Girth-based Administered Activity for Pediatric ^{99m}Tc-DMSA SPECT

Ye Li[1,2], Justin L. Brown[3], Jingyan Xu[2], Junyu Chen[1,2], Michael Ghaly[7], Monet Dugan[4,5], Xinhua Cao[4,5], Yong Du[2], Frederic H. Fahey[4,5], Wesley Bolch[3], George Sgouros[2], and Eric C. Frey[1,2,6]

[1] Department of Electrical and Computer Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD
[2] The Russell H Morgan Department of Radiology and Radiological Science, School of Medicine, Johns Hopkins University, Baltimore, MD
[3] J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL
[4] Department of Radiology, Boston Children's Hospital, Boston, MA
[5] Department of Radiology, Harvard Medical School, Boston, MA
[6] Sidney Kimmel Comprehensive Cancer Center, School of Medicine, Johns Hopkins University, Baltimore, MD
[7] Radiopharmaceutical Imaging and Dosimetry(Rapid), LLC., Baltimore, MD

Corresponding author: Ye Li, Ph.D.

E-mail: bettergary@gmail.com

Telephone number: 410-960-8271

Fax number: N/A

Address: 100 Cambridge St, Boston, MA 02114

Word count: 6022

Financial support: This work was supported by National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health 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.

Short running title: Girth-based Dosing for Pediatric DMSA

Abstract

Background: Pediatric molecular imaging requires a balance between administering an activity that will yield sufficient diagnostic image quality while maintaining patient radiation exposure at acceptable levels. In current clinical practice, this balance is arrived at by the current North American Consensus Guidelines in which patient weight is used to recommend the administered activity (AA).

Purpose: We have previously demonstrated that girth (waist circumference at the level of the kidneys) is better at equalizing image quality than patient weight for pediatric Tc-99m DMSA renal function imaging. However, the correlation between image quality (IQ), AA, and patient girth has not been rigorously and systematically developed. In this work, we generate a series of curves showing the tradeoff between AA and IQ as a function of patient girth, providing the data for standard bodies to develop the next generation dosing guideline for pediatric DMSA SPECT.

Methods: An anthropomorphic phantom series that included variations in age (5, 10 and 15 years), gender (M, F), local body morphometry (5, 10, 50, 90, and 95th girth percentiles), and kidney size (±15% standard size), were used to generate realistic SPECT projections. A fixed and clinically challenging defect-to-organ volume percentage (0.49% of renal cortex value) was used to model a focal defect with zero uptake (i.e., full local loss of renal function). Task-based IQ assessment methods were used to rigorously measure IQ in terms of renal perfusion defect detectability. This assessment was performed at multiple count levels (corresponding to various AAs) for groups of patients that had similar girths and defect sizes. Receiver-operating characteristics (ROC) analysis was applied; the area under the ROC curve (AUC) was used as a figure-of-merit for task performance. Curves showing the tradeoff between AUC and AA were generated for these groups of phantoms.

Results: Overall, the girth-based dosing method suggested different amounts of AA compared to weight-based dosing for the phantoms who have a relatively large body weight but a small girth or phantoms with relatively small bodyweight but large girth. Differences of AA up to 62.9% compared to weight-based dosing guidelines can potentially be saved while maintaining a baseline (AUC=0.80) IQ for certain 15-year-olds who

have a relatively small girth and large defect size. Note that the task-based IQ results are heavily dependent on the simulated defect size for the defect detection task and the appropriate AUC value must be decided by the physicians for this diagnostic task. These results are based purely on simulation and are subject to future clinical validation.

Conclusions: The study provides simulation-based IQ-AA data for a girth-based dosing method for pediatric renal SPECT, suggesting that patient waist circumference at the level of kidneys should be considered in selecting the AA needed to achieve an acceptable IQ. This paper provides standard bodies the IQ-AA data for detecting a clinically challenging defect.

Keywords

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

I. INTRODUCTION

In pediatric nuclear medicine, current guidelines [1-4] for calculating the administered activity (AA) for each patient are mainly based on patient weight. Before the guidelines were developed, there was little consensus on recommended AA in pediatrics. As a consequence, there was great variation (from 5 to 20 fold) in the actual AA used in hospitals in North America and worldwide [5]. This led to a substantial variation in pediatric AAs and consequently radiation absorbed dose (RD) in pediatric patients undergoing nuclear medicine examinations. The development of expert consensus guidelines has led to dosing consistency across hospitals, a reduction in AA and a corresponding reduction in RD [6, 7]. However, in a recent study, we showed that weight alone may not be the optimal method to determine an appropriate patient-specific AA. Other factors, such as patient local body habitus, have been shown to be more closely correlated with AA required to give desired defect detectability than weight in prescribing AA [8].

In theory, an increase or decrease in patient weight can be manifested by changes in height or size in areas that are distant from the organ being imaged and thus have very little effect on the factors that affect image

quality (IQ), i.e., attenuation, scatter, and spatial resolution from the change in the radius of rotation. In this work, we estimated a set of tradeoff curves for AUC as a function of AA as a function of patient girth for ^{99m}Tc-DMSA imaging using a simulated projection image database [9] developed from a realistic pediatric phantom population [10]. Human observer studies were conducted to calibrate the model observer to the task performance equivalent to an ensemble of human observers. These curves provide standard bodies IQ vs. AA tradeoff data that can be used to develop improved girth-based dosing guidelines for DMSA SPECT. These guidelines will better equalize IQ, as measured by the AUC, across patients with different body morphometries.

II. Materials and Methods

A. Population of realistic digital phantoms

This study used a dedicated series of computational phantoms developed at the University of Florida (UF) [10]. The population focused on modeling one anatomical variation – girth– that has been found to be the factor most correlated with diagnostic IQ in pediatric renal imaging using ^{99m}Tc-DMSA [11]. The phantom population is based on demographic and morphometric data from the CDC's National Health and Nutrition Examination Survey (NHANES) data. It consisted of 90 phantoms that varied in age, gender, waist circumference (girth), and kidney volume. For each gender, three age groups were modeled: 5, 10, and 15 years. At a given age, five girth percentiles were modeled: 5th (small), 25th, 50th (reference), 75th, and 95th (large). The masses and heights of the phantoms were determined by correlating waist circumference with the subject's mass to height ratio (from CDC data). First, nominal values of kidney volume, length, and depth were used for each phantom. In order to increase anatomical variations in the simulated data, we modeled two additional kidney sizes for each girth percentile: -15% and +15% of the standard mass, resulting in three phantoms for each specific combination of patient age, gender, and girth percentile. The phantoms were digitized using 0.1 cm cubic voxels. As pyelonephritis is not the primary target of disease for newborns and 1-year-old, we chose not to perform task-based image quality assessment for these two ages. The phantoms for each girth percentile of the 10-year-old female group are shown in Fig. 1.

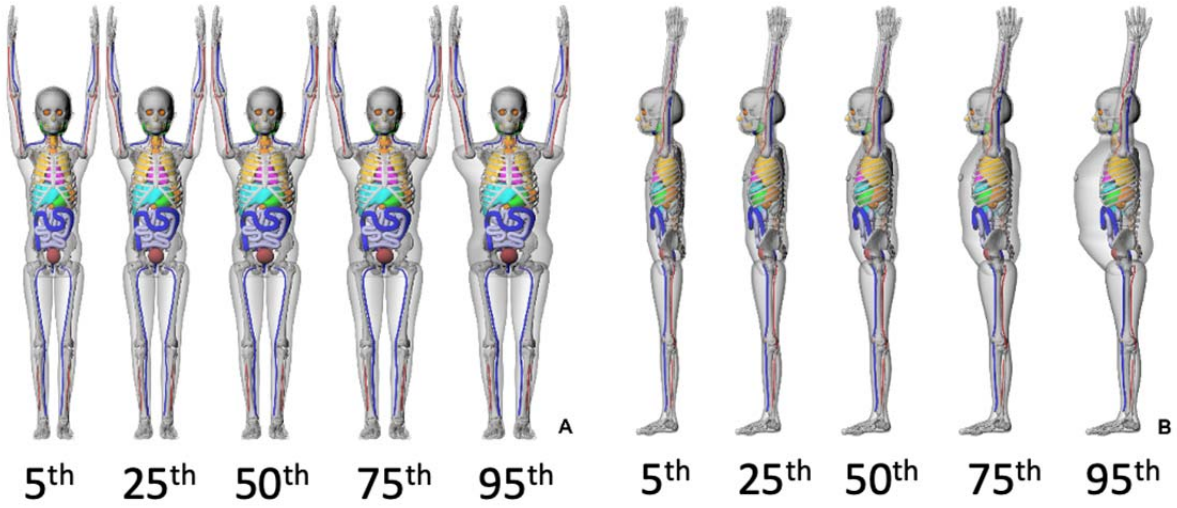

Fig. 1. **(A)** Frontal and **(B)** lateral views of the five 10-year female phantoms.

B. Model of organ uptake

The fraction of the administered activity in kidney, pelvis, and medulla was estimated from pediatric imaging studies used to develop a new pharmacokinetics model for ^{99m}Tc-DMSA [12]. Data from a total of 47 patients were used to calculate the mean, coefficient of variation, minimum, and maximum of the fractional tracer uptake in the kidneys, spleen, liver, and body remainder, at 3 hours post injection. These parameters are summarized in Table I. To model activity uptake variation in patients, we randomly sampled the renal activity and relative activity concentrations for structures inside the kidney (the renal cortex, medulla, and pelvis) from truncated Gaussian distributions with these parameters to generate different activity uptake realizations.

Table I. 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

C. *Defect model*

The model previously described in [11, 13] was used to simulate defects in the cortical wall of the left kidney and modeled volumes of reduced uptake consistent with focal, acute pyelonephritis. To simulate a defect severity that was clinically relevant across patients of different age and kidney size, we set the defect volume to be a fixed percentage of the kidney cortex volume. The used percentage was 0.49%, which was deemed by an experienced pediatric nuclear medicine specialist as clinically relevant and at the limits of clinical detectability in a 5-year-old phantom with reference girth. The same percentage was used in the other simulated ages (10- and 15-year-old) for which a defect detection task was considered. Using this model, we created four randomly located focal transmural renal defects in each of the following general regions in the right kidney cortex: upper pole, lower pole, and lateral. There were a total of 12 random locations for the defects generated in this study, modeling a (signal-known-statistically) SKS task.

D. *Projection data simulation*

We simulated noise-free projection data for each of six regions in each phantom - the renal cortex, medulla, and pelvis, liver, spleen, and background (including all other organs), modeling the physics and acquisition parameters appropriate for ^{99m}Tc-DMSA renal SPECT. An extensively validated analytic projection code that modeled attenuation, spatially varying collimator-to-detector response [14], and object-dependent scatter [15] was used to generated the projections. To model uptake variability inside the kidneys, we randomly sampled kidney activity and relative activity concentrations for structures inside the kidney from the above-mentioned truncated Gaussian distributions. These parameters for those distributions are given in [11].

We simulated a low-energy, ultra-high-resolution collimator as this collimator is most commonly used for these studies at Boston Children's Hospital (BCH), and is likely most appropriate for resolving the smaller structures present in pediatric patients. The projections were generated at 120 projection views over a 360° body-contouring orbit and a 0.2-cm projection bin size. Prior to simulation, the phantom was placed on a simulated patient bed obtained from a CT scan of a Siemens Symbia camera bed. This bed constrained the orbit, especially for small phantoms.

The final noise-free projections were obtained by scaling and summing the individual organ projections by their relative uptake values and the product of AA, acquisition duration, and scanner sensitivity. The simulated, noise-free projections were then scaled to represent different absolute AAs. Poisson noise was simulated using a Poisson distributed random number generator applied to the scaled, noise-free data.

E. Girth and kidney volume grouping

As discussed above, in [11], we showed that patient girth at the level of the kidneys is likely a better index to use in setting patient AA than weight. This can be understood by considering that an increase in patient weight can be manifested by changes in in height or size in areas of the body that are distant from the organ being imaged and that thus has little effect on diagnostic IQ. Defect size, though not measurable a priori, also affects the defect detectability due to the effects of spatial resolution. Thus, we grouped phantoms by a given range of girth and defect volume. The ensembles of phantoms in these groups were expected to have similar defect detectability (for a given AA). The groups were created by selecting phantoms with girths that are less than 6 cm apart and their defect volume within ±50% of the defect volume for the 50^{th} percentile girth phantom in that group. For each group, we investigated 4 to 6 AAs for the proposed girth-based dosing approach, aiming to cover a performance range (AUC= [0.75,0.95]) that is of clinical interest. For each phantom, 1152 activity uptake realizations were generated, resulting in 576 defect-present and defect-absent images for each AA simulated. In total we generated approximately 518,400 (90 phantoms×1152 realizations per phantom × 5 AAs) projections and reconstructions.

F. Image reconstruction and post-reconstruction processing

We followed the clinical reconstruction protocol routinely used at BCH, i.e., reconstructed using OS-EM iterative reconstruction algorithm with compensation for the geometric collimator-detector response and post-filtered with a Gaussian filter using a 5 mm FWHM. The reconstructed images had cubic voxels with a side length of 0.2 cm. We extracted images containing the defect centroid from the coronal, transaxial, and sagittal slices and used them in the IQ evaluation. The extracted images had a size of 128 × 128 pixels.

G. *Model observer*

A close relative of the channelized Hotelling observer (CHO), named multi-template channelized linear discriminant observer (MTCLDO) [16], was used in this study. The CHO, first proposed by Myers and Barrett for image quality evaluation [17], has been shown to provide good predictions of human performance on detection tasks for a variety of nuclear medicine imaging applications [18-21]. The MTCLDO is known to be more efficient than the CHO in handling data that are not described by multivariate normal (MVN) distributions from a population of phantoms with background and signal variations.

In the MTCLDO, the entire ensemble of input data vectors is decomposed into sub-ensembles that are approximately MVN. A different linear discriminant is then used to analyze each sub-ensemble. For each phantom, images that have the same defect location were grouped into the same sub-ensemble. The set of test statistics from each sub-ensemble was then pooled and analyzed using ROC analysis. A key element of this strategy is the use of the linear discriminant rather than the Hotelling Observer (HO) [16].

The test statistic of the MTCLDO is

$$\lambda_{MTCLDO}(\boldsymbol{v}) = \Delta\overline{\boldsymbol{v}}^{T}\boldsymbol{K}_{v}^{-1}\,\boldsymbol{v} + \frac{1}{2}(\overline{\boldsymbol{v}_0}^{T}\boldsymbol{K}_{v}^{-1}\overline{\boldsymbol{v}_0} - \overline{\boldsymbol{v}_1}^{T}\boldsymbol{K}_{v}^{-1}\overline{\boldsymbol{v}_1}) \quad (1)$$

where $\boldsymbol{v} \in \mathbb{R}^{S\times 1}$ represents an S dimensional (channel output) feature vector $\overline{v_m}$ and $\boldsymbol{K}_v$are the mean feature vector and covariance matrix, respectively, over the sub-ensemble data and under hypothesis $H_{m,}(m = 0,1)$.

In the MTCLDO strategy, channel output vectors were sorted into sub-ensembles based on defect location, age, gender, girth percentile, and kidney size. This decomposition was used so that the distributions of channel output vectors in each sub-ensemble were approximately MVN. It was verified visually that the resulting channel output data were not multi-modal and were nearly MVN distributed.

Training the linear discriminate means computing the means and co-variances for some set of input data vectors. After training, the discriminant is applied to an (ideally different) ensemble of data, the testing set, to

produce a test statistic for each member of the testing set. We used a leave-one-out training-testing strategy [16]. In this strategy, one feature vector is left out (i.e., not used in the training), and the remaining vectors are used to train the observer. The observer is then applied to the left-out vector to produce a test statistic. This process was repeated with each vector in the ensemble being left out once. This process was applied to each sub-ensemble. The resulting test statistics for each sub-ensemble were pooled, internal noise was added, and ROC analysis was performed as described below. The same non-overlapping rotationally symmetric difference-of-mesa (DOM) channels described in [11] were used in this study. The input images were those described in Section F and had the defect centroids at the center of the image. These feature vectors served as inputs to the MTCLDO described above.

H. Human observer study for calibration of MTCLDO

We conducted a limited number of human observer studies in an attempt to calibrate the MTCLDO's performance to that of an average nuclear medicine radiologist. Note that the calibration method used for the MTCLDO has not been fully validated as representative of human observers. A more comprehensive study is needed to fully validate the MTCLDO's ability to emulate human observer performance.

The human observer studies were conducted using volumetric displays of images very similar to the display used clinically at the BCH (Fig 2). Two experiments were performed in the human observer study using a 5-year-old male phantom and the same AA (35.15 MBq (0.95mCi)). The first used a phantom with the reference girth and the second used one with the largest girth. The result of the first experiment was used to calibrate the MTCLDO. Internal noise as described in [22] was added to the MTCLDO's test statistic to calibrate the performance of the MTCLDO to that of the human observer for the first experiment. This final test statistic can be expressed using the following equation:

$$\lambda_{Final}(\boldsymbol{v}) = \lambda_{MTCLDO}(\boldsymbol{v}) + n, \qquad (2)$$

where n is a normally distributed random variable with zero mean and a standard deviation of $\alpha \times \sigma_{\lambda\_bkg}$. Here α is a constant that controls the magnitude/amount of the additive internal noise and σ_{λ_bkg} is the standard deviation of test statistics from the defect-absent images.

The calibrated amount of internal noise was determined by comparing model observer AUC value to the average AUC of the human observers on the 5-year-old reference girth male phantom injected with weight-based AA recommended by the North American Consensus Guidelines. It was found that $\alpha = 1.39$ provided the smallest difference between the average AUC of the human observers and the AUC of the model observer. The value of α was then used for the model observer calculations at all other AA levels and defect sizes. The goal of the second experiment was to examine whether the amount of internal noise from the first experiment was able to capture the correct amount of human observer performance drop due to the girth difference and the answer was positive.

In the study, the observer was asked to rate their confidence that a defect was present on a continuous scale ranging from 1 to 5, with the highest number representing the greatest confidence that a defect was present. To familiarize themselves with the display program and the nature of the clinical defect detection task, all observers participated in an initial training session comprised of 24 images. In the training session, phantom images of the kidney cortex were provided as ground truth to the observers once their rating value was recorded.

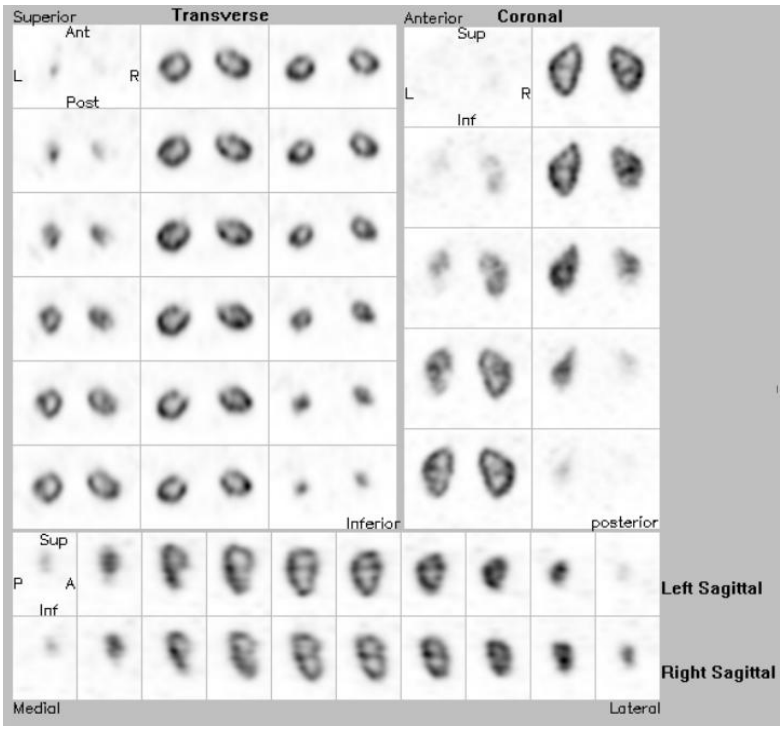

Fig. 2. A sample 48-slice image shown in the volumetric display format routinely used in clinical practice at the Boston Children's Hospital.

One nuclear medicine radiologist, one research professor in Radiology, and two senior medical imaging physics Ph.D. students participated in the human observer study. A total of 384 composite images were used. To simulate an SKS detection task, the training and test datasets did not use the exact same set of defect locations. Thus, the test dataset could contain defect locations that were not present in the initial training dataset. The images were divided into an initial training set and three test blocks. The block layout for each observer is shown in Table II. At the beginning of each test block, a refresher set of 24 images was provided to refresh the observer's memory about the task. A total of 288 rating values were collected from each observer. The AUC values from the four observers were averaged to determine the average AUC of the human observers.

Table II. Summary of human observer study session partition

Session	Initial training images	Images/session	Total images
	24	24 training	24
1	0	24 training/96 test	120
2	0	24 training/96 test	120

3	0	24 training/96 test	120

I. ROC and statistical analysis

We applied the MTCLDO to feature vectors in the sub-ensembles described above. For each phantom, there were 96 images (1152 realizations for each phantom divided by 12 defect locations) in a sub-ensemble, including both defect-present and absent images. We first pooled the test statistics for different defect locations for each phantom and then sampled the pooled test statistics by bootstrapping with replacement; normally distributed internal noise was added to the test statistics [22]. For each phantom group, as defined above, we pooled each phantom's test statistics (with noise) and calculated one AUC value. The pooled test statistics were analyzed using the LABROC4 code [23] to estimate ROC curves. Bootstrapping and nonparametric analysis were used to compute 95% confidence intervals of the mean AUC. Plots of the AUC as functions of absolute AA for different groups of phantoms are shown in the results section and Appendix. We fitted the AUC vs AA data with a previously-derived empirical relationship [11]:

$$AUC = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\sqrt{\frac{AA \times K_1}{AA \times K_2 + K_3}}}{2}\right). \qquad (3)$$

where K_1, K_2, K_3 represents the mean signal difference, the object variability noise, and the quantum noise, respectively.

III. RESULTS

A. Pilot study on how defect volume and girth together affect task-based IQ

In the pilot study, we picked a 5-year-old and 10-year-old phantom with almost identical girths but drastically different weights, heights, and kidney volumes. Fig. 3.a shows the central sagittal plane of both phantoms. We injected the phantoms with a fixed AA of 38.48 Mbq which is the AA recommended by the North American Consensus Guidelines for the 5-year-old phantom. The biometric properties of the phantoms are summarized in Table III.

Table III. Summary of biometric properties of phantoms in the pilot study

	Girth (cm)	Weight(kg)	Kidney volume(cm^3)	Height(cm)
5-year-old	54.77	20.71	86	130.3
10-year-old	55.85	26.06	101	171.9
Δ	2%	26%	17%	32%

To study the effect of defect volumes on task-based IQ, we simulated three defect volumes on each of the phantoms: 0.2, 0.15, and 0.1 cm^3, resulting in a total of 6 different phantoms. 1152 activity uptake realizations were generated for each phantom. No internal noise was added in the pilot study. The ROC and statistical analysis method described in section I was used. The results of the CHO study and area under the ROC curve (AUC) values for the 6 phantoms are shown in Fig. 3.b. The plot showed that, the 5- and 10-year-old phantoms, which had very similar girths but large differences in weight, heights and kidney volumes had very similar (non-statistically significantly different) AUC values for all three defect volumes. Overall, the pilot study showed that girth and defect volume are the two most significant factors affecting task-based IQ, and confirmed that dosing based on girth may be more effective at equalizing image quality than dosing based on weight [10, 11, 24].

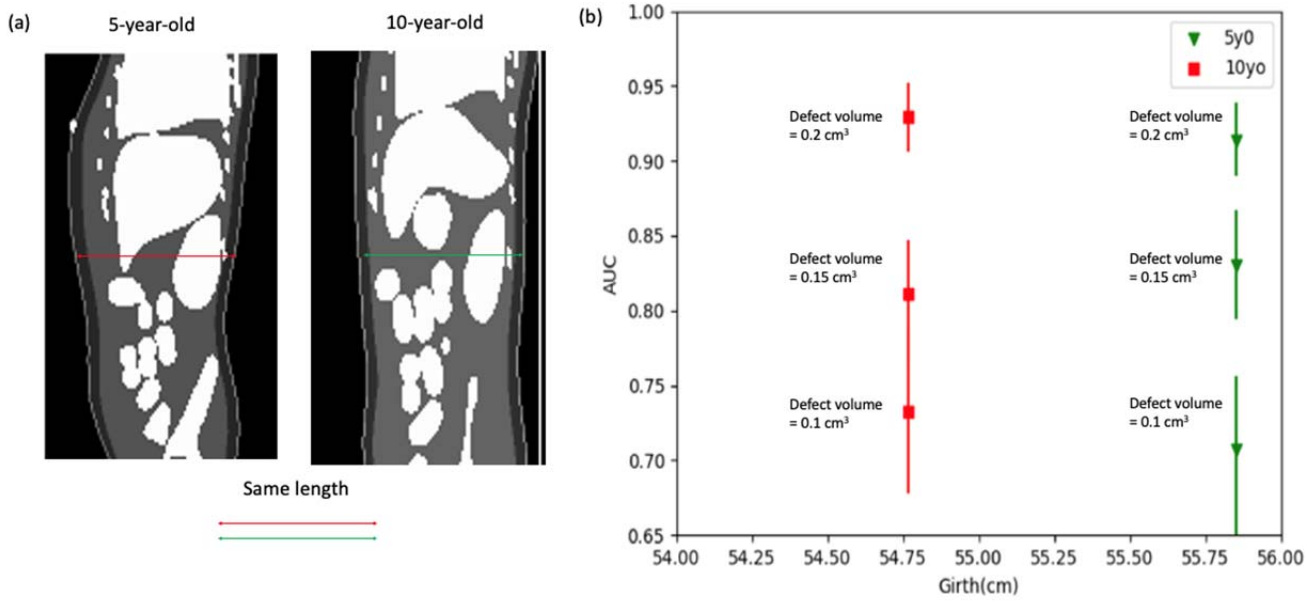

Fig. 3. Central sagittal view of the 5-year-old and 10-year-old phantom in (a) and in (b) AUC values for all three defect volumes.

B. The correlation between task-based IQ and AA for patients with different girth and defect volume

The results of the CHO study and area under the ROC curve (AUC) values for some selected ~~phantom~~ girth groups as a function of absolute AAs are included in Fig. 4.a and in the Appendix. The data points in each subplot were fitted to the empirical AUC vs. AA relationship above. To provide a quantitative comparison between the proposed girth-based dosing method and the current weight-based dosing method (North American Consensus Guidelines), we calculated the AAs recommended by these two dosing methods, for providing the same task-based image qualities (AUC = 0.8) for a realistic pediatric phantom population. The baseline AUC chosen here was to provide a quantitative comparison. Selecting an AA requires taking into account the risk of misdiagnosis and the radiation exposure risk. Thus, the decision on IQ for a given defect size and corresponding AA is left to the nuclear medicine physician.

In general, our results showed that the proposed local body morphometry-based dosing method suggested significantly different AAs than the current weight-based dosing method for the phantoms who have a relatively large body weight but a small girth or phantoms with relatively small bodyweight but large girth. In

Fig. 4.a, the results demonstrate that, using the proposed girth-based dosing method, that the baseline IQ can be achieved using an AA of 32.65 MBq for all the phantoms in the girth group specified in Fig. 4.a, which is 62.9% less than the AA (88.10 MBq) recommended by the current weight-based dosing guideline for the heaviest phantom in this girth group.

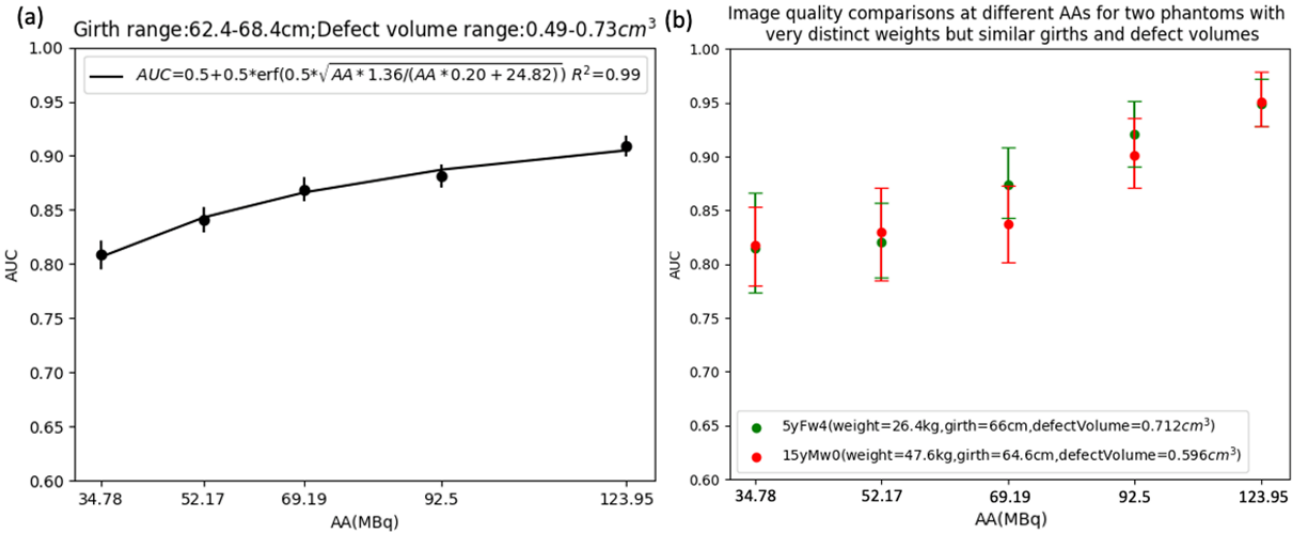

Fig. 4. AUC vs. AA curves and the fitted functions for the patients in (a) the girth range of 62.4 – 68.4 cm and defect volume range of 0.49 – 0.73 cm^3 and in (b) AUC vs. AA plot for the heaviest and lightest phantom in the corresponding girth bin.

We computed the difference in weight between the heaviest and lightest phantom in each girth group. For example, in the girth range used in Fig. 4, the heaviest and lightest phantoms were a 15-year-old male and a 5-year-old female phantom. These two phantoms have weights that differ by almost a factor of 2 (47.6 kg vs 26.4 kg), but similar girth and kidney sizes. In Fig. 4.b, we show that the IQs from these two phantoms are not statistically significantly different at multiple AA levels. This is in agreement with the findings in [24] and is consistent with the hypothesis that the girth at the location of the kidney would have a more direct effect than weight on attenuation, scatter, and spatial resolution, and thus IQ, given the same AA. Fig. 5 shows that qualitatively the defect visibility degrades significantly as the AA decreases for both the 15- and 5-year-old, but is relatively equal for the two phantoms, supporting the quantitative IQ results.

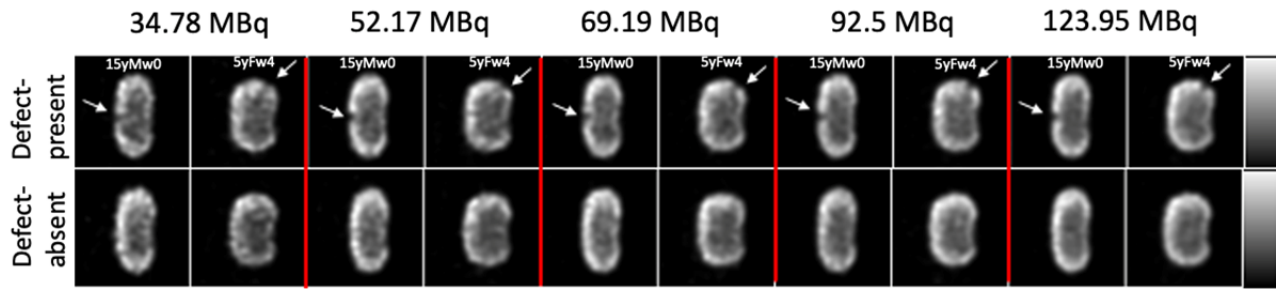

Fig. 5. Coronal views of the reconstructed images at different AAs for the heavist (15yMw0: 15 yo male with 5th girth percentile) and lightest (5yFw4: 5 yo female with 95th girth percentile) phantoms in the girth range of 62.8-68.4 cm. The top and bottom rows show the defect-present and absent images, respectively. White arrows indicate the defect locations.

A summary of AAs needed to achieve the baseline AUC of 0.80 is included in Table IV for patients of different girth ranges and defect volumes. However, physicians must ultimately decide on the appropriate AUC value to use for this diagnostic task. This is because the AUC can be increased by increasing the AA or the time required to perform the study; increasing AA results in increased radiation dose to the patient and increased acquisition time increases the chance of motion or might necessitate anesthesia. Thus, person who is knowledgeable about the tradeoff between diagnostic IQ, radiation dose and the cost of diagnostic errors must ultimately decide on the tradeoff between IQ and AA. Overall, these results suggest that substantial reductions in AA (as compared to the current weight-based dosing method) for heavy patients with waist circumferences within this girth range. The results again confirmed that patient weight may not be the optimal factor to use to recommend AA. Instead, as demonstrated in this study, body morphometry factors that are local to the target imaging organ should be used.

Table IV. Summary of AAs for different girth ranges and defect volume ranges to achieve a desired task-based image quality as measured by the AUC

Girth range (cm)	Defect volume range (cm^3)	AA for the average baseline IQ (AUC=0.80) (MBq)
50.4-56.4	0.41-0.62	28.85
56.4-62.4	0.54-0.81	24.9
62.4-68.4	0.49-0.73	32.65
62.4-68.4	0.74-1.11	31.45

74.4-80.4	0.85-1.27	37.92
86.4-92.4	1.10-1.51	53.55

321

322 To compare the proposed dosing method with weight-based dosing method, we listed the AAs
323 recommended by the North American Guidelines (NAG, subject to their maximum and minimum) for the
324 heaviest (red) and lightest (green) phantoms in each girth and defect volume range in Table V. Their
325 corresponding task-based image quality (AUC) were calculated using the fitted equation in the form of eqn. 3.
326 The results showed that the NAG recommended AAs produced variable image quality for patients of different
327 weights, consistent with our previous finding [11]. In addition, besides for the lightest 5-year-old phantom in
328 the girth range of 50.4-56.4, the AUCs for the AAs recommended by the NAG surpassed the baseline IQ
329 (AUC=0.80) consistently, indicating that a significant amount of AA can be saved if the desired IQ chosen is
330 consistent with the baseline IQ. Based on the results of this study, different amounts of decreases were
331 indicated for some of the 5-year-old phantoms and all the 10- and 15-year-old phantoms, respectively. A
332 reduction of AA up to 62.9% compared to weight-based dosing could be used (while maintaining an AUC of
333 0.80) for a 15-year-old phantom with relatively large bodyweight but small amount of weight distributed
334 around the organ being imaged. In contrast, up to 8.8% less AA was indicated for certain a 5-year-old phantom
335 with relatively small bodyweight but more weight distributed around the kidneys.

Table V. Comparisons of image quality for AAs recommended by NAG and by the proposed method

Girth range (cm)	Defect volume range (cm^3)	Girth-based AA for baseline IQ (AUC=0.8) (MBq)	Gender	Age (yr)	Waist Circumference (cm)	Body Mass (kg)	Simulated Defect Volume (cm^3)	NAG (MBq)	Task-based IQ for AA by NAG (AUC)	% Difference of Girth-based AA over AA by NAG
50.4-56.4	0.41-0.62	28.85	m	10	54.33	26.39	0.49	48.82	0.85	40.9%
			m	5	50.40	17.91	0.49	33.14	0.81	8.8%
56.4-62.4	0.54-0.81	24.9	f	10	60.00	30.62	0.63	56.66	0.88	56.1%
			m	10	58.90	30.22	0.67	55.90	0.88	55.5%

62.4-68.4	0.49-0.73	32.65	m	15	64.60	47.62	0.60	88.10	0.88	62.9%
			f	5	66.00	26.42	0.71	48.88	0.84	33.2%
62.4-68.4	0.74-1.11	31.45	f	10	66.40	36.20	0.88	66.98	0.87	53.0%
			f	5	66.00	26.42	0.82	48.88	0.85	35.7%
74.4-80.4	0.85-1.27	37.92	f	15	77.20	59.96	1.04	100.00	0.89	62.1%
			m	10	75.60	44.17	1.00	81.72	0.88	53.6%
86.4-92.4	1.10-1.51	53.55	m	15	89.30	76.11	1.43	100.00	0.84	46.5%
			m	10	91.20	57.21	1.18	100.00	0.84	46.5%

IV. DISCUSSION

We have examined the implications of using a local body morphometry metric with a well-defined DMSA diagnostic imaging task for pediatric patients to recommend AA. The results showed that substantial (up to 62.9%) reductions in AA will yield an image quality consistent with that resulting from NAG AA if circumference around the waist is used rather than patient weight in determining AA. This result is heavily dependent on assumed defect size. This is illustrated by the results obtained for a 5-year-old wherein the small kidney size yields a defect size far smaller than typically encountered clinically. This corresponds to a 0.41cm^3 defect. More reduction of AA could potentially be achieved if a less challenging/larger and clinically more relevant defect percentage was modeled in the phantom population. Highlighted in Table IV are two identical phantoms that were simulated with different defect size. The results showed 3.8% of AA can be saved (while maintaining the same diagnostic performance) for the phantom with a larger defect (0.82 cm^3) compared to the same phantom with a smaller defect (0.71 cm^3).

There are three limitations of the work: 1) a single defect kind (focal defect simulating pyelonephritis) was modeled in this work so the model of the defect may not reflect the complexity of defect shapes seen in the clinic, 2) a single fractional defect size was used, and the size selected was deliberately challenging by one expert pediatric nuclear medicine physician, which might not represent the consensus opinion of a group of average nuclear medicine physicians, and 3) the MTCLDO has not been fully validated against humans, and so

that the results may not be accurate in an absolute sense. These limitations suggest the need for identifying defect sizes and kinds (other than focal defects) in pediatric molecular imaging that are more clinically relevant as well as the need for validating (in terms of reproducing human observer performance) numerical observers that can handle non-multivariate normal data, which was the case of this data as there were a lot of anatomical variations in the phantom library.

On a practical level, the results show that a better risk-benefit balance will be achieved by measuring the patient's waist circumference in conjunction with a table of AAs to meet a given IQ (AUC) value for the most likely defect size. Table IV provides an example of such a scheme. However, note that the comparison between the girth- and weight-based dosing approach shown in Table V can be somewhat limited in its fairness as 1) IQ in the weight-based dosing method is not measured quantitatively but by expert consensus, and 2) the results are solely based on simulation data instead of real clinical practice. Further investigation and validation of the results present in the study are eagerly needed in future clinical studies. Importantly, the decision on IQ for a given size and corresponding AA is left to the nuclear medicine physician, as selecting an AA requires taking into account both the risk of misdiagnosis and the radiation exposure risk. This work provides data on the misdiagnosis part of the equation, which allows the physician to weigh IQ/defect detectability with other, patient-specific considerations in selecting an AA. The data on patient radiation risk is provided in [10]. Together, we present a complete IQ-Risk-AA relationship to physicians to select an AA based on local-body-morphometry measures of the patient.

V. CONCLUSIONS

A local-body-morphometry (girth)-based index for selecting DMSA AA for pediatric renal SPECT shows substantial reductions in AA relative to a weight-based AA specification method. A girth-based dosing scheme may be implemented by replacing the body weight with a measure of waist circumference, selecting a desired IQ (AUC) and selecting the AA from a tradeoff curve corresponding to the patient's girth. With an adequate representation of expected defect size and future clinical validation of the results present in the paper, this

approach may be used to inform task-based guidelines in the broader context of the clinical practice of pediatric SPECT.

Acknowledgements

This work was supported by National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health 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 statement

No potential conflicts of interest relevant to this article exist.

References

1. Lassmann, M. and S.T. Treves, *Pediatric Radiopharmaceutical Administration: harmonization of the 2007 EANM Paediatric Dosage Card (Version 1.5.2008) and the 2010 North American Consensus guideline.* Eur J Nucl Med Mol Imaging, 2014. **41**(8): p. 1636.
2. Gelfand, M.J., et al., *Pediatric radiopharmaceutical administered doses: 2010 North American consensus guidelines.* J Nucl Med, 2011. **52**(2): p. 318-22.
3. Treves, S.T., M. Lassmann, and E.S.P.D.H.W. Group, *International guidelines for pediatric radiopharmaceutical administered activities.* J Nucl Med, 2014. **55**(6): p. 869-70.
4. Treves, S.T., et al., *2016 Update of the North American Consensus Guidelines for Pediatric Administered Radiopharmaceutical Activities.* J Nucl Med, 2016. **57**(12): p. 15N-18N.
5. Treves, S.T., R.T. Davis, and F.H. Fahey, *Administered radiopharmaceutical doses in children: a survey of 13 pediatric hospitals in North America.* J Nucl Med, 2008. **49**(6): p. 1024-7.
6. Fahey, F.H., et al., *Effects of Image Gently and the North American guidelines: administered activities in children at 13 North American pediatric hospitals.* J Nucl Med, 2015. **56**(6): p. 962-7.
7. Fahey, F.H., et al., *Dose Estimation in Pediatric Nuclear Medicine.* Semin Nucl Med, 2017. **47**(2): p. 118-125.
8. Sgouros, G., et al., *An approach for balancing diagnostic image quality with cancer risk: application to pediatric diagnostic imaging of 99mTc-dimercaptosuccinic acid.* J Nucl Med, 2011. **52**(12): p. 1923-9.
9. Li, Y., et al., *A projection image database to investigate factors affecting image quality in weight-based dosing: application to pediatric renal SPECT.* Phys Med Biol, 2018. **63**(14): p. 145004.
10. Justin L. Brown, B.S.-S., Ye Li, Eric C. Frey, S. Ted Treves, Frederic and D.P. H. Fahey, Xinhua Cao, George Sgouros, Wesley E. Bolch, *Body morphometry appropriate computational phantoms for dose and risk optimization in pediatric renal imaging with Tc-99m DMSA and MAG3.* Physics in Medicine and Biology, 2020.
11. Li, Y., et al., *Current pediatric administered activity guidelines for (99m) Tc-DMSA SPECT based on patient weight do not provide the same task-based image quality.* Med Phys, 2019. **46**(11): p. 4847-4856.

12. Plyku, D., et al., *Renal (99m)Tc-DMSA pharmacokinetics in pediatric patients.* EJNMMI Phys, 2021. **8**(1): p. 53.
13. Li, Y., et al., *; Development of a Defect Model for Renal Pediatric SPECT Imaging Research.* 2015 Ieee Nuclear Science Symposium and Medical Imaging Conference (Nss/Mic), 2015.
14. Frey, E.C., Z.W. Ju, and B.M.W. Tsui, *A Fast Projector-Backprojector Pair Modeling the Asymmetric, Spatially Varying Scatter Response Function for Scatter Compensation in Spect Imaging.* Ieee Transactions on Nuclear Science, 1993. **40**(4): p. 1192-1197.
15. Frey, E.C. and B.M.W. Tsui, *A new method for modeling the spatially-variant, object-dependent scatter response function in SPECT.* 1996 Ieee Nuclear Science Symposium - Conference Record, Vols 1-3, 1997: p. 1082-1086.
16. Li, X., et al., *Use of Sub-Ensembles and Multi-Template Observers to Evaluate Detection Task Performance for Data That are Not Multivariate Normal.* IEEE Trans Med Imaging, 2017. **36**(4): p. 917-929.
17. Myers, K.J. and H.H. Barrett, *Addition of a channel mechanism to the ideal-observer model.* J Opt Soc Am A, 1987. **4**(12): p. 2447-57.
18. Sankaran, S., et al., *Optimum compensation method and filter cutoff frequency in myocardial SPECT: A human observer study.* Journal of Nuclear Medicine, 2002. **43**(3): p. 432-438.
19. Gifford, H.C., et al., *Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging.* J Nucl Med, 2000. **41**(3): p. 514-21.
20. Yao, J. and H.H. Barrett, *Predicting Human-Performance by a Channelized Hotelling Observer Model.* Mathematical Methods in Medical Imaging, 1992. **1768**: p. 161-168.
21. He, X., J.M. Links, and E.C. Frey, *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.* Physics in Medicine and Biology, 2010. **55**(17): p. 4949-4961.
22. Leng, S., et al., *Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain.* Medical Physics, 2013. **40**(8).
23. Metz, C.E., B.A. Herman, and J.H. Shen, *Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data.* Stat Med, 1998. **17**(9): p. 1033-53.
24. Ye Li, J.B., Shannon O'Reilly, S Ted Treves, Frederic Fahey, Xinhua Cao, Briana Sexton-Stallone, Wesley Bolch, George Sgouros, Eric Frey. *Patient Girth is Better than Weight for Selecting Administered Activity in Renal Pediatric Imaging*. in *SNMMI*. 2019. Society of Nuclear Medicine.

Appendix

The figures below show the plots in absolute AA in the units of MBq for a variety of different girth and defect volume range combinations; the theoretical relationship that we derived (eqn.3) fitted well. These fitted functions provide an analytic relationship between AA and AUC and could be used to compute the AA required to yield a desired AUC for a given patient weight. To find out how much AA to give for a desired AUC, we horizontally sliced through a fixed AUC and recorded the AAs that are needed to achieve the AUC from the fitted functions. This IQ-dependent AA selection method may be applicable to other data for other imaging agents.

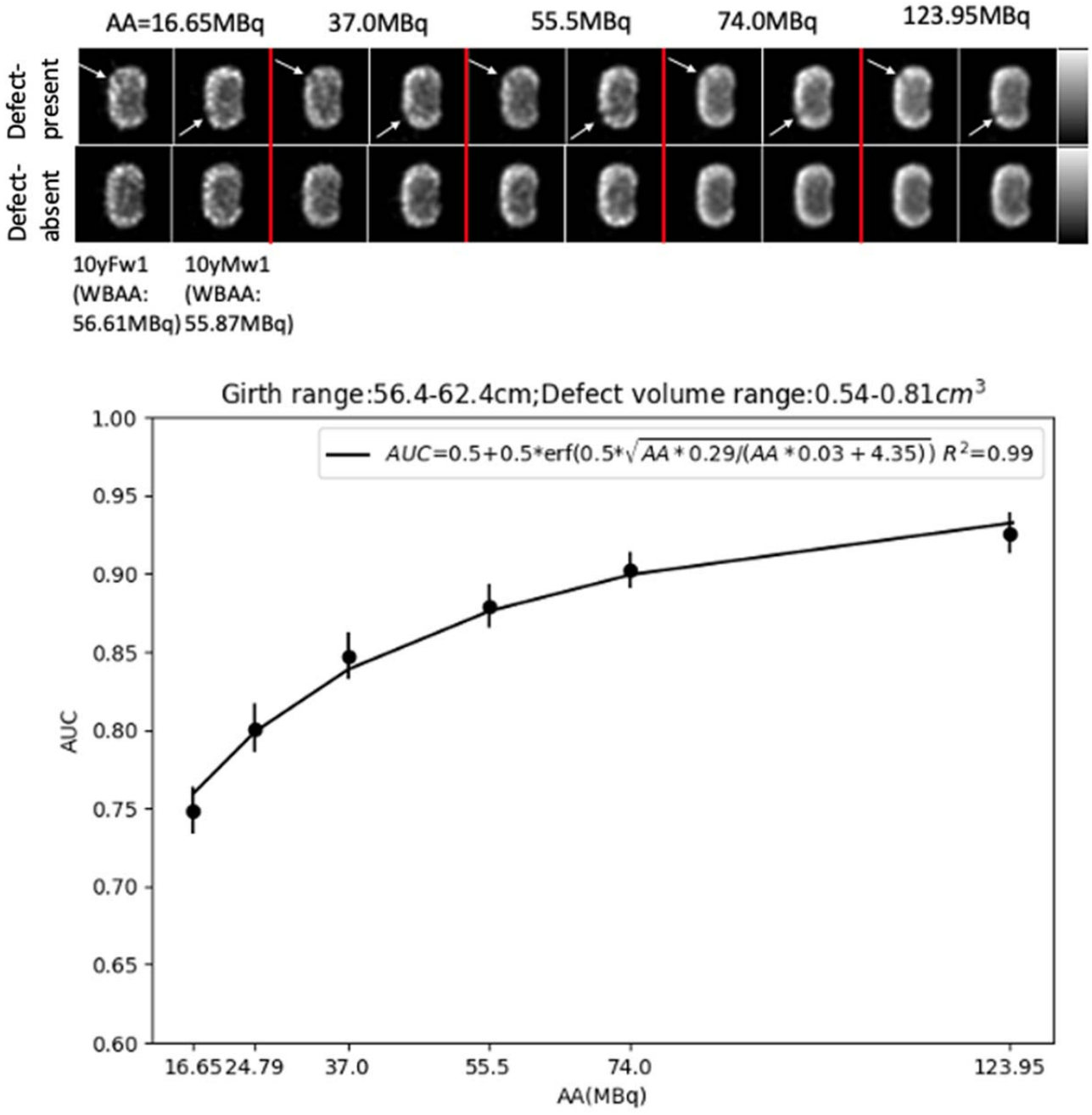

Fig. A.1. From left to right the images show coronal view of images generated using different AAs from the heaviest (10yFw1: 10 yo female with 25th girth percentile, 30.6 kg) and lightest (10yMw1: 10 yo male with 25th girth percentile, 30.2 kg) phantom in the girth bin. AUC vs. AA curves and the fitted functions for the patients in the girth range of 56.4 – 62.4 cm and defect volume range of 0.54 – 0.81 cm^3

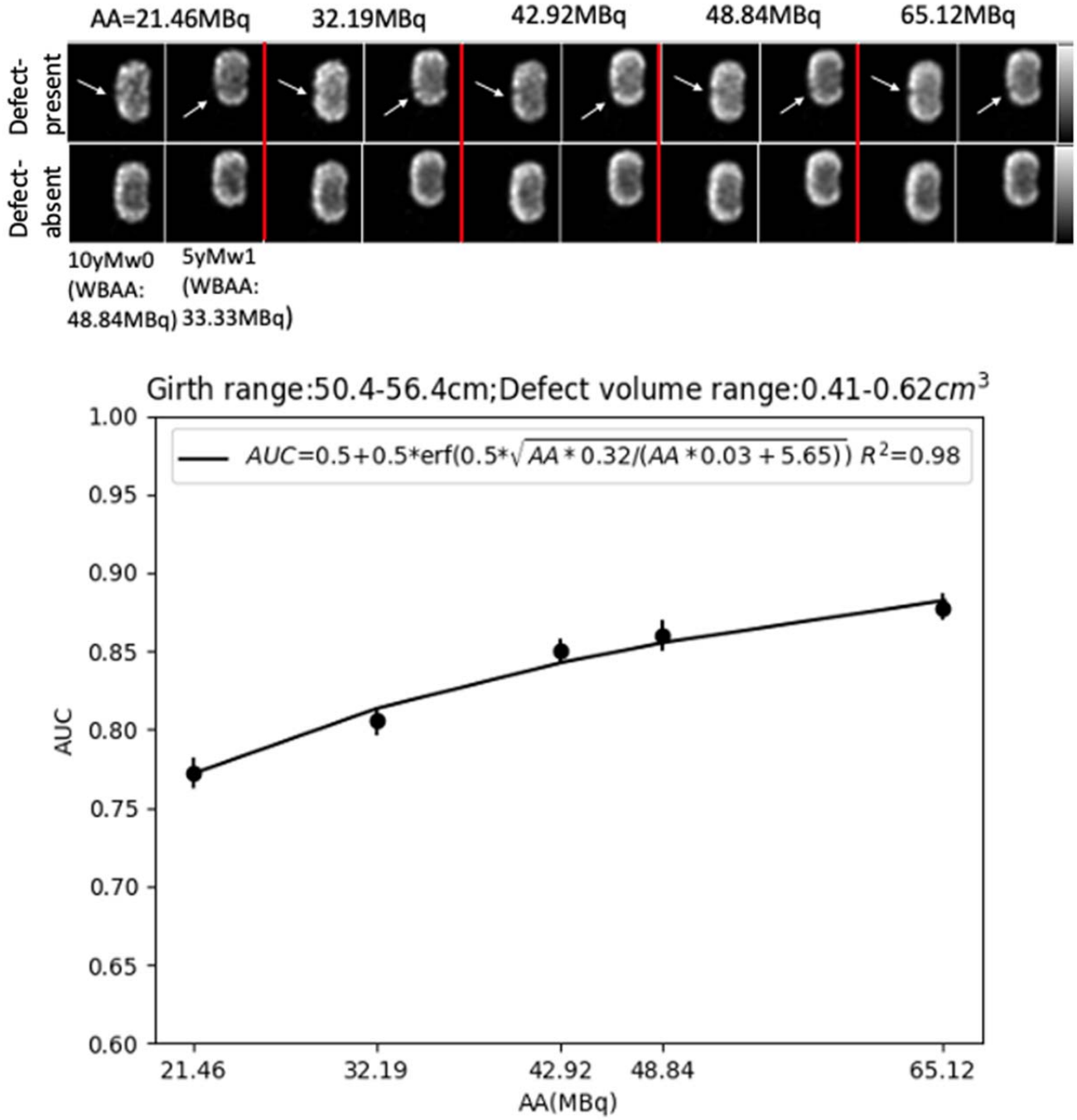

Fig. A.2. From left to right the images show coronal view of images generated using different AAs from the heaviest (10yMw0: 10 yo male with 5th girth percentile, 26.4 kg) and lightest (5yMw1: 5yo male with 25th girth percentile, 18.0 kg) phantom in the girth bin. AUC vs. AA curves and the fitted functions for the patients in the girth range of 50.4 – 56.4 cm and defect volume range of 0.41 – 0.62 cm^3.

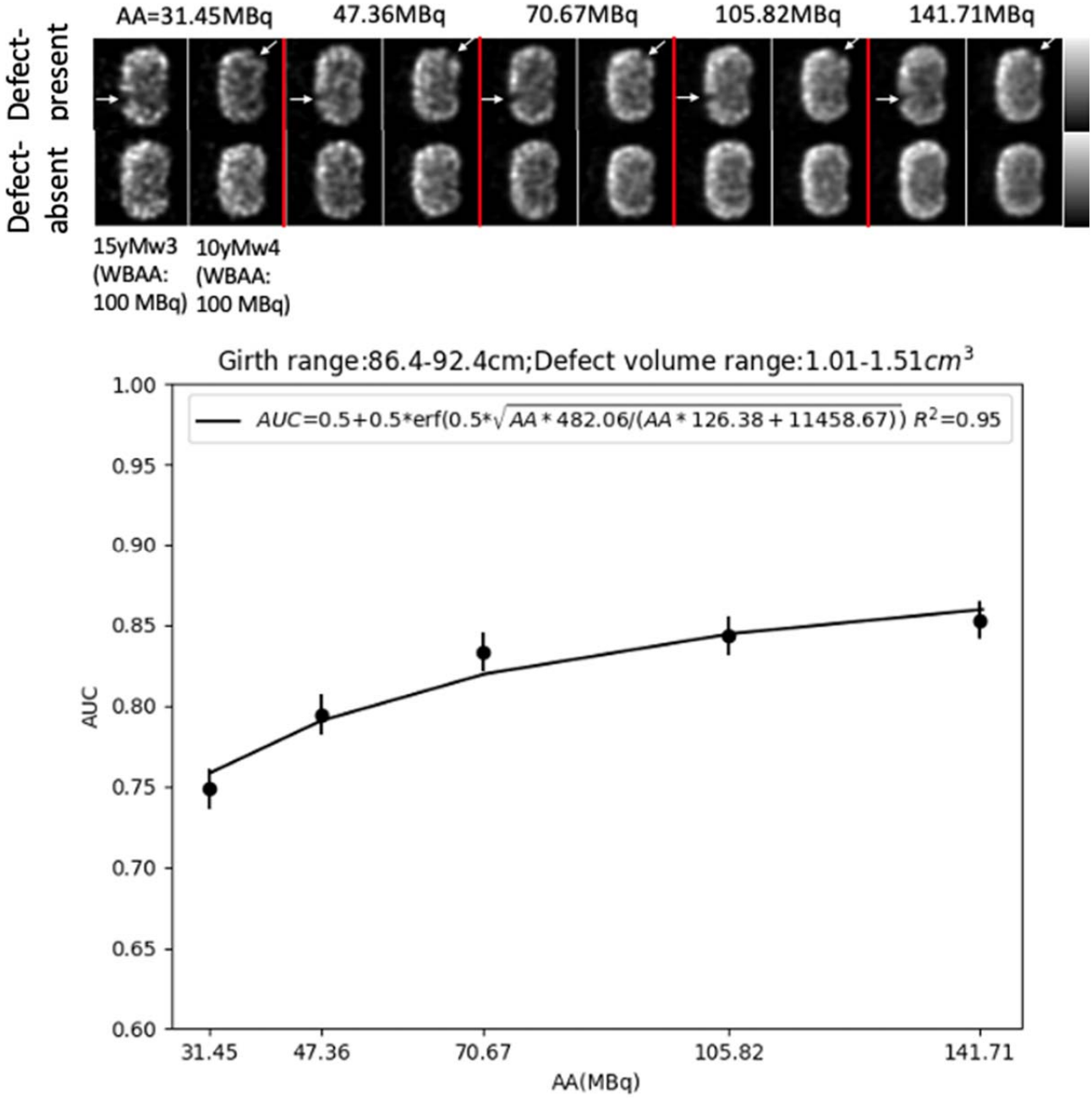

Fig. A.3. From left to right the images show coronal view of images generated using different AAs from the heaviest (15yMw3: 15 yo male with 75th girth percentile, 76 kg) and lightest (10yMw4: 10 yo male with 95th girth percentile, 57 kg) phantom in the girth bin. AUC vs. AA curves and the fitted functions for the patients in the girth range of 86.4 – 92.4 cm and defect volume range of 1.01 – 1.51 cm^3.

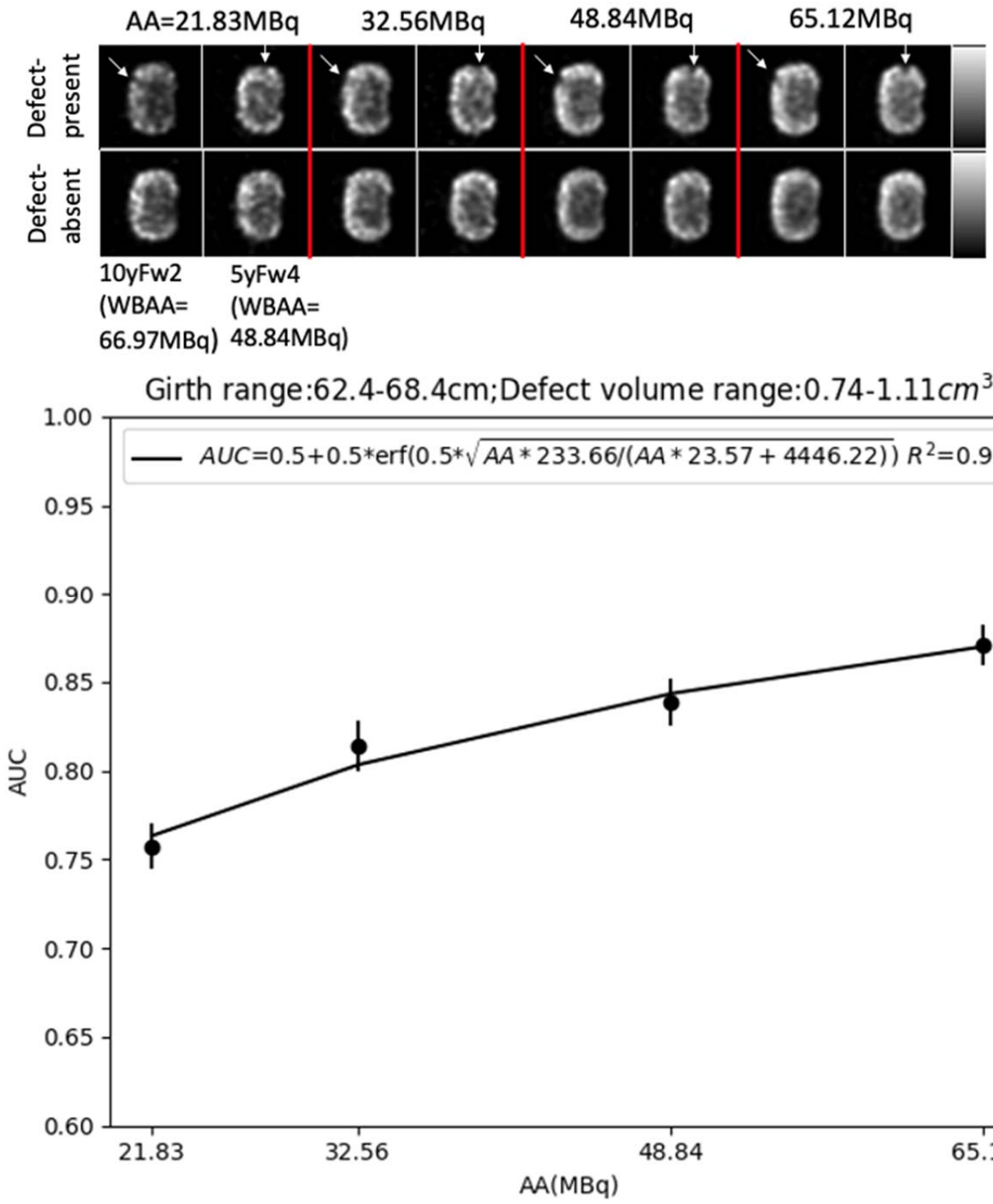

Fig. A.4. From left to right the images show coronal view of images generated using different AAs from the heaviest (10yFw2: 10 yo female with 50th girth percentile, 36.2 kg) and lightest (5yFw4: 5 yo female with 95th girth percentile, 26.4 kg) phantom in the girth bin. AUC vs. AA curves and the fitted functions for the patients in the girth range of 62.4 – 68.4 cm and defect volume range of 0.74 – 1.11 cm^3.

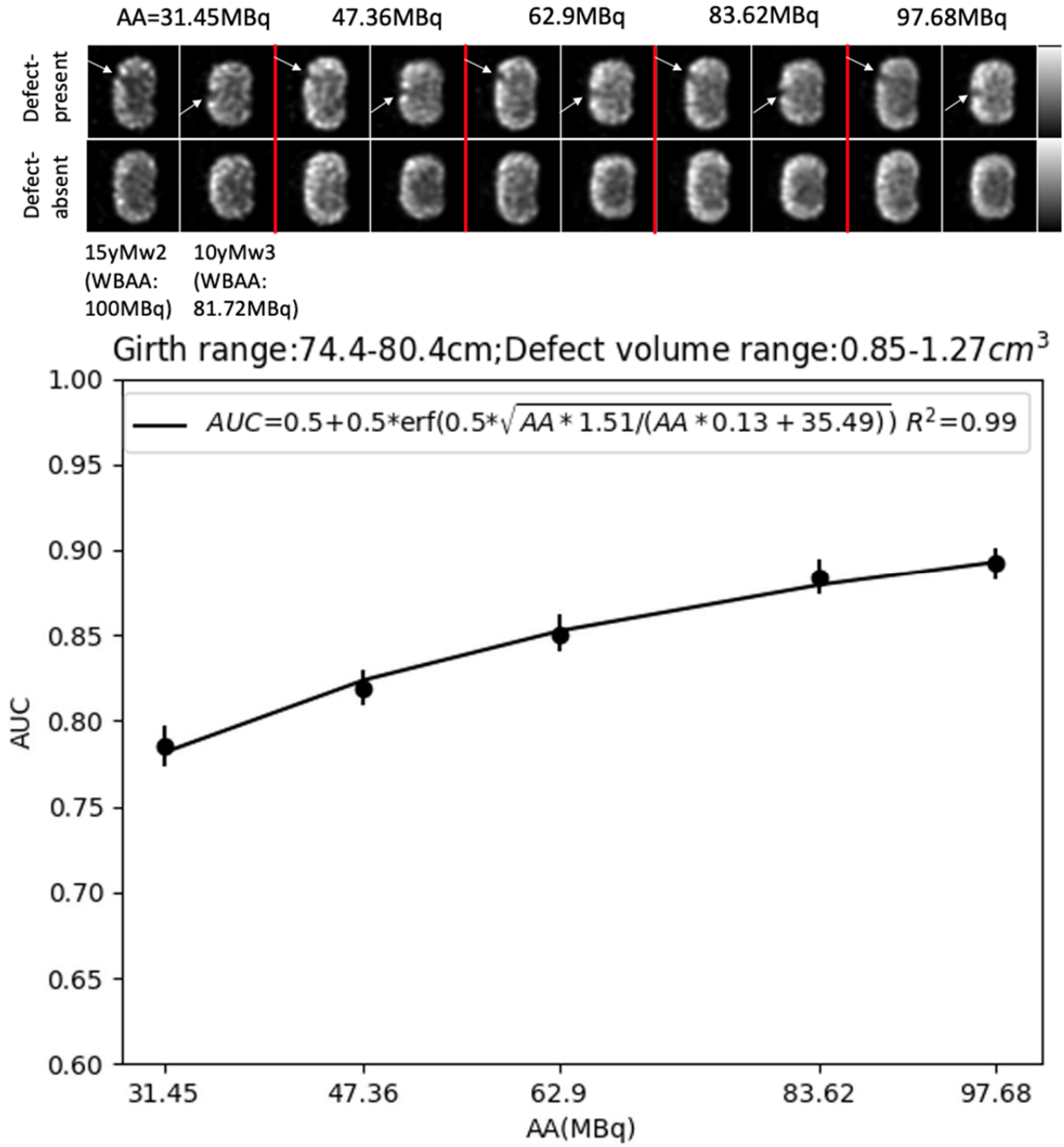

Fig. A.5. From left to right the images show coronal view of images generated using different AAs from the heaviest (15yMw2: 15 yo male with 50th girth percentile, 61.2 kg) and lightest (10yMw3: 10 yo male with 75th girth percentile, 44.17 kg) phantom in the girth bin. AUC vs. AA curves and the fitted functions for the patients in the girth range of 74.4 – 80.4 cm and defect volume range of 0.85 – 1.27 cm^3.