Physics in Medicine & Biology

PAPER

A projection image database to investigate factors affecting image quality in weight-based dosing: application to pediatric renal SPECT

To cite this article: Ye Li *et al* 2018 *Phys. Med. Biol.* **63** 145004

IOP Publishing Phys. Med. Biol. 63 (2018) 145004 (13pp) https://doi.org/10.1088/1361-6560/aacbf0

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
26 January 2018

REVISED
23 May 2018

ACCEPTED FOR PUBLICATION
12 June 2018

PUBLISHED
9 July 2018

A projection image database to investigate factors affecting image quality in weight-based dosing: application to pediatric renal SPECT

Ye Li[1,2], Shannon O'Reilly[3], Donika Plyku[2], S Ted Treves[5,6], Yong Du[2], Frederic Fahey[4,6], Xinhua Cao[4,6], Abhinav K Jha[7,8], George Sgouros[2,9], Wesley E Bolch[3] and Eric C Frey[1,2,9]

[1] Department of Electrical and Computer Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, United States of America
[2] The Russell H Morgan Department of Radiology and Radiological Science, School of Medicine, Johns Hopkins University, Baltimore, MD 21287, United States of America
[3] J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States of America
[4] Department of Radiology, Boston Children's Hospital, Boston, MA, United States of America
[5] Department of Radiology, Brigham and Women's Hospital, Boston, MA, United States of America
[6] Department of Radiology, Harvard Medical School, Boston, MA, United States of America
[7] Department of Biomedical Engineering, Washington University in St. Louis, MO, United States of America
[8] Mallinckrodt Institute of Radiology, Washington University in St. Louis, MO, United States of America
[9] Sidney Kimmel Comprehensive Cancer Center, School of Medicine Johns Hopkins University, MD 21287, United States of America

E-mail: yli192@jhu.edu

Keywords: dose reduction/optimization, dosing guidelines, DMSA, pediatric imaging, SPECT, task-based image quality

Abstract

Balancing the tradeoff between radiation dose, acquisition duration and diagnostic image quality is essential for medical imaging modalities involving ionizing radiation. Lower administered activities to the patient can reduce absorbed dose, but can result in reduced diagnostic image quality or require longer acquisition durations. In pediatric nuclear medicine, it is desirable to use the lowest amount of administered radiopharmaceutical activity and the shortest acquisition duration that gives sufficient image quality for clinical diagnosis. However, diagnostic image quality is a complex function of patient factors including body morphometry. In this study, we present a digital population of 90 computational anatomic phantoms that model realistic variations in body morphometry and internal anatomy. These phantoms were used to generate a large database of projection images modeling pediatric SPECT imaging using a ^{99m}Tc-DMSA tracer. We used an analytic projection code that models attenuation, spatially varying collimator-detector response, and object-dependent scatter to generate the projections. The projections for each organ were generated separately and can be subsequently scaled by parameters extracted from a pharmacokinetics model to simulate realistic tracer biodistribution, including variations in uptake, inside each relevant organ or tissue structure for a given tracer. Noise-free projection images can be obtained by summing these individual organ projections and scaling by the system sensitivity and acquisition duration. We applied this database in the context of ^{99m}Tc-DMSA renal SPECT, the most common nuclear medicine imaging procedure in pediatric patients. Organ uptake fractions based on literature values and patient studies were used. Patient SPECT images were used to verify that the sum of counts in the simulated projection images was clinically realistic. For each phantom, 384 uptake realizations, modeling random variations in the uptakes of organs of interest, were generated, producing 34 560 noise-free projection datasets (384 uptake realizations times 90 phantoms). Noisy images modeling various count levels (corresponding to different products of acquisition duration and administered activity) were generated by appropriately scaling these images and simulating Poisson noise. Acquisition duration was fixed; six count levels were simulated corresponding to projection images acquired using 25%, 50%, 75%, 100%, 125%, and 150% of the original weight-based administrated activity as computed using the North American Guidelines (Gelfand *et al* 2011 *J. Nucl. Med.* **52** 318–22). Combined, a total number of 207 360 noisy projection images were generated, creating a realistic projection database

for use in renal pediatric SPECT imaging research. The phantoms and projection datasets were used to calculate three surrogate indices for factors affecting image quality: renal count density, average radius of rotation, and scatter-to-primary ratio. Differences in these indices were seen across the phantoms for dosing based on current guidelines, and especially for the phantom modeling the newborn. We also performed an image quality study using an anthropomorphic model observer that demonstrates that the weight-based dose scaling does not equalize image quality as measured by the area under the receiver-operating characteristics curve. These studies suggest that a dosing procedure beyond weight-based scaling of administered activities is required to equalize image quality in pediatric renal SPECT.

1. Introduction

In nuclear medicine imaging, the product of acquisition duration and administered activity (AA) determines the level of quantum noise present in the image. Quantum noise can have a direct impact on diagnostic image quality, and, for the purposes of maximizing image quality, reducing AA, or reducing acquisition duration, it is desirable to study the relationship between these factors.

Over the past decade, there has been an increased interest in reducing patient radiation exposure in diagnostic imaging studies that use ionizing radiation. Therefore, there has been significant interest in the nuclear medicine community in establishing universally accepted and optimized dosing guidelines for pediatric nuclear medicine studies. The European Association of Nuclear Medicine (EANM) and Society of Nuclear Medicine and Molecular Imaging (SNMMI) have, respectively, published the European pediatric dosage card and the North American consensus guidelines for pediatric AA (Lassmann *et al* 2007, Treves *et al* 2016). The goal of these guidelines is to provide a balance between radiation risk and image quality. However, these guidelines were developed either based on a consensus of best practices or a simple estimate of image quality and not on a rigorous evaluation of diagnostic image quality relative to AA.

A second concern in pediatric imaging is the acquisition duration. Sedation is often required, especially for longer acquisitions. Longer acquisition durations increase the chance of patient motion, which can degrade image quality. Shorter acquisition durations are thus desirable.

All else being equal, reducing the product of AA times acquisition duration will increase the Poisson noise in the image. However, the effect of changes in quantum image noise on diagnostic performance are complicated (He *et al* 2010). Similarly, decreasing quantum noise in the images requires increasing AA, acquisition duration, or both. Increasing the AA above that needed to provide acceptable image quality violates the principle of as low exposure as reasonably possible (ALARA). Consequently, appropriate guidelines for pediatric AAs are of significant interest (Treves *et al* 2008). Similarly, increasing the acquisition duration in pediatric patients to compensate for reduced AA may not be acceptable. Thus, understanding the tradeoff between image quality and the product of AA and acquisition duration is an important problem.

In 2008, the Dosimetry and Pediatrics Committees of the EANM published the first version of the EANM pediatric dosage card to better standardize the AAs in pediatric nuclear medicine procedures. The dosage card was based on data from a publication by Jacobs *et al* (2005). In that study, count rates and effective doses were computed as a function of body weight for 10 radionuclides and 95 radiopharmaceuticals, respectively, using seven hermaphrodite anthropomorphic computational phantoms (Cristy and Eckerman 1987). Count rate was used as the only surrogate for image quality; a discussion of the details and limitations of that aspect of that work are provided in the discussion section.

A second effort at standardization of pediatric dosages was the 2010 North America Consensus Pediatric Dosing Guidelines (Gelfand *et al* 2011). The AAs recommended in that report were slightly lower for infants and small children as compared to the EANM guidelines, compensating for the higher radiation risk in early childhood. Those guidelines were based on a combination of experience and retrospective analysis of clinical data, taking into account the patient's weight and count rate density per unit area or volume, and using these as the surrogates for radiation risk and count rates as the surrogate for image quality.

In 2011, Sgouros *et al* proposed a rigorous method to balance diagnostic image quality with cancer risk using ^{99m}Tc-DSMA as an example (Sgouros *et al* 2011). The study showed that weight alone may not be sufficient for optimally scaling AA in children. In that study, nonuniform rational B-spline (NURBS)-based anatomic phantoms, realistic organ uptakes and models of the image formation process, and task-based measures of image quality were used to objectively compare image quality of ^{99m}Tc-DMSA SPECT images. Two 10-year-old females of the same weight but different heights, respectively representing short-stout and tall-thin patients, were used in that study. Several different AAs (25%, 50%, 75%, 100%, 125%, and 150%), defect locations, and lesion severities with different target-to-background activity concentration ratios were simulated to represent clinical imaging. Channelized Hoteling observer (CHO) methodology was used in a receiver-operating-characteristic (ROC) analysis of lesion detectability to study the relationship between AA and the area under the ROC curve (AUC).

Figure 1. Sample coronal slices of the body reminder, cortex, medulla, pelvis, liver and spleen (from left to right) of a newborn 50th height percentile male phantom.

The results of the study showed that the same AUC could be obtained for the tall-thin phantom with approximately half the AA as for the short-stout phantom (Sgouros *et al* 2011).

In this present study, we have built upon the Sgouros *et al* work by developing a realistic pediatric phantom population including variations in age, gender, kidney size, and height. We have also proposed a novel method that produces contrast-matched, clinically-relevant defects in all of the phantoms across different ages, gender, body morphometries, and kidney sizes. The combination of these methods allows application of task-based image quality methods to rigorously assess current dosing guidelines in terms of their effectiveness for equalizing image quality across patients with different ages and body morphometries.

Toward this end, we simulated realistic projections of the pediatric patient population in preparation for future detailed investigations of the tradeoffs between image quality, the product of AA and acquisition duration, patient weight and height, and reconstruction method for ^{99m}Tc-DMSA renal imaging. Using this realistic phantom population and projection database, we investigated the effects of scatter, count density, and radius of rotation as a function of patient morphometry. These studies provide insight into the changes in these surrogate indices for factors affecting image quality and how they change with patient weight and body morphometry and the limitations of weight-based scaling of AA. We also performed a model observer study to investigate the impact of patient weight on image quality to study the validity of weight-based dose scaling for ^{99m}Tc- DMSA imaging.

2. Methods

2.1. Population of realistic digital phantoms

The database of projection data for this study was generated using the Advanced Laboratory for Radiation Dosimetry Studies (ALRADS) UF NHANES-based phantom series (O'Reilly *et al* 2016). The phantom population realistically models pediatric heights, weights, organ sizes and anatomies for both genders at five ages. The phantoms were adjusted to model variations in height and organ size prior to voxelization. The ages modeled were newborn, 1-, 5-, 10-, and 15-years old. For each age, we modeled the 50th percentile weight and 10th, 50th, and 90th percentile heights, simulating patients having the average weight at each age with varying body habitus. The 10th, 50th and 90th height percentile phantoms are referred to as short-stout, reference, and tall-thin patients, respectively.

For each age and height percentile, we modeled three kidney masses: -15%, average, and +15%, where average is the International Commission on Radiological Protection (ICRP) standard mass for a patient with the corresponding age and the percentiles are the change relative to this standard mass. The variations in kidney mass model variations in patient kidney size; for newborn patients, the dosimetric impact of these sizes on risk has been previously studied (O'Reilly *et al* 2016).

In addition to anatomic variability, we simulated variations in uptake in 6 tissues: cortex, medulla, pelvis, spleen, liver, and body reminder (the remaining soft tissues of the phantom). Figure 1 shows sample coronal slices of these different objects (organs and renal sub-structures) in a newborn phantom of average height (50th percentile height). Projections of each object were generated separately assuming a uniform activity distribution. The individual projections could then be scaled and summed to represent the count level that would be obtained in projections for an arbitrary AA, acquisition duration, or set of relative uptakes. By individually generating and scaling these projections, we were able to adjust the uptakes in each individual object to simulate uptake variability.

Each phantom was digitized prior to simulation into 0.1 cm cubic voxels and truncated in the axial direction to exclude regions more than 5 cm below the bottom or above the top of the kidneys. This truncation was done in order to reduce simulation time and data storage requirements.

2.2. Organ uptake model

Uptake in the kidneys was estimated using data from a single imaging time point, which varied slightly across patients, from datasets of 47 patients with ages ranging from 1 to 16 years acquired at the Boston Children's Hospital (BCH). We did not attempt to develop an age-specific pharmacokinetics model from this data, and considered the data from all patients as a single mixed-age population sample for estimating uptake of activities in the kidneys. CT scans of these patients were not available for attenuation compensation as they were not acquired as part of the patient's clinical study. Instead, attenuation maps were estimated based on automated intensity

Table 1. Summary of population parameters.

	Kidney uptake fraction	Cortex-to-medulla + pelvis act. conc. ratio
Maximum	0.393	2.00
Minimum	0.329	1.36
Sample mean	0.361	1.68
Sample standard deviation	0.025	0.25

thresholding of images reconstructed from scatter windows. The data were reconstructed using five iterations with eight subsets per iteration of an ordered-subsets expectation-maximization (OS-EM) reconstruction method that included attenuation, scatter and collimator-detector response compensation. Reconstructed images were converted to units of activity concentration using the measured camera sensitivity. The kidneys were segmented automatically using intensity thresholding, and the reasonability of the kidney VOIs and body contours were reviewed manually. The percent of the decay-corrected AA in the kidneys in these VOIs is referred to as the kidney uptake fraction. In addition, we used thresholding to segment the kidney cortex and pelvis/medulla regions. From these we computed the ratio of activity concentrations (sum of activity values divided by volume of the VOI) in the cortex to the medulla/pelvis. This ratio is referred to as the cortex-to-medulla plus pelvis activity concentration ratio. The results obtained from the above procedure are summarized in table 1, and were used as estimates for percent tracer uptakes in the patient's kidneys.

The fractional uptakes in the spleen and liver at the imaging time were obtained from Evans *et al* (1996); the values used were 4.3% and 1.7%, respectively. We validated these percentages against the real patient data from BCH and found that the uptake variations for these organs were small across the patient datasets and small compared to the uptake in the kidneys. Therefore, we used constant uptake percentages for liver and spleen in the simulations.

To model the differences in uptake of the fine structures inside the kidney (renal cortex, medulla, and pelvis), we quantified the relative uptakes inside these renal structures using the data from the aforementioned 47 patient images. The relative uptake values were estimated using the reconstructed images described above. We used threshold-based segmentation to separate the cortex from the medulla plus pelvis and created two separate VOIs for the two entities. These VOIs and the activity values inside them were used to estimate activity concentration of each entity. The mean and standard deviation of the cortex-to-medulla plus pelvis activity concentration ratios were calculated and are summarized in table 1.

The resolution in the images was not sufficient to estimate accurately the activity concentration ratio between medulla and pelvis. Thus, we based the activity concentration ratio in these two structures on input from our clinical collaborators. Images were generated with a variety of concentration ratios; images having a medulla-to-pelvis concentration ratio of 1:1 were deemed most realistic, and that ratio was thus used in the study.

2.3. Organ uptake variations

We modeled random variations in the uptake of the kidneys as a whole and in the cortex relative to the medulla plus pelvis using truncated Normal distributions. The values of the minimum, maximum, mean and standard deviation of these distributions were obtained from the 47 patients described above, and are given in table 1. For each phantom anatomy, we randomly sampled 384 values each of the fraction of injected activity in the entire kidney and the cortex-to-medulla plus pelvis activity concentration ratio. From these data combinations, the weight-based AA, and the kidney volume, we calculated the activity concentrations in the cortex, medulla and pelvis for each of the 384 uptake realizations.

2.4. Projection data simulation

The projections were simulated using an analytic projection code that models attenuation, spatially varying detector-to-collimator response (Frey *et al* 1993), and object-dependent scatter (Frey and Tsui 1997). This code has been extensively validated for imaging of a variety of radionuclides by comparison to Monte Carlo simulations and experimental projection measurements. We modeled a low-energy, ultra high-resolution (LEUHR) collimator, a 360° body-contouring orbit, 120 projection views, a 15% wide energy window centered at 140 keV, an energy resolution of 9% at 140 keV, and a 0.1 cm projection bin size. After each simulation, the projections were collapsed isotropically by a factor of 2 to simulate a 0.2 cm projection bin size. Attenuation maps used in the projection operation were constructed by assigning the attenuation coefficient of the organ in the phantom containing the voxel center to the entire voxel. The attenuation coefficients were evaluated at 140 keV for the material composition of each organ based on ICRP organ composition data (ICRP 2002). Figure 2 shows sample transaxial images of attenuation distributions that illustrate variations in body habitus of the population.

Projections were generated using the above methods individually for the kidney cortex, medulla, pelvis, liver, spleen, and the body reminder, by assigning unit intensity to the phantom voxels in each of these regions. The

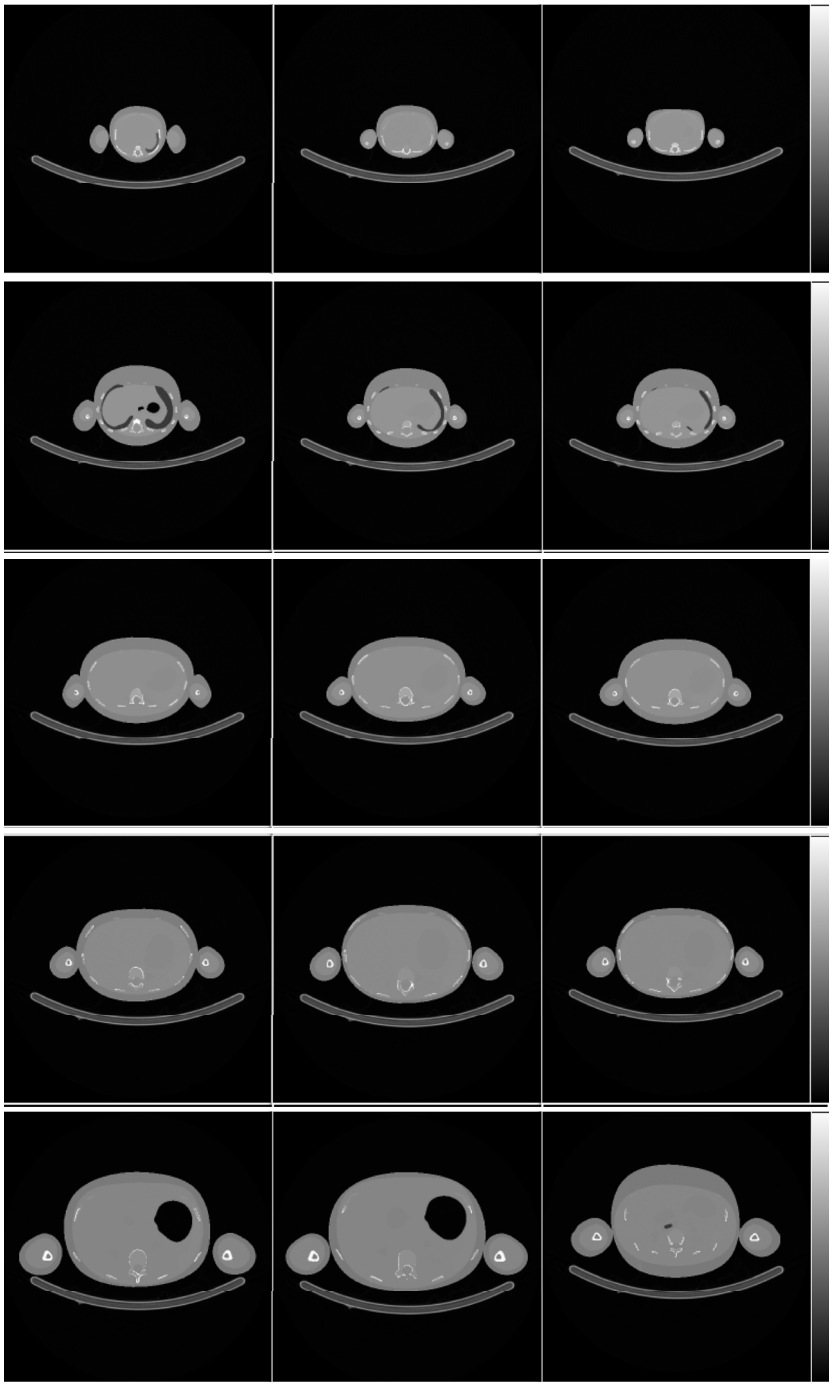

Figure 2. Sample transaxial images of the attenuation distribution for the (left to right) 10th, 50th, and 90th height percentile versions of the male phantom for ages (top to bottom) 0 (newborn), 1, 5, 10, and 15 years showing variations in body habitus.

organ projections were then scaled by the randomly-sampled uptake scaling factors needed to obtain the desired activity concentrations. They were then summed and scaled by the camera sensitivity to obtain the raw projections per unit injected activity per acquisition duration. These raw projections were next scaled by the acquisition duration and desired AA to give the mean projections in units of counts. Using these as inputs to a Poisson random noise generator gave the noisy projections.

2.5. Simulated projection data with variation in injected activity

Six count levels were simulated: 25%, 50%, 75%, 100%, 125%, and 150%. Here, a count-level indicates the fraction of AA relative to the AA obtained from the 2010 North American Consensus Dosing Guidelines. Note that the suggested minimum and maximum AAs in these guidelines were only enforced for the clinical (100%) count level. Figures 3 and 4 show a sample set of noise-free projections of the organs for the newborn and sample noisy projection images from the various count levels.

Figure 3. Noise-free projection images of the kidney cortex, medulla, spleen, liver, pelvis, and body remainder for a male, reference-height, newborn phantom.

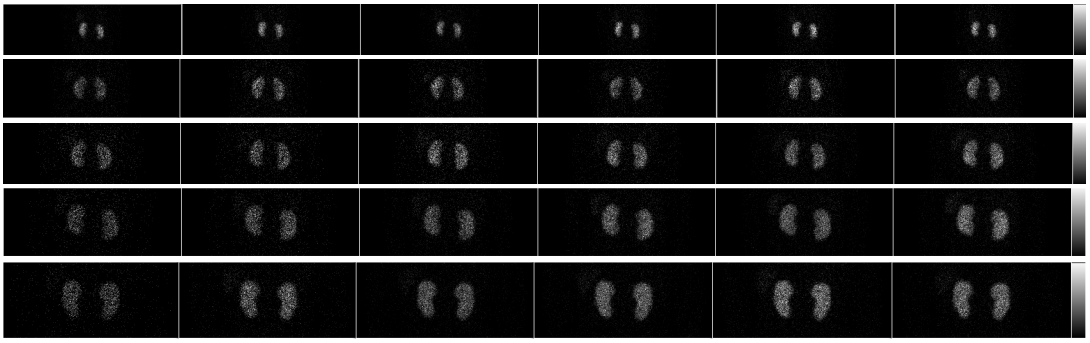

Figure 4. Sample noisy posterior projection images from the various count levels. From top to bottom shows kidneys for the 0, 1, 5, 10, and 15-year-old phantoms. From left to right, the simulated count levels were 25%, 50%, 75%, 100%, 125%, and 150% of those of the 2010 North American Consensus Dosing Guidelines.

2.6. Validation of simulated projection image

To validate simulation process, we chose patient images from each of the five ages. The patient and simulated projections were reconstructed using two iterations of eight subsets of the OS-EM iterative reconstruction with detector-to-collimator response compensation only, followed by filtering with a 5 mm FWHM Gaussian filter. Sample reconstructed image slices are shown in figure 5. Note that the kidney model in the phantom does not model the detailed structure of the medulla and pelvis. We computed the total of the reconstructed voxel values in volumes of images containing the kidneys. Table 2 shows a comparison of the counts in the simulated and patient images for the various ages.

2.7. Defect model

Assessing image quality should be done with respect to the task that will be performed with the images. In the case of DMSA SPECT, the task is to detect functional defects in the renal cortex. Thus, it is necessary to create defects in the simulated images. Since the ultimate goal of the project is to provide guidance data for selecting minimum AAs commensurate with detecting clinically relevant defects, we developed defects for each age that were challenging but clinically relevant (Li *et al* 2015). Challenging defects will tend to be ones that are small, where partial volume effects produce low contrast defects in the reconstructed images. Since the distance from the collimator face to the kidney depends on patient size, resolution will tend to be worse for large patients compared to small ones. In addition, since the thickness of the cortex tends to be greater for larger patients, and partial volume effects depend on the uptake in tissues surrounding the defect that are within approximately two times the full width at half maximum (FWHM), the same physical defect size would be harder to visualize in a large patient than in a smaller patient. Also, when the kidney is larger, the same size defect would tend to be of less consequence in terms of total renal function. Thus, we chose to create defects with sizes that varied depending on the patient size.

Based on input from an experienced pediatric nuclear-medicine specialist (S.T. Treves), we selected a defect volume of 0.3 cm^3 as the defect size that is clinically relevant and challenging for a newborn. To create a realistic defect, we used an ellipsoid with one major axis length equal to approximately the thickness of the cortex. The center of the ellipsoid was positioned along the line extending through the center-of-mass of the cortex at the point where it intersects the outer cortical surface. The half-length of the axis of the ellipsoid in this direction was equal to the cortical thickness along this line, meaning that the apex of the ellipse was at the inner surface of the cortex. The half-lengths of the other two axes were the same; for the newborn, the length of these remaining axes was set so that the intersection with the kidney cortex had a volume of 0.3 cm^3. This was verified numerically by creating a voxelized version of the defect where the voxel values were set to unity using sub-voxel sampling by a factor of 2, taking the product of this ellipsoid with the cortical VOI, and summing the values in the

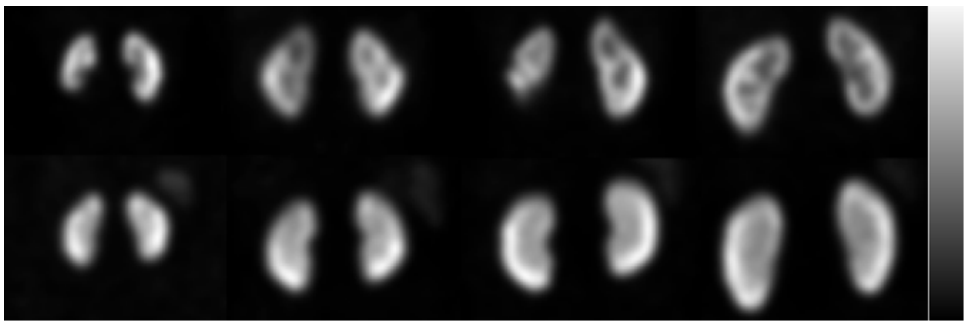

Figure 5. From left to right, the top row shows patient images from 1.2, 5, 9, and 16 year olds reconstructed using two iterations of eight subsets of the OS-EM reconstruction with detector response compensation followed by a Gaussian filter with a FWHM of 0.5 mm. The bottom row shows simulated images from 1, 5, 10, and 15 year olds reconstructed using the same methods.

Table 2. Comparison of total counts in clinical and simulated projections.

Phantom age	Corresponding patient age	Total counts in patient image	Counts in simulated image	Percent difference
1	1.2	328 821	373 482	−12.718
5	5	543 850	540 371	0.642
10	9	831 381	711 803	15.498
15	16	1100 752	1215 463	−9.905

resulting product image. A trans-cortical defect with a realistic shape can be created by scaling the defect image by the desired contrast and subtracting this from the cortical VOI. Note that, due to the linearity of the projection operation, the subtraction can be performed in either the image or projection domain. The position along the cortex was defined by an angular coordinate in the coronal slice containing the defect center of mass. Gradual transitions of function can be modeled by blurring the ellipsoid with a Gaussian kernel prior to the multiplication described above. Sample images containing defects for phantoms representing various ages in figure 6.

2.8. Reconstruction and post-reconstruction processing

SPECT images were reconstructed from the simulated projections using filtered backprojection (FBP) reconstruction and a ramp filter with no apodization. The reconstructed images had cubic voxels with a side length of 0.2 cm. The reconstructed images were post-filtered with 3D Butterworth filters of order 8 and cutoff frequency 0.12 cycles/pixel. Figure 7 shows a sample set of reconstructed images for the five ages.

2.9. Quantitative measures of image quality

Defect detectability depends, in principle and among other factors, on the contrast of the defect and the amount of noise in the reconstructed image. The contrast of the defect depends on the intrinsic contrast of the defect relative to surrounding tissues and the size of the defect with respect to the resolution of the imaging system. In addition, contrast is degraded by the effects of scattered photons from surrounding tissues. In the following we present physical measures of image quality that quantify the noise, image resolution, and scatter, all of which together affect image quality.

We measured the contribution of scattered photons in the kidney region by the scatter-to-primary ratio obtained from projection images that only contain the kidneys. The images of the kidneys resulting from detected scattered photons were obtained by subtracting the kidney projection generated with attenuation, collimator-to-detector response, and scatter from the same projection generated with attenuation and collimator-to-detector response. The numbers of scattered and primary photons were obtained by summing the counts in the resulting images, respectively.

Lastly, we quantified the noise in the kidney region by the kidney count density, i.e., the number of detected primary photons emitted in the kidneys divided by the kidney volume. We used the primary-photon-only projection images of the kidneys, generated as described above, for this calculation. We performed this calculation for all 90 phantoms using the same mean kidney uptake fraction (0.361) and cortex-to-medulla plus pelvis activity concentration ratio (1.68) in all cases. The count density was averaged over the 3 kidney sizes for each phantom.

In SPECT, the image resolution at the center of rotation is proportional to the radius of rotation. Thus, we quantified the system resolution by the distance from the collimator face to the center of rotation averaged over the (body contouring) camera orbit. Note that the phantoms were placed on a camera bed measured using a CT scan, so the camera orbit, especially for the small phantoms, was constrained in some views by the size of the bed.

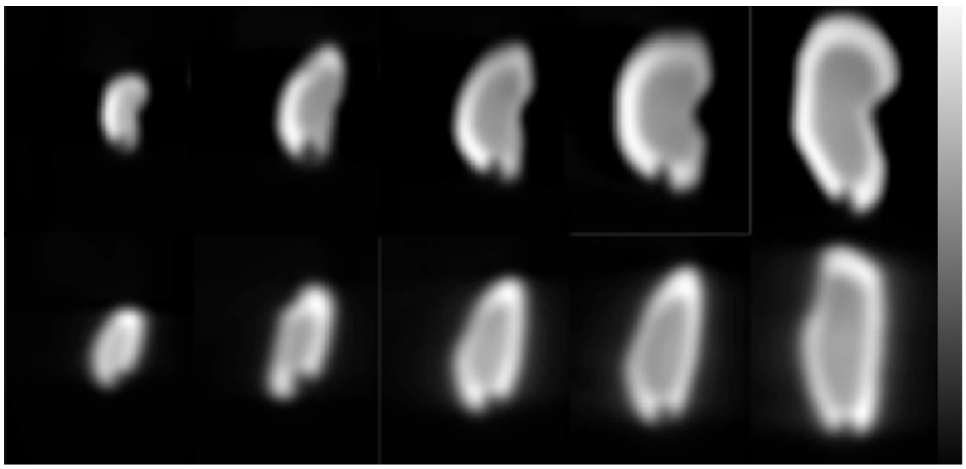

Figure 6. Sample lower pole defects in noise-free reconstructed images for newborn, 1-, 5-, 10-, and 15-year-old male phantoms with reference heights in coronal and sagittal views. The defect volumes for ages 1, 5, 10 and 15 were determined by matching their contrasts to the newborn.

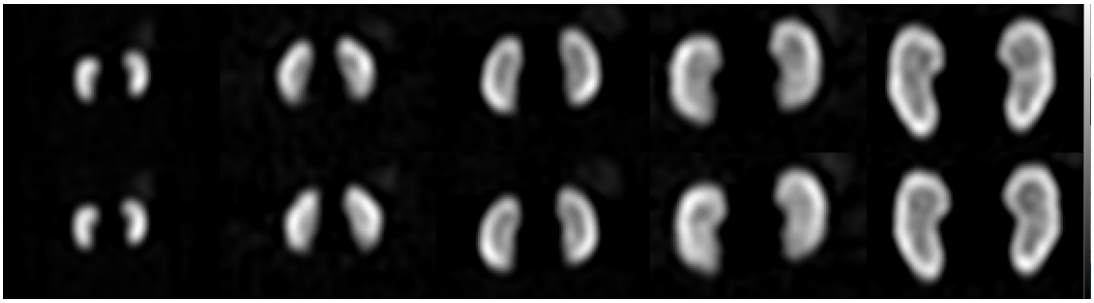

Figure 7. Sample reconstructed images from noisy projection data using FBP reconstruction followed by a post-reconstruction 3D Butterworth filter with an order of eight and cutoff frequency of 0.12 cycle/pixel. Negative values were mapped to zero in the display. From left to right, the bottom and top rows shows coronal images with and without, respectively, a (lower pole) defect for the newborn, 1-, 5-, 10-, and 15-year-old male phantoms at the 50th height percentile. The volumes of these defects were chosen to be near the limits of clinical relevance and to have the same defect contrast.

2.10. Model observer study

The model observer study was performed using methods similar to those previously described in Sgouros *et al* (2011). In summary, seven 2D frequency-domain bandpass difference-of-mesa channels were used to approximately model the human visual system. The starting frequency and width of the first channel was 0.5 cycles per pixel, and subsequent channels had widths that doubled and abutted the previous channel. These channels were applied to the three orthogonal slices (transaxial, sagittal, and coronal) that contained the defect centroid. The output of this procedure was a 21-element feature vector.

In previous work, feature vectors were analyzed using a Hotelling observer (HO) methodology, a combination often referred to as the CHO. However, the projections in the database and resulting reconstructed images reflect variations in anatomy and uptake. This resulted in features vectors that were often multi-modal and thus not multivariate normally (MVN) distributed. We have previously demonstrated that this can result in difficulty for the HO (Elshahaby *et al* 2016). Thus, instead of a traditional CHO, we used a multi-template strategy proposed by Li *et al* (2017) to handle the non-MVN data. In this strategy, the entire ensemble of input data vectors is decomposed into sub-ensembles that are approximately MVN. A linear discriminant trained using the data for that sub-ensemble is then used to analyze the set of feature vectors for that sub-ensemble. The set of test statistics from each sub-ensemble is then pooled and analyzed using ROC analysis. In this work, the channel output vectors were sorted into sub-ensembles based on defect location, age, and height percentile. It was verified qualitatively that the resulting channel output data were not multi-modal and were nearly MVN distributed. We used a leave-one-out sampling methodology to generate the subsets for each sub-ensemble (Elshahaby *et al* 2017). We pooled the test statistics for each defect location and height percentile and applied ROC analysis. The AUC was used as a figure-of-merit for task performance.

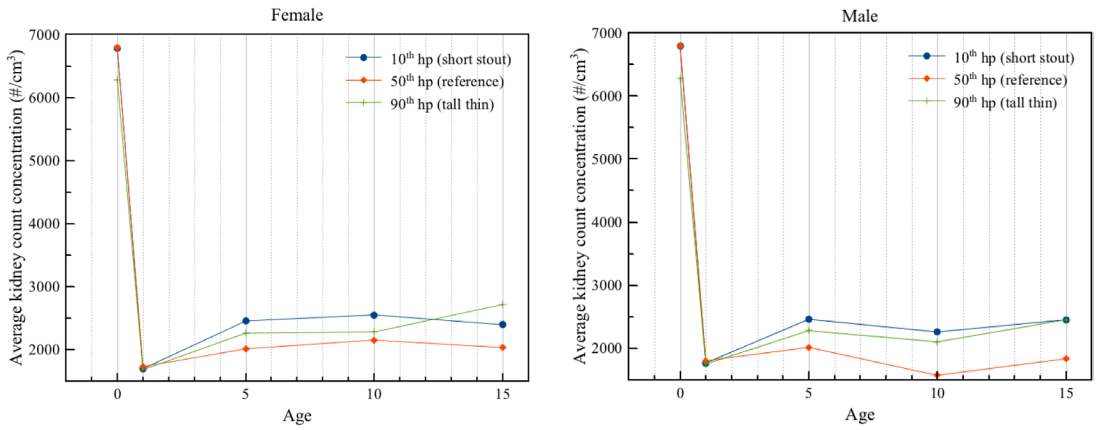

Figure 8. Average kidney count density obtained for three different height percentiles as a function of phantom age for male and female phantoms.

3. Results

3.1. Quantification of noise by renal count density

Figure 8 shows plots of the average kidney count density as a function of patient age for the different height percentiles for male and female phantoms, respectively. Overall, the plots demonstrate that the weight based AA produced nearly equal kidney count densities for all ages except for the newborn. The data also show that gender did not affect count density in patients less than 10 years old. This indicates that there is not much difference in the overall attenuation in the kidney region between the male and female phantoms for these ages. Note that, in theory, we expect to see rankings of the values of the count density in the order 90th $>$ 50th $>$ 10th height percentile: we would expect the short-stout phantoms to allow fewer photons to escape the body than for the reference or the tall-thin phantoms. However, this was not observed in all cases. For example, we see that, for the 10 year olds, the largest difference was between the 10th and 50th height percentile. This difference in count density was approximately 30% in the male phantoms and 15% in the female phantoms. For the 15 year olds, the differences were 25% for the male phantom and 15% for the female phantoms, respectively. There was essentially no difference in count density between 10th and 90th height percentile phantoms for both the 10 and 15 year olds for both genders.

Figure 9 shows transaxial images of phantom slices at the mid-kidney level for the 10th, 50th, and 90th height percentile from male phantoms with ages 0, 1, 5, 10, and 15. From these images, we can see that there was not a significant difference in body circumference (girth) among phantoms for different heights nor a strong correlation between girth and phantom height rankings. That is, the 10th height percentile (short-stout) phantom did not necessarily have a larger girth than the 90th height percentile (tall-thin) phantom or the reference phantom, at the mid-kidney level. These images provide a pictorial illustration of the reason that the observed count density did not vary as expected with patient height.

3.2. Quantification of scatter by scatter-to-primary ratio

Figure 10 shows plots of the average scatter-to-primary ratio for the three different height percentiles as a function of patient age for male and female phantoms, respectively. It is clear that dosing by weight did not equalize the effects of scatter. These data also demonstrate that the scatter-to-primary ratios depended to a varying degree on height. Just as for count density, the expected rankings were not consistently observed. For example, one would expect the tall-thin patient to have a smaller scatter-to-primary ratio than the short-stout phantom. However, this was not observed for the 15-year-old male, though the differences in the ratios for the different heights were generally small. In any event, these data suggest that weight and height alone may not sufficient for predicting the effects of scatter, and thus their degrading effects on image quality.

3.3. Quantification of resolution by camera radius of rotation

Figure 11 shows plots of the average camera radius of rotation for the three different height percentiles as a function of patient age for male and female phantoms, respectively. Again, the expected trend, larger height percentiles having smaller average radii of rotations, was not always observed. The cause is likely for the same reason that differences in count density did not vary as expected with height: the maximum girth of the patient, which determines the distance from the camera to the body for a body contouring orbit, did not vary directly with height percentile. This again indicates that height and weight are not sufficient to predict resolution effects.

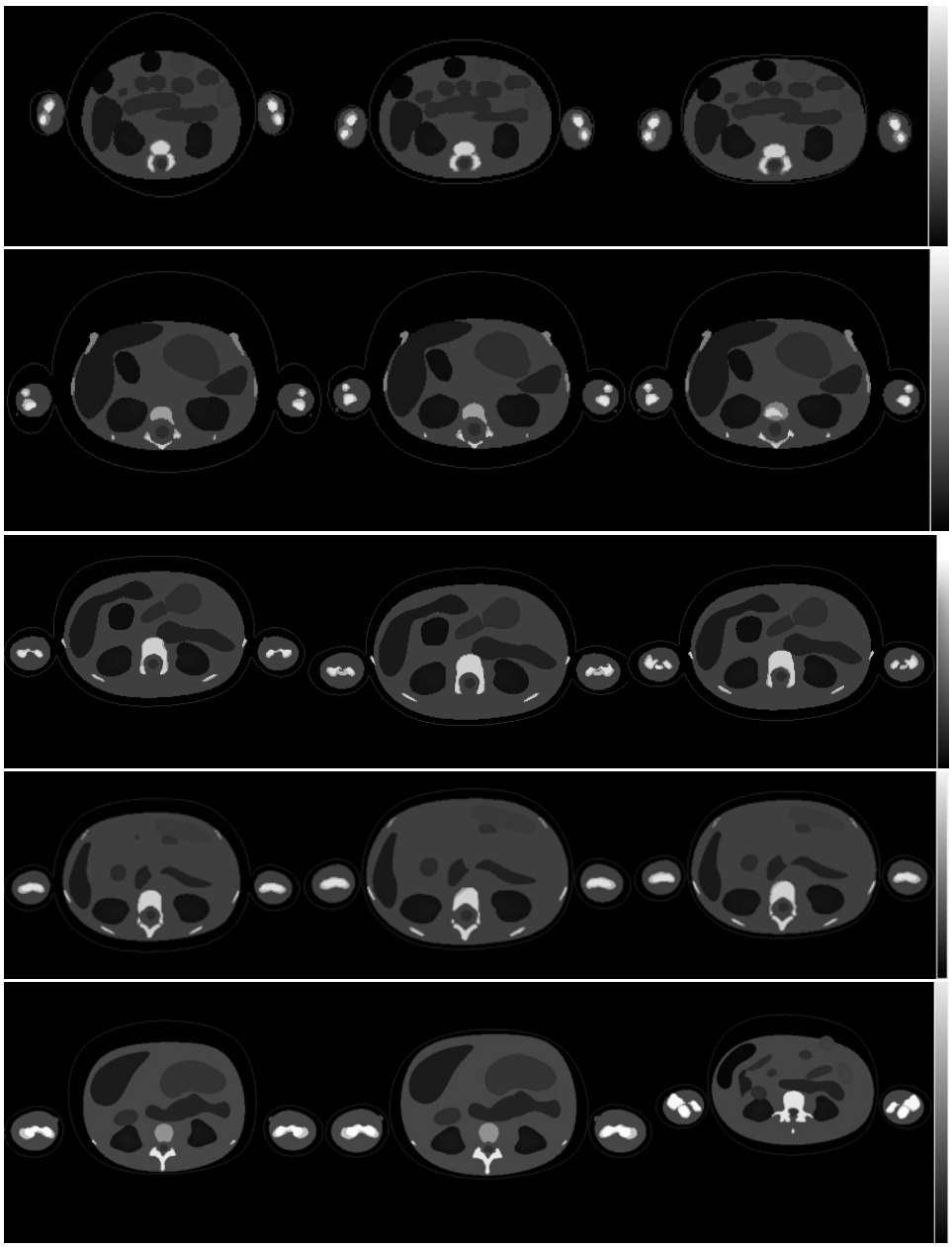

Figure 9. Sample transaxial phantom images at mid-kidney level in 10th, 50th, and 90th height percentile (from left to right) from the male phantom of age 0, 1, 5, 10, and 15 (from top to bottom) showing variations in body habitus.

This is especially true for the small phantoms as the radius of rotation was limited by the size of the imaging bed in the lateral direction.

3.4. Model observer study results

Figure 12 shows a plot of AUC as function of the product of percent AA and acquisition duration for the 1- and 15-year-old on a defect detection task. Image quality was studied as a function of count level using the developed database with a fixed acquisition duration of 960 s. The saturation of the AUC values indicates that there is a decreasing benefit of increasing the AA, indicating that detection is limited by background variation. These results show that there was a monotonic but modestly saturating increase in AUC with AA. The fact that there was modest saturation indicates that image quality was limited by quantum noise and the effects of object variability were modest (He *et al* 2010). More importantly, the results show that the AUCs for an AA of 100% of the weight-based, consensus dosing guideline were not equal, indicating that the ability to detect a defect with the same contrast, was not the same for the AA recommended by the North American Consensus Guidelines. This, combined with the results of the other image quality surrogates reflecting noise, scatter, and resolution effects, suggests that weight-based scaling is not sufficient to equalize image quality.

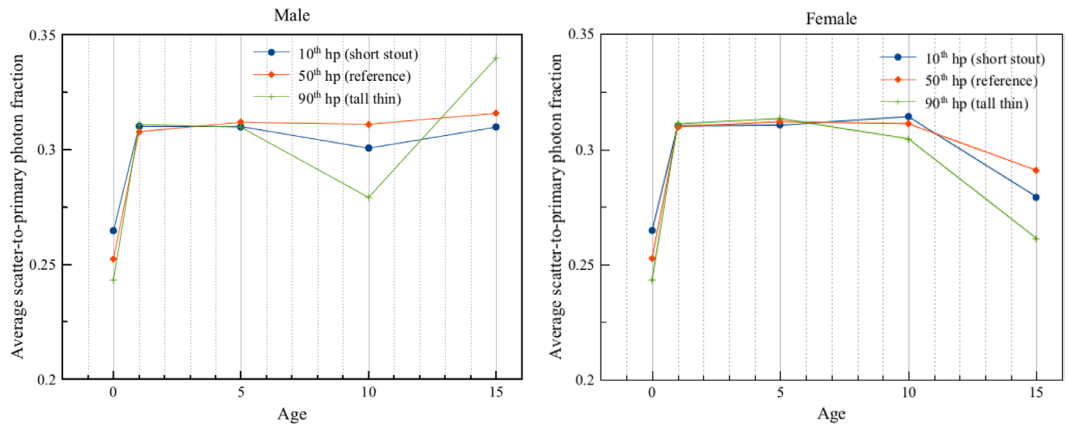

Figure 10. Average scatter-to-primary ratio obtained from three different height percentiles as a function of phantom age for male and female.

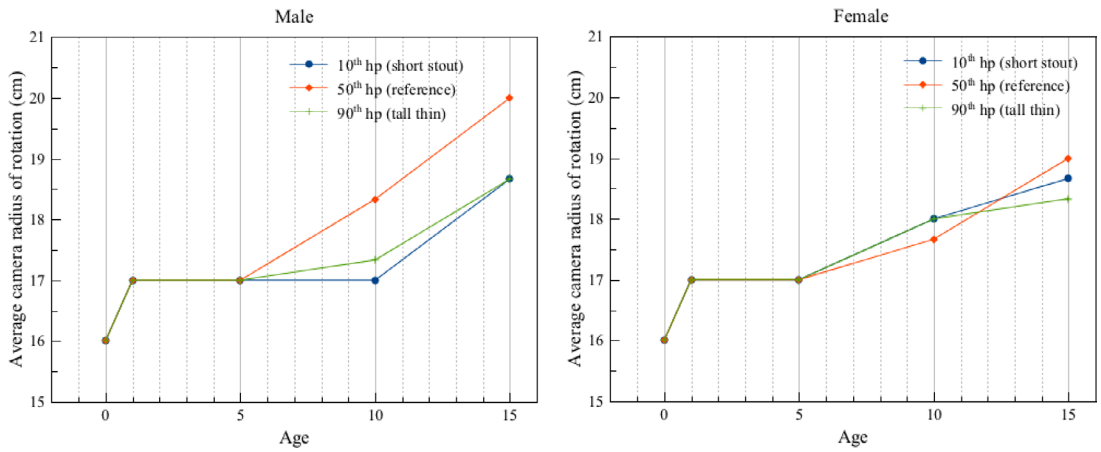

Figure 11. Average camera radius of rotation obtained from three different height percentiles as a function of phantom age for male and female.

4. Discussion

This paper describes the design and simulation of a realistic projection database for use in pediatric renal SPECT research. The population included variability in age (and thus weight), gender, kidney size, and height. The specific dataset generated here was focused on matching the defect detectability across the population for a challenging and relevant defect detection task. We thus designed a set of defects, one for each phantom and kidney size, that is clinically relevant but at the limits of what is likely to be detectable. Thus the curve describing the tradeoff between image quality and AA is for a difficult case; larger defects are likely to be easier to detect and thus not as affected by reductions in AA.

Using the established projection database, we investigated the AAs based on the current North American consensus weight-based dosing guidelines in terms of impact on surrogates for factors that affect image quality (image noise, resolution, and contrast). As compared to the approach by Jacobs *et al* (2005), the present study provides a more rigorous evaluation of image quality by adopting a more realistic phantom database, imaging simulations, pharmacokinetic model, and task-based image quality evaluation method. In the Jacob *et al* study, seven phantoms were used, representing newborns and children (male only) of 1, 5, 10, and 15 years, adult females and adult males, corresponding to their reference weights. Though these phantoms included seven organs, the radionuclide was assumed uniformly distributed in the phantom (no inter-organ uptake variability) for the purposes of estimating count rates. The fraction of energy absorbed by the target organs at the emitted photon energy was computed using Monte Carlo simulations. The count rates that would be obtained with gamma camera imaging were assumed to be proportional to the average number of photons (at energies useful for imaging) that exited the body. These count rates were considered to have potentially contributed to the image. The fraction of exiting photons was computed as a weighted sum of the non-absorbed fractions (1 minus the absorbed fraction) at the energy of each emitted photon. The average number of photons emitted per disintegration for each emitted photon energy was used to weight the absorbed fractions, and only emitted photons with

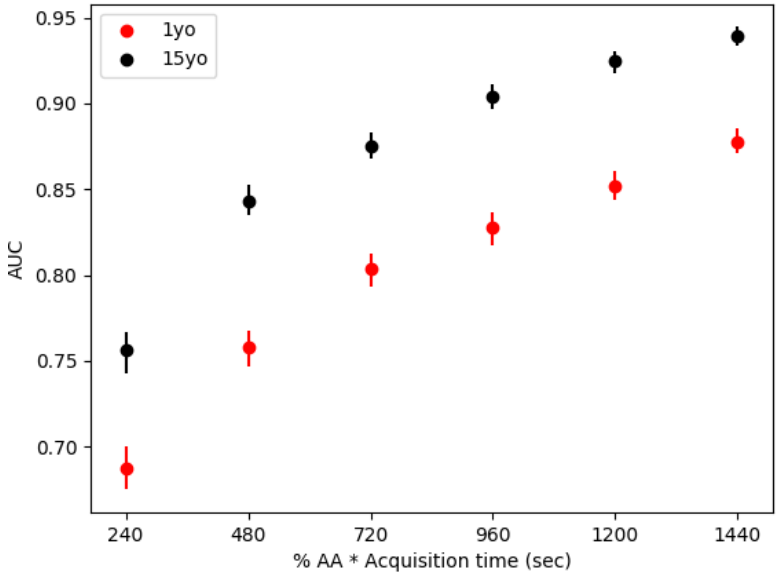

Figure 12. Image quality result on a defect detection task for the 1- and 5-year-old phantoms. A 20% defect contrast was modeled for these patients.

abundances greater than or equal to 10% and energies suitable for imaging were included in the calculation. Normalization factors for the count rates were obtained by dividing corresponding count rates by that for an adult male of 70 kg (normalization factor = 1.0).

Using the count rates estimated by this method as the sole surrogate for image quality has several limitations. First, it ignores the effects of other factors, such as scatter and resolution, which can vary with body size. Related to this, energy exiting the body is an incomplete surrogate for primary photon counts. It equates, for example, two or more scattered photons with total energy equal to the primary photon energy to a primary photon. In other words, the use of total energy as a surrogate for primary photon counts is valid if the scatter-to-primary photon fraction is a constant across the entire patient population. However, this is not the case as was demonstrated above. Second, the method assumed uniform radionuclide distribution in the body. This is less than ideal, especially for agents such as DMSA that concentrates in a small number of tissues. Third, there was no variation in organ size or patient height for a given weight. Finally, the suggested minimum AAs were purely based on effective dose and not image quality. The authors did point out in the discussion that the suggested AAs calculated using this method could possibly lead to impractical scanning times or unusable images.

5. Conclusion

A realistic projection database has been generated for investigation of relationship between image quality and patient morphometry in ^{99m}Tc-DMSA renal SPECT. A total of 207 360 projection images was generated, encompassing 6 different administrated activities for 90 phantom anatomies. This projection database can be used to study the relationship between the product of AA and acquisition duration and image quality in a way that is impossible with either real patients or via experimental phantoms. The database generated in this work is immediately applicable to other pharmaceuticals labeled with ^{99m}Tc used in pediatric imaging such as ^{99m}Tc-MAG3 or ^{99m}Tc-MDP; only scaling and summing of the organ projections with appropriate scaling factors reflecting agent biokinetics. Further, the methods used in this study are applicable to studying these tradeoffs for other diagnostic and/or therapeutic radiopharmaceuticals in both pediatric and adult patients.

Using this projection database, we conducted a quantitative analysis of three factors that affect image quality: noise, as measured by kidney count density; scatter, as measured by the scatter-to-primary ratio for photons emitted from the kidneys; and resolution, as measured by the average radius of rotation. The results of this study showed that weight-based dosing was partially able to offset losses in count density due to variations in patient weight. However, it suggested that the kidney count density for newborns was higher than for other ages using weight-based dosing. The results also demonstrated variations in scatter and resolution that depend on body morphometry, but were not well correlated with phantom height. We also performed a task-based image quality study using an anthropomorphic model observer that demonstrated that weight-based scaling of the AA did not equalize image quality as measured by the AUC. This, combined with the image quality surrogate data on noise, scatter, and resolution, suggests that weight-based scaling is not sufficient, suggesting that a dosing procedure beyond simple weight-based scaling of AA is required to equalize image quality in pediatric renal SPECT.

Further, the results also suggest the need for more detailed task-based studies of image quality, and that variables beyond height and weight are needed in order to prescribe AAs that optimize image quality in order to achieve as low as reasonably possible dosing.

Acknowledgment

This work was supported by NIBIB under grant number R01-EB013558. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure

The iterative reconstruction software used to generate projections and reconstruct images in this work has been licensed by Johns Hopkins to GE Healthcare. One of the authors, E C Frey, receives a portion of the licensing revenue. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

ORCID iDs

Abhinav K Jha https://orcid.org/0000-0002-4795-4878
Wesley E Bolch https://orcid.org/0000-0002-7175-9968

References

Cristy M and Eckerman K F 1987 Specific absorbed fractions of energy at various ages for internal photon sources *Oak Ridge National Laboratory* ORNL/TM-8381

Elshahaby F E A, Jha A K, Ghaly M and Frey E C 2017 A comparison of resampling schemes for estimating model observer performance with small ensembles *Phys. Med. Biol.* **62** 7300–20

Elshahaby F E, Ghaly M, Jha A K and Frey E C 2016 Factors affecting the normality of channel outputs of channelized model observers: an investigation using realistic myocardial perfusion SPECT images *J. Med. Imaging* **3** 015503

Evans K, Lythgoe M F, Anderson P J, Smith T and Gordon I 1996 Biokinetic behavior of technetium-99m-DMSA in children *J. Nucl. Med.* **37** 1331–5

Frey E C and Tsui B M W 1997 A new method for modeling the spatially-variant, object-dependent scatter response function in SPECT *1996 IEEE Nuclear Science Symp.—Conf. Record* vols 1–3 pp 1082–6

Frey E C, Ju Z W and Tsui B M W 1993 A fast projector–backprojector pair modeling the asymmetric, spatially varying scatter response function for scatter compensation in SPECT imaging *IEEE Trans. Nucl. Sci.* **40** 1192–7

Gelfand M J, Parisi M T, Treves S T and Pediatric Nuclear Medicine Dose Reduction Workgroup 2011 Pediatric radiopharmaceutical administered doses: 2010 North American consensus guidelines *J. Nucl. Med.* **52** 318–22

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

ICRP 2002 Basic anatomical and physiological data for use in radiological protection: reference values. A report of age- and gender-related differences in the anatomical and physiological characteristics of reference individuals. ICRP Publication 89 *Ann. ICRP* **32** 5–265

Jacobs F, Thierens H, Piepsz A, Bacher K, Van de Wiele C, Ham H, Dierckx R A and European Association of Nuclear Medicine 2005 Optimised tracer-dependent dosage cards to obtain weight-independent effective doses *Eur. J. Nucl. Med. Mol. Imaging* **32** 581–8

Lassmann M, Biassoni L, Monsieurs M, Franzius C, Jacobs F and EANM Dosimetry and Paediatrics Committees 2007 The new EANM paediatric dosage card *Eur. J. Nucl. Med. Mol. Imaging* **34** 796–8

Li X, Jha A K, Ghaly M, Elshahaby F E, Links J M and Frey E C 2017 Use of sub-ensembles and multi-template observers to evaluate detection task performance for data that are not multivariate normal *IEEE Trans. Med. Imaging* **36** 917–29

Li Y, O'Reilly S, Plyku D, Cao X, Fahey F, Bolch W E, Treves S T, Sgouros G and Frey E C 2015 Development of a defect model for renal pediatric SPECT imaging research *2015 IEEE, Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC)* (IEEE)

O'Reilly E S, Plyku D, Sgouros G, Fahey F H, Ted Treves S, Frey E C and Bolch W E 2016 A risk index for pediatric patients undergoing diagnostic imaging with (99m)Tc-dimercaptosuccinic acid that accounts for body habitus *Phys. Med. Biol.* **61** 2319–32

Sgouros G, Frey E C, Bolch W E, Wayson M B, Abadia A F and Treves S T 2011 An approach for balancing diagnostic image quality with cancer risk: application to pediatric diagnostic imaging of ^{99m}Tc-dimercaptosuccinic acid *J. Nucl. Med.* **52** 1923–9

Treves S T, Davis R T and Fahey F H 2008 Administered radiopharmaceutical doses in children: a survey of 13 pediatric hospitals in North America *J. Nucl. Med.* **49** 1024–7

Treves S T, Gelfand M J, Fahey F H and Parisi M T 2016 2016 Update of the North American Consensus guidelines for pediatric administered radiopharmaceutical activities *J. Nucl. Med.* **57** 15N–18N