections per detector (32 to 68 in total). However, variation was almost dichotomized between angular sampling of around 3 degrees and 6 degrees. The majority of hospitals maintained a fine angular sampling of around 3 degrees. Similarly, while minimum and maximum time per projection exhibited wide ranges (14 to 60 seconds), a substantial number of hospitals consistently employed approximately 20 seconds per projection, including all, with one exception, around 6 degrees angular sampling. Since there is no outspoken higher effective activity for these hospitals one would expect that by choosing a lower angular sampling the time per projection would have been increased to compensate for a resulting lower total counts per scan. This, however, was not the case. The consistency of time per projection is reflected in the narrow interquartile range of cardiac SPECT/CT scans (Table 3.12, 20-25 s).

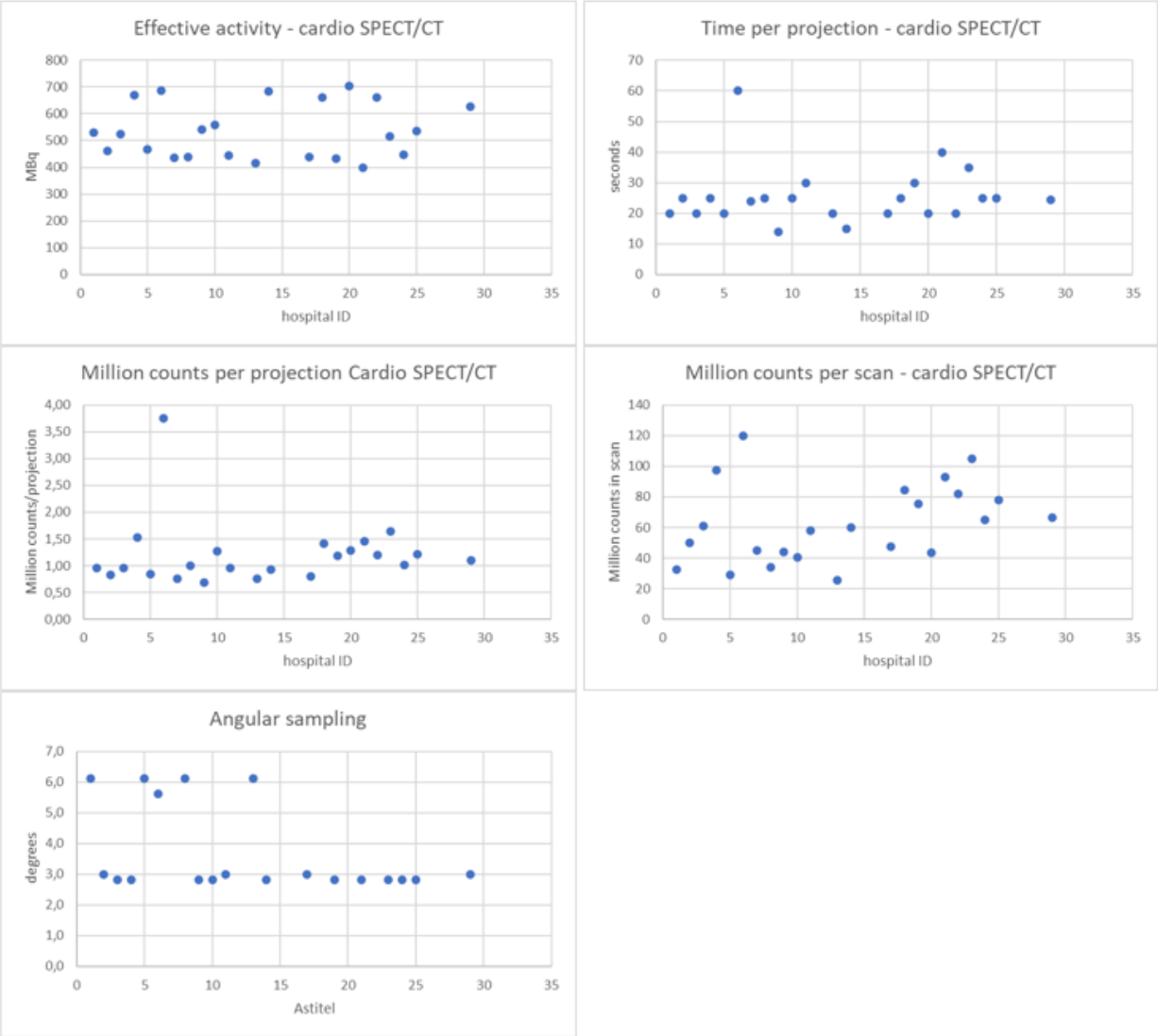

Figure 3.11. Overview of acquisition parameters for cardio SPECT/CT.

In addition a wide range in total scan counts (ranging from 25.7 to 120 million counts per scan) across participating hospitals was observed. Despite this variation, total scan counts did not

exhibit a clear predominant correlation with effective activity in the patient (R=0.38), number of projections used (R=0.44), or time per projection (R=0.7), although it showed a weak to moderate dependency on all these factors.
However, when examining the total counts per projection, a consistent trend was noticed. Despite significant variations in patient activity, time per projection and angular sampling, this value maintained a narrow interquartile range (0.89-1.28 million counts/projection). Delving deeper into the parameters influencing count statistics per projection, it appeared that this metric was primarily determined by the time per projection (R=0.88), and that it exhibited a weak correlation with patient activity (R=0.44) and a very weak inverse relationship with the angular sampling used (R=-0.24). Overall, these findings suggest that hospitals aimed for a consistent quality of projection images, while the total number of projections and patient activity levels remained largely dependent on location. To standardize image quality, adjustments were predominantly made to the duration of projection images.

3.1.2.1.2 DAT-scan SPECT

An overview of quantitative acquisition parameters for SPECT/CT DAT-scans is shown in Figure 3.12, while actual parameter values are listed in Table 3.12.
With patients consistently administered uniform levels of activity during scans (interquartile range: 109-126 MBq), nearly identical angular sampling (interquartile range: 2.8-3.0 degrees), and a consistent number of projections per detector (interquartile range: 60-64 projections), the overall scan duration was predominantly determined by the time per projection (R = 0.95). While variations in patient activity (R = -0.38) and number of projections (R = 0.41) did play a role, their impact was comparatively weaker. Consequently, both total scan counts and counts per projection were most significantly affected by time per projection (R = 0.66 and 0.65, respectively).
It is likely that this noticeable uniformity could be the result of strict criteria set by the radiopharmaceutical provider for DAT-scans (General Electric) to enable the use of established quantitative parameters that can be derived from the DAT-scan. Unfortunately, the questionnaire did question this issue and, consequently, no definite explanation can be provided.

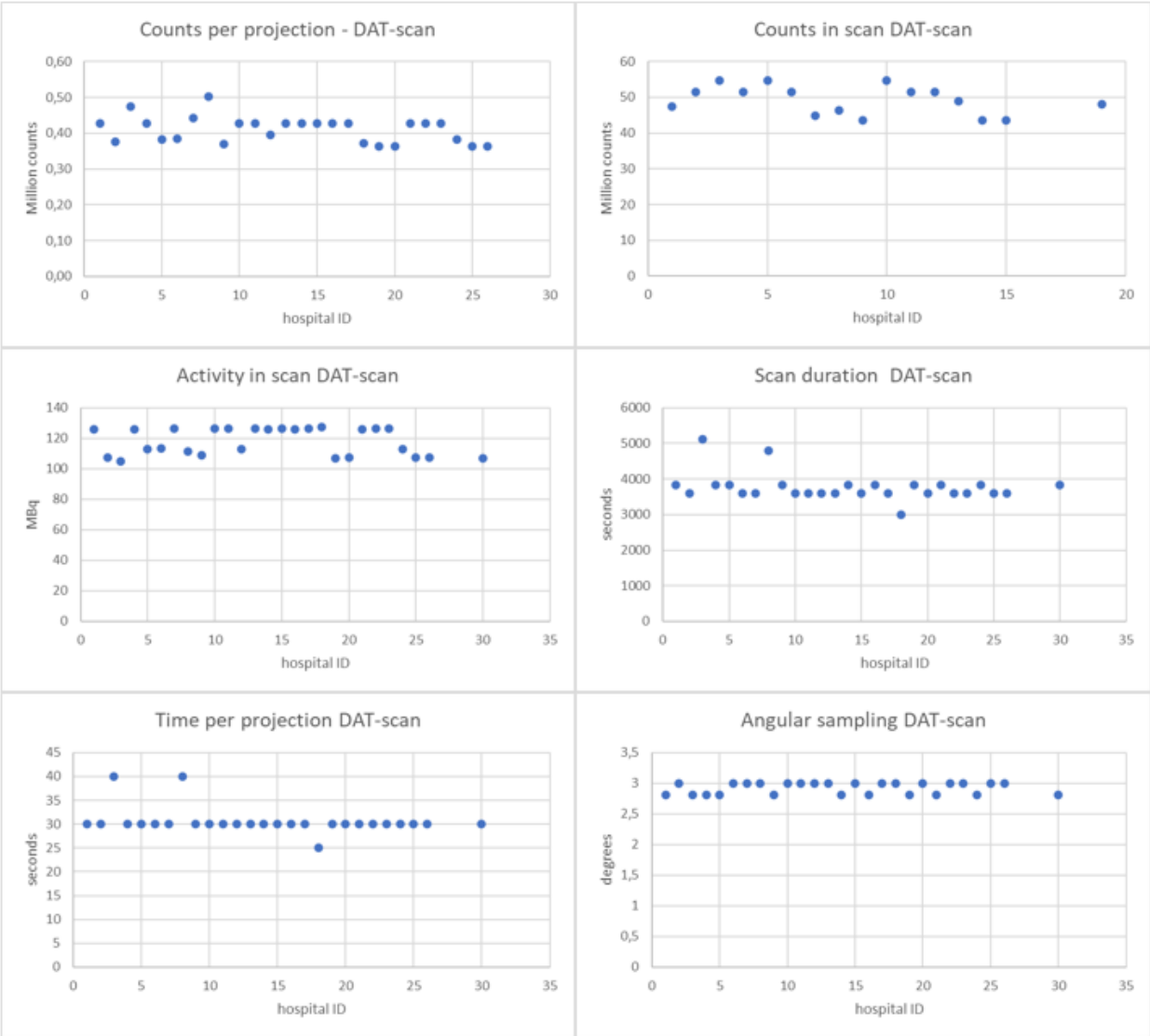

Figure 3.12. Overview of acquisition parameters for DAT-scan SPECT/CT.

3.1.2.2 Gamma camera

Table 3.11 illustrates that among the five gamma camera examination types surveyed, only three garnered a sufficient response or consistency of shared information. Consequently, the detailed analysis focuses on bone scan, DMSA, and stomach gamma camera imaging. Table 3.13 offers a comprehensive overview of the quantitative parameters for these three examination types.

Parameter	**Bone**	**DMSA**	**Stomach**
Scan duration [s]	821±196 [700, 875]*#	389±147 [298,600]*	77±59 [60, 60]*
Motion speed [mm/s]	2.2± 0.5 [2.0, 2.5]	n/a	n/a
Effective activity [MBq]*	272±47 [255, 294]	71±17 [58, 75]	15.0±7.5 [10,20]
Million counts/cm	0.11±0.03 [0.09,0.12]	n/a	n/a
Total million counts/scan	19.4±5.8 [16.4, 21.1]	2.0± 1.4 [1.4,3.0]	0.09± 0.11 [0.05, 0.11]

*mean ± standard deviation [Q1, Q3]
#calculated based on the bed motion speed for a scan of 175 cm
Table 3.13. Overview of quantitative gamma camera acquisition parameters.

3.1.2.2.1 Bone scintigraphy

Figure 3.13 provides a visual representation of quantitative acquisition parameters bone gamma camera scans, while Table 3.13 summarises these parameters. Similar to the case of bone SPECT/CT discussed in chapter 3.1.2.1.1, a moderate variation in activity levels across hospitals was noted, with a mean value around 272±47 MBq (min=170 MBq, max=398 MBq). Upon closer examination of the data acquisition process, it is interesting to note that the speed of total body bone gamma camera acquisitions was consistent across participating hospitals, averaging at 2.2 ± 0.5 mm/s (min=1 mm/s, max=3 mm/s). Further analysis showed that average scan duration and, therefore, also the speed of bed motion for total body imaging, exhibited a moderate correlation with effective activity in the patient (R=0.59).
In summary, these findings suggest that the average activity in the patient during a scan moderately affects the definition of the total body bone gamma camera protocol, a factor that was absent in the case of SPECT/CT (i.e. projection time to activity in the patient).

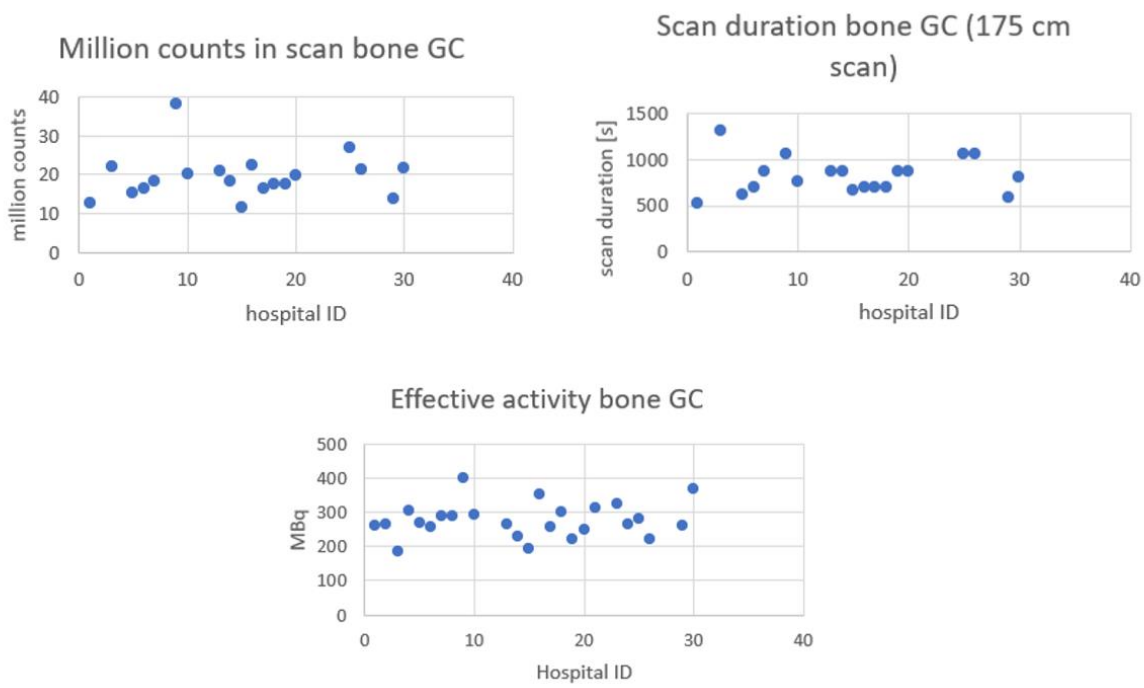

Figure 3.13. Overview of acquisition parameters for bone gamma camera scans.

3.1.2.2.2 DMSA scintigraphy

Figure 3.14 provides a visual representation of quantitative acquisition parameters for DMSA renal gamma camera scans, while Table 3.13 summarises these parameters.

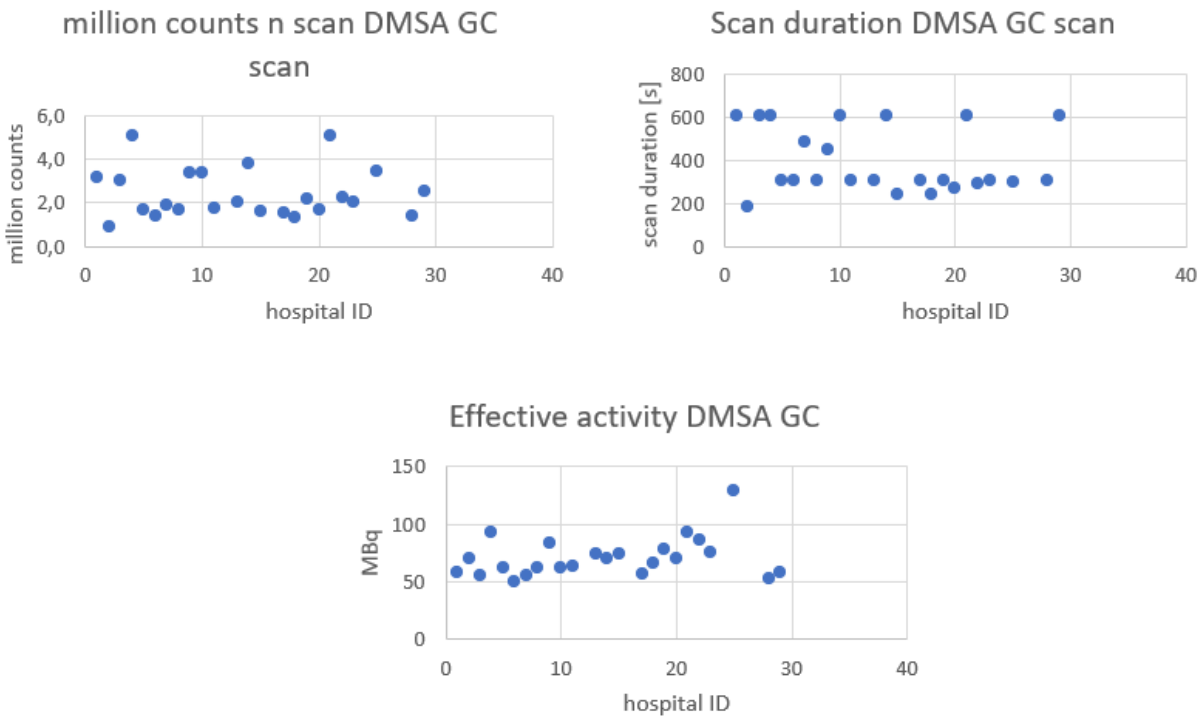

Figure 3.14. Overview of acquisition parameters for DMSA gamma camera scans.

It is evident that a relatively narrow range of activity levels was used for the scans, with the average effective activity in the patient during scanning being 71 ±17 MBq (min=50 MBq, max=130 MBq). Despite scan durations varying by more than a factor of 2 across hospitals, actually ranging from 180 to 600 seconds, almost half of the hospitals opted for a scan duration of approximately 5 minutes (11 hospitals). The scan duration exhibited considerable variability across hospitals, with a mean value of 389 ± 147 seconds (Q1=298 s, Q3=600 s). Surprisingly, there was no strong correlation between scan duration and the effective activity in the patient (R=-0.05), indicating that activity levels were not considered in designing the imaging protocol. Consequently, the total number of counts in a kidney DMSA gamma camera scan was determined predominantly by scan duration (R=0.8), with only a moderate effect of activity in the patient (R=0.52).

3.1.2.2.3 Stomach scintigraphy

Figure 3.15 provides a visual representation of quantitative acquisition parameters for stomach gamma camera scans, while Table 3.13 summarises these parameters.

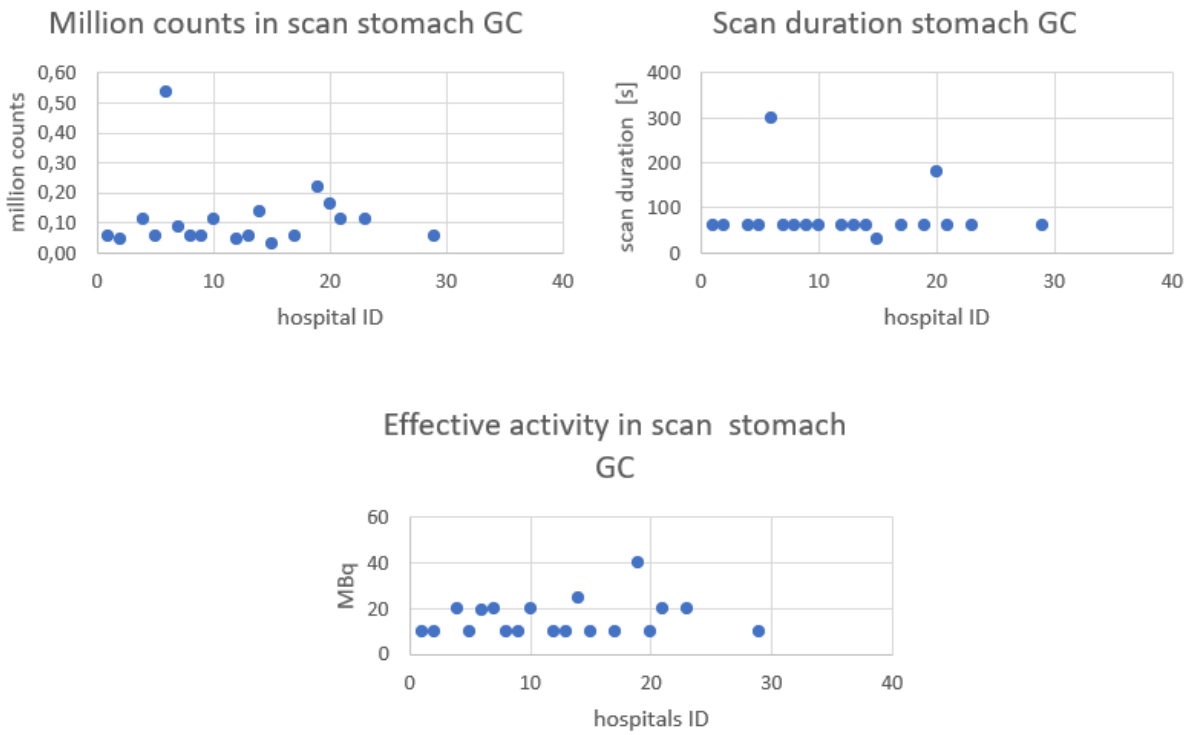

Figure 3.15. Overview of acquisition parameters for stomach gamma camera scan.

Visualising stomach emptying can be performed both as a dynamic scan, and as multiple static images. The comparison here is based on a single static image for a multiple static protocol. Except for three hospitals, all others opted for a fixed scan duration of 60 seconds (Q1=60, Q3=60). Effective activity in the patient ranged from 10 to 40 MBq, with an average of 15 ± 7.7 MBq. No correlation was evident between scan duration and the effective activity level in the patient (R=0.06), which means that effective activity level was not the defining factor while designing the protocol. Consequently, similar to DMSA kidney gamma camera scans, the total counts in stomach gamma camera acquisitions was affected primarily by scan duration (R=0.89), with the patient's activity level only playing a secondary role (R=0.46).
It is interesting to note that, among the two hospital outliers conducting 180 and 300 seconds stomach acquisitions, the latter consistently pursued significantly higher total scan statistics compared with the national average for SPECT, gamma camera, and PET/CT scans. It is hypothesized that this behaviour reflects a preference among local nuclear medicine physicians for enhanced image quality.

3.2 Systematic literature review for selection of Model Observer

Deze paragraaf is geschreven met hulp van de methodologisch ondersteuner van het Kennis Instituut Medisch Specialisten in het begin van het traject van dit project. Het besluit om het rapport in het Engels te schrijven is later genomen. Vanwege de niet begrote kosten voor het herschrijven in het Engels is besloten om deze paragraaf te laten staan in het Nederlands.

Uitgangsvraag

Welke model observers (MO's) zijn bruikbaar en betrouwbaar voor de objectieve beoordeling van beeldkwaliteit ten behoeve van evaluatie[1] en optimalisatie van nucleair geneeskundige beelden?

Subvragen:

1. Welke model MO's zijn beschikbaar binnen de nucleaire geneeskunde en voor welke taak/scenario (typisch classificatie en detectie) kunnen ze worden ingezet?
2. Welke MO met een bewezen goede performance is het meest doelmatig en praktisch implementeerbaar?

Inleiding

Voor een kwalitatief hoogwaardig diagnostisch proces is het vaststellen, bewaken, en optimaliseren van beeldkwaliteit van essentieel belang. Om beeldkwaliteit door een menselijke waarnemer te objectiveren, kan gebruik worden gemaakt van rekenkundige modellen, ook wel MO's genoemd. Deze modellen worden al decennia onderzocht en ingezet voor evaluatie en optimalisatie van beeldkwaliteit in de radiologie [27]. De vraag is welke model observers er toepasbaar zijn voor de beoordeling van nucleaire beelden en hoe deze zich verhouden tot menselijke waarneming bij verschillende klinische taken.

Beeldkwaliteit is feitelijk de mate waarin de in een beeld opgesloten informatie kan worden ontsloten, zodat de beoordelaar in staat is om zijn of haar waarnemingstaak uit te kunnen voeren. Optimale beeldkwaliteit is essentieel voor het kunnen beoordelen van medische beelden door radiologen en nucleair geneeskundigen. Het vaststellen van (optimale) beeldkwaliteit met behulp van radiologen en nucleair geneeskundigen is echter een tijdsintensief en daarmee ook een kostenintensief proces. Bovendien is inter-observer variabiliteit een mogelijke bron van onzekerheid bij het bepalen van optimale beeldkwaliteit.

Traditioneel wordt beeldkwaliteit daarom vaak gekarakteriseerd met behulp van de klassieke beeldkwaliteitsparameters zoals resolutie, contrast en ruis van beelden. Deze parameters zijn efficiënt en reproduceerbaar vast te stellen. Anderzijds is de menselijke waarneming en beoordeling van beeldkwaliteit te complex om met deze enkele parameters samen te vatten.

MO's zijn mathematische modellen die zijn ontwikkeld om de beoordelingsprestaties van beeldvormend specialisten met betrekking tot medische beelden te simuleren. Hiertoe worden trainings- en validatiebeelden gebruikt die representatief zijn voor de medische beelden voor een gegeven klinische taak (bijvoorbeeld laesie detectie). De MO doet een

[1] Bij 'evaluatie' kan bijvoorbeeld gedacht worden aan benchmarking van beeldkwaliteit tussen verschillende ziekenhuizen, maar ook aan routinematige kwaliteitscontrole van de nucleair geneeskundige beelden.

uitspraak over de mate waarin de taak met succes kan worden uitgevoerd. Onder de aannames dat de gebruikte trainings- en validatiedata representatief zijn voor de klinische praktijk en dat de gemodelleerde MO vergelijkbare prestaties heeft als de menselijke waarnemer kan dit model vervolgens ingezet worden voor protocoloptimalisatie of beoordeling van nieuwe beelden ten opzichte van een geldende standaard. Het gebruik van MO's voor het beoordelen van beeldkwaliteit sluit daardoor beter aan bij de klinische praktijk dan de klassieke beeldkwaliteitsparameters dat doen. MO's zijn nadrukkelijk niet bedoeld voor de medische beoordeling van individuele beelden.

Zoeken en selecteren

Voor de eerste subvraag is gebruik gemaakt van reeds bekende literatuur.
Om de tweede subvraag te kunnen beantwoorden is er een systematische literatuuranalyse verricht naar de volgende zoekvraag:
P: PET, SPECT, FDG, nuclear CT, breast imaging;
I: model observer;
C: human observer;
O: Area under the curve (AUC) van de receiver operating characteristic (ROC).

Zoeken en Selecteren (Methoden)

In de databases Medline (via OVID) en Embase (via Embase.com) is op 19-5-2021 met relevante zoektermen gezocht naar systematische reviews en vergelijkend onderzoek. De zoekverantwoording is weergegeven onder het tabblad Verantwoording. De literatuurzoekactie leverde 91 treffers op. Studies werden geselecteerd op grond van de volgende selectiecriteria: vergelijking tussen verschillende MO's of een vergelijking tussen een MO en een menselijke waarnemer in nucleaire medische beeldvorming. Op basis van titel en abstract werden in eerste instantie 20 studies voorgeselecteerd. Na raadpleging van de volledige tekst, werden vervolgens 11 studies geëxcludeerd (zie exclusietabel onder het tabblad Verantwoording), en 9 studies geselecteerd. Na verdere selectie op de praktische toepasbaarheid van de MO, werden uiteindelijk 2 studies definitief geïncludeerd.

Type model observers

De Amerikaanse fysicus Albert Rose (1910 –1990) heeft een basis gelegd voor theoretische analyse van menselijke signaaldetectie in medische beelden met de introductie van zijn Rose-model. In dit model is het "signaal" de gemiddelde waarde over het gebied van interesse (het object) minus een vergelijkbare meting over eenzelfde gebied in de achtergrond. De "ruis" wordt beschreven als de fluctuatie in de gemiddelde achtergrond in vergelijkbare gebieden [28]. Met deze definities van signaal en ruis kan vervolgens een signaal-ruis-verhouding (signal-to-noise ratio - SNR) worden bepaald. Rose vond als detectiegrens voor menselijke waarnemers een SNR van circa 5. Het Rose model beschrijft een eerste en eenvoudige variant van een model observer - en ligt daarmee aan de basis van later ontwikkelde en hieronder beschreven model observers.

In de latere literatuur kunnen we kortweg de volgende type MO's onderscheiden: nonprewithening model observer (NPW), nonprewithening model observer with eyefilter (NPWE), Hotelling en Channelized Hotelling observer (CHO) [27]. De modellen NPW en NPWE tonen de meeste verwantschap met het hierboven beschreven Rose-model. Meer recent zijn in aanvulling op het genoemde rijtje ook visual-search gebaseerde model observers ontwikkeld en model observers die gebruik maken van deep learning.

Net als het Rose-model, houden Non-prewithening model (**NPW**) observers geen rekening met de correlatiestructuur van ruis in beelden. Deze modellen houden alleen rekening met eigenschappen van het signaal (bijvoorbeeld de te detecteren laesie). Het signaal wordt in het model beschreven door een voor-gedefinieerde template van het signaal. Zie onderstaand schema, waarbij in dit geval - voor deze NPW toepassing - het eye filter achterwege moet worden gelaten.

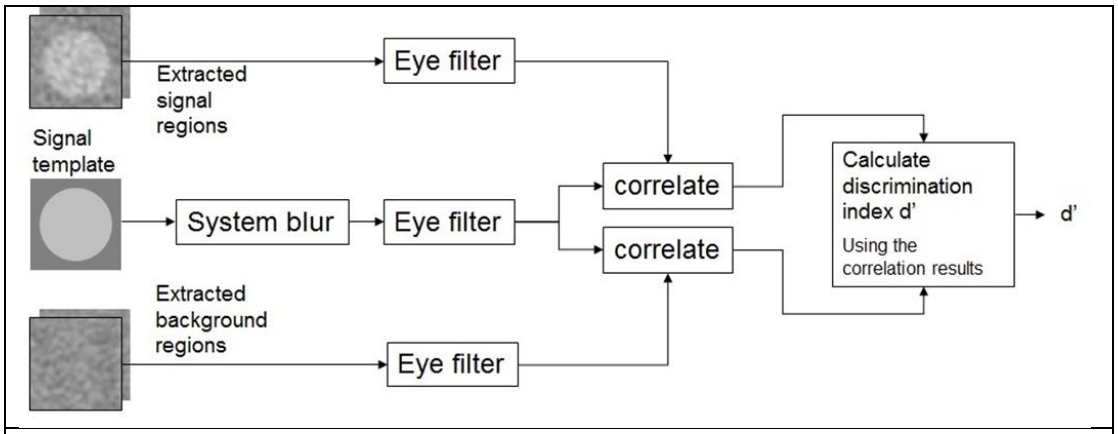

Fig 1: Vereenvoudigd stroomschema van de NPWE model observer. Een dataset van 'regions of interest' (ROIs) met een signaal (object) wordt vergeleken met een dataset van ROIs met alleen achtergrond (geen object aanwezig). Om een mate van detecteerbaarheid te verkrijgen, gebruikt de model observer een vooraf gedefinieerde template van het signaal. De template wordt gefilterd om rekening te houden met blur van het beeldvormingssysteem en om rekening te houden met eigenschappen van het menselijke visuele systeem (oogfilter). Het template wordt zowel gecorreleerd met de beelddataset die het object bevat als de beelddataset die alleen de achtergrond bevat (beide ook gefilterd door het oogfilter). Uit de correlatieresultaten kan een detectie-index d' worden berekend. Intuïtief kan worden aangevoeld dat een sterke correlatie van template en objectafbeeldingen een goede detecteerbaarheid van het object betekent (hogere d').

Het feit dat de correlatie van pixelruis niet wordt meegenomen in deze modellen maakt ze suboptimaal in beelden waar de ruis in zekere mate gecorreleerd is. Bepalende onderzoekers in het vakgebied [29] hebben aangetoond dat de detectieprestaties met name niet consistent zijn met mensen, in beelden waar laagfrequente ruis sterker aanwezig is. Onder andere bij beeldvorming binnen nucleaire geneeskunde is bekend dat laagfrequente ruis een rol speelt en de NPW MO minder consistente resultaten geeft [30,31].

Toch zijn NPW MO's nuttig gebleken, bijvoorbeeld bij vergelijkingen waarbij het frequentiespectrum van de ruis in beelden niet of nauwelijks verandert, of in situaties waar laag frequente ruis geen belangrijke rol speelt. Het grote voordeel van deze MO's is de eenvoudige implementatie.

De inconsistentie bij laagfrequente beeldruis bracht Burgess ertoe om een 'eye filter' toe te passen in de NPW MO en dit resulteerde in de **NPWE** MO (Fig. 1). Het 'menselijk' oogfilter onderdrukt de lage frequenties en maakt de model observer daardoor consistenter in vergelijk met mensen bij sterkere aanwezigheid van lage frequenties in de beeldruis.

Tegenover de NPW MO's staan prewithening model observers die wel rekening houden met correlaties in de beeldruis. Een voorbeeld hiervan is de **Hotelling observer** die gebruik maakt

van de covariantiematrix die de correlaties in de beeldruis beschrijft. Een probleem hierbij is de rekenkracht die nodig is voor dit type model observer (met name bij het inverteren van de covariantie-matrix in het model).

Een oplossing voor dit probleem vormen de channelized Hotelling observers (**CHOs**), die een compromis geven tussen prewithening en non-prewithening. De channels zorgen ervoor dat de hoeveelheid data hanteerbaar wordt (in het kader van de covariantie-matrix). Ze geven feitelijk een compactere, maar uiteraard minder complete, weergave van de beeldinformatie. De kanalen zijn daarnaast een poging om het menselijke visuele systeem te modelleren. Er zijn in de literatuur veel verschillende channels beschreven. De keuze van channels kan helpen om de model observer te optimaliseren voor een bepaalde detectie-taak (afhankelijk van beeldkarakteristieken en het type object/signaal). Het tot stand brengen van de covariantie-matrix vindt plaats op basis van een training-set aan beelden.

De laatste jaren zijn er naast bovenstaande model observers ook algoritmes ontwikkeld die op deep learning zijn gebaseerd [32]. Deze nieuwe variant van model observers is nog geen gemeengoed maar toont wel potentie om tekortkomingen van de meer conventionele model observers te ondervangen. Zo is de verwachting dat deep learning MO's beter geschikt zijn voor detectie-taken in beelden met complexe anatomische achtergronden.

De bovenstaande MO's beoordelen doorgaans samples van beelden die wel of niet een signaal bevatten. Hieruit volgt dan een mate van detecteerbaarheid: hoe goed is het onderscheid tussen signaal-samples en achtergrond-samples. Er wordt hier geen 'search' in meegenomen.

Als laatste variant kunnen ook nog de zogenaamde scan- en visual-search-algoritmen worden genoemd: deze model observers behelzen een zoektocht ('search') naar potentiële signaallocaties en de berekening van een detectie-parameter als functie van signaallocatie. Deze MO's zijn een variant op de bovengenoemde model observers (bijvoorbeeld Hotelling) en verkeren nog in een ontwikkelingsstadium. Search algoritmen zijn in de literatuur frequent onderzocht in relatie tot beeldvorming binnen nucleaire geneeskunde [30].

Resultaten systematische literatuuranalyse

Uit de literatuur search kwamen de volgende MO's naar voren binnen de context van beeldvorming in de nucleaire geneeskunde:

- Channelized hoteling observer (CHO)
  - [33-38]
- (Scanning) channelized nonprewhitening observer (CNPWO)
  - [33-35,39,40]
- Non-prewithening model observer (NPWO)
  - [35]
- Visual-search (VS)
  - [39,40]
- Ideal observer (IO)
  - [36]
- Deep learning based MO
  - [41]

Na het beooordelen van de literatuur, is de werkgroep tot de conclusie gekomen dat de Channelized hoteling observer (CHO) praktisch het best implementeerbaar is voor het doel van deze richtlijn. In het algemeen wordt in de geïncludeerde literatuur de CHO model observer als gouden standaard beschreven. Daarnaast zijn er verschillende voorbeelden in

de literatuur te vinden waarbij CHO modellen worden gebruikt voor de beoordeling van beeldkwaliteit optimalisatie taken. Daarom is er voor gekozen alleen in te gaan op de prestaties van de CHO.

Twee van de negen geïncludeerde studies vergelijken de prestaties van een CHO met menselijke waarnemers. Deze studies zijn gebruikt voor de beantwoording van deelvraag 2. De belangrijkste studie resultaten zijn samengevat in de evidence-tabellen.

Gifford (2005) [33] onderzocht het gebruik van model observers die multi-slice lokalisatie analyse kunnen doen. Er werd een vergelijking gemaakt tussen menselijke beoordelaar en twee verschillende channelized MO's (CHO en CNPWO). Gebruikte data waren gesimuleerde SPECT data van een digitaal antropomorf cardiac torso fantoom waarin op random locaties tumoren (Ga-67 avide lymphomas) zijn geplaatst. Het fantoom was representatief voor een patiënt van 170 cm en 70 kg. Tumoren 1 cm groot. De taak was tumordetectie. Verschillende range van tumor background contrasten werden gescand. De totale set omvatte 640 (50/50 met en zonder tumor) beelden voor training en testen, de helft (320) werd gebruikt voor training. Er waren vier menselijke beoordelaars. 100 beelden per reconstructie/display format.
Resultaten: De resultaten werden gepresenteerd aan de hand van oppervlakte onder de localisation ROC curve (LROC curve)[2], als maat voor detectieperformance. De resultaten van mens en model worden in grafieken getoond. De CHO geeft goede kwantitatieve overeenkomst met menselijke waarnemers voor het geval er interne ruis in het model wordt meegenomen.
Conclusie: CHO met interne ruis benadert het beste de menselijke beoordelaar. Voor enkele SPECT data uit iteratieve reconstructies kan de CNPW ook gebruikt worden met als voordeel dat de benodigde training set kleiner is. Kanttekening is dat beide MO's nog getuned moesten worden in keuze van frequentiekanalen om het beste aan te sluiten bij de menselijke beoordelaars.

Sen en Gifford (2016) [38] vergeleken twee verschillende MO's (CHO en visual-search (VS)) als methodes om menselijke beoordeling te simuleren in realistische detectietaken, waarbij de locatie van de te detecteren laesies alsmede de anatomische en kwantumruis als variabelen werden meegenomen. De collimator sensitiviteit werd bovendien gevarieerd in deze studie. De gebruikte data betroffen gesimuleerde SPECT In-111 planaire beelden van een digitaal bekken-abdomenfantoom XCAT. De taak was detectie van lesies in de prostaat en in lymfeklierregio's in het bekken. De resultaten van de MO en VS zijn vergeleken met die van een menselijke beoordelaar. Er werd een set van 450 afbeeldingen gebruikt, waarvan 100 om de MO en VS te trainen en 300 beelden om te beoordelen. Vier menselijke beoordelaars (geen radiologen) kregen 25 beeldparen om te trainen en 50 om werkelijk te beoordelen.
Resultaten: De detectie resultaten werden gepresenteerd aan de hand van oppervlakte onder de LROC-curve. De resultaten van mens en model worden in grafieken getoond. De CHO met interne ruis geeft goede kwantitatieve overeenkomst met menselijke waarnemers, met name geldt dit bij lagere collimatorsensitiviteit. Het VS model met interne ruis geeft goede overeenkomst met menselijke waarnemers, zonder restrictie van de collimatorsensitiviteit.
Conclusie: Zowel de MO als de VS komen overeen met de menselijke beoordelaar. Het VS model is echter wel flexibeler en komt beter overeen met de menselijke beoordelingen, tevens is de VS MO sneller dan de CHO. Als kanttekening kan worden geplaatst dat in deze studie bij het VS model naast interne ruis ook een 'search threshold' is geïntroduceerd als

[2] De ROC-methodologie is niet geschikt voor studies waar de lokalisatie van een signaal of object een belangrijke rol speelt ('search'). Voor dit doel is de zogenaamde lokalisatie ROC (LROC) methodologie geïntroduceerd.

parameter om het model te tunen. Deze tweede parameter heeft het VS model mogelijk een voordeel gegeven boven CHO (waar een dergelijke parameter niet werd geïntroduceerd).

Bewijskracht van de literatuur

Gezien de technische insteek van bovenstaande artikelen is er geen GRADE beoordeling uitgevoerd. Wel is op een aantal punten gekeken naar de kwaliteit van de studies, zie de tabel. Het is echter niet mogelijk hier een gradering van de bewijslast aan te koppelen.

Geen GRADE	Beoordelingen door de channelized hotelling observer (CHO) lijken overeen te komen met resultaten door een menselijke beoordelaar in de context van de nucleaire geneeskunde. Bron: [33,38]

Overwegingen – van bewijs naar aanbeveling

In de gevonden literatuur worden diverse MO's besproken. Veelal betreft het verbeterde varianten of volledige nieuwe modellen. De artikelen doen dit in de context van de nucleaire geneeskunde. Hieruit blijkt dat het gebruik van een MO voor beeldoptimalisatie in de nucleaire geneeskunde een beproefde methode is.

De mate waarin de gepresenteerde MO's goed ingezet kunnen worden als alternatief voor een menselijke beoordelaar varieert echter sterk als functie van het model type en de detectie taak. Een optimale keuze hierin maken is voor dit project niet zozeer een keuze maken voor het best presterende MO in termen van het benaderen van een menselijke beoordelaar, eerder een keuze voor de gevestigde standaard die het meest praktisch te implementeren is en een 'proven commodity' is.

De channelized hoteling observer (CHO) is één van de standaard MO's, dit blijkt ook uit de frequentie waarin de geïncludeerde studies dit model beschrijven. Het betreft een gangbaar model, wat reeds doorontwikkeld is en veel wordt gebruikt als referentiemodel in vergelijking tot andere nieuwere modellen. Hoewel dit model tekortkomingen heeft, zoals mogelijk de mindere prestaties in beelden met complexe anatomische achtergronden [32], benadert het wel in voldoende mate de menselijke beoordelaar voor een beeldoptimalisatie taak en de implementatie ervan is reeds veelvuldig gedaan. Sen en Gifford [38] laten in hun studie zien dat de Visual-search (VS) beter presteert dan de CHO, maar het aantal implementaties op basis van de literatuur is beperkt en de richtlijnwerkgroep kent ook geen ongepubliceerde praktische toepassing van deze MO en acht de VS daarom nu nog te experimenteel om breed in te zetten in de praktijk.

Ook in de radiologie is het gebruik van het CHO model een geaccepteerde standaard. In het kwaliteitsprotocol, "Quality control in cone-beam computed tomography (CBCT)" EFOMP-ESTRO-IAEA (2019) staat het onderwerp MO expliciet opgenomen als kwaliteitscontrole middel. Het NPWE model en CHO model worden hierin benoemd als modellen, welke de menselijke beoordelaar reproduceren in laag-contrast detectietaken [42-45].

Daarnaast wordt in een rapportage van de American Association of Physics in Medicine task group 233 "Performance Evaluation of Computed Tomography Systems (2019) beschreven dat MO's een krachtig alternatief zijn voor menselijke beoordelaars in CT data. Er worden artikelen benoemt die de CHO en NPWE MO's beschrijven als erg goede voorspeller voor menselijke beoordelaars [44-50].

Voortkomende uit de geselecteerde artikelen in de literatuurstudie is tevens een artikel gevonden [2] welke een gelijkwaardige casus beschrijft als de taakstelling in dit project. Deze studie richt zich volledig op het kwantificeren van de detecteerbaarheid van laesies in PET met behulp van een 3D geprint fantoom om acquisitie en reconstructie parameters te kunnen optimaliseren. De objectieve automatische analyse van de detecteerbaarheid van de laesies wordt gedaan met behulp van een CHO model. Het artikel van Wollenweber [2] is daarmee één van de weinige artikelen die een vergelijkbare optimalisatie stapsgewijs beschrijft zoals beoogd in dit project.

Op basis van de bovenstaande overwegingen is gekozen voor het CHO model als meest geschikte en praktisch uitvoerbare model observer voor de doelstellingen in dit project.

De chanellized hotelling observer (CHO) is een toepasbaar en bewezen rekenkundig model om het presteren van een menselijke beoordelaar te voorspellen. De CHO kan daarom gebruikt worden ten behoeve van evaluatie en optimalisatie van beeldkwaliteit.

Kennislacunes

De gangbare standaard model observer voert de detectie taak uit in de afzonderlijke afbeeldingen (slices) van een 3D dataset van beelden. De model observer gebruikt hier een 2D benadering waarbij de detectie in de afzonderlijke beelden niet afhangt van informatie uit aangrenzende beelden. Of deze 2D benadering voldoende goed werkt of dat een 3D benadering de menselijke waarnemer beter zou voorspellen in 3D beeld-datasets, is nog onvoldoende uitgezocht. Christiana Balta heeft in haar promotie-onderzoek aan de Radboud universiteit (promotie in 2019 - https://arxiv.org/abs/2011.01405) een aanzet gedaan met haar onderzoek naar de detectie-eigenschappen van menselijke waarnemers in 2D beelden versus de menselijke detectie-eigenschappen in 3D datasets van beelden. Als 3D afbeeldingstechniek werd hier tomosynthese van de borst gekozen. Het onderzoek had als doel aanwijzingen te vinden voor het gewenste design van een model observer in 3D beeld-datasets. Er werd in dit onderzoek geen aanwijzing gevonden voor de noodzaak van een 3D benadering van de model observer, maar diepgaander onderzoek lijkt nodig.

Gradering van de bewijskracht

Author	Type data (reeel > fantoom > simulatie)	Voldoende variatie in beelden?	In geval van fantoom: humanoid fantoom > simpel fantoom; achtergrondactiviteit ja/nee, vorm van laesie (a)symmetrisch	Model observer: Grootte van de trainingsset. Verificatie middels humane observers?	Menselijke beoordelaars (aantal, ervaring, trainig)	Effectschatter en spreiding	Industrie/ meerdere studies van zelfde onderzoeksgroep	Eindoordeel
Sen en Gifford, 2016 [38]	simulatie	XCAT fantoom data voor de bekken en abdomen regio's gebruikt. Klinkt realistisch. Er zijn 16 combinaties van contrast en collimator type gebruikt. Resulterend in een dataset van 450 beelden waarin laesies wel of niet aanwezig waren en er 255 verschillende laesie locaties zijn gebruikt.	Betreft een digitaal bekken-abdomen fantoom XCAT.	255 beelden training	3 human observers. Niet radiologen maar wel image science onderzoekers. 25 training en 50 testbeelden per collimator waarvan er 5 zijn gebruikt.	Niet te beoordelen	n.v.t.	Kwaliteit van de studie lijkt voldoende
Gifford, 2005 [33]	simulatie	Gebruikte randvoorwaarde 1 of 0 tumoren. 320 verschillende locaties. Het aantal verschillende ruis en contrast variaties onbekend.	Digitaal antropomorf cardiac torso fantoom waarin op random locaties tumoren zijn geplaatst. Representatief voor pt. 170 cm, 70 kg. Tumoren 1cm groot. Verschillende range van tumor background contrasten gescand.	totale set van 640 (50/50 met en zonder tumor) beelden voor training en testen. 320 werden gebruikt voor training.	4 human observers. 100 beelden per reconstructie/ display format.	Niet te beoordelen	n.v.t.	Kwaliteit van de studie lijkt voldoende

3.3 Evaluation of applicability of standard PET phantoms

3.3.1 Introduction

As we have shown in the previous chapters, protocols used for the visual evaluation of PET/CT exams significantly differ between hospitals. This divergence arises because hospitals equipped with newer generation PET/CT scanners can obtain higher image quality or perform faster scans at lower patient doses. Consequently, variations in image appearance and quality parameters exist across different hospitals, driven by the individual preferences of nuclear medicine physicians regarding visual examination. What is needed is a modus to objectively test these clinical images for different detection tasks suited to the quality of the images these cameras can produce.

3.3.2 Method

We used a Biograph Vision 600 (Siemens Healthineers, Knoxville, TN, United States) digital PET/CT system and scanned the NEMA Image Quality phantom for a range of target to background concentration ratios and from this we derived contrast-detail curves.
The same formula (1) described in chapter 2 was used for determining the acquisition count statistic at the time of acquisition.

3.3.2.1 Phantom preparation

A standard NEMA image quality phantom was used to collect data. This phantom contains a body compartment that is at least 18 cm in interior length, six fillable spheres with internal diameters of 10, 13, 17, 22, 28, and 37 mm with a wall thickness of not more than 1 mm, and a cylindrical insert with 5 cm outside diameter (see figure 3.17) that is filled with low atomic number material that mimics lung attenuation and is centred inside the body compartment and extends axially through the entire

phantom.

Fig 3.17. NEMA Image Quality phantom.

3.3.2.2 Phantom analysis

From each acquisition we used the CT series to derive the exact location of the spheres by using a thresholding method. The resulting template was used to measure the contrast of the spheres in the attenuation corrected PET series by first up-sampling the PET images to the resolution of the CT scan (see figure 3.18).

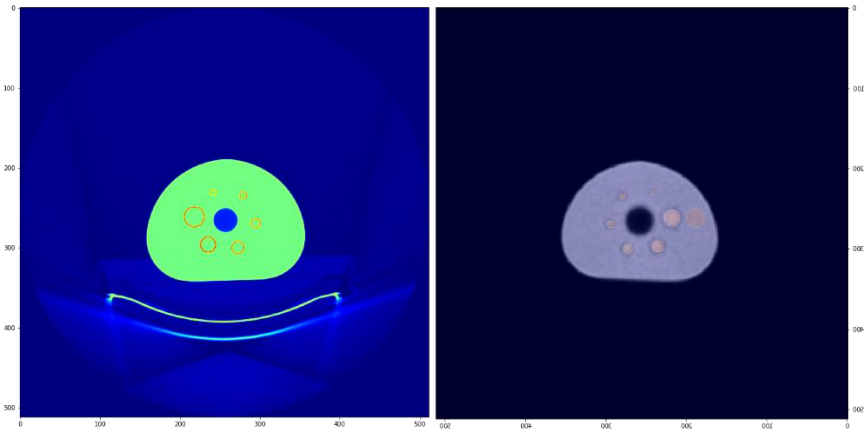

Fig. 3.18. Images of the phantom from CT (left) and CT with fused PET (right)

A background ROI was drawn to measure the contrast with respect to this ROI. Each CNR curve is fitted

$$CNR = \frac{Cmean(sphere) - Cmean(background)}{SD(background)} \qquad (12)$$

where the contrast C is defined with respect to a background region and given by C = (Nl-Nbg)/Nbg, where Nl can be either the maximum or average number of counts per pixel measured in the ROI that was drawn around the hotspot [51]. Because most of the sphere diameters were comparable to or smaller than the pixel size of the camera, the maximum ROI value was used for analysis. Nbg is the average number of the background counts. The background ROIs for each row were chosen to be as large as possible without overlapping with the spheres in each row. The coefficient of variation (COV) is given by COV = σbg/Nbg, where σbgis the standard deviation of the background ROI.
The detectability of a hotspot by an observer was correlated with the CNR value of that hotspot.

3.3.3 Results

3.3.3.1 CNR curves of standard NEMA Phantom

Figure 3.19 shows the measured CNR curve for one of the spheres in the NEMA IQ phantom as a function of acquisition counts for different values of the target to background activity concentration ratio C'.

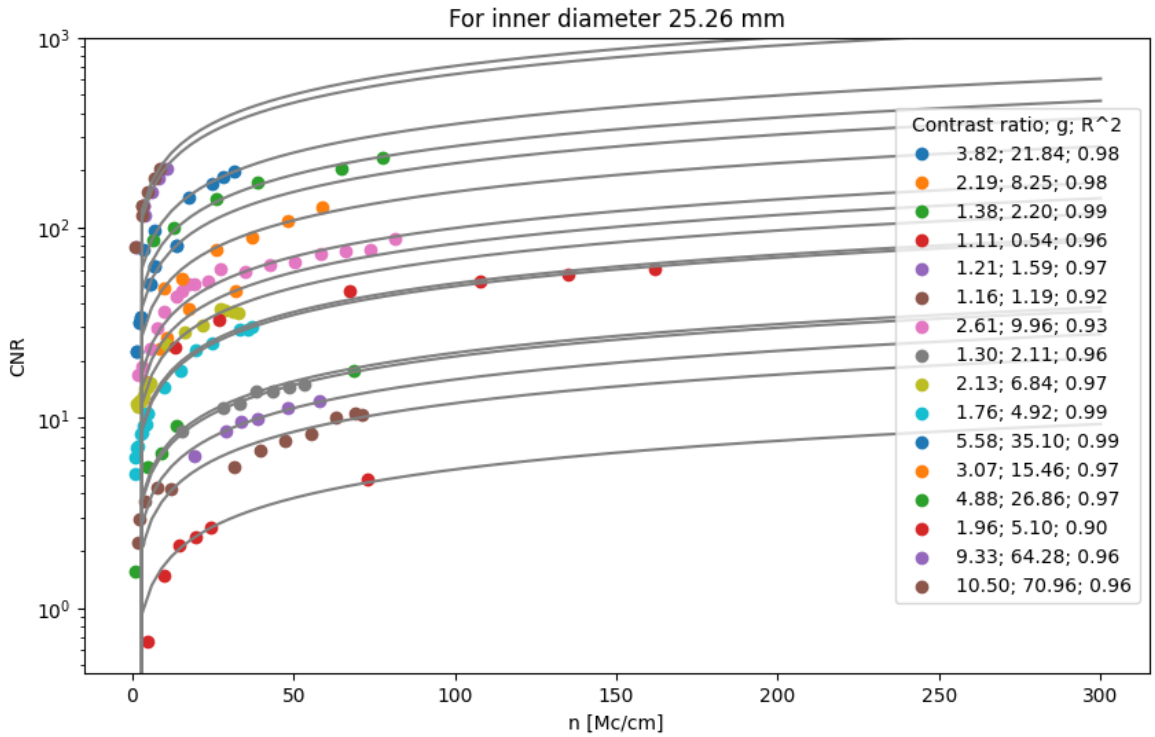

Figure 3.19: CNR curves for different sphere / background activity ratios as a function of acquisition counts for different values of the target to background activity concentration ratio C'.

For a given value of the acquisition counts the relationship can be derived between target to background activity concentration ratio C' and the CNR value for a given object size. Figure 3.20 shows an example of such a relationship.

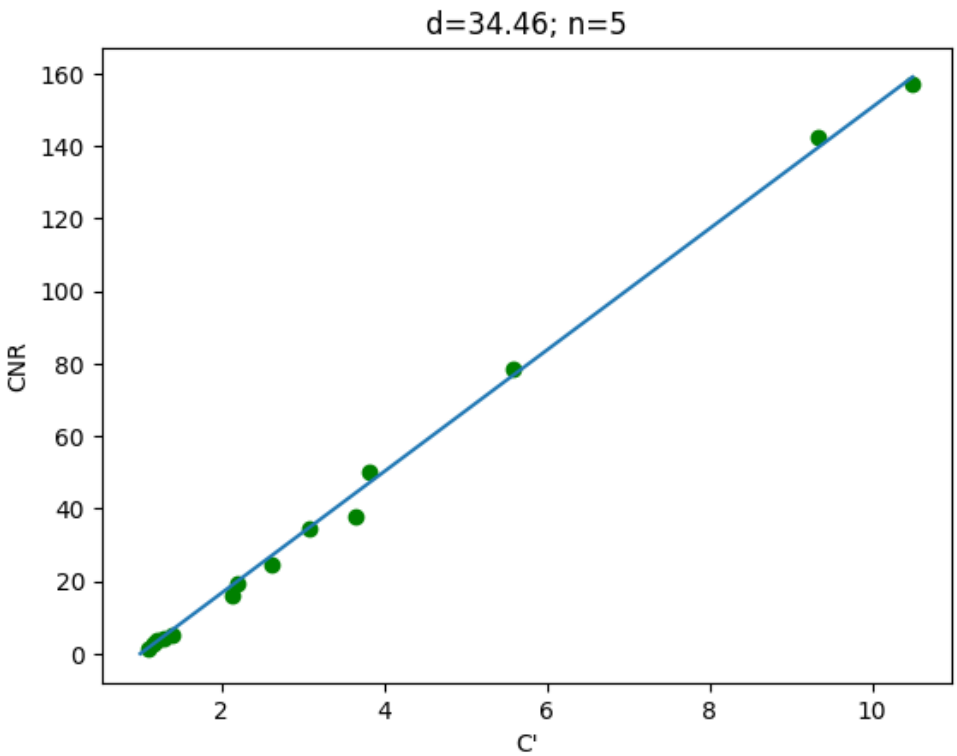

Figure 3.20. CNR as function of target to background activity concentration ratio for a sphere with a measured diameter on the CT image of 34.46 mm.

For each sphere we determined the relation between the CNR and the activity concentration ratio C' by fitting the data. By applying the Rose criterion we derived the value of the activity concentration C' which can still be seen by a human observer. The observed contrast C of a lesion of a given size d is related to the activity concentration C':

$$C/C' = RC \qquad (13)$$

where RC is the recovery coefficient. This now results in the contrast – detail curve for a specific choice of reconstruction settings. The recovery coefficients for our measurements are shown in figure 3.21 and are comparable with measurements by Delcroix et al. [52].

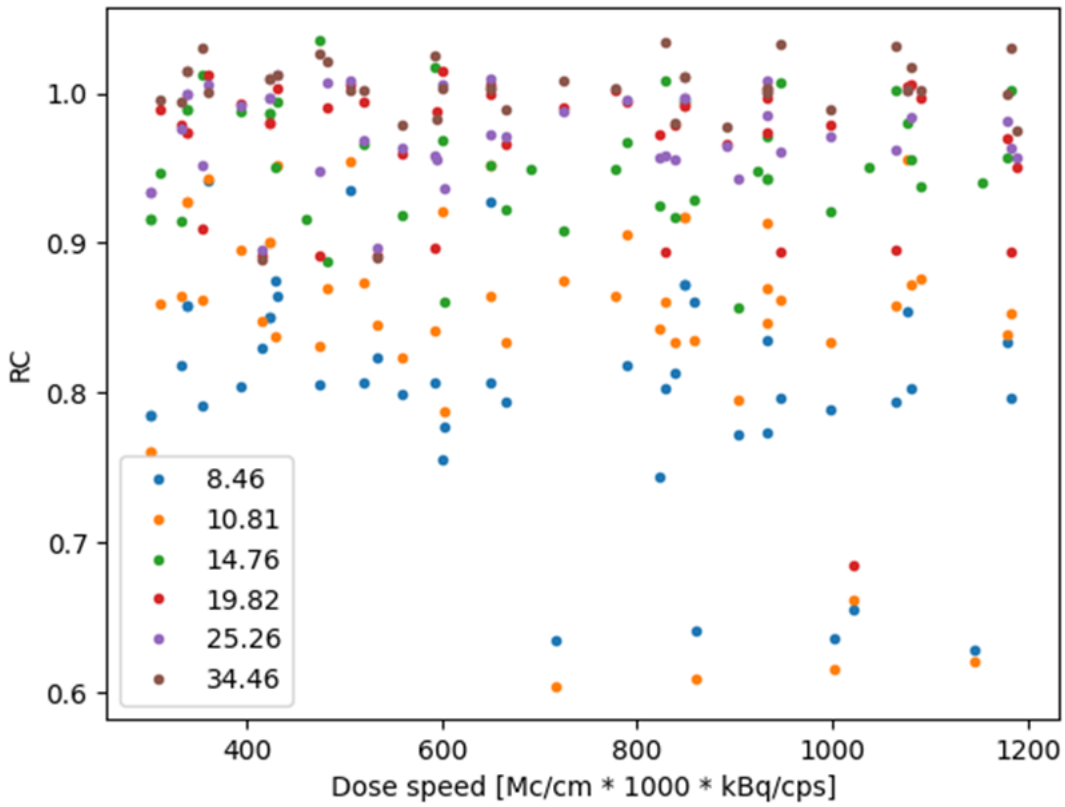

Figure 3.21. The recovery coefficients as a function of object size.

To derive the full contrast-detail curve we approximated the RC value with its average value over the range of dose speeds shown in figure 5. The resulting contrast – detail curve is shown in figure 3.22.

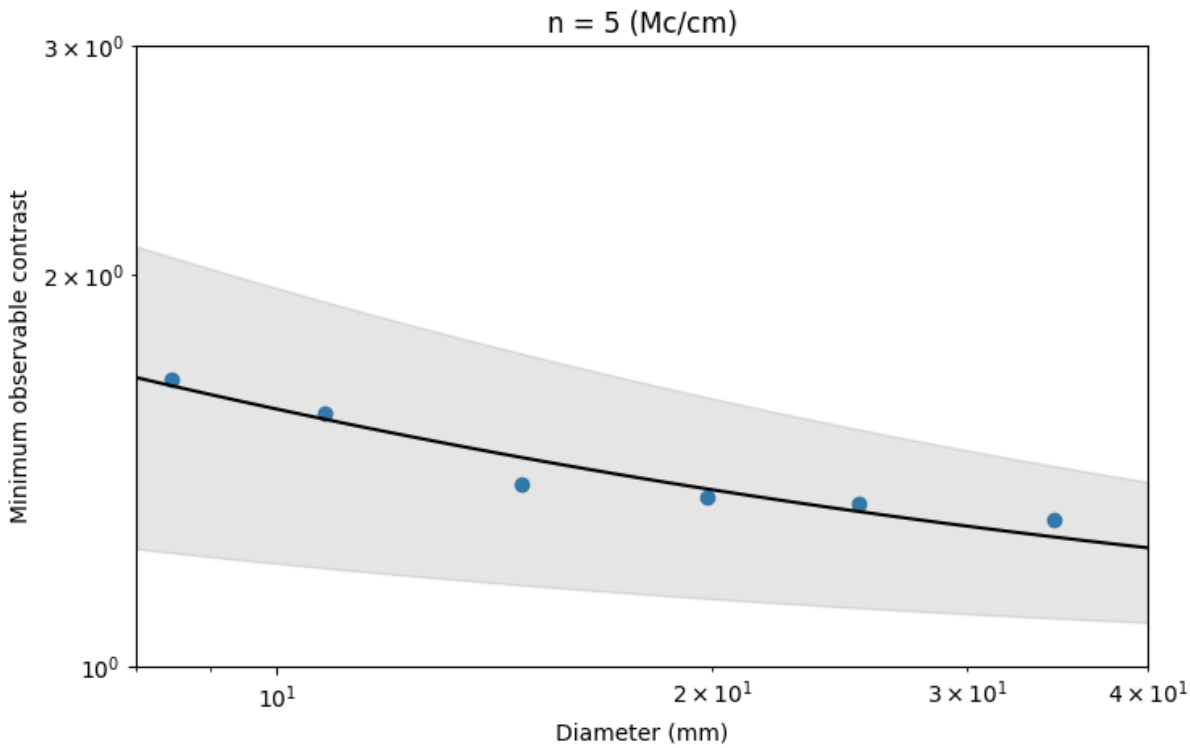

Figure 3.22. The contrast – detail curve for a specific value of the count statistics n.

The data points of the CD curve have been fitted using a power law $y = a\,X^b$ [59].

3.3.3.2 CNR curves of NEMA Phantom with microspheres insert

To validate the extent to which the fit of the CD curve can be extrapolated to smaller size objects we performed measurements with the NEMA phantom with microspheres insert. The microspheres insert has 6 spheres with outer diameter sizes of 5.95, 6.95, 8.23, 9.86, 11.89 and 14.43 mm. The volumes of these spheres are 790, 1150, 2540, 5610, 10610 and 25510 µL, respectively.

From our measurements we derived the minimum observable contrast as a function of sphere diameter, as shown in figure 3.23.

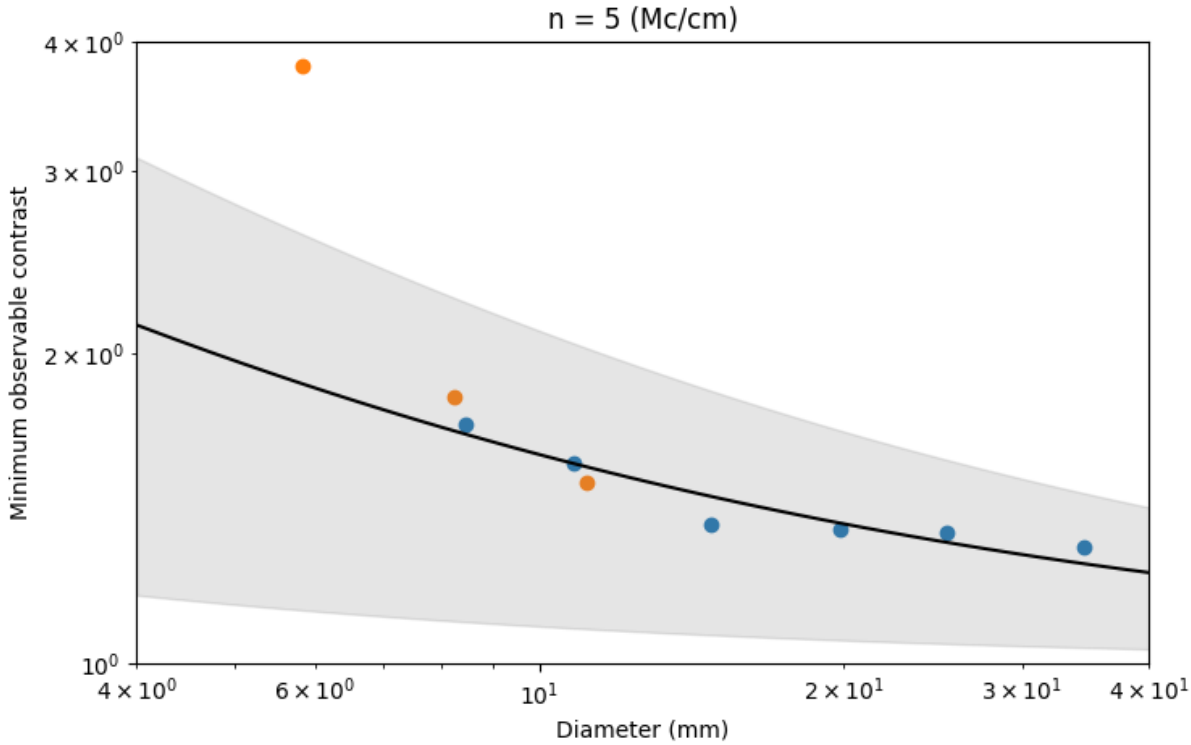

Figure 3.23. The contrast – detail curve for a specific value of the count statistics n. The regular NEMA phantom data points are shown in blue. The microsphere insert measurement is shown in orange.

We were unable to analyse the smallest spheres of the phantom using our automated analysis tools. There were two spheres with diameters similar to two standard NEMA phantom spheres and for these spheres the minimum observable contrast was similar to our measurements with the standard phantom. The smallest size sphere that we could analyse with our method had a diameter of about 6 mm. We can also see that the minimum observable contrast for this sphere lies outside the 95% confidence interval.

4. References

1. Mattoli MV, Calcagni ML, Taralli S, Indovina L, Spottiswoode BS, Giordano A. How often do we fail to classify the treatment response with [^{18}F]FDG PET/CT acquired on different scanners? Data from clinical oncological practice using an autmoatic tool for SUV harmonization. Mol Imaging Biol. 2019;21:1210-1219.
2. Wollenweber SD, Alessio AM, Kinahan PE. A phantom design for assessment of detectability in PET imaging. Med Phys. 2016;43:5051-5062.
3. Kadrmas DJ, Casey ME, Black NF, Hamill JJ, Panin VY, and Conti M. Experimental comparison of lesion detectability for four fully-3d pet reconstruction schemes. IEEE transactions on medical imaging, 2009;28:523–534.
4. van Sluis J, de Jong J, Schaar J, Noordzij W, van Snick P, Dierckx R, Borra R, Willemsen A, Boellaard R. Performance Characteristics of the Digital Biograph Vision PET/CT System. J Nucl Med. 2019;60:1031-1036.
5. Carlier T, Ferrer L, Conti M, Bodet-Milin C, Rousseau C, Bercier Y, Bendriem B, Kraeber-Bodéré F. From a PMT-based to a SiPM-based PET system: a study to define matched acquisition/reconstruction parameters and NEMA performance of the Biograph Vision 450. EJNMMI Phys. 2020;7:55.
6. Rausch I, Cal-González J, Dapra D, Gallowitsch HJ, Lind P, Beyer T, Minear G. Performance evaluation of the Biograph mCT Flow PET/CT system according to the NEMA NU2-2012 standard. EJNMMI Phys. 2015;2:26.
7. Jakoby BW, Bercier Y, Conti M, Casey ME, Bendriem B, Townsend DW. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. Phys Med Biol. 2011;56:2375-89.
8. Stefan Eberl, Roger Fulton, Armin Mohamed, Jordan Verschuer and Michael Fulham. Comparison of 3 ring and 4 ring Biograph TruePoint PET/CT Systems. JNM 2008;49 (Suppl 1):409P.
9. Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. J Nucl Med. 2007;48:471-80.
10. Kolthammer JA, Su KH, Grover A, Narayanan M, Jordan DW, Muzic RF. Performance evaluation of the Ingenuity TF PET/CT scanner with a focus on high count-rate conditions. Phys Med Biol. 2014;59:3843-59.
11. Rausch I, Ruiz A, Valverde-Pascual I, Cal-González J, Beyer T, Carrio I. Performance Evaluation of the Vereos PET/CT System According to the NEMA NU2-2012 Standard. J Nucl Med. 2019;60:561-567.
12. Chicheportiche A, Marciano R, Orevi M. Comparison of NEMA characterizations for Discovery MI and Discovery MI-DR TOF PET/CT systems at different sites and with other commercial PET/CT systems. EJNMMI Phys. 2020;7:4.
13. Pan T, Einstein SA, Kappadath SC, Grogg KS, Lois Gomez C, Alessio AM, Hunter WC, El Fakhri G, Kinahan PE, Mawlawi OR. Performance evaluation of the 5-Ring GE Discovery MI PET/CT system using the national electrical manufacturers association NU 2-2012 Standard. Med Phys. 2019;46:3025-3033.
14. Wagatsuma K, Miwa K, Sakata M, Oda K, Ono H, Kameyama M, Toyohara J, Ishii K. Comparison between new-generation SiPM-based and conventional PMT-based TOF-PET/CT. Phys Med. 2017;42:203-210.
15. Vandendriessche D, Uribe J, Bertin H, De Geeter F. Performance characteristics of silicon photomultiplier based 15-cm AFOV TOF PET/CT. EJNMMI Phys 2019;6:8.

16. Khobragade P. Effect of radiotracer on NEMA NU-2 2018 PET-CT performance for Biograph Horizon scanner. J Nucl Med. 2022;63 (Suppl 2): 3313.
17. Caribé PRRV, Koole M, D'Asseler Y, Deller TW, Van Laere K, Vandenberghe S. NEMA NU 2–2007 performance characteristics of GE Signa integrated PET/MR for different PET isotopes. EJNMMI Phys. 2019;6:11 .
18. Hohberg M, Kobe C, Krapf P, Täger P, Hammes J, Dietlein F, Zlatopolskiy BD, Endepols H, Wild M, Neubauer S, Heidenreich A, Neumaier B, Drzezga A, Dietlein M. Biodistribution and radiation dosimetry of [18F]-JK-PSMA-7 as a novel prostate-specific membrane antigen-specific ligand for PET/CT imaging of prostate cancer. EJNMMI Res. 2019;9:66.
19. Piron S, De Man K, Van Laeken N, D'Asseler Y, Bacher K, Kersemans K, Ost P, Decaestecker K, Deseyne P, Fonteyne V, Lumen N, Achten E, Brans B, De Vos F. Radiation Dosimetry and Biodistribution of 18F-PSMA-11 for PET Imaging of Prostate Cancer. J Nucl Med. 2019;60:1736-1742.
20. Giesel FL, Hadaschik B, Cardinale J, et al. F-18 labelled PSMA-1007: biodistribution, radiation dosimetry and histopathological validation of tumor lesions in prostate cancer patients. Eur J Nucl Med Mol Imaging. 2017;44:678–688.
21. Rogasch, J.M., Hofheinz, F., Lougovski, A. et al. The influence of different signal-to-background ratios on spatial resolution and F18-FDG-PET quantification using point spread function and time-of-flight reconstruction. EJNMMI Phys. 2014;1:12.
22. Martí-Climent JM, Prieto E, Domínguez-Prado I, García-Velloso MJ, Rodríguez-Fraile M, Arbizu J, Vigil C, Caicedo C, Peñuelas I, Richter JA. Contribution of time of flight and point spread function modeling to the performance characteristics of the PET/CT Biograph mCT scanner. Rev Esp Med Nucl Imagen Mol. 2013;32:13-21.
23. A. M. Morey, F. Noo and D. J. Kadrmas. Effect of using 2 mm voxels on observer performance for PET lesion detection. IEEE Trans Nucl Sci. 2016;63:1359-1366.
24. Øen SK, Aasheim LB, Eikenes L, Karlberg AM. Image quality and detectability in Siemens Biograph PET/MRI and PET/CT systems—a phantom study. EJNMMI Phys. 2019;6:16.
25. Bieszk JA, Hawman EG. Evaluation of SPECT angular sampling effects: continuous versus step-and-shoot acquisition. J Nucl Med. 1987;28:1308-1314.
26. Chen X, Zhou B, Xie H, Miao T, Liu H, Holler W, Lin MD, Miller EJ, Carson RE, Sinusas AJ, Liu C. DuDoSS: deep-learning-based dual-domain sinogram synthesis from sparsely sampled projections of cardiac SPECT. Med Phys. 2023;50:89–103.
27. Samei E, Krupinsky EA. Handbook of Medical Image Perception and Techniques. Cambridge University Press. 2018.
28. Wagner R F. Lessons from my dinners with the giants of modern image science. Radiat Prot Dosimetry. 2005;114:4-10.
29. Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. J Opt Soc Am A. 1992 9:649-658.
30. Vaishnav JY, Jung WC, Popescu LM, Zeng R, and Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. Med Phys. 2014;41:071904.
31. Burgess AE. Statistically defined backgrounds: performance of a modified nonprewhitening observer model. J Opt Soc Am A Opt Image Sci Vis. 1994;11:1237-42.
32. Gong H, Hu Q, Walther A, Koo CW, Takahashi EA, Levin DL, Johnson TF, Hora MJ, Leng S, Fletcher JG, McCollough CH, Yu L. Deep-learning-based model observer for a lung nodule detection task in computed tomography. J Med Imaging (Bellingham). 2020;7:042807.
33. Gifford HC, King MA, Pretorius PH and Wells RG. A comparison of human and model observers in multislice LROC studies. IEEE Trans Med Imaging. 2005;24:160-169.

34. Khurd P, Gindi G. Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers. Phys Med Biol. 2005;50:1519-32.
35. Murase K, Mukai W, Fukami S, Matsumoto K, Yang Z. Performance of model observers for lesion detection in SPECT imaging. Igaku Butsuri. 2006;26:108-17.
36. Ghaly M, Links JM, Frey E. Optimization of energy window and evaluation of scatter compensation methods in myocardial perfusion SPECT using the ideal observer with and without model mismatch and an anthropomorphic model observer. J Med Imaging (Bellingham). 2015;2:015502.
37. Elshahaby FE, Ghaly M, Jha AK, Frey EC. Factors affecting the normality of channel outputs of channelized model observers: an investigation using realistic myocardial perfusion SPECT images. J Med Imaging (Bellingham). 2016;3:015503.
38. Sen A, Gifford HC. Accounting for anatomical noise in search-capable model observers for planar nuclear imaging. J Med Imaging (Bellingham). 2016;3:015502.
39. Gifford HC. A visual-search model observer for multislice-multiview SPECT images. Med Phys. 2013;40:092505.
40. Gifford HC. Efficient visual-search model observers for PET. Br J Radiol. 2014;87:20140017.
41. Li Y, Chen J, Brown JL, Treves ST, Cao X, Fahey FH, Sgouros G, Bolch WE, Frey EC. DeepAMO: a multi-slice, multi-view anthropomorphic model observer for visual detection tasks performed on volume images. J Med Imaging (Bellingham). 2021;8:041204.
42. Hernández-Girón I, Geleijns J, Calzado A, Veldkamp WJH. Automated assessment of low contrast sensitivity for CT systems using a model observer. Med Phys. 2011;38:S25–35.
43. Hernández-Girón I, Calzado A, Geleijns J, Joemai RMS, Veldkamp WJH. Low contrast detectability performance of model observers based on CT phantom images: kVp influence. Phys Med. 2015;31:798–807.
44. Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. Med Phys. 2013;40:041908.
45. Leng S, Yu L, Zhang Y, Carter RE, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. Med Phys. 2013;40:081908.
46. Zhang Y, Leng S, Yu L, Carter RE, and McCollough CH. Correlation between human and model observer performance for discrimination task in CT. Phys Med Biol. 2014;59:3389–404.
47. Solomon JB and Samei E. Correlation between human detection accuracy and observer model-based image quality metrics in computed tomography. J Med Imaging' 2016;3:12.
48. Abbey CK and Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am A Opt Image Sci Vis. 2001;18:473–88.
49. Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, and Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. Phys Med Biol. 2015;60:R1–R75.
50. Favazza CP, Ferrero A, Yu L, Leng S, McMillan KL, and McCollough CH. Use of a channelized Hotelling observer to assess CT image quality and optimize dose reduction for iteratively reconstructed images. J Med Imaging (Bellingham). 2017;4:031213.
51. Dieckens D, Lavalaye J, Romijn L, Habraken J. Contrast-noise-ratio (CNR) analysis and optimisation of breast-specific gamma imaging (BSGI) acquisition protocols. EJNMMI Res. 2013;3:21.

52. Delcroix O, Bourhis D, Keromnes N, Robin P, Le Roux PY, Abgral R, Salaun PY, Querellou S. Assessment of Image Quality and Lesion Detectability With Digital PET/CT System. Front Med (Lausanne). 2021;8:629096.