

# Objective comparison of lesion detectability in low and medium-energy collimator iodine-123 mIBG images using a channelized Hotelling observer

Rebecca A Gregory<sup>1</sup>, Iain Murray<sup>1</sup>, Jonathan Gear<sup>1</sup>,  
Matthew D Aldridge<sup>2</sup>, Daniel Levine<sup>3</sup>, Lucy Fowkes<sup>3</sup>,  
Wendy A Waddington<sup>2</sup>, Sue Chua<sup>3</sup> and Glenn Flux<sup>1</sup>

<sup>1</sup> Department of Physics, the Royal Marsden NHS Foundation Trust and Institute of Cancer Research, Downs Road, Sutton SM2 5PT, UK

<sup>2</sup> UCL Institute of Nuclear Medicine and UCL Hospitals NHS Foundation Trust, 235 Euston Road, London NW1 2BU, UK

<sup>3</sup> Department of Nuclear Medicine and PET/CT, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM2 5PT, UK

E-mail: [rebecca.gregory@icr.ac.uk](mailto:rebecca.gregory@icr.ac.uk)

Received 18 September 2015, revised 13 October 2016

Accepted for publication 9 November 2016

Published 13 December 2016



CrossMark

## Abstract

Iodine-123 mIBG imaging is widely regarded as a gold standard for diagnostic studies of neuroblastoma and adult neuroendocrine cancer although the optimal collimator for tumour imaging remains undetermined. Low-energy (LE) high-resolution (HR) collimators provide superior spatial resolution. However due to septal penetration of high-energy photons these provide poorer contrast than medium-energy (ME) general-purpose (GP) collimators. LEGP collimators improve count sensitivity. The aim of this study was to objectively compare the lesion detection efficiency of each collimator to determine the optimal collimator for diagnostic imaging.

The septal penetration and sensitivity of each collimator was assessed. Planar images of the patient abdomen were simulated with static scans of a Liqui-Phil™ anthropomorphic phantom with lesion-shaped inserts, acquired with LE and ME collimators on 3 different manufacturers' gamma camera systems (Skylight (Philips), Intevo (Siemens) and Discovery (GE)). Two-hundred normal and 200 single-lesion abnormal images were created for



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

each collimator. A channelized Hotelling observer (CHO) was developed and validated to score the images for the likelihood of an abnormality. The areas under receiver-operator characteristic (ROC) curves,  $A_z$ , created from the scores were used to quantify lesion detectability. The CHO ROC curves for the LEHR collimators were inferior to the GP curves for all cameras. The LEHR collimators resulted in statistically significantly smaller  $A_z$ s ( $p < 0.05$ ), of on average  $0.891 \pm 0.004$ , than for the MEGP collimators,  $0.933 \pm 0.004$ . In conclusion, the reduced background provided by MEGP collimators improved  $^{123}\text{I}$  mIBG image lesion detectability over LEHR collimators that provided better spatial resolution.

Keywords: iodine-123, neuroblastoma, collimator, channelized Hotelling observer, lesion detectability

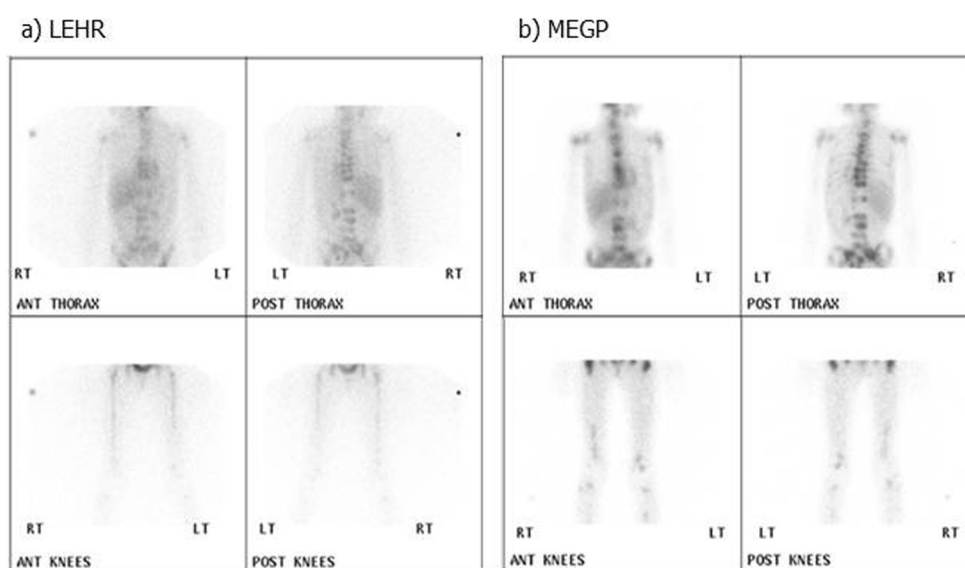
(Some figures may appear in colour only in the online journal)

## Introduction

Radionuclide imaging with Metaiodobenzylguanidine (mIBG) has a fundamental role in the diagnosis, staging and evaluation of treatment response in childhood neuroblastoma and adult neuroendocrine tumours (NETs) (Gelfand 1993, Shapiro *et al* 2001, Brisse *et al* 2011). Scintigraphy is routinely undertaken using  $^{123}\text{I}$ -mIBG in preference to  $^{131}\text{I}$ -mIBG, because better quality images can be obtained from 159 keV  $^{123}\text{I}$  gamma photons and twenty times more  $^{123}\text{I}$  activity can be administered for diagnostic scans (Shapiro *et al* 1987, Matthay *et al* 2010). Iodine-123 is also advocated for treatment planning prior to  $^{131}\text{I}$  mIBG therapy (Monsieurs *et al* 2002), despite the fact that the images obtained have been shown to be less sensitive in detecting lesions than post-therapy  $^{131}\text{I}$ -mIBG scans (Yang *et al* 2012). In response an alternative positron emission tomography tracer,  $^{124}\text{I}$ , has been used in attempts to improve diagnostic image quality and quantitative accuracy for dosimetry (Huang *et al* 2014, Koopmans *et al* 2014). As  $^{124}\text{I}$ -mIBG is not widely available and its clinical efficacy remains unproven,  $^{123}\text{I}$ -mIBG remains the standard.

Although  $^{123}\text{I}$  predominantly emits gamma photons at 159 keV (83% abundance), photons above 500 keV (2.3% abundance) are also emitted. These high energy photons penetrate the septa of low-energy (LE) collimators and down-scatter into the main photo peak energy window, producing background counts that reduce image contrast (Dobbeleir *et al* 1999). It was demonstrated almost 40 years ago that medium-energy general purpose (MEGP) collimators provide better signal detection than the a LE high-resolution (HR) collimators (Bolmsjo *et al* 1977). A number of groups now recommend the use of MEGP collimators for  $^{123}\text{I}$ -mIBG imaging based on a qualitative assessment (Macey *et al* 1986, Dobbeleir *et al* 1999, Snay *et al* 2011, Gelfand *et al* 2013). However the spatial resolution provided by a MEGP collimator is typically poorer than that provided by a LEHR collimator and its count sensitivity lower than its LE general-purpose (GP) counterpart. Consequently the choice of collimator may significantly impact upon image quality and ultimately interpretation of the  $^{123}\text{I}$ -mIBG study. Figure 1 demonstrates the difference in appearance of 10 min static scans for the same patients acquired with Philips' high-resolution and MEGP collimators.

Whilst there are well established guidelines on the use of  $^{123}\text{I}$ -mIBG imaging in neuroblastoma and NET cases, explicit guidance with regard to collimator use is lacking. The choice of collimator is left to the discretion of the imaging centre (Olivier *et al* 2003, Bombardieri *et al* 2010, Matthay *et al* 2010, Taieb *et al* 2012).



**Figure 1.** Static images of a 7 year old child with metastatic neuroblastoma deposits A acquired with a high-resolution collimator then B with a MEGP collimator.

The aim of this study was to determine which collimator optimises lesion detection on planar  $^{123}\text{I}$  mIBG images. The channelized Hotelling observer (CHO) model can be used to rapidly score larger sets of images than humans for the presence of an abnormality providing good statistical quality receiver operator characteristic (ROC) curves, that can be used to assess the efficacy of lesion detection for different imaging methods. This model has been shown to correlate well with human observers in lesion detection tasks and has been widely used in the optimization of several nuclear medicine imaging techniques (Hanley *et al* 1982, Gifford *et al* 2000, Gifford *et al* 2005, He *et al* 2010, El Fakhri *et al* 2011, Bal *et al* 2014, Yang *et al* 2014). In this study our implementation of the CHO model was first validated against human observations of 100 images for 2 collimators. The validated model was then applied to score 400 test images generated from 8 separate collimators from 3 different manufacturers. ROC curves were produced from these scores for each collimator. The areas under the ROC curves were used to provide an objective measure of the lesion detectability. These values were compared to objectively demonstrate the collimator that optimised lesion detection.

## Methods

Collimator design varies widely between manufacturers. Therefore collimators on a Discovery NMCT (GE), a Skylight gamma camera (Philips) and a Symbia Intevo Excel SPECT/CT (Siemens) were assessed (table 1). It should be noted that the Skylight camera had thicker crystals (5/8") than the Discovery and Intevo (3/8") models.

### Collimator characterisation

Each collimator's septal penetration fraction and system sensitivity were measured for  $^{123}\text{I}$  imaging in accordance with the NEMA standards (NEMA 2007).

**Table 1.** Collimator characteristics with NEMA septal penetration and system sensitivity.

Camera	Collimator	Hole size <sup>a</sup> (mm)	Septal thickness <sup>a</sup> (mm)	Length (mm)	Spatial resolution (mm) <sup>a</sup>	Septal penetration %	System sensitivity (cps/MBq) $\pm$ error <sup>b</sup>
Discovery (GE)	MEGP	3.0	1.05	58.0	9.4	0.0	70.3 $\pm$ 1.2
	LEHR	1.50	0.20	35.0	7.4	0.2	96.6 $\pm$ 1.7
Skylight (Philips)	MEGP	2.95	1.14	48.0	11.3	0.0	101.1 $\pm$ 1.9
	LEHR	1.40	0.15	32.8	7.4	1.1	109.3 $\pm$ 2.1
	LEGP	1.40	0.18	24.7	8.8	5.8	204.3 $\pm$ 3.9
Intevo (Siemens)	MELP	2.94	1.14	40.64	12.5	0.0	115.6 $\pm$ 0.5
	LEHR	1.11	0.16	24.05	7.5	16.3	185.9 $\pm$ 0.9
	LEAP	1.45	0.2	24.05	9.4	12.0	225.9 $\pm$ 1.0

MEGP/LP = medium-energy general purpose/low penetration, LEG/AP = low-energy general/all purpose and LEHR = low-energy high-resolution.

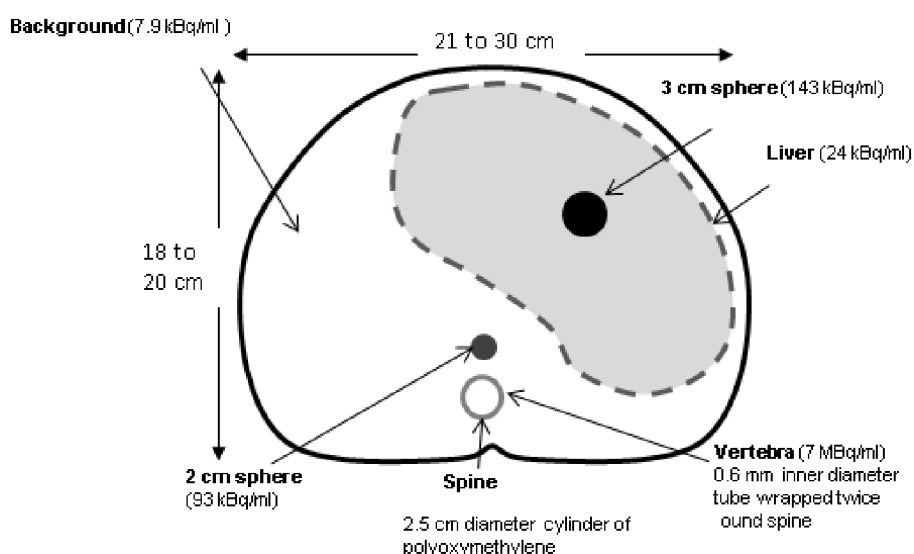
<sup>a</sup> Spatial resolution is supplied for collimated photons, as the FWHM at 10cm from the collimator external surface in the manufacturers specification, for a 3/8" crystal. The thicker (5/8") Skylight crystal will provide poorer spatial resolution, however this value is not available from the manufacturer.

<sup>b</sup> Error propagated from the uncertainty in the measured activity and square root of the region of interest counts, used in the NEMA calculation of sensitivity.

#### *Physiological background and lesion present patient image simulation with anthropomorphic phantoms*

Images were created using an anthropomorphic phantom to simulate the clinical challenge of lesion detection for each collimator/camera combination.

The Liqui-Phil<sup>TM</sup> anthropomorphic abdomen-shaped phantom was filled with 9585 ml inactive water. The dimensions of this phantom, indicated in figure 2, correspond to an older child or adult. The phantom contained an inactive 2.5 cm diameter polyoxymethylene cylinder to simulate the attenuation of the spine. Lesion-mimicking-inserts filled with <sup>123</sup>I were positioned in the phantom at the positions shown in figure 2. These were a 2 cm diameter sphere adjacent to the anterior surface of the spine, a 3 cm diameter sphere near the centre of the liver and (1 mm internal diameter) Polyethylene tubing wrapped twice around the spinal cylinder to represent a single mIBG avid vertebra. Three separate anterior and posterior pairs of planar scans were acquired with the abdomen-shaped phantom containing only one lesion at a time. In addition the Liqui-Phil<sup>TM</sup> liver insert filled with <sup>123</sup>I was imaged separately within the abdomen-shaped phantom. Finally <sup>123</sup>I was added to the abdomen-shaped phantom itself and it was imaged independently to produce an abdominal background image. The activity concentrations used for all inserts and the background are shown in figure 2. The phantom was positioned on the standard imaging couch. Planar scans were acquired with each collimator on each gamma camera with a 20% wide energy window centred on 159 keV, a 256  $\times$  256 matrix and uniformity correction. The uniformity correction on the Skylight was based on intrinsic <sup>123</sup>I maps, to correct for photon detector tube edge enhancement artefacts. No such artefacts were observed in these phantom images. This was not necessary for the other 2 systems and an extrinsic Cobalt-57 uniformity map was used. The pixel dimensions were 4.6  $\pm$  0.2 mm<sup>2</sup> as the detector field-of-view varies for each camera. The Siemens Intevo has the largest 614 mm FOV, the Philips Skylight has a 597 mm FOV and the GE Discovery has the smallest 565 mm



**Figure 2.** A cross-section of the abdomen-shaped phantom. The positions and activity concentrations of the fillable inserts are shown. These were scanned individually within the abdomen shaped shell alongside the spine insert. The shell dimensions vary within the range indicated along the axial length of the phantom.

FOV. The collimator faces were positioned as close as possible to the phantom posterior and anterior surfaces. The objects were scanned for as long as possible depending on each camera's availability. The frame duration used for each collimator image set was adjusted to account for the physical decay of the  $^{123}\text{I}$  between each image set, so that the image count statistics remained comparable. Therefore these high-count images reflect the relative variation in sensitivity, spatial resolution and septal penetration between the collimators. A minimum of 3k counts was collected for the inserts and 347k counts for the abdomen. For each image set acquired the abdomen-shaped phantom position was accurately reproduced using markings on the camera couch.

The acquired images were combined using an interactive data language (IDL) (release 8.2.3, ITT Visual Information Solutions) user script. This program first reduced the liver and abdomen image counts to values representing clinical image frame times between 400 and 600 s. The count in each pixel was then randomly selected from a Poisson distribution, with a mean value set to the rescaled count. A different random seed was used to create 200 different statistical realisations of the liver and abdomen. These images were then summed to produce 200 background images per collimator of the liver and physiological abdominal background images with liver-to-background ratios (LBR) ranging from 3.6 to 5.6.

For each collimator 200 images containing a low-contrast single-lesion were created by adding lesion-insert-images to the background images. The lesion-insert-image counts were rescaled prior to Poisson resampling to simulate varied tumour-to-background ratios (TBR). The tumour images were also shifted axially to simulate a variety of lesion locations. The resampled-shifted-lesion-insert images were then summed to the normal scans. In this way 17 images with the 3 cm sphere in the liver, 79 with the 2 cm sphere next to the spine and 104 images with a vertebra were created, all with different TBRs. The resulting liver-lesion TBRs ranged from 2.8 to 7.2, the spine-lesion TBRs were 29 to 255 and the vertebra TBRs were 179–1785. The vertebra required a high TBR to be visible. The range of simulated activity

**Table 2.** Simulated activity concentration ranges in the resampled phantom images.

Object	Simulated Range (kBq ml <sup>-1</sup> )
Abdomen	0.53–0.79
Liver	1.6–2.4
3 cm sphere in liver	36–143
2 cm sphere by spine	23–202
Vertebra	1000–7000

concentrations are shown in table 2. These lesion-to-background activity concentration ratios and positions were chosen to provide a broad range of challenging lesion detection tasks, from clearly visible to barely perceivable lesions.

A total of 3200 posterior images were used for CHO scoring. The resultant average signal-to-noise ratios (SNR) in 3 cm circular ROIs in the hottest part of the liver and coldest part of the background were compared to those measured in 6 recent patient scans. Three of these 10 min static patient scans were acquired with MEGP collimators and 3 were acquired with LEHR collimators. The SNR was calculated as the mean divided by the standard deviation in the mean ROI counts.

#### Channelized Hotelling observer

A mathematical model, the channelized Hotelling observer (CHO) was used to score the images for the possibility of an abnormality. The CHO model was written in IDL following the method published by Shidahara *et al* (2006). Test statistics were produced from the Fourier transformed images,  $\mathbf{F}(\rho)$ , filtered into different frequency channels. These channels have a psychophysiological basis in reflecting the frequency selective channels of the human visual system. Radially symmetric channels were applied in the frequency domain, described by;

$$u_c(\rho) = \begin{cases} 1 & \|\rho\| \in [\rho_0 2^{c-1}, \rho_0 2^c] \\ 0 & \text{Otherwise} \end{cases}, \quad (1)$$

where  $c \in \{1, 2, 3, 4, 5\}$   $\rho_0$  pixel<sup>-1</sup> is the low end cut-off frequency. The average value in each channel was calculated to create a matrix of values,  $r_i$ , for each of the 5 channels,  $i$ ,  $\mathbf{r} = [r_1, r_2, r_3, r_4, r_5]$  for each image  $f$ . The difference of the means of the channel matrices per camera-collimator combination, for the lesion present  $\bar{\mathbf{r}}_a$  and lesion absent  $\bar{\mathbf{r}}_n$  channel matrices were used to define a matching filter  $\bar{\mathbf{r}}_a - \bar{\mathbf{r}}_n$ .

Pre-whitening was used to de-correlate the noise in the channels by multiplication with  $\mathbf{K}^{-1}$ , which is the inverse of the average of the 2 noise covariance matrices for the entire abnormal and normal ensemble. Then a test statistic  $\lambda_f$  for each image was calculated as;

$$\lambda_f = [\bar{\mathbf{r}}_a - \bar{\mathbf{r}}_n]^T \mathbf{K}^{-1} \mathbf{r}. \quad (2)$$

In effect CHO provides a score for the possibility of a lesion presence, as the weighted sum of the counts in the frequency channels of the tested image.

#### Channelized Hotelling observer validation

In order to validate the CHO model, 4 experienced human observers scored 240 of the posterior phantom images for the presence of a lesion, using the scoring system in table 3. The observers were 2 medical physicists and 2 radiologists all with over 10 years' experience at interpreting nuclear medicine scans. The images consisted of 60 of the background

**Table 3.** Human observer scoring system.

Score	Decision
1	Definitely normal
2	Probably normal
3	Equivocal
4	Maybe abnormal
5	Definitely abnormal

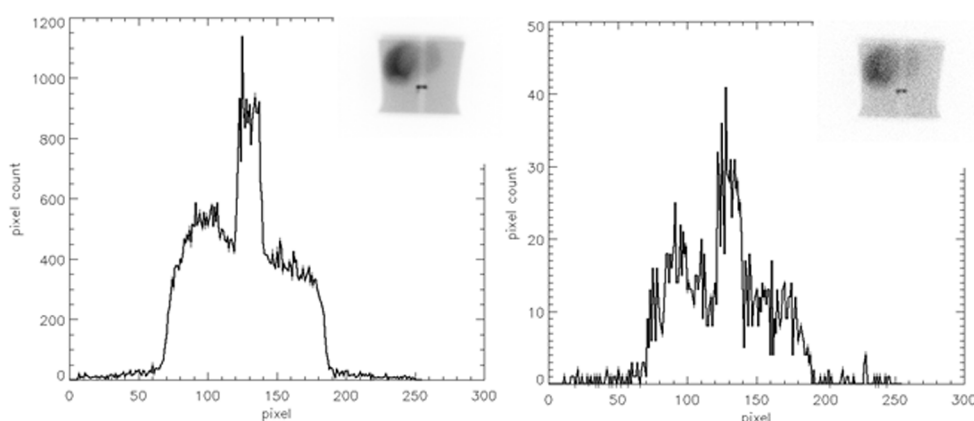
and 60 images containing lesion inserts generated for each of the Philips LEHR and MEGP collimators. These were viewed on the Hermes work station (Hermes Medical Solutions) that all the observers were familiar with. The images from the different collimators were randomly interspersed so the observers were not informed of the collimator used. The first 20 images from each collimator were used as training sets as the observer learnt the scoring system and adjusted their own criteria to apply it. The following 100 images were used to form ROC curves. For each score for each of the background and lesion present images, the fraction of the scores equal to and above a threshold score was calculated. The threshold that was used to assign an image as positive or negative for the presence of a lesion, was incremented from the lowest to the highest score (1–5). The true positive fraction (TPF) was calculated from the lesion present images and the false positive fraction (FPF) was calculated from the background images. The average TPF and FPF for each human score threshold was used to create an average human ROC curve against which the CHO curves were compared.

CHO was used to score the same 100 (non-training) images for each collimator. CHO scores were binned into 12 score thresholds equally covering the range of scores generated by CHO for each data set. This gave comparable discrete FPF sampling frequency for the CHO and human ROC curves for validation. The area under these ROC curves,  $A_z$ , was then calculated using a trapezoidal method. Standard deviations on the data points were calculated according to Metz (Metz 1978). These standard deviations were propagated to estimate errors on the area under the ROC curves. The CHO cut-off frequency was adjusted to achieve the best agreement between the CHO and average human observer  $A_z$ . The validated cut-off frequency was used for the further comparison of collimators.

#### *Collimator lesion detectability comparison*

The full set of 3200 posterior anthropomorphic phantom images was scored using the validated CHO model with the optimised cut-off frequency. The anterior images were discarded as neither observer was able to identify the lesions in the images leading to ineffectual ROC curves with an  $A_z$  of 0.5.

The 400 CHO scores for each collimator were used to generate ROC curves for near continuous thresholds using ROckit 0.9B (Metz 1998). This software also fitted a binormal distribution to the ROC curve, calculated  $A_z$ , compared the correlated ROC curves and provided a two-tailed p-value to quantify the statistical significance of the difference between them. It had not been possible to apply this software to the human scores due to limited number of data points associated with the 5 human scores.



**Figure 3.** An example of a high count image prior to Poisson resampling left, with a profile through the insert image overlaying the spine to mimic one active vertebra. The image on the right is the same geometry following Poisson resampling.

## Results

### *Collimator characterisation*

Table 1 shows the NEMA test results alongside the other collimator properties. No septal penetration was found for any of the medium-energy collimators. The long thick septa of these collimators appeared to stop the high energy photons. However the ME collimators were the least sensitive for each camera. The LEGP collimators were the most sensitive and gave double the sensitivity of the MEGP collimators but with over 5% septal penetration. The most septal penetration (16.3%) occurred for the Siemens LEHR collimator which has the shortest thin septa of all the collimators tested. Star artefacts were evident in the images from this collimator. The septal penetration of the Siemens' LEHR collimator was 15 times higher than that of the Philips, due to the shorter Siemens' collimator holes. The GE collimators provided the lowest septal penetration of the three manufacturers.

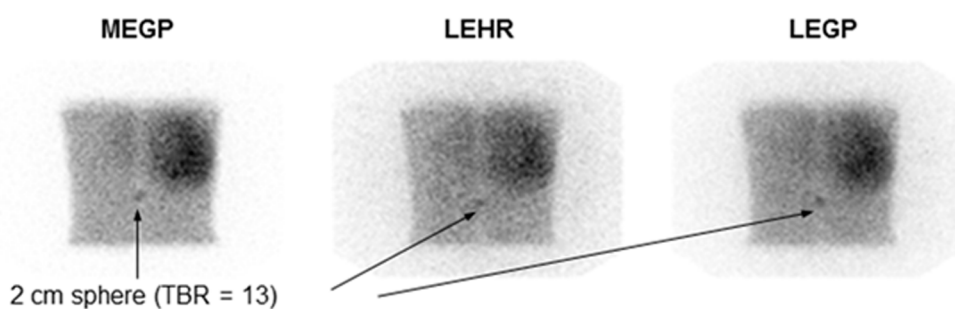
The sensitivity between camera collimators could not be compared because the thicker 5/8" crystal of the Skylight (Philips) will bias these sensitivity results.

### *Physiological Background and Lesion Present Patient Image Simulation with Anthropomorphic Phantoms*

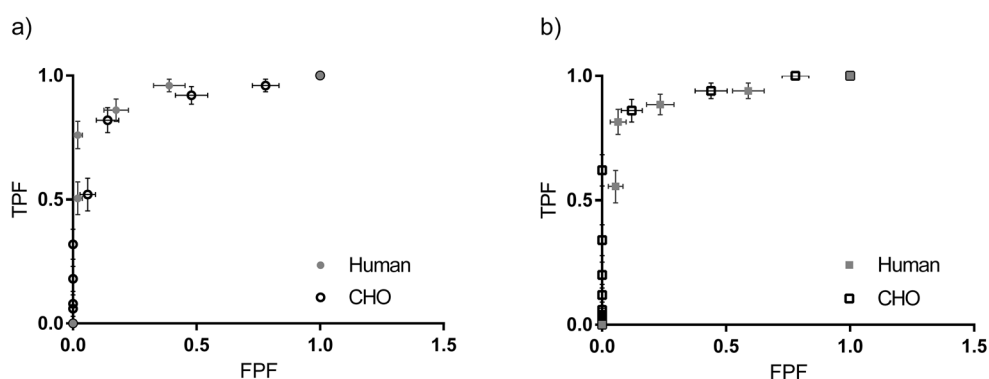
Images before and after Poisson resampling are shown in figure 3. These are inset on a plot of a single pixel wide profile across the vertebra insert.

The resultant simulated total image counts ranged from 245k to 378k counts, which are typical of those acquired from 10 min abdominal mIBG planar scans. Examples of the generated images are shown in figure 4. The average pixel count in the background of the 6 patient scans was  $20 \pm 5$  compared to  $15 \pm 5$  in the background of the phantom scans. The average liver counts were  $45 \pm 7$  and  $30 \pm 8$  in the patient and phantom images respectively. The average liver-to-background ratio (LBR) counts were  $2.26 \pm 0.89$  and  $1.96 \pm 1.59$  in the patient and phantom images respectively. The SNR in the patient images background was  $4.16 \pm 0.59$  and the liver was  $6.49 \pm 1.38$ . In the phantom images the SNRs were  $3.47 \pm 0.36$  in the background and  $4.73 \pm 0.44$  in the liver. All errors on these values are the standard deviations in the means. The count densities and SNRs were therefore lower in the phantom than the patient images, but a similar order of magnitude.





**Figure 4.** Example abnormal posterior images from the Skylight (Philips) gamma camera, corresponding to each of the available collimators.



**Figure 5.** CHO with a 0.008/cycle cut-off frequency versus the average human ROC curves for (a) the LEHR and (b) MEGP posterior images generated for the Philips Skylight gamma camera. The errors on the human curves are the standard deviation in the FPFs and TPFs.

#### *Channelized Hotelling observer validation*

The least difference between the human observer and CHO ROC  $A_z$  scores was observed at a low end cut-off frequency of  $0.008 \text{ pixel}^{-1}$ . This was the lowest achievable frequency at 2 per 256 pixels. This value of  $\rho_0$  was used for all subsequent CHO experiments. The validation ROC curves are shown in figure 5, the areas under these ROC curves are shown in table 4. The values are within the standard errors of one another.

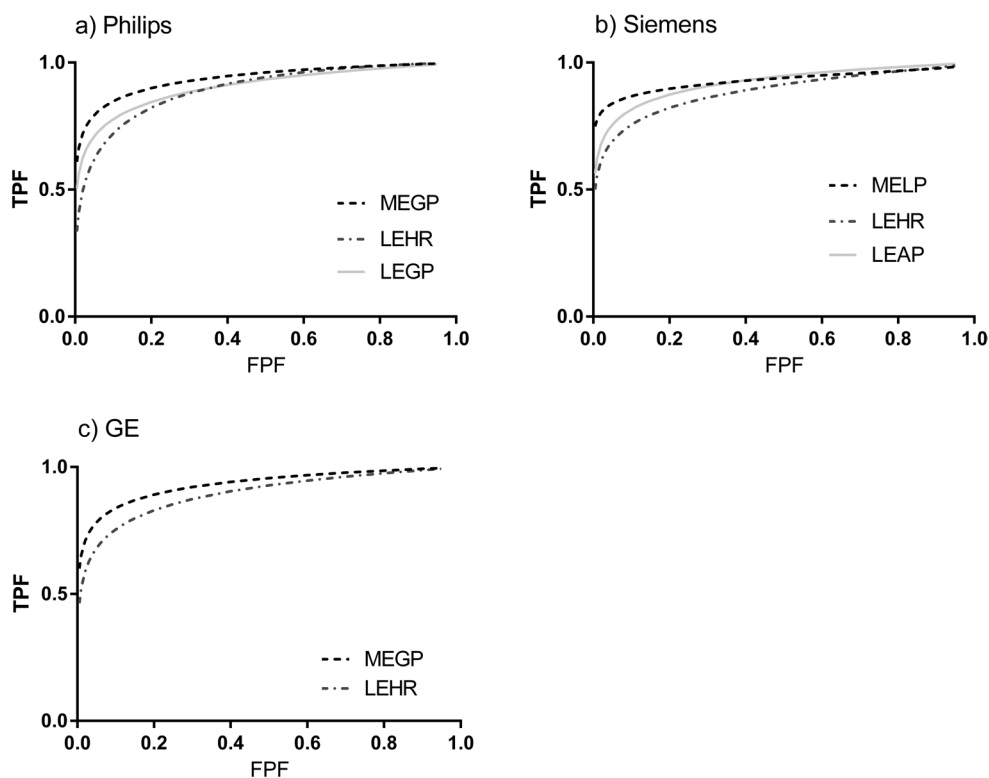
#### *Collimator lesion detectability comparison using receiver operator characteristic curves*

The ROC curves generated using CHO from 200 normal and 200 abnormal images for each collimator are shown in figure 6. The LEHR ROC curves consistently demonstrated lower TPF values than for the GP curves indicating poorer lesion detectability with LEHR collimators than for the other collimators.

The areas under these curves are given in table 5. On average the LEHR collimators resulted in smaller  $A_z$ s of  $0.891 \pm 0.004$  than  $0.933 \pm 0.004$  for the MEGP collimators. This difference was statistically significant ( $p < 0.05$ ) see table 6. There was no statistically significant difference between the Philips LEGP and LEHR collimators. However there was a statistical difference between the  $A_z$  for the equivalent Siemens collimators. In contrast there

**Table 4.** CHO Validation (120 images for each collimator): areas under the ROC curves (figure 5)  $A_z \pm$  standard error.

Collimator	Human	CHO
MEGP	$0.89 \pm 0.11$	$0.93 \pm 0.12$
LEHR	$0.92 \pm 0.11$	$0.87 \pm 0.12$

**Figure 6.** The binormal fit to the ROC curves for each manufacturer and tested collimator.

was a statistically significant difference between the Philips MEGP and LEGP collimators but not for the corresponding Siemens collimators.

## Discussion

Eight collimators from the three major manufacturers were characterised and assessed for  $^{123}\text{I}$  imaging in this study. These cover the majority of the collimators currently used for  $^{123}\text{I}$  imaging. Collimators designed specifically for  $^{123}\text{I}$  imaging are not in wide use and were not available for this study.

Septal penetration was measured for all the low-energy collimators. The Siemens LEHR collimators gave the highest levels of 16.3% septal penetration. At acceptance testing of the Intevo system (Siemens) it was noted that this collimator similarly suffered from high (9.4%) septal penetration with Tc-99m. This may be expected as the Siemens LEHR collimators have the smallest holes and shortest thin septa of the collimators tested.

**Table 5.** CHO collimator comparison (400 images for each collimator): areas under the ROC curves (figure 6)  $A_z \pm$  standard error.

Manufacturer	Collimator	$A_z$
GE	MEGP	$0.932 \pm 0.013$
	LEHR	$0.894 \pm 0.016$
Philips	MEGP	$0.937 \pm 0.012$
	LEHR	$0.893 \pm 0.016$
	LEGP	$0.903 \pm 0.016$
Siemens	MELP	$0.929 \pm 0.015$
	LEHR	$0.887 \pm 0.017$
	LEAP	$0.920 \pm 0.014$

**Table 6.** Two tailed  $p$ -values of correlation between collimators  $A_z$ .

Manufacturer	Collimators	$p$ -value
GE	MEGP versus LEHR	0.029
Philips	MEGP versus LEHR	0.012
	LEHR versus LEGP	0.657
	MEGP versus LEGP	0.039
Siemens	MELP versus LEHR	0.005
	LEHR versus LEAP	0.011
	MELP versus LEAP	0.563

The MEGP collimators allowed no measurable septal penetration, therefore the count rates did not vary with source-collimator distance. Future work will investigate the optimal protocol for quantitative imaging to provide standard uptake values for early assessment of treatment response and dosimetry calculations. Quantification should be simpler using MEGP collimators as the counts did not vary with source-collimator distance. The count rates for the low-energy collimators did vary with this distance due to septal penetration, so the quantitative accuracy for these images would vary with the activity distribution. Therefore benefits from using MEGP over low-energy collimators for quantitative imaging are expected and have been demonstrated for cardiac imaging (Inoue *et al* 2003).

The areas under the CHO and human ROC curves agree within the standard errors of each other. Therefore the CHO model was successfully validated against human scoring (figure 5 and table 4). The smallest possible low end cut-off frequency ( $0.008 \text{ pixels}^{-1}$ ) for a 256 matrix gave the best agreement between the CHO and human observer areas under the ROC curves,  $A_z$ . This corresponds to scoring images at the lowest usable 5 frequency bands, where statistical noise is effectively filtered. Therefore the variations in count concentrations correspond to structural rather than statistical changes. The human observer would also concentrate on these lower frequencies to identify structures, such as lesions. Therefore this finding is as expected.

The areas under each of the collimator's human ROC curves were similar at  $0.92 \pm 0.11$  and  $0.89 \pm 0.11$  for the MEGP and LEHR collimators respectively. However ROCKit could not be used to fit a ROC curve to the human's scores, due the limited number of data points. The ROC curve trapezoidal  $A_z$  was therefore calculated manually. These curves were not compared to assess the statistical significance of the difference in  $A_z$  between the collimators from human scores. However the overlapping error intervals on the areas under the ROC curves in table 4 indicate the difference in  $A_z$  between collimators is insignificant.

The ROC curves for CHO achieved from large image sets (200) of images per collimator provided improved statistical quality ROC curves compared to that attained with an ensemble of 50 images and human scores. The low errors on the  $A_z$  from these ROC curves allow demonstration of the significant differences in  $A_z$  for some of the collimators with confidence in 2-tailed  $p$ -values less than 0.05 (table 6). The variations in collimator and couch designs and pixel size between the manufacture's systems led to a 5% range in  $A_z$ , of 0.05 compared to the average  $A_z$  of 0.91.

CHO scores are based on the counts within the low frequency channels. The CHO ROC curves obtained from the MEGP collimators were consistently highest due to the larger number of low frequency counts in the MEGP images corresponding to lesions (figure 6). Septal penetration causes a higher proportion of counts from the lesions to appear at higher frequencies within the LEHR collimator images, which additionally increases the image noise and reduces image contrast. Therefore the areas under the ROC curves (table 5) objectively demonstrate improved lesion detectability with MEGP compared to LEHR collimators, due to these factors. These results are in agreement with published findings that MEGP collimators demonstrate improved  $^{123}\text{I}$  image quality (Bolmsjo *et al* 1977, Snay *et al* 2011, Gelfand *et al* 2013). The improvement in lesion visibility with general purpose collimators is also evident in the example images of figure 3.

A limitation of this work was that test data were necessarily generated from phantom images, as it is impractical to acquire 200 patient images of normal physiology and 200 images with a single lesion on multiple systems. The solution in this study was to simulate these using Poisson resampled phantom images. The generated images in figure 3 have similar appearance to the abdominal region of figure 1. Although the count densities and SNRs are lower in the phantom-based simulations than in the average patient images, they are the right order of magnitude and therefore realistic. The liver-to-background ratios were varied to simulate the physiological differences in patients and were on average similar to those measured in a sample of clinical scans. However using the phantoms available for this study it was not possible to simulate variations in patient geometries. Therefore the training images were similar to those used to produce the ROC curves.

This assessment was performed for static scans as lesions tend to be identified on planar images and these images may be used for disease staging alone. If SPECT images are used the lesions are first located in the planar scans to identify the region where the SPECT scan is performed. These 3D images are then generally used to confirm the lesion location after it has been detected (Barwick *et al* 2010). This work may be extended to optimise the many acquisition and reconstruction parameters for SPECT imaging to improve lesion detection and therefore localisation.

## Conclusions

In this study a channelized Hotelling observer model was developed, validated and used to objectively quantify lesion detectability as the area under ROC curves. This demonstrated that the benefit of reduced background noise, provided by the general-purpose collimators, outweighs the benefit of the superior spatial resolution of LEHR collimators for mIBG avid lesion detection using  $^{123}\text{I}$  mIBG planar images.

## Acknowledgments

We acknowledge the Cancer Research UK Centre for Drug Development and Rising Tide funding and NHS funding to the NIHR Biomedical Research Centre and Clinical Research Facility in Imaging at The Royal Marsden and Institute of Cancer Research.

## References

- Bal H, Guerin L, Casey M E, Conti M, Eriksson L, Michel C, Fanti S, Pettinato C, Adler S and Choyke P 2014 Improving PET spatial resolution and detectability for prostate cancer imaging *Phys. Med. Biol.* **59** 4411–26
- Barwick T, Murray I, Megadmi H, Drake W M, Plowman P N, Akker S A, Chew S L, Grossman A B and Avril N 2010 Single photon emission computed tomography (SPECT)/computed tomography using Iodine-123 in patients with differentiated thyroid cancer: additional value over whole body planar imaging and SPECT *Eur. J. Endocrinol.* **162** 1131–9
- Bolmsjo M S, Persson B R and Strand S E 1977 Imaging 123I with a scintillation camera. A study of detection performance and quality factor concepts *Phys. Med. Biol.* **22** 266–77
- Bombardieri E *et al* 2010 I-131/I-123-Metaiodobenzylguanidine (mIBG) scintigraphy: procedure guidelines for tumour imaging *Eur. J. Nucl. Med. Mol. Imaging* **37** 2436–46
- Brisse H J *et al* 2011 Guidelines for imaging and staging of neuroblastic tumors: consensus report from the International Neuroblastoma Risk Group Project *Radiology* **261** 243–57
- Dobbeleir A A, Hambye A S E and Franken P R 1999 Influence of high-energy photons on the spectrum of iodine-123 with low- and medium-energy collimators: consequences for imaging with I-123-labelled compounds in clinical practice *Eur. J. Nucl. Med.* **26** 655–8
- El Fakhri G, Surti S, Trott C M, Scheuermann J and Karp J S 2011 Improvement in lesion detection with whole-body oncologic time-of-flight PET *J. Nucl. Med.* **52** 347–53
- Gelfand M, Sharp S and Weiss B 2013 Collimator choice for I-123-MIBG imaging in children undergoing treatment with I-131-MIBG *J. Nucl. Med. Mol. Imaging* **54** 2007
- Gelfand M J 1993 Metaiodobenzylguanidine in children *Semin. Nucl. Med.* **23** 231–42
- Gifford H C, King M A, de Vries D J and Soares E J 2000 Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging *J. Nucl. Med.* **41** 514–21
- Gifford H C, King M A, Pretorius P H and Wells R G 2005 A comparison of human and model observers in multislice LROC studies *IEEE Trans. Med. Imaging* **24** 160–9
- Hanley J A and McNeil B J 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve *Radiology* **143** 29–36
- He X, Links J M and Frey E C 2010 An investigation of the trade-off between the count level and image quality in myocardial perfusion SPECT using simulated images: the effects of statistical noise and object variability on defect detectability *Phys. Med. Biol.* **55** 4949–61
- Huang S Y, Bolch W E, Lee C, Van Brocklin H F, Pampaloni M H, Hawkins R A, Sznewajs A, DuBois S G, Matthay K K and Seo Y 2014 Patient-specific dosimetry using pretherapy [<sup>125</sup>I]m-iodobenzylguanidine ([<sup>125</sup>I]mIBG) dynamic PET/CT imaging before [<sup>125</sup>I]mIBG targeted radionuclide therapy for neuroblastoma *Mol. Imaging Biol.* **17** 284–94
- Inoue Y, Suzuki A, Shirouzu I, Machida T, Yoshizawa Y, Akita F, Ohnishi S, Yoshikawa K and Ohtomo K 2003 Effect of collimator choice on quantitative assessment of cardiac iodine 123 MIBG uptake *J. Nucl. Cardiol.* **10** 623–32
- Koopmans K P and Glaudemans A W 2014 Other PET tracers for neuroendocrine tumors *PET Clin.* **9** 57–62
- Macey D J, Denardo G L, Denardo S J and Hines H H 1986 Comparison of low-energy and medium-energy collimators for spect imaging with iodine-123-labeled antibodies *J. Nucl. Med.* **27** 1467–74 (PMID: 3489084)
- Matthay K K, Shulkin B, Ladenstein R, Michon J, Giammarile F, Lewington V, Pearson A D J and Cohn S L 2010 Criteria for evaluation of disease extent by I-123-metaiodobenzylguanidine scans in neuroblastoma: a report for the International Neuroblastoma Risk Group (INRG) Task Force *Brit. J. Cancer* **102** 1319–26
- Metz C E 1978 Basic principles of ROC analysis *Semin. Nucl. Med.* **8** 283–98
- Metz C E 1998 Metz ROC Software <http://metz-roc.uchicago.edu/MetzROC/software>

- Monsieurs M, Brans B, Bacher K, Dierckx R and Thierens H 2002 Patient dosimetry for <sup>131</sup>I-MIBG therapy for neuroendocrine tumours based on <sup>123</sup>I-MIBG scans *Eur. J. Nuc. Med. Mol. Imaging* **29** 1581–7
- NEMA 2007 *NEMA Standards Publication NU 2-2007 Performance Measurements of Gamma Cameras* (Rosslyn, VA: National Electrical Manufacturers Association)
- Olivier P *et al* 2003 Guidelines for radioiodinated MIBG scintigraphy in children *Eur. J. Nuc. Med. Mol. Imaging* **30** BP45–50
- Shapiro B and Gross MD 1987 Radiochemistry, biochemistry, and kinetics of <sup>131</sup>I-metaiodobenzylguanidine (MIBG) and <sup>123</sup>I-MIBG: clinical implications of the use of <sup>123</sup>I-MIBG *Med. Pediatr. Oncol.* **15** 170–7
- Shapiro B, Gross M D and Shulkin B 2001 Radioisotope diagnosis and therapy of malignant pheochromocytoma *Trends Endocrinol. Metab.* **12** 469–75
- Shidahara M *et al* 2006 Predicting human performance by channelized Hotelling observer in discriminating between Alzheimer's dementia and controls using statistically processed brain perfusion SPECT *Ann. Nucl. Med.* **20** 605–13
- Snay E R, Treves S T and Fahey F H 2011 Improved quality of pediatric <sup>123</sup>I-MIBG images with medium-energy collimators *J. Nucl. Med. Technol.* **39** 100–4
- Taieb D *et al* 2012 EANM 2012 guidelines for radionuclide imaging of pheochromocytoma and paraganglioma *Eur. J. Nuc. Med. Mol. Imaging* **39** 1977–95
- Yang J, Codreanu I, Servaes S and Zhuang H 2012 I-<sup>131</sup>I MIBG post-therapy scan is more sensitive than I-<sup>123</sup>I MIBG pretherapy scan in the evaluation of metastatic neuroblastoma *Nucl. Med. Commun.* **33** 1134–7
- Yang L, Zhou J, Ferrero A, Badawi R D and Qi J 2014 Regularization design in penalized maximum-likelihood image reconstruction for lesion detection in 3D PET *Phys. Med. Biol.* **59** 403–19