Assessment of different patient-to-phantom matching criteria applied in Monte Carlo–based computed tomography dosimetry

Elliott J. Stepusin

J Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611-6131, USA

Daniel J. Long

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

Emily L. Marshall and Wesley E. Bolch^{a)}

J Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611-6131, USA

(Received 9 February 2017; revised 14 June 2017; accepted for publication 25 July 2017; published 1 September 2017)

Purpose: To quantify differences in computationally estimated computed tomography (CT) organ doses for patient-specific voxel phantoms to estimated organ doses in matched computational phantoms using different matching criteria.

Materials and methods: Fifty-two patient-specific computational voxel phantoms were created through CT image segmentation. In addition, each patient-specific phantom was matched to six computational phantoms of the same gender based, respectively, on age and gender (reference phantoms), height and weight, effective diameter (both central slice and exam range average), and water equivalent diameter (both central slice and exam range average). Each patient-specific phantom and matched computational phantom were then used to simulate six different torso examinations using a previously validated Monte Carlo CT dosimetry methodology that accounts for tube current modulation. Organ doses for each patient-specific phantom were then compared with the organ dose estimates of each of the matched phantoms.

Results: Relative to the corresponding patient-specific phantoms, the root mean square of the difference in organ dose was 39.1%, 20.3%, 22.7%, 21.6%, 20.5%, and 17.6%, for reference, height and weight, effective diameter (central slice and scan average), and water equivalent diameter (central slice and scan average), respectively. The average magnitude of difference in organ dose was 24%, 14%, 16.9%, 16.2%, 14%, and 11.9%, respectively.

Conclusion: Overall, these data suggest that matching a patient to a computational phantom in a library is superior to matching to a reference phantom. Water equivalent diameter is the superior matching metric, but it is less feasible to implement in a clinical and retrospective setting. For these reasons, height-and-weight matching is an acceptable and reliable method for matching a patient to a member of a computational phantom library with regard to CT dosimetry. © 2017 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.12502>]

Key words: computational phantoms, computed tomography, Monte Carlo, organ dose, phantom matching

1. INTRODUCTION

When Monte Carlo simulations are performed for computed tomography (CT) organ dosimetry, a computational phantom must be employed as a virtual model of patient anatomy. Computational phantoms presently come in one of three format types, as well as in one of four morphometric categories. Format types include stylized (mathematical equation-based), voxel (segmented CT/MR images), and hybrid (NURBS and polygon mesh surfaces). Morphometric categories include reference (small library of phantoms by age at 50th height/weight percentile), patient-dependent (larger library of phantoms at various combinations of height/weight), patient sculpted (phantoms altered to match the patient's unique body contour with proportional changes to internal anatomy), and finally, patient-specific (an exact representation of the patient with respect to both body contour and internal

anatomy).^{1–6} A given computational phantom may thus come in different combinations of both format type and morphometric category, each with its own strengths and weaknesses. Furthermore, computational phantoms have varying degrees of anatomical accuracy, ease of use, and morphometric scaling flexibility, all of which can impact dose accuracy.

Patient-specific phantoms are near exact computational representations of individual patients, created through CT image segmentation. They would be the ideal choice to provide CT dose estimates for clinical patients, as they capture the patient's unique anatomical information regarding muscle and adipose tissue distribution, and internal organ size, depth, and shape. However, the process for creating a phantom for each individual clinical patient to the level of detail necessary for organ-level dosimetry is laborious and infeasible at present. Alternatively, for clinically viable CT dose estimates, patients must be matched to other computational phantoms (e.g.,

members of a patient-dependent phantom library) for which CT organ doses have previously been computed. The purpose of this study is to investigate the impact of various patient-phantom matching metrics on the accuracy of patient organ dose utilizing a previous validated CT dosimetry methodology, one that accounts for tube current modulation (TCM).^{7,8}

2. MATERIALS AND METHODS

2.A. Patient-specific phantom creation

For this study, 52 patient CT image sets were obtained and segmented using 3D-DOCTOR™ (Able Software Corporation, Lexington, MA, USA). Of these patients, 13 were adult females (previously segmented), 13 were adult males (previously segmented), 13 were pediatric females, and 13 were pediatric males.⁹ Eight organs of interest (bladder, heart wall, kidney, liver, lung, pancreas, spleen, and stomach) and homogeneous bone were manually segmented and imported into 3D-modeling software Rhinoceros™ (Robert McNeel & Associates, Seattle, WA, USA). Next, the outer body contour and internal organ boundaries were converted into either a polygon mesh or NURBS (nonuniform rational Basis spline) surface. Each organ was assigned an integer tag number so that age- and gender-specific material and density information (International Commission on Radiological Protection Report No. 89) could be assigned during Monte Carlo transport.¹⁰ Each patient-specific phantom was then converted to a discrete array of voxel elements using an in-house MATLAB (MathWorks, Natick, MA, USA) script. Finally, to be usable in the radiation transport code MCNPX v2.70 (Los Alamos National Laboratory, Los Alamos, NM, USA) for the previously validated CT dosimetry methodology, the voxel elements were converted to a 1-D array, referred to as a lattice file.⁷ Table I shows each of the patients used in this study, including their height, weight, body mass index (BMI), BMI classification, and age (for pediatric patients only). It should be noted that pediatric female 1 (PF1) was a newborn, and pediatric male 2 (PM2) had only seven organs segmented (e.g., the urinary bladder was not discernable in the original CT image).

2.B. Reference phantom matching

When matching a patient to a computational phantom that has been previously created, a reference phantom is the simplest choice. For this study, the previously created University of Florida (UF) family of reference hybrid phantoms was utilized for patient matching.¹¹ These 12 phantoms span six ages (newborn, 1-year-old, 5-year-old, 10-year-old, 15-year-old, and adult) and both sexes. They were modeled after the corresponding age and gender-specific information given in ICRP Publication 89 pertaining to reference organ masses, and total body heights and weights. In this context, reference refers to 50th-percentile and is based on an internationally defined population (not specifically US). Matching the patient-specific phantoms to a reference phantom was straightforward as

the patient was matched to the phantom with gender and the closest corresponding age.

2.C. Patient-dependent phantom matching

To provide more specificity to the phantom matching process, each patient-specific phantom was further matched to a phantom of the same gender from the UF/NCI library of hybrid computational phantoms (350+ hybrid computational phantoms scaled to specific age–height–weight combinations based on CDC data) in each of three ways: (a) based on height and weight, (b) based on effective diameter, and (c) based on water equivalent diameter.^{12–14} The latter two matching schemes were further divided into two methods: (a) based on the central slice of the exam range and (b) based on the average value of all slices within the exam range. Figures 1 and 2 visually compare the patient-specific phantoms to the reference and patient-dependent phantoms.

For the height-and-weight matching, each patient-specific phantom was first matched to a phantom in the UF/NCI library of hybrid computational phantoms based on the nearest height. Of the phantoms in the library with that height, the phantom with the closest weight was then used as the match. It should be noted that for pediatric female one (PF1), the reference newborn phantom was used as there are no newborn models in the UF/NCI library (that library was based on CDC body circumference data for individuals 2 years and up).

The second matching parameter was effective diameter which is defined in the American Association of Physicists in Medicine (AAPM) Report No. 204 as the square-root of the lateral and anterior–posterior (AP) dimensions (as shown in Equation 1).¹³

$$\begin{aligned} \text{Effective Diameter (cm)} \\ = \sqrt{\text{Diameter}_{\text{Lateral}}(\text{cm}) \times \text{Diameter}_{\text{AP}}(\text{cm})} \end{aligned} \quad (1)$$

The effective diameter (D_{eff}) was calculated for each z-slice (cranial-caudal direction) within each phantom in the UF/NCI library of hybrid computational phantoms as well as each patient-specific phantom. This was accomplished using a custom MATLAB™ script that counted the voxels in each dimension and multiplied by the corresponding voxel size. Just like the height-and-weight matching, each patient-specific phantom was first matched based on height, then matched based on the closest effective diameter within that height range. The first match was based on the central slice of the exam range (regardless of the exam type). The motivation for using only the central slice is that it is much more clinically feasible to calculate on a patient's image set. In addition, the phantom was matched based on the closest D_{eff} averaged across all slices within the exam range.

The final matching parameter was water equivalent diameter (WED) which is defined in AAPM Report No. 220.¹⁴ Unlike effective diameter, WED takes into account patient attenuation by utilizing the average CT number (Hounsfield Units) of each image in a CT image set as shown in Equation 2:

TABLE I. Relevant patient data for each of the 52 patient whose CT image sets were used for creating the patient-specific phantoms.

Age-gender	Patient ID	Height (cm)	Weight (kg)	BMI (kg·m ⁻²)	BMI classification ^a	Age-gender	Patient ID	Height (cm)	Weight (kg)	BMI (kg·m ⁻²)	Age (yr)	BMI classification ^b
Adult female	AF1	152.4	66.2	28.5	Overweight	Pediatric female	PF1 ^c	53.3	5	17.5	< 2	No classification ^d
	AF2	154.9	47.6	19.8	Healthy weight		PF2	88.9	11.8	14.9	< 2	No classification ^d
	AF3	154.9	69.9	29.1	Overweight		PF3	88.9	13.6	17.2	2	Healthy weight
	AF4	154.9	98	40.8	Obese		PF4	124.5	21.8	14.1	6	Healthy weight
	AF5	160	51.3	20	Healthy weight		PF5	134.6	26.8	14.8	7	Healthy weight
	AF6	160	51.7	20.2	Healthy weight		PF6	144.8	44.9	21.4	14	Healthy weight
	AF7	160	60.8	23.7	Healthy weight		PF7	154.9	59.9	24.9	13	Overweight
	AF8	163.8	59	22	Healthy weight		PF8	160	50.8	19.8	17	Healthy weight
	AF9	162.6	80.3	30.4	Obese		PF9	160	52.6	20.5	13	Healthy weight
	AF10	162.6	117.5	44.5	Obese		PF10	160	70.3	27.5	18	Overweight
	AF11	165.1	62.6	23	Healthy weight		PF11	167.6	56.7	20.2	16	Healthy weight
	AF12	172.7	82.1	27.5	Overweight		PF12	170.2	69.4	24	15	Overweight
	AF13	175.3	135.6	44.2	Obese		PF13	175.3	68	22.2	16	Healthy weight
Adult male	AM1	157.5	43.5	17.6	Underweight	Pediatric male	PM1	104.1	13.2	12.1	3	Underweight
	AM2	165.1	74.4	27.3	Overweight		PM2 ^c	104.1	15	13.8	4	Underweight
	AM3	167.6	78.5	27.9	Overweight		PM3	114.3	24	18.4	6	Overweight
	AM4	172.7	74.4	24.9	Healthy weight		PM4	144.8	35.8	17.1	8	Healthy weight
	AM5	172.7	98	32.8	Obese		PM5	152.4	46.7	20.1	12	Healthy weight
	AM6	175.3	66.2	21.6	Healthy weight		PM6	154.9	38.6	16.1	11	Healthy weight
	AM7	175.3	80.7	26.3	Overweight		PM7	154.9	45.4	18.9	14	Healthy weight
	AM8	177.8	73.5	23.2	Healthy weight		PM8	162.6	63.5	24	18	Healthy weight
	AM9	177.8	99.8	31.6	Obese		PM9	172.7	64.9	21.7	14	Healthy weight
	AM10	180.3	81.6	25.1	Overweight		PM10	177.8	63.5	20.1	17	Healthy weight
	AM11	182.9	85.7	25.6	Overweight		PM11	180.3	89.8	27.6	17	Overweight
	AM12	182.9	112.5	33.6	Obese		PM12	182.9	68.9	20.6	15	Healthy weight
	AM13	182.9	119.7	35.8	Obese		PM13	185.4	94.8	27.6	16	Obese

^aAdult BMI classifications (CDC): underweight (< 18.5), healthy weight (18.5 ≤ ... ≤ 24.9), overweight (25.0 ≤ ... ≤ 29.9), and obese (≥ 30.0).

^bPediatric BMI classifications (CDC): underweight (< 5th percentile), healthy weight (5th percentile ≤ ... < 85th percentile), overweight (85th percentile ≤ ... < 95th percentile), and obese (≥ 95th percentile).

^cMatched to reference newborn phantom.

^dNo BMI classification for pediatric patients less than 2-yr-old.

^eBladder could not be segmented.

$$\text{Water Equivalent Diameter(cm)} = 2\sqrt{\left[\frac{1}{1000}\overline{CT(x,y)_{ROI}} + 1\right]\frac{A_{ROI}(\text{cm}^2)}{\pi}}. \quad (2)$$

In this equation, $\overline{CT(x,y)_{ROI}}$ is the average CT number in the region of interest (ROI) and A_{ROI} is the area of the ROI. The issue with using this metric for phantom matching is that each computational phantom does not have a corresponding CT image that contains CT numbers defined as follows:

$$CT(x,y) = \left(\frac{\mu(x,y) - \mu_{\text{water}}}{\mu_{\text{water}}}\right) \times 1000 \quad (3)$$

In this equation, μ represents the linear attenuation coefficient for that voxel element. In order to estimate this value in the computational phantoms, the elemental composition of each voxel and the energy spectra for the exam must be known, which they are in this study. The energy spectra of a

Toshiba Aquilion ONE scanner was previously computationally defined and is stored in discrete 1 keV bins.⁷ The methods for obtaining these spectra are described in Turner et al.^{15,16} The elemental composition for each voxel is defined for Monte Carlo transport and is based on data given in ICRP Publication 89.¹⁰ These data can be used to estimate the linear attenuation coefficient ($\tilde{\mu}$) as shown in Eq. 4:

$$\tilde{\mu} = \rho \times \sum_i^{N_c} \left[\sum_j^{N_e} \left(w_i \left(\frac{\mu}{\rho} \right)_{ij} p_j \right) \right] \quad (4)$$

In this equation, ρ is the density of the tissue, summation index i is for each elemental constitute (N_c is the total number of elemental constitutes), summation index j is for each energy bin in the respective spectrum (N_e is the total number of energy bins for a given spectrum), w_i is the weight fraction for elemental constitute, i , $\left(\frac{\mu}{\rho}\right)_{ij}$ is the mass attenuation coefficient for element i at energy j (interpolated from NIST

data), and p_j is the normalized probability of energy j .¹⁷ Once $\tilde{\mu}$ is calculated for each voxel in a respective phantom, it can be converted into a CT number using Equation 3 and can then be used to calculate the WED for any slice in any computational phantom.

To further validate this methodology, the CT image set of a scanned reference anthropomorphic 15-year-old female phantom was compared with that of the computationally equivalent phantom.¹⁸ The WED of the central slice and the average WED of the entire exam range are compared to the

FIG. 1. Visual comparison in the anterior–posterior direction of the three phantom types: (a) patient-specific, (b) reference phantom, and (c) patient-dependent phantom. In this visual comparison, the outer body contour of phantom (c) is reduced to better match phantom (a). [Color figure can be viewed at [wileyonlinelibrary.com](#)]

FIG. 2. Visual comparison in the lateral direction of the three phantom types: (a) patient-specific, (b) reference phantom, and (c) patient-dependent phantom. In this visual comparison, the outer body contour of phantom (c) is reduced to better match phantom (a). [Color figure can be viewed at [wileyonlinelibrary.com](#)]

TABLE II. Comparison of the water equivalent diameter calculated from the CT image set to the computational estimate for six common torso exams for the UF 15-year-old female reference phantom at 120 kVp.

Exam	Central slice			Entire exam range average		
	WED _{CT Image-Set} (cm)	WED _{Phantom} (cm)	%Diff	WED _{CT Image-Set} (cm)	WED _{Phantom} (cm)	%Diff
Chest–abdomen–pelvis	24.4	24.6	−0.9	25.3	25.4	−0.7
Chest–abdomen	28.8	28.4	1.4	24.9	24.8	0.4
Abdomen–pelvis	23.3	24.0	−2.8	25.7	25.9	−0.8
Chest	29.2	27.3	7.0	25.4	25	1.4
Abdomen	24.1	25.0	−3.8	25.3	25	1
Pelvis	27.3	28.0	−2.5	26	26.7	−2.7

computational WED for six common torso exams as shown in Table II and discussed below.

For the phantom matching, each patient-specific phantom was matched to the phantoms of closest height to be consistent with the other matching methods. Next, the patient-specific phantom was matched to the phantom with the closest WED for the central slice of the exam of average or the average WED for the entire exam range.

2.D. CT exam dosimetry

A set of standard body CT exams were simulated for each patient-specific phantom and each of its five matched phantoms as well as the entire reference family. The exams were modeled using the previously validated methodology that accounts for TCM.^{7,8} The six torso exams modeled were as follows: chest–abdomen–pelvis (CAP), chest–abdomen (CA), abdomen–pelvis (AP), chest (C), abdomen (A), and pelvis (P). These exams and their parameters were derived from The University of Florida's Department of Radiology Standard Names for Imaging Procedures (SNIP) database and are described in Table III (see <http://protocols.xray.ufl.edu>). The anatomical landmarks listed in Table III were computationally defined for each phantom, and were additionally used as the basis for the D_{eff} and WED calculation. For each exam, only in-field organs were used for each of the six exam types: CAP (bladder, heart wall, kidney, liver, lung, pancreas, spleen, and stomach), CA (heart wall, kidney, liver, lung, pancreas, spleen, and stomach), AP (bladder, kidney, liver, pancreas, spleen, and stomach), C (heart wall, liver, lung, spleen, and stomach), A (kidney, liver, pancreas, spleen, and stomach), and P (bladder only).

3. RESULTS

For each patient-specific phantom, organ doses were compared with their corresponding organ doses in each of the six matched phantoms (reference, height-and-weight, effective diameter — Central slice, effective diameter — Entire exam

range, water equivalent diameter — Central slice, and water equivalent diameter — Entire exam range). The magnitude of percent difference was calculated for each organ as follows:

$$|Organ\ Dose\ \%Difference| = 100\% \left| \frac{Organ\ Dose_{matched} - Organ\ Dose_{patient-specific}}{OrganDose_{patient-specific}} \right| \quad (5)$$

Table II provides the comparison of the image-set-derived WED (using HU values for each pixel from an actual CT image set) to that of the computationally derived WED (Equation 4 on the computationally equivalent phantom). These data show that Equation 4 is sufficient in estimated WED for a computational phantom for which no CT image set exists.

FIG. 3. Boxplots comparing all organ dose percent differences for each of the six matching parameters. The vertical lines extend at most 1.5 times the interquartile range (in which no outlier beyond that range is visually shown). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE III. Exam parameters for each of the six simulated CT torso exams.

Parameter	Chest–abdomen–pelvis	Chest–abdomen	Abdomen–pelvis	Chest	Abdomen	Pelvis
Tube current modulation	Yes	Yes	Yes	Yes	Yes	Yes
Collimation	0.5 mm × 64	0.5 mm × 64	0.5 mm × 64	0.5 mm × 64	0.5 mm × 64	0.5 mm × 64
Energy (kVp)	120	120	120	120	120	120
Exam start	Thoracic inlet	Thoracic inlet	Dome of diaphragm	Thoracic inlet	Dome of diaphragm	Iliac crest
Exam end	Lesser trochanter	2 cm below iliac crest	Lesser trochanter	Top of kidneys	2 cm below iliac crest	Lesser trochanter
Filter	Large	Large	Large	Large	Large	Large
Gantry tilt (°)	0	0	0	0	0	0
Average mA	300	300	300	300	300	300
mA (min, max)	(100, 500)	(100, 500)	(100, 500)	(100, 500)	(100, 500)	(100, 500)
Pitch	0.828	0.828	0.828	1.484	0.828	0.828
Rotation time (s)	0.5	0.5	0.5	0.5	0.5	0.5

Due to the large number of organ doses (1661 total organ comparisons), boxplots were used to compare the different matching parameters. It should be noted that outliers (which are data points that extend beyond the interquartile range of the data, as described in Equations 6 and 7) are not shown.

$$\text{Interquartile Range}(IQR) = (q_3 - q_1) \quad (6)$$

$$\text{Outlier} : q_1 - 1.5 * IQR \text{ OR } q_3 + 1.5 * IQR \quad (7)$$

Figure 3 shows the comparison of all organs across the different matching parameters. Figures 4 and 5 show each of the six matching parameters compared across the different BMI classifications (described in the footer of Table I) for adults and pediatrics, respectively. Figure 6 shows each of the six matching parameters compared across the six different exam types. Table IV provides the average magnitude, median magnitude, as well as the root mean square (RMS) of difference for each matching metrics split based on overall, exam type, and BMI classification.

The full complement of organ doses computed in this study are provided in an electronic Annex in MS Excel format (Data S1 - Organ Dose Comparison Study.xlsx). The spreadsheet includes 52 tabs — One for each of the patient-specific computational phantoms, and labeled as either af, am, pf, or pm for adult female, adult male, pediatric female, or pediatric male, respectively. Organ doses are reported for each patient-specific phantom, as well for each of the matched phantoms (mGy/mAs). Percent differences are reported for each organ, along with average percent differences for each matching criterion and each CT exam type.

4. DISCUSSION

4.A. Comparison of organ doses

For the six matching metrics, water equivalent diameter matching was shown to be superior in terms of percent difference of organ dose estimates. These results support the motivation of AAPM Report No. 220 to account for the

FIG. 4. Boxplots comparing organ dose percent difference for each of the six matching parameters based on CDC BMI classifications for adult patients. The vertical lines extend at most 1.5 times the interquartile range (in which no outlier beyond that range is visually shown). [Color figure can be viewed at wileyonlinelibrary.com]

attenuation of internal anatomy within the patient.¹⁴ The other benefit to this metric (as well as effective diameter matching) is that a phantom is matched for a specific exam range, and not just across the entire patient anatomy. This approach allows a patient to be matched to a phantom with similar local, not entire, body morphometry. As expected, a reference phantom was the inferior choice as only patient age is taken into account. Interestingly, height-and-weight matching, which does not account for exam type or local patient size, was superior to effective diameter matching. This is believed to be caused by the effective diameter misrepresenting the true shape of the patient when they lie on a CT table. The computational phantoms are ideal nongravity representations of human anatomy. The patient-specific phantoms are lying down and their shape is distorted, which changes the effective diameter. This effect also appears with the female breast, as the AP extent is larger for the computational models as they are not being simulated lying on a table. This effect is not present in water equivalent diameter as shape is not a factor, just simply the area and voxel composition (and therefore attenuation) of the patient anatomy.

For both effective diameter and water equivalent diameter, the exam range average is superior to that of the central slice. This is expected as far more information is considered when you average across the entire exam range. This finding is supported by the comparison of the scanned anthropomorphic

phantom to that of its computationally equivalent phantom (Table II). On the contrary, in a clinical setting or with regard to a retrospective epidemiology study, it is far more time consuming to calculate D_{eff} or WED for every slice in an image set. In order to achieve this, a methodology for calculating D_{eff} and WED would need to be applied to each image in a CT image set. If this was fully automated, it could be easily achieved and applied to this study as exam range average was used as matching criteria. On the contrary, some healthcare institutions only calculate this value for the central slice and the additional information needed for the exam range average would be unknown, especially in a retrospective study where the CT image sets are no longer available. It would be even less expected to have this information when conducting a retrospective study, especially for a large patient cohort. This result, along with the comparable results of the height-and-weight matching, supports its use for retrospective studies. In the event that a patient's height and weight is not known, either this information can be interpolated from CDC growth curves, or a reference phantom can be used, both based on age. When considering healthy weight individuals, a reference phantom is comparable to the matched phantoms for both pediatrics and adults. Even underweight and overweight patients had respectable organ dose estimates using a reference phantom. The major downside to using a reference phantom is with obese patients, where organ dose estimates

FIG. 5. Boxplots comparing organ dose percent difference for each of the six matching parameters based on CDC BMI classifications for pediatric patients. The vertical lines extend at most 1.5 times the interquartile range (in which no outlier beyond that range is visually shown). [Color figure can be viewed at wileyonlinelibrary.com]

are poor (average magnitude, median magnitude, and RMS of 76.1%, 65.2%, and 87.2%, respectively, for adults). This is more exhibited in adults as obesity is more common, especially in a clinical setting when compared with that of a pediatric patient population.

In terms of exam range, the overall trend stays the same as WED matching is superior with height-and-weight being comparable. It is important to ensure that there is no trend with regard to exam type as this methodology would be generic and could be applied to any torso exam with TCM. The chest and pelvis exams are overall inferior, and this is believed to be caused by both the presence of lung (patients have varying lung volumes and inhalation) and the female breast (which has a larger effect on D_{eff} and WED in this region) for lung exams and the varying pelvic bone size for pelvis exams.

The annex provides dosimetry data for this entire study. As noted in Equation 5, the magnitude of organ dose was used as the basis for comparing the patient-to-phantom matching metrics. Overall, the matched phantoms tended to overestimate (provide conservative estimates of) organ doses to that of the patient-specific phantoms. There are no significant trends with regard to organ, exam type, nor gender with respect to overall organ dose. Each of the phantom-exam combinations were on average positive, except for the adult male's effective diameter matched phantoms. This is possibly caused by the male phantoms

having more muscle tissue with respect to their patient-specific counterparts during this matching process. This observation, along with that of the female breast (as previously mentioned), is the only two gender-specific trends, which make effective diameter in general a less reliable metric.

4.B. Sources of error and study limitations

There are various sources of error and limitations in this study including patient segmentation and the UF/NCI library of hybrid computational phantoms. In terms of the patient-specific phantoms, outer body contour and internal anatomy are segmented manually, where best judgment is exercised for organ boundaries and locations. In addition, converting these phantoms into a voxel phantom relies on altering the shapes of each organ slightly during 3D modeling. Another source of error with the patient-specific phantoms is the assumption made in material compositions and density. All anatomy that was not segmented is treated as residual soft tissue or homogenous bone (ICRP Publication 89 specifications). This varies from the computational phantoms as all organs are modeled, muscle and lymph nodes are implicitly modeled, as well as bone is separated into cortical and trabecular spongiosa.¹⁹

With regard to the CT examinations, exam ranges were based on anatomical landmarks, which is harder to define

FIG. 6. Boxplots comparing organ dose percent difference for each of the six matching parameters based on exam type. The vertical lines extend at most 1.5 times the interquartile range (in which no outlier beyond that range is visually shown). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE IV. Average magnitude, median magnitude, and root mean square of difference for each of the six matching metrics. Data shown below are gender averaged.

		Difference (%)					
		Reference (age and gender)	Height-and-weight	Effective diameter		Water equivalent diameter	
				Central slice	Exam range	Central slice	Exam range
Overall	Average ^a	24	14	16.9	16.2	14	11.9
	Median ^a	12.7	9.9	13.3	12.5	9.8	8.4
	RMS ^b	39.1	20.3	22.7	21.6	20.5	17.6
Chest–abdomen–pelvis	Average ^a	23.3	12.3	14.0	14.5	11.2	10.9
	Median ^a	12.4	9.2	11.1	11.0	8.7	8.4
	RMS ^b	39.0	16.7	18.1	18.7	14.5	14.2
Chest–abdomen	Average ^a	23.1	14.8	19.9	16.2	16.5	11.5
	Median ^a	12.7	11.2	16.0	12.2	11.4	7.9
	RMS ^b	37.2	20.9	25.9	21.9	23.6	17.3
Abdomen–pelvis	Average ^a	22.2	10.2	12.8	14.1	10.6	9.3
	Median ^a	10.6	8.0	10.2	11.0	8.7	7.3
	RMS ^b	36.7	13.9	16.4	18.1	13.8	12.4
Chest	Average ^a	29.2	22.8	22.9	21.1	24.1	19.6
	Median ^a	19.8	16.1	18.1	15.5	17.8	12.7
	RMS ^b	43.0	31.5	31.2	28.9	33.2	29.1
Abdomen	Average ^a	21.5	11.2	14.8	15.6	8.6	9.0
	Median ^a	10.3	7.4	11.0	12.6	6.0	6.9
	RMS ^b	35.2	16.1	20.0	20.2	12.7	13.0
Pelvis	Average ^a	34.6	14.8	23.3	21.3	17.6	13.1
	Median ^a	17.4	11.1	21.2	21.9	16.5	11.3
	RMS ^b	60.3	19.5	27.0	24.5	21.1	15.5
Underweight adult	Average ^a	20.8	8.1	11.6	16.5	10.4	6.0
	Median ^a	20.9	6.8	11.2	15.5	9.8	4.2
	RMS ^b	21.7	9.8	12.6	18.3	12.1	9.2
Healthy weight adult	Average ^a	11.3	13.1	15.6	16.1	13.5	11.1
	Median ^a	8.7	9.3	11.3	11.9	8.9	7.9
	RMS ^b	16.2	18.9	20.9	20.6	19.6	16.3
Overweight adult	Average ^a	21.6	16.2	21.4	21.4	16.6	12.0
	Median ^a	17.7	13.5	19.7	21.0	13.5	9.2
	RMS ^b	27.3	21.1	25.6	25.0	22.5	16.3
Obese adult	Average ^a	76.1	24.7	21.3	20.8	21.2	18.2
	Median ^a	65.2	20.9	15.0	14.9	15.7	12.7
	RMS ^b	87.2	32.8	30.9	30.1	30.6	27.9
Underweight pediatric	Average ^a	5.5	5.0	5.0	5.0	5.0	5.0
	Median ^a	4.5	3.9	4.0	3.9	3.9	3.9
	RMS ^b	8.4	6.8	6.9	6.8	6.8	6.8
Healthy weight pediatric	Average ^a	10.3	11.3	16.0	14.0	12.3	11.1
	Median ^a	7.7	8.7	14.2	11.8	9.8	8.9
	RMS ^b	14.8	16.0	20.2	17.9	17.1	15.3
Overweight pediatric	Average ^a	23.8	11.2	17.0	16.1	12.2	10.9
	Median ^a	21.0	8.7	12.3	12.6	8.2	8.2
	RMS ^b	30.4	14.7	21.5	20.2	16.5	14.4
Obese pediatric	Average ^a	35.2	14.3	18.4	14.3	13.8	11.7
	Median ^a	35.9	11.9	13.4	11.9	9.0	9.0
	RMS ^b	37.6	18.3	24.2	18.3	18.4	15.0

^aAverage and median are of the magnitude of organ dose percent difference.^bRoot mean square.

in the patient-specific phantoms. These are being compared to equivalent exam ranges in the computational phantoms that are also being manually computationally defined. Due to the relative comparison of organ doses (all Monte Carlo simulations), it can be argued that any errors within the TCM CT dosimetry methodology are not biasing organ doses as it would be the same for all patients in this study.

Another overall error with matching any patient to a computational phantom is the variations in organ size, density, and spatial location. Patient anatomy varies internally and a computational phantom is designed based on ideal geometry and anatomy location. This issue can only be avoided by manually segmenting organs and creating a patient-specific phantom. In addition, patients have varying body morphometries and degrees of lean body mass. Patients that are matched based solely on height-and-weight do not directly take lean body mass into account, and may be matched to a member of the computational phantom library with equivalent weight but not necessarily an equivalent distribution of muscle and adipose tissue. One other source of error with regard to phantom matching is that some patients lie outside of the height/weight range of the presently constructed UF/NCI library of hybrid computational phantoms. Some of the patients in this study were larger than the largest phantoms present in the existing library. This could potentially be addressed in the future by expanding the library and/or extrapolating organ doses beyond the range of the library. These inherent errors in patient matching exist not only in CT dosimetry but any medical dosimetry scenario (radiography, nuclear medicine, or fluoroscopy) where computational phantoms are applied.

5. CONCLUSIONS

The purpose of this study was to quantify differences in CT organ dose estimation for six different criteria for matching a patient to a member of a computational phantom library for which CT organ doses are precomputed. Although using a reference phantom is acceptable in most cases (although not for obese adults), matching a patient based on height-and-weight is an acceptable and clinically feasible method. Height-and-weight matching is also more favorable for retrospective studies as patient information is much more readily available than are CT image sets as required for phantom matching based on either effective diameter or water equivalent diameter. If, however, the CT image set is available, matching the patient based on the average WED across the entire exam range is a method superior to all others explored in this study. As this approach might be labor intensive in the clinical setting, assessment of the WED for the central slice is acceptable and provides roughly the same degree of dosimetry accuracy as that seen for height-and-weight matching. When considering the magnitude of the number of CT exams, this approach does not outweigh the simplicity and quickness of height-and-weight matching. In addition, creating a patient-specific model of each patient is not feasible

although it would provide the best organ dose estimates as the patient's outer body contour and internal anatomy would be fully modeled. Overall, this study concludes that with regard to computational CT dosimetry, matching a patient to a computational phantom can be easily and effectively accomplished by simply considering the patient's height and weight.

ACKNOWLEDGMENTS

Study was supported, in part, by the National Cancer Institute through contract HHS-N2612-0110-0510P and by the National Institute for Biomedical Imaging and Bioengineering through grant R01 EB013558.

^{a)}Author to whom correspondence should be addressed. Electronic mail: wbolch@ufl.edu; Telephone: +1 (352) 846 1361.

REFERENCES

1. Stabin MG, Xu XG, Emmons MA, Segars WP, Shi C, Fernald MJ. RADAR Reference Adult, Pediatric, and Pregnant Female Phantom Series for Internal and External Dosimetry. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2012;53:1807–1813.
2. Segars WP, Bond J, Frush J, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Med Phys*. 2013;40:043701.
3. Cassola VF, Lima VJ, Kramer R, Khoury HJ. FASH and MASH: female and male adult human phantoms based on polygon mesh surfaces: I. Development of the anatomy. *Phys Med Biol*. 2010;55:133–162.
4. Cassola VF, Milian FM, Kramer R, de Oliveira Lira CA, Khoury HJ. Standing adult human phantoms based on 10th, 50th and 90th mass and height percentiles of male and female Caucasian populations. *Phys Med Biol*. 2011;56:3749–3772.
5. Ding A, Mille MM, Liu T, Caracappa PF, Xu XG. Extension of RPI-adult male and female computational phantoms to obese patients and a Monte Carlo study of the effect on CT imaging dose. *Phys Med Biol*. 2012;57:2441–2459.
6. Stabin MG, Emmons MA, Segars WP, Fernald MJ. Realistic reference adult and paediatric phantom series for internal and external dosimetry. *Radiat Prot Dosimetry*. 2012;149:56–59.
7. Stepusin EJ, Long DJ, Ficarrota K, Hintenlang D, Bolch WE. Physical validation of a Monte Carlo-based phantom-derived approach to computed tomography organ dosimetry under tube current modulation. *Med Phys*. 2017; <https://doi.org/10.1002/mp.12461>.
8. Long DJ. Monte Carlo calculations of patient organ doses in Toshiba computed tomography examinations with automatic tube current modulation: a feasibility study. Dissertation, University of Florida; 2013.
9. Johnson PB, Geyer A, Borrego D, Ficarrota K, Johnson K, Bolch WE. The impact of anthropometric patient-phantom matching on organ dose: a hybrid phantom study for fluoroscopy guided interventions. *Med Phys*. 2011;38:1008–1017.
10. ICRP. ICRP Publication 89: basic anatomical and physiological data for use in radiological protection – Reference values. *Ann ICRP*. 2002;32:1–277.
11. Lee C, Lodwick D, Hurtado J, Pafundi D, Williams JL, Bolch WE. The UF family of reference hybrid phantoms for computational radiation dosimetry. *Phys Med Biol*. 2010;55:339–363.
12. Geyer AM, O'Reilly S, Lee C, Long DJ, Bolch WE. The UF/NCI family of hybrid computational phantoms representing the current US population of male and female children, adolescents, and adults—application to CT dosimetry. *Phys Med Biol*. 2014;59:5225–5242.
13. American Association of Physicists in Medicine. Size-specific dose estimates (SSDE) in pediatric and adult body ct examinations. Report No. 204; 2011.

14. American Association of Physicists in Medicine. Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT. Report No. 220; 2014.
15. Turner AC, Zhang D, Kim HJ, et al. A method to generate equivalent energy spectra and filtration models based on measurement for multidetector CT Monte Carlo dosimetry simulations. *Med Phys*. 2009;36:2154–2164.
16. Siewerdsen JH, Waese AM, Moseley DJ, Richard S, Jaffray DA. Spektr: a computational tool for x-ray spectral analysis and imaging system optimization. *Med Phys*. 2004;31:3057–3067.
17. Berger MJ, Hubbell JH, Seltzer SM, et al. *XCOM: Photon cross section database (version 1.5)*. Gaithersburg, MD: National Institutes of Standards and Technology.
18. Winslow JF, Hyer DE, Fisher RF, Tien CJ, Hintenlang DE. Construction of anthropomorphic phantoms for use in dosimetry studies. *J Appl Clin Med Phys*. 2009;10:2986.
19. Lee C, Lamart S, Moroz BE. Computational lymphatic node models in pediatric and adult hybrid phantoms for radiation dosimetry. *Phys Med Biol*. 2013;58:N59–N82.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Data S1. Organ Dose Comparison Study by Patient-Specific Voxel Phantom.