

Are age and gender suitable matching criteria in organ dose reconstruction using surrogate childhood cancer patients' CT scans?

Ziyuan Wang,^{a)} Irma W. E. M. van Dijk, and Jan Wiersma

Department of Radiation Oncology, Academic Medical Center (AMC), Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Cécile M. Ronckers

Department of Pediatric Oncology, Emma Children's Hospital/AMC, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Foppe Oldenburger and Brian V. Balgobind

Department of Radiation Oncology, Academic Medical Center (AMC), Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Peter A. N. Bosman

Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG, Amsterdam, The Netherlands

Arjan Bel and Tanja Alderliesten

Department of Radiation Oncology, Academic Medical Center (AMC), Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

(Received 28 November 2017; revised 19 March 2018; accepted for publication 27 March 2018; published xx xxxx xxxx)

Purpose: The purpose of this work was to assess the feasibility of using surrogate CT scans of matched patients for organ dose reconstructions for childhood cancer (CC) survivors, treated in the past with only 2D imaging data available instead of 3D CT data, and in particular using the current literature standard of matching patients based on similarity in age and gender.

Methods: Thirty-one recently treated CC patients with abdominal CT scans were divided into six age- and gender-matched groups. From each group, two radiotherapy plans for Wilms' tumor were selected as reference plans and applied to the age- and gender-matched patients' CTs in the respective group. Two reconstruction strategies were investigated: S1) without field adjustments; S2) with manual field adjustments according to anatomical information, using a visual check in digitally reconstructed radiographs. To assess the level of agreement between the reconstructed and the reference dose distributions, we computed (using a collapsed cone algorithm) and compared the absolute deviation in mean and maximum dose normalized by the prescribed dose (i.e., normalized errors $|NE_{\text{mean}}|$ and $|NE_{2\text{cc}}|$) in eight organs at risk (OARs): heart, lungs, liver, spleen, kidneys, and spinal cord. Furthermore, we assessed the quality of a reconstruction case by varying acceptance thresholds for $|NE_{\text{mean}}|$ and $|NE_{2\text{cc}}|$. A reconstruction case was accepted (i.e., considered to pass) if the errors in all OARs are smaller than the threshold. The pass fraction for a given threshold was then defined as the percentage of reconstruction cases that were classified as a pass. Furthermore, we consider the impact of allowing to use a different CT scan for each OAR.

Results: Slightly smaller reconstruction errors were achieved with S2 in multiple OARs than with S1 ($P < 0.05$). Among OARs, the best reconstruction was found for the spinal cord (average $|NE_{\text{mean}}|$ and $|NE_{2\text{cc}}| \leq 4\%$). The largest average $|NE_{\text{mean}}|$ was found in the spleen (18%). The largest average $|NE_{2\text{cc}}|$ was found in the left lung (26%). Less than 30% of the reconstruction cases (i.e., pass fraction) meet the criteria that $|NE_{\text{mean}}| < 20\%$ and $|NE_{2\text{cc}}| < 20\%$ in all OARs when using age and gender matching and a single CT to do reconstructions. Allowing other matchings and combining reconstructions for OARs from multiple patients, the pass fraction increases substantially to more than 60%.

Conclusions: To conclude, reconstructions with small deviations can be obtained by using CC patients' CT scans, making the general approach promising. However, using age and gender as the only matching criteria to select a CT scan for the reconstruction is not sufficient to guarantee sufficiently low reconstruction errors. It is therefore suggested to include more features (e.g., height, features extracted from 2D radiographs) than only age and gender for dose reconstruction for CC survivors treated in the pre-3D radiotherapy planning era and to consider ways to combine multiple reconstructions focused on different OARs. © 2018 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.12908]

Key words: childhood cancer, dose comparison, dose reconstruction, late adverse effects

1. INTRODUCTION

Childhood cancer (CC) survivors who historically received radiotherapy (RT) are prone to late adverse effects (LAEs).^{1,2} In contrast with adults, CC patients have a lower tolerance to radiation and a longer life expectation. The treatment of CC patients requires specific care and knowledge.³ Understanding how the dose delivered to organs at risk (OARs) relates to LAEs in CC survivors is key information needed for the design of RT plans with which a good cure rate of cancer can be achieved while the LAEs are reduced as much as possible, offering CC survivors a better quality of life.^{1,3,4} Retrospective studies on dose–effect relationships in CC survivors with long-term follow-up using both clinical information (e.g., age, gender, tumor type, metastasis status, LAEs) and treatment information (e.g., treatment modality, intended (prescribed) treatment dose, organ mean dose) have provided valuable information.^{5–9} The sensitivity to radiation-induced second cancers was found to vary with tissue type and dose–response relationships for various cancer sites were reported.^{8,9} However, the relation between radiation dose to (subvolumes of) OARs and LAEs is still not fully understood.

Associating 3D organ dose to LAEs is hampered by the fact that 3D dose distributions of the RT plans used on historical patients are often not available. For patients with long-term follow-up, CT scans are absent, because when these patients were treated, 2D imaging data (i.e., simulator films) were used for RT field localization and dose calculation.¹⁰ Moreover, even these historic simulator films are often no longer available.

To address the absence of 3D images of patients, various phantom-based methods have been developed for dose reconstruction.^{11,12} Anthropomorphic phantoms that are constructed of tissue-equivalent material provide possibilities to directly measure dose to organs. However, the accuracy of these methods is limited by the rough estimation of the human anatomy inside these phantoms. Moreover, creating anthropomorphic phantoms is expensive, time-consuming, and limited to phantoms of a few sizes.¹¹ Computational phantoms, using mathematical representations of human anatomy, originate from stylized phantoms that use simple 3D geometrical shapes (e.g., spheres and cylinders) to represent human anatomy.¹³ As tomographic medical imaging techniques became widely available, voxel-based models were developed with which better anatomical realism can be achieved.^{12–15} In recent studies, so-called boundary representation phantoms were introduced that combine voxel-based models with mathematical surface equations, to permit organ reshaping and repositioning.^{14,16–18} This allows phantoms to be deformed to simulate individual anatomy.

Computational phantoms for children have generally been based on medical images (e.g., CT, MRI) of a patient that has been considered to be the representative of a group of patients of a certain age and gender. Features like sitting height and nonabnormal anatomy are sometimes also taken into account in the selection of the representative.¹⁹ Average

anthropometric and reference organ mass data provided by public health institutes for specific age and gender groups were also used to adjust the representative CT-based model of the respective groups.^{20,21}

Existing computational phantoms do not sufficiently capture the variations in organ shape, size, and location that are observed in the population at large.^{22,23} In a study, phantoms that represent average male and female anatomies at different ages (newborn, 1, 5, 15 yr) were adjusted by using image deformation techniques to form new phantoms based on the CT scans of individuals of varying age, height, and weight.^{24,25} Large anatomical variations, in organ shape and size, were observed among phantoms representing the same age and gender.²⁵ Such a phantom library provides some variation in anatomies of different age, gender, height, and weight categories, but for the task of providing a good resemblance of a historical patient's anatomy, how to match the historical patient with an existing phantom remains a question. In a recent study, several patient-to-phantom matching methods were tested with Monte Carlo-based dose calculation for CT.²⁶ The results indicated that (a) the water equivalent diameter of the phantom is the superior matching metric, though the method is less feasible to implement in a retrospective setting; and (b) height-and-weight matching is superior to age-and-gender matching. However, it is unknown if the same conclusion goes for RT dosimetry, since the beam energy is higher, and radiation is focused to a tumor in the patient with a steep dose gradient in the field boundary.

A more individualized 2D-to-3D dose reconstruction approach was reported in which 2D imaging of the patients were used.²⁷ A reference 3D organ model was built and modified based on patient-specific structural information extracted from the 2D images. The approach was tested on 15 Hodgkin lymphoma patients with 3D CT planning data available to analyze the differences between the normal tissue doses derived from the actual treatment plans and the reconstructed doses. For the reconstruction approach 2D digitally reconstructed radiographs (DRRs) generated from the available 3D CT scans were used to simulate the availability of only 2D imaging. The median organ mean dose difference was ≤ 1 Gy ($\leq 5\%$ of the prescribed dose), which indicates that this approach is promising. The study was nevertheless limited to adult patients treated for one type of cancer.

The availability of CT scans of recently treated CC patients provides possibilities for individualized 3D dose reconstruction. In particular, it might enable the use of a CT scan from a patient that closely resembles (i.e., matches) the historical patient's anatomy. However, similar to the challenge for dose reconstruction using computational phantoms, it is yet unknown how such a CT scan should be selected based on the data available from historically treated patients so as to get the best match. In this pilot study, in line with the previously published phantom-based reconstruction methods, we therefore tested the suitability of only using age and gender as selection criteria.^{11,26,28} We focused on Wilms' tumor (WT) plans.³ To be able to quantify the quality of the dose reconstructions (i.e., the agreement with the original dose

distribution), we used recently treated patients for whom CT scans and reference plans were available for dose comparison. Our ultimate goal was to develop an individualized 2D-to-3D dose reconstruction strategy (3D dose reconstruction for patients with only 2D imaging data available) to be applied to data from historically treated patients with long-term follow-up, to obtain a better understanding of the relation between treatment and LAEs. As a first step, the aim of this study was to assess the usability of age- and gender-matched patients' CT scans for organ dose reconstructions by analyzing the level of agreement that the approach can achieve between the reconstructed and the original dose distributions.

2. MATERIALS AND METHODS

2.A. Data preparation

Clinical data of 31 CC patients, who received RT at the Academic Medical Center/Emma Children's Hospital in Amsterdam ($n = 28$) or at the University Medical Center Utrecht/Princess Máxima Center for Pediatric Oncology in Utrecht ($n = 3$) from 2009 to 2016, with CT scans of the abdomen available were included in our study. Patients' ages at the time of scanning ranged from 2.5 to 5.3 yr. The patients were divided into six age- and gender-matched groups such that the age range within each group is ≤ 1 yr and the average age of the groups is around 3, 4, or 5 yr. Detailed information can be found in Table I.

Most patient data included the original RT plan(s), except for the three patients who received RT in Utrecht, for which only CT scans were provided for this study. The patients were diagnosed with different types of CC and WT was the largest disease category.

We selected two original WT plans as reference RT plans from each group. The selected reference plans were typical flank irradiation treatments with an anterior–posterior–posterior–anterior field setup. According to the clinical SIOP WT 2001 protocol, the treatment field of a typical right-sided RT plan covers the tumor region including the vertebral column, the iliac crest, and major parts of the right liver, whereas the treatment field of a typical left-sided RT plan covers the major part of the left hemiabdomen with the vertebral column, the left part of the liver, and the spleen.²⁹ There was only one exception: the reference plan of one female in the 4.5–5.5 yr age group (i.e., patient 14, reference plan number 6) included partial irradiation of the flank on both sides.

Each reference RT plan was reconstructed by use of the other patients' CT scans in the respective age- and gender-matched group; resulting in a total of 50 dose reconstructions (two reference plans in group F1 have five reconstructions each and the other 10 reference plans have four reconstructions each, see Table I). The 12 reference plans were retrieved from our archive and together with all CT scans imported to the Oncentra treatment planning system (TPS) (version 4.3, Elekta, Stockholm, Sweden). The patient/CT scan on which the plan was originally calculated is referred to as the *reference patient/CT*, and the calculated dose is referred to as the

reference dose. Other patients/CT scans and the calculated doses within the age- and gender-matched group are referred to as *matched patients/CT scans* and *reconstructed doses* for that reference plan, respectively. Since for each age- and gender-matched group, there are two reference patients, a *reference patient/CT* is at the same time a *matched patient/CT* for another *reference patient/CT*.

To be able to assess errors in organ dose reconstruction, (partially) in-field OARs were delineated by clinical experts, including the heart, lungs, liver, spleen, kidney(s), and spinal cord. For 19 patients, the CT scans did not include full heart and/or lungs. Furthermore, all the patients diagnosed with WT (21 out of 31) underwent (partial) nephrectomy. Thus, the (partially) removed kidneys were not delineated for these patients.

2.B. Plan reconstruction strategies

In the Oncentra TPS, reference plans were reconstructed by applying them to the CT scans in the same age- and gender-matched group. We investigated two strategies to reconstruct reference plans (see Fig. 1). In both strategies, DRRs were utilized to a different extent, to simulate the realistic situation that only 2D simulator films of the historical patients would be available. The DRRs were derived from the CT scans using a setting that resulted in an enhancement of bone structures.

2.B.1. Strategy 1 (S1): identical plan strategy

In this strategy, we duplicated the treatment field without any adjustments of field size or beam settings. We manually positioned the reference treatment field on a matched CT, by using DRRs of the CT scans for visual confirmation. Collimator angles of the beams were also slightly adapted when necessary to properly align the field when the vertebral column is bent in different directions compared to the reference. Out of 50 cases, 26 collimator angles were adapted with a maximum adaptation of 8°.

2.B.2. Strategy 2 (S2): adjusted-plan strategy

When applying S1, anatomical differences in bone structures near the field boundaries between reference and matched CT scans were often visible in the DRRs. Therefore, in the so-called adjusted-plan strategy (S2), we adjusted the treatment field, where needed, to correct for the anatomical differences as visible in the DRR (Fig. 1). The adjusted plans were checked by an experienced pediatric radiation oncologist (BVB). In total, 46 of 50 reconstructed plans were adjusted. In the remaining four cases, no obvious anatomical differences near the field boundaries were observed.

2.C. Dose reconstruction details

As some of the data dated back to 2006, the original doses of reference plans were calculated in a previous TPS using treatment units that are no longer available in

TABLE I. Overview of patient characteristics and grouping (i.e., females were grouped in F1, F2, or F3; males were grouped in M1, M2, or M3). Furthermore, it is indicated which patients were selected as reference patients. Associated reference plan details (i.e., prescribed dose, number of fractions) are given, and it is indicated whether the CT scans included full heart/lungs.

Group	Patient Number	Age (yr)	Weight (kg)	Height (cm)	Wilms' tumor (Y/N)	Tumor site	Used as reference (Y/N) If Y: prescribed dose (Gy)/fractions	Reference plan number	Heart/Lungs complete in CT (Y/N)
F1	1	2.5	14	93	Y	Left kidney	Y: 14.4/8	1	N/N
	2	2.6	13	92	N	Left suprarenal gland	N		N/N
	3	2.8	12	NA	N	Left suprarenal gland	N	N/N	
	4	2.9	13	98	Y	Right kidney	N	Y/Y	
	5	3.1	16	99	Y	Right kidney	Y: 14.4/8	2	Y/Y
	6 ^a	3.2	13	95	Y	Left kidney	N		N/N
F2	7	3.6	15	NA	N	Left kidney	N	Y/Y	
	8	3.6	14	98	Y	Right kidney	N	N/N	
	9	3.9	18	108	Y	Right kidney	Y: 14.4/8	3	N/N
	10	4.2	20	115	Y	Left kidney	Y: 14.4/8		4
	11	4.2	16	101	Y	Left kidney	N	Y/Y	
F3	12	4.8	15	110	Y	Right kidney	Y: 14.4/8	5	Y/Y
	13	4.8	NA	NA	N	Pelvic region	N		N/N
	14	4.9	22	110	Y	Para-aortic metastasis of WT left kidney	Y: 10.8/6	6	N/N
	15	5.3	22	117	Y	Left kidney	N		N/N
	16 ^a	5.3	19	118	Y	Left kidney	N		N/N
	17 ^a	2.8	18	98	Y	Left kidney	N		N/N
M1	18	3.1	10	92	Y	Left kidney	Y: 14.4/8	7	Y/Y
	19	3.1	13	92	Y	Right kidney	Y: 25.2/14		8
	20	3.3	15	96	Y	Right kidney	N	Y/N	
	21	3.3	17	106	N	Ductus choledochus	N	Y/N	
	22	3.7	14	101	N	Right lower abdomen	N	Y/Y	
M2	23	3.8	15	102	Y	Right kidney	Y: 14.4/8	9	Y/Y
	24	3.9	18	104	N	Left suprarenal gland	N		N/N
	25	3.9	17	100	N	Retroperitoneal	N	N/N	
	26	4.2	16	106	Y	Left kidney	Y: 14.4/8	10	Y/Y
M3	27	4.7	23	119	N	Left suprarenal gland	N		Y/Y
	28	4.9	18	105	N	4th ventricle	N	Y/Y	
	29	4.9	28	123	Y	Right kidney	Y: 14.4/8	11	N/N
	30	5.1	17	109	N	4th ventricle	N		Y/Y
	31	5.2	18	116	Y	Left kidney	Y: 14.4/8	12	N/N

^aPatient received RT at the University Medical Center Utrecht/Princess Máxima Center for Pediatric Oncology.

the current TPS. To enable a consistent comparison, we calculated all the doses for both reference and matched CTs in the current TPS using a treatment unit (Elekta Linac with a multileaf collimator (MLC) beam limiting device, energy: 6 MV) that is similar to the previous ones. For reference RT plans that included the use of blocks that are no longer available in the current TPS, we used an MLC (leaf width 0.5 cm) as a substitute to shape the block contour. A collapsed cone algorithm was used to calculate the dose for all plans.³⁰

The reference doses and the reconstructed doses for S1 were calculated using the original monitor units (MUs) of the beams as stored in the reference plan. For S2, however, the field size adjustments influence the collimator and phantom scatter factors.³¹ Therefore, after field adjustment, the MUs were scaled to keep the dose point in the middle

of the field (isocenter) similar to its value before field adjustment. We observed that the required change in MUs was less than 1%.

2.D. Error quantification of reconstructions

We compared the reconstructed dose relative to the reference dose in individual OARs to assess the reconstruction error. We analyzed two common dose metrics: mean dose (D_{mean}) and maximum dose ($D_{2\text{cc}}$, the minimal dose to the maximally 2 cc exposed volume of the OAR).

To assess the expected error in dose reconstruction, irrespective of the prescribed dose, we considered normalized dose metric errors calculated as the difference in dose between reconstructed and reference plans normalized by the prescribed dose (NE_{mean} and $NE_{2\text{cc}}$):

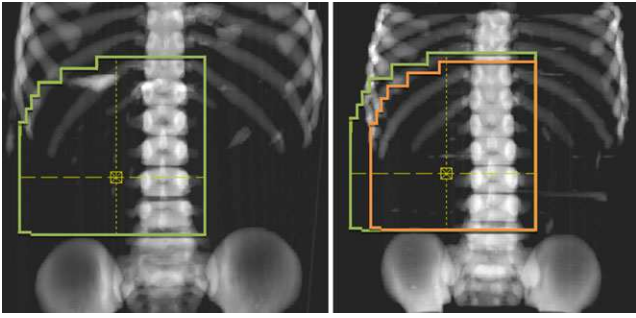


FIG. 1. Left: DRR of reference CT scan and reference RT plan field (green) of patient 29. Right: DRR of matched CT scan of patient 31 together with the reference RT plan field (green) as used in strategy 1 (S1) and an RT plan field adjusted according to the anatomical information (bony structures and outer body contour) observed in the DRRs (orange) as used in strategy 2 (S2).

$$NE_x = \left(\frac{D_x(\text{rec}) - D_x(\text{ref})}{D_{\text{pre}}(\text{ref})} \right) \times 100\%, x \in \{\text{mean}, 2\text{cc}\}, \quad (1)$$

where $D_x(\text{rec})$ and $D_x(\text{ref})$ refer to the reconstructed and reference dose metrics, respectively, and $D_{\text{pre}}(\text{ref})$ refers to the prescribed dose of the reference plan. $|NE_x|$ quantifies how large the deviation is compared to the value of dose metric x in the reference dose relative to the prescribed dose of the reference plan. The positive or negative sign of NE_x indicates whether it is an overestimation or underestimation of the dose metric.

Similarly, the relative dose metric errors RE_{mean} and $RE_{2\text{cc}}$ in an OAR were calculated as

$$\text{Pass fraction}(T_{\text{mean}}, T_{2\text{cc}}) = \frac{\sum_{i=1}^N \text{Pass}_i}{N} \times 100\%, T_{\text{mean}}, T_{2\text{cc}} \in \{5\%, 10\%, \dots, 100\%\},$$

where

$$\text{Pass}_i = \begin{cases} 1, & \text{if for each OAR} \in \{\text{heart, lungs}\} \text{ in patient } i: |NE_{2\text{cc}}(\text{OAR})| < T_{2\text{cc}} \quad \text{AND} \\ & \text{if for each OAR} \in \{\text{liver, spleen, kidneys, spinal cord}\} \text{ in patient } i: |NE_{\text{mean}}(\text{OAR})| < T_{\text{mean}} \text{ and } |NE_{2\text{cc}}(\text{OAR})| < T_{2\text{cc}} \\ 0, & \text{otherwise} \end{cases}$$

$$RE_x = \left(\frac{D_x(\text{rec}) - D_x(\text{ref})}{D_x(\text{ref})} \right) \times 100\%, x \in \{\text{mean}, 2\text{cc}\}. \quad (2)$$

D_{mean} can only be computed based on complete organs. Because 19 out of 31 CT scans did not contain complete heart and/or lungs, for heart and lungs only $D_{2\text{cc}}$ was investigated. For matched patients that underwent (partial) nephrectomy, dose in the (partially) resected kidney could not be reconstructed. Furthermore, for right-sided flank irradiation reference plans, the spleen is an out-of-field organ which received low dose (<1 Gy). In this pilot study, we excluded the spleen as an

OAR when comparing organ dose of right-sided RT plans.

2.E. Reconstruction evaluation

We created graphical summaries in the form of boxplots and bee swarm plots of the dose comparison results in each OAR, by plotting NE_x and RE_x of all reconstructions using both strategies. Per OAR, the average and range of the magnitude of the normalized and relative errors (i.e., $|NE_x|$ and $|RE_x|$) were calculated as well.

In order to assess whether a better reconstruction accuracy can be achieved with S2 than with S1, we used the paired Wilcoxon signed-rank statistical hypothesis test. Specifically, for each dose metric, the H_0 hypothesis was that S1 and S2 result in similar average error magnitudes, and the alternative hypothesis was that S1 results in larger average error magnitudes than S2. All the statistical analyses were performed using R .³² The significance level was set to 0.05.

The previous paragraph focused on the dose reconstruction error for each individual OAR. To gain insight into the overall reconstruction quality associated with the use of a matched CT scan, we considered all OARs in that matched CT scan jointly. We assessed the quality of a reconstruction case by varying acceptance thresholds for $|NE_{\text{mean}}|$ and $|NE_{2\text{cc}}|$ (denoted by T_{mean} and $T_{2\text{cc}}$, respectively). A reconstruction case was accepted (i.e., considered to pass) if the errors in all OARs are smaller than the threshold. Moreover, the pass fraction for a given threshold was defined as the percentage of the reconstruction cases of all the reference plans that were classified as a pass:

N represents the total number of reconstructions, and Pass_i is a Boolean value that indicates if the i_{th} reconstruction case is considered a pass under thresholds $(T_{\text{mean}}, T_{2\text{cc}})$. Pass fractions were calculated for S1 and S2 separately.

Furthermore, to investigate what is achievable in terms of overall dose reconstruction quality if we do not limit ourselves to the use of only one matched CT scan, we calculated pass fractions by selecting per reference plan for each OAR the best reconstruction (i.e., the reconstruction with smallest $|NE_{\text{mean}}|$, or, in case of the heart and lungs, the smallest $|NE_{2\text{cc}}|$). Combined, these form the best reconstruction possible for the reference plan. For all the 12 reference plans, we calculated the pass fraction of these best reconstructions.

3. RESULTS

The distributions of NE_x and RE_x per OAR for all reconstructions are visualized in boxplots and bee swarm plots in Figs. 2 and 3. The different colors of the bee swarm points represent the corresponding reference plans that were reconstructed, i.e., each color represents findings associated with a specific reference plan. It can be observed that the reconstruction deviations have different distributions in different OARs. Mostly, the findings associated with the different reference plans are highly mixed in these distributions which indicate that the observed variations of dose errors are a general characteristic for the WT plans.

The averages and ranges of $|NE_x|$ and $|RE_x|$ are summarized in Tables II and III, respectively. The average $|NE_{mean}|$ is smallest for the left kidney, 4% and 3% ($n = 13$) for S1 and S2, respectively. For the right kidney, located slightly more caudally than the left kidney to accommodate the liver, a larger average $|NE_{mean}|$ was found: 9% ($n = 16$) for both strategies. For the spinal cord, the average $|NE_{mean}|$ was 6% and 4% ($n = 50$) for S1 and S2, respectively. Furthermore, relatively small ranges of $|NE_{mean}|$ (0%–18% for S1, 0%–13% for S2) were found for the spinal cord. The largest average (range) $|NE_{mean}|$ was found for the spleen: 18% (0%–48%) and 18% (0%–42%) for S1 and S2, respectively. The liver had $|NE_{mean}|$ ranges comparable to the spleen but smaller averages (12% and 11% for S1 and S2).

For the lungs, the heart, and the kidneys, large averages ($>10\%$) and ranges (up to 0%–90%) in $|NE_{2cc}|$ were found, while for the liver, the spleen, and the spinal cord, smaller average values for $|NE_{2cc}|$ ($<10\%$) were found.

RE_x indicates the relative errors compared to the reference dose. The spinal cord had the smallest average $|RE_{mean}|$: 8% and 5% for S1 and S2, respectively. For the other four OARs, compared to the $|NE_{mean}|$ findings, larger averages and ranges were found for $|RE_{mean}|$: the average varied from 26% to 54% for S1, and from 20% to 48% for S2.

Large $|RE_{2cc}|$ values were found for the heart and the lungs, with averages $\geq 120\%$. For the kidneys, an average $|RE_{2cc}|$ around 20% was observed. For the liver, the spleen, and the spinal cord, the average $|RE_{2cc}|$ values were small (mostly $<5\%$, except for the spleen with S1), with only several outliers (Fig. 3).

Based on both the NE_x and RE_x findings, the most favorable results (i.e., best level of agreement, between reconstructed dose and reference dose and smallest variations among reconstructions) were obtained for the spinal cord.

The results of the one-sided paired Wilcoxon signed-rank test indicated that S1 results in larger average $|NE_{mean}|$ than S2 in the liver ($P < 0.004$) and the spinal cord ($P < 0.001$). For $|NE_{2cc}|$, significantly larger average values were found for S1 than S2 in the spleen ($P < 0.001$) and the spinal cord ($P < 0.04$). Furthermore, significantly larger average $|RE_{mean}|$ values for S1 than S2 were found in the liver ($P < 0.005$) and

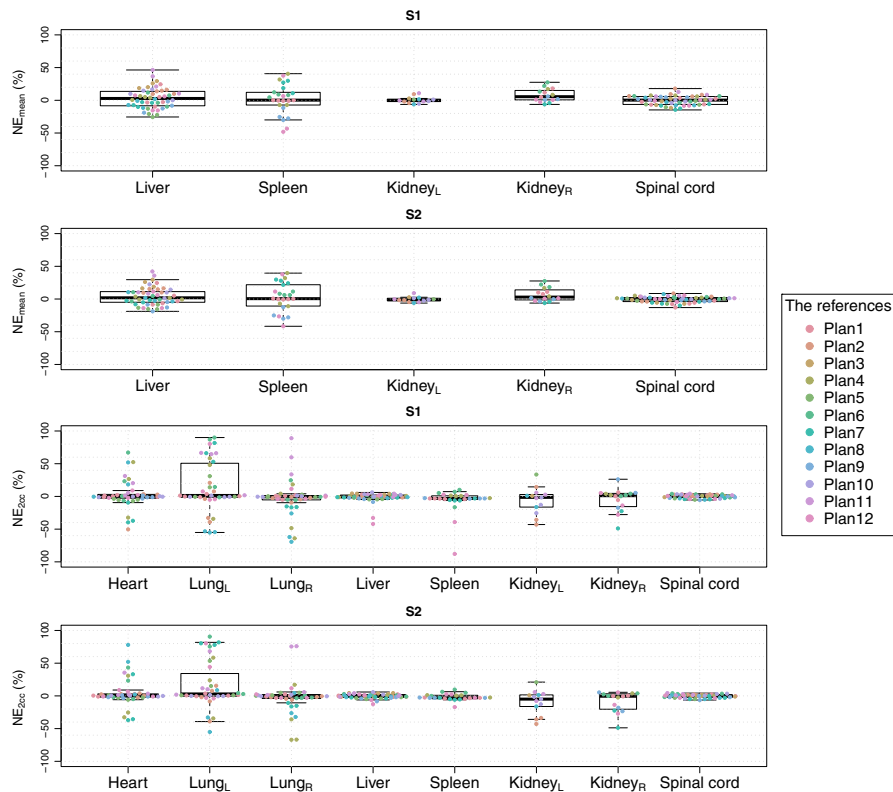


FIG. 2. Boxplots and bee swarm plots of deviations in mean dose and maximum dose normalized by prescribed dose (NE_{mean} and NE_{2cc}) of the reference plan to OARs between reconstructed plans and reference plans for the two dose reconstruction strategies (S1: applying the identical plan; S2: adjusting the plans according to the anatomical information observed in the DRRs). The whiskers extend to the 10th and 90th percentiles, the boxes extend from the 25th to 75th percentiles, and the thick horizontal line inside the box indicates the 50th percentile (i.e., the median).

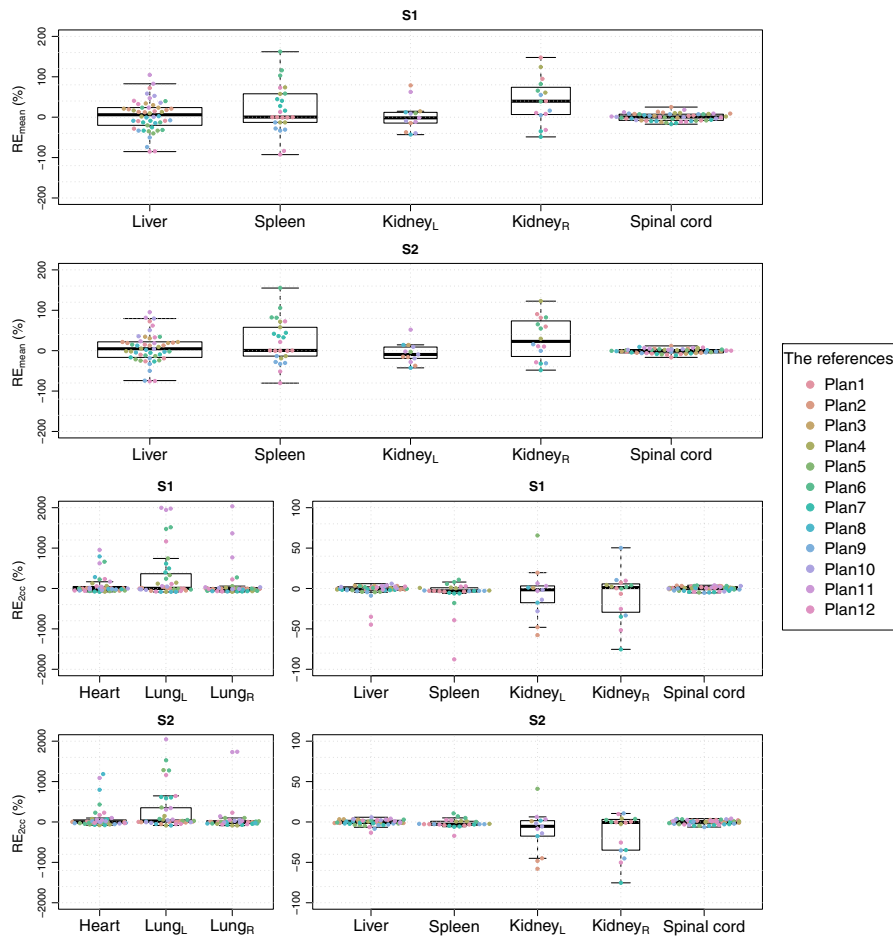


FIG. 3. Boxplots and bee swarm plots of relative deviations in mean dose and maximum dose (RE_{mean} and RE_{2cc}) to OARs, between reconstructed plans and reference plans for the two dose reconstruction strategies (S1: applying the identical plan; S2: adjusting the plans according to the anatomical information observed in the DRRs). The whiskers extend to the 10th and 90th percentiles, the boxes extend from the 25th to 75th percentiles, and the thick horizontal line inside the box indicates the 50th percentile (i.e., the median).

TABLE II. Average and range of the magnitude of deviations in mean dose and maximum dose normalized by the prescribed dose of the reference plan ($|NE_{mean}|$ and $|NE_{2cc}|$) for two dose reconstruction strategies in OARs.

OARs	No. of cases	$ NE_{mean} $ (%)		$ NE_{2cc} $ (%)	
		Average (range)		Average (range)	
		S1	S2	S1	S2
Heart	43			12 (0–67)	12 (0–78)
Left lung	39			30 (0–90)	26 (0–90)
Right lung	40			15 (0–89)	13 (0–76)
Liver	50	12 (0–46)	11 (1–42)	4 (0–42)	3 (0–13)
Spleen	25	18 (0–48)	18 (0–42)	9 (0–88)	5 (0–17)
Left kidney	13	4 (0–11)	3 (0–9)	15 (1–43)	14 (1–43)
Right kidney	16	9 (0–28)	9 (0–28)	12 (0–49)	11 (0–49)
Spinal cord	50	6 (0–18)	4 (0–13)	2 (0–5)	2 (0–6)

TABLE III. Average and range of the magnitude of relative deviations in mean dose and maximum dose ($|RE_{mean}|$ and $|RE_{2cc}|$) for two dose reconstruction strategies in OARs.

OARs	No. of cases	$ RE_{mean} $ (%)		$ RE_{2cc} $ (%)	
		Average (range)		Average (range)	
		S1	S2	S1	S2
Heart	43			120 (0–957)	122 (0–1190)
Left lung	39			362 (0–1998)	300 (0–2044)
Right lung	40			144 (0–2033)	125 (0–1738)
Liver	50	30 (1–105)	28 (1–95)	4 (0–45)	2 (0–13)
Spleen	25	44 (0–162)	43 (0–155)	9 (0–88)	4 (0–17)
Left kidney	13	26 (2–79)	20 (2–52)	21 (1–66)	20 (1–58)
Right kidney	16	54 (5–148)	48 (0–123)	20 (0–75)	19 (0–75)
Spinal cord	50	8 (0–25)	5 (0–17)	2 (0–5)	2 (0–6)

the spinal cord ($P < 0.001$). For $|RE_{2cc}|$ significantly larger average values were found for S1 than S2 in the spleen ($P < 0.001$) and the spinal cord ($P < 0.04$). In summary, the

use of S2 yielded better results than the use of S1 for both the mean dose (in the liver and the spinal cord) and the maximum dose (the spleen and the spinal cord).

Heat maps illustrating the pass fraction results under different acceptance thresholds are presented in Fig. 4. It can be observed that for both S1 and S2, pass fractions of <30% were found when both T_{mean} and $T_{2\text{cc}}$ are smaller than 20% (i.e., the dose deviation in mean and maximum dose is less than 20% of the prescribed dose). S2 resulted in slightly higher pass fractions than S1. For the same threshold values, T_{mean} was easier to pass than $T_{2\text{cc}}$. Not being restricted to the use of only one matched CT scan (i.e., combining the best reconstruction results obtained for each OAR) resulted in substantially higher pass fractions (e.g., pass fractions of >60%, when both T_{mean} and $T_{2\text{cc}}$ are $\geq 20\%$).

4. DISCUSSION

In this study, we investigated the usability of CT scans of matched recently treated CC patients, for 3D dose reconstructions for CC survivors for whom only 2D imaging data were available at the time of treatment, and in particular using the current literature standard of matching patients based on

similarity in age and gender. The use of CT scans from recently treated patients in combination with a set of reference treatment plans allowed us to evaluate the deviations between the doses to the OARs from the reference treatment plans and the reconstructed doses.

On one hand, our results, with a focus on Wilms' tumor patients, indicate that using only age and gender as features to select CT scans does not guarantee a useful reconstruction of dose to all OARs (i.e., a reconstruction with small errors in dose to all OARs). The magnitude of the organ dose metric deviations using two reconstruction strategies (S1 and S2: without and with field adaptation, respectively) varied with a broad range: 0%–48% (S1), 0%–42% (S2) of the prescribed dose in mean dose ($|NE_{\text{mean}}|$) and 0%–90% (S1 and S2) of the prescribed dose in maximum dose ($|NE_{2\text{cc}}|$) among reconstructions and among OARs. For less than 30% of the reconstructions, an overall reconstruction with deviations smaller than 20% of the prescribed dose in both mean and maximum dose in all OARs was achieved.

On the other hand, for individual OARs, more than half of the reconstructions had a deviation smaller than 20% of the prescribed dose. When selecting for each OAR, the best possible reconstruction based on all reconstructions obtained

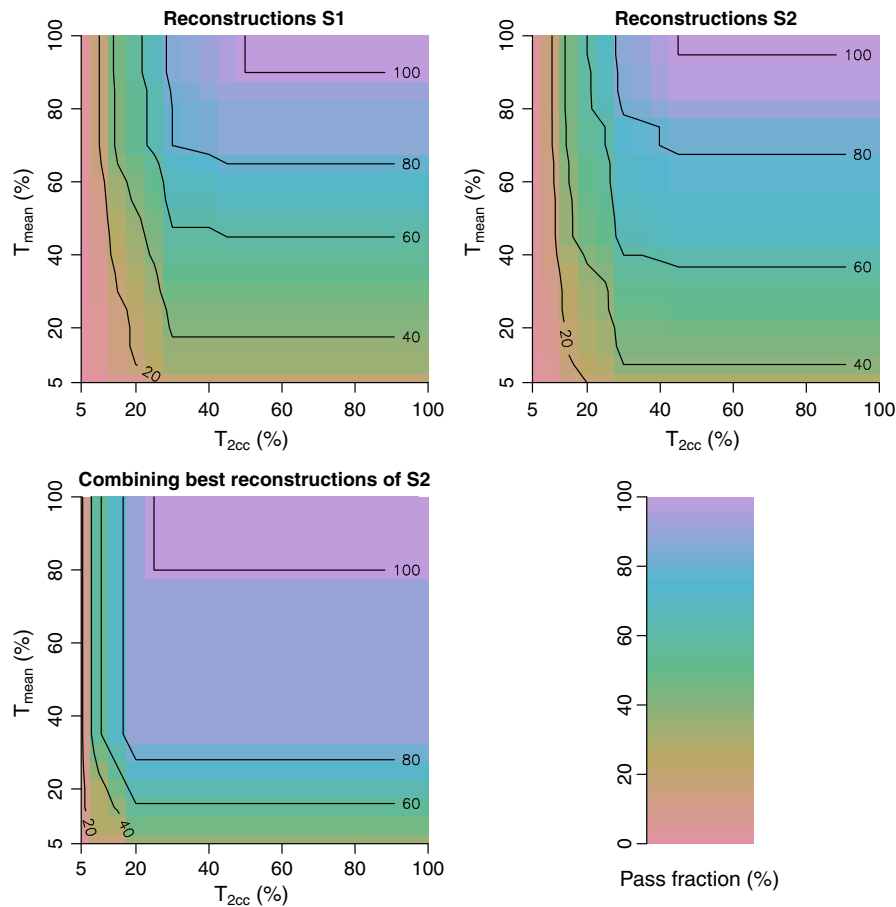


FIG. 4. Heat map of pass fractions for reconstructions of all reference plans obtained with the two strategies (S1 and S2) and the best reconstruction possible (i.e., when we do not limit ourselves to the use of only one matched CT scan) under different acceptance thresholds for normalized errors in mean dose and maximum dose ($|NE_{\text{mean}}|$ and $|NE_{2\text{cc}}|$).

within the specific age- and gender-matched group (i.e., not being limited to one matched CT scan), the overall reconstruction quality therefore improved a lot. These results indicate the potential of using CT scans for 3D dose reconstruction. However, a more sophisticated selection strategy (i.e., not solely based on age and gender) is needed. Moreover, next to focusing on how to find one good matching CT scan, the focus could be on finding multiple CT scans that each provide a good match for a different OAR. Considering that only four or five age- and gender-matched patients with CT scans were included in this study, we expect that the chance of finding a good match will increase if we enlarge our data set.

The selection of the surrogate CT scan will always depend on the available data of the historically treated patients. Sometimes this could mean that no more information other than the patient's age and gender is available. It is then important to be aware of the potential deviations in dose reconstructions, as presented in this study.

Two dose reconstruction strategies were investigated: using the identical plan without field adjustments and, if needed, manually adjusting the plan according to the anatomy visible in the DRRs. We considered the adjusted-plan strategy in order to see if we could, to some extent, compensate for the anatomical differences between the matched and reference CT scans. Our results indicate that the adjusted-plan strategy slightly improves the reconstruction quality. A statistically significant improvement was found in the spinal cord, the liver, and the spleen. For the spinal cord, which is contained within the vertebral column, an overall good reconstruction accuracy was achieved after field adaptation based on bony structure (average errors < 5%). For organs that are not fixated due to proximity of bony structure (e.g., spleen), the field adjustment provided limited improvements, which indicates that the internal organs' shape/location is not sufficiently related to the bony structure. Furthermore, organ motion that is independent of bony structure, can also contribute to organ variations among patients.³³

A good dose reconstruction quality was achieved for the spinal cord, where, compared to the other OARs, a high dose was delivered. For the kidneys, the average $|NE_{\text{mean}}|$ was smaller than 10%, however, much larger $|RE_{\text{mean}}|$ values were found (average $|RE_{\text{mean}}| > 20\%$). This can be explained by the low reference dose in the remaining kidneys for WT plans. Similarly, very large $|RE_{2\text{cc}}|$ values were reported for the lungs and the heart, these OARs are located near the boundary of the RT field where there is a steep dose gradient. Thus, the maximum dose values obtained for these OARs were very much influenced by the organ location relative to the RT field. In one case for the right lung, the reference $D_{2\text{cc}}$ was 0.63 Gy, and in a reconstruction 13.44 Gy; this led to an $|RE_{2\text{cc}}|$ of 2033%.

The collapsed cone algorithm used to calculate the dose in this study was reported to underestimate the dose far away (10–15 cm) from the field boundary.³⁴ However, the OARs included in this study are either in field or in the vicinity of the field boundary (< 5 cm), where the agreement between

the calculations and measurements obtained in the dose is reported to be between 1% and 2% of the prescribed dose.³⁵ For studies interested in OARs that are further away from the field boundary, a more advanced dose calculation method that can accurately estimate the out-of-field (> 10 cm) dose is needed.

A key question is what constitutes a proper notion of accuracy for dose reconstruction and what level of accuracy is needed for clinical research purposes. There is no clear answer currently. However, in the study about an individualized 2D-to-3D dose reconstruction approach we mentioned in the introduction, it was claimed that a median mean dose difference of 1 Gy (5% of the prescribed dose) between the reconstructed and actual RT plan would not be the weak link in the chain of estimating dose–risk relationships given the dose precision used in existing modeling studies.²⁷ In our study, the prescribed dose for 10 out of 12 reference plans was 14.4 Gy, which means that 20% of the prescribed dose is around 3 Gy. Based on the pass fraction findings associated with 20% thresholds, we believe that using solely age- and gender-matched anatomies for 2D-to-3D dose reconstruction could make an undesired difference in the dose–effect estimation model. Also, it is hard to compare our current results with other published methods because the reported dose reconstruction errors used different dose metrics and focused on different tumor types and patients' age groups. The previously mentioned 2D-to-3D dose reconstruction approach utilized more information (i.e., 2D radiographs) than just age and gender for anatomy matching.²⁷ It is therefore understandable that it yielded more accurate dose reconstruction results.

We quantified the reconstruction quality by assessing the pass fractions associated with different thresholds on dose metric deviations normalized by prescribed dose. In the pass fraction calculation, we combined the thresholds of the mean dose and the maximum dose; however, it is yet unclear if the error in estimating the mean dose and the maximum dose will lead to similar errors in dose–effect estimation models.

There are limitations to this study. First, the number of included patients for this pilot study is small. With only five or six patients in each age- and gender-matched group and six groups in total, we cannot achieve highly confident estimates of the variation in dose reconstruction deviations. However, even with our small numbers, negative conclusions on using solely age and gender as matching features can clearly be drawn. Secondly, it is hard to delineate gender-specific organs such as the glandular breast tissue and the ovaries that are not clearly visible in the CT scans, thus in this pilot study, these OARs were not included. Thirdly, the focus on Wilms' tumor reference plans only provides insights on dose reconstruction quality for this specific type of RT plan.

In this study, we directly compared dose metrics. Detailed spatial information of the 3D distribution is not taken into account. It would be more informative if we could perform a voxel-to-voxel 3D dose comparison. However, this requires a reliable and accurate way to compensate for differences in shape and location of organs among patients to make direct

3D dose comparison possible. One approach is deformable image registration, but this would introduce new uncertainties to the dose, as current solutions do not guarantee anatomical correctness of deformations. Without anatomical landmarks or fiducial implanted markers available in the imaging data, it is difficult to report the accuracy of registration of organs.³⁶

Nevertheless, 3D/2D deformable image registration between the matched CT scan and the available 2D image(s) of the reference patient can potentially (partially) compensate for residual mismatches.³⁷ However, since such a registration will mostly be guided by the outer body contour and bony structures visible in the 2D images, this compensation will most likely mainly be in size and/or shape of the matched patient and not in organ locations.

5. CONCLUSIONS

Using age and gender as matching features (i.e., the current literature standard) for 2D-to-3D dose reconstruction for CC survivors, does not result in a sufficiently robust reconstruction. Adapting the RT field based on the bony structure visible in DRRs slightly reduces reconstruction deviations, but this is not substantial for all OARs. However, there are combinations of patients for which a dose reconstruction with sufficiently small errors is achievable, especially when OARs may be selected from different patients. As such, a dose reconstruction method based on similar patients' CT scans can be said to be promising to provide accurate organ dose reconstructions for CC survivors.

In future studies, we therefore aim to develop a method that will be able to automatically select the most similar CT scan for dose reconstruction, based on a more extensive set of patients' characteristics and features (extracted from e.g., the 2D simulator films). A larger patient data library is currently being built up. As one of the next steps, we will investigate the anatomical variation among CC patients and relate it to the dose reconstruction quality. Ultimately, our goal was to develop a method that provides sufficiently accurate 2D-to-3D dose reconstructions so that these can be used in the investigation of dose–effect relationships in CC survivors with long-term follow-up.

ACKNOWLEDGMENTS

Financial support of this work was provided by Stichting Kinderen Kankervrij (KiKa; project no. 187). Dr. C.M. Ronckers was supported by the Dutch Cancer Society (KWF; grant no. UVA2012-5517). From the department of Radiation Oncology, AMC, Amsterdam, the Netherlands, the authors thank Koen F. Crama and Rianne (M.)A.J. de Jong for their help with organ delineation, and Prof.dr. Coen R.N. Rasch for sharing his expertise. We further thank Raquel Dávila Fajardo (department of Radiation Oncology, UMC Utrecht Cancer Center, Utrecht, the Netherlands) for sharing the data of three patients treated at the UMC Utrecht/Princess

Máxima Center for Pediatric Oncology for inclusion in this study.

CONFLICTS OF INTEREST

Dr. Alderliesten, Dr. Bel, and Dr. Bosman are involved in projects supported by Elekta. KiKa and Elekta had no involvement in the study design; on the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

^{a)}Author to whom correspondence should be addressed. Electronic mail: z.wang@amc.uva.nl; Telephone: +31 (0) 644861780.

REFERENCES

- Merchant TE, Hodgson D, Laack NN, Wolden S, Indelicato DJ, Kalapurakal JA. Children's Oncology Group's 2013 blueprint for research: radiation oncology. *Pediatr Blood Cancer*. 2013;60:1037–1043.
- Hudson MM, Mertens AC, Yasui Y, et al. Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *JAMA*. 2003;290:1583–1592.
- Olch AJ. *Pediatric Radiotherapy Planning and Treatment*. Boca Raton: CRC Press; 2013.
- Simon SL, Linet MS. Radiation-exposed populations: who, why, and how to study. *Health Phys*. 2014;106:182–195.
- van Dijk IW, Cardous-Ubbink MC, van der Pal HJ, et al. Dose-effect relationships for adverse events after cranial radiation therapy in long-term childhood cancer survivors. *Int J Radiat Oncol Biol Phys*. 2013;85:768–775.
- Stokkeväg CH, Engeseth G-M, Ytre-Hauge KS, et al. Estimated risk of radiation-induced cancer following paediatric cranio-spinal irradiation with electron, photon and proton therapy. *Acta Oncol*. 2014;53:1048–1057.
- Paulino AC, Wen B-C, Brown CK, et al. Late effects in children treated with radiation therapy for Wilms' tumor. *Int J Radiat Oncol Biol Phys*. 2000;46:1239–1246.
- Newhauser WD, de Gonzalez AB, Schulte R, Lee C. A review of radiotherapy-induced late effects research after advanced technology treatments. *Front Oncol*. 2016;6:13.
- Inskip PD, Sigurdson AJ, Veiga L, et al. Radiation-related new primary solid cancers in the Childhood Cancer Survivor Study: comparative radiation dose response and modification of treatment effects. *Int J Radiat Oncol Biol Phys*. 2016;94:800–807.
- Thwaites DI, Tuohy JB. Back to the future: the history and development of the clinical linear accelerator. *Phys Med Biol*. 2006;51:R343.
- Stovall M, Weathers R, Kasper C, et al. Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies. *Radiat Res*. 2006;166:141–157.
- Caon M. Voxel-based computational models of real human anatomy: a review. *Radiat Environ Biophys*. 2004;42:229–235.
- Xu XG, Eckerman KF. *Handbook of Anatomical Models for Radiation Dosimetry*. Routledge: Taylor & Francis; 2009.
- Bolch W, Lee C, Wayson M, Johnson P. Hybrid computational phantoms for medical dose reconstruction. *Radiat Environ Biophys*. 2010;49:155–168.
- Zankl M, Eckerman K, Bolch W. Voxel-based models representing the male and female ICRP reference adult—the skeleton. *Radiat Prot Dosi-met*. 2007;127:174–186.
- Cassola V, De Melo Lima V, Kramer R, Khoury H. FASH and MASH: female and male adult human phantoms based on polygon mesh surfaces: I. Development of the anatomy. *Phys Med Biol*. 2009;55:133.
- Xu XG. An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history. *Phys Med Biol*. 2014;59:R233.

18. Segars WP, Tsui B, Lalush D, Frey E, King M, Manocha D. *Development and application of the new dynamic Nurbs-based Cardiac-Torso (NCAT) phantom* [Ph.D. Dissertation], The University of North Carolina, Chapel Hill; 2001.
19. Lee C, Lodwick D, Williams JL, Bolch WE. Hybrid computational phantoms of the 15-year male and female adolescent: applications to CT organ dosimetry for patients of variable morphometry. *Med Phys.* 2008;35:2366–2382.
20. Valentin J. Basic anatomical and physiological data for use in radiological protection: reference values: ICRP Publication 89. *Ann ICRP.* 2002;32:1–277.
21. Shefer A, Atkinson W, Friedman C, et al. Immunization of health-care personnel: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recommendations and reports: Morbidity and mortality weekly report Recommendations and reports. 2011;60 (RR-7):1.
22. Na YH, Zhang B, Zhang J, Caracappa PF, Xu XG. Deformable adult human phantoms for radiation protection dosimetry: anthropometric data representing size distributions of adult worker populations and software algorithms. *Phys Med Biol.* 2010;55:3789.
23. Cassola V, Milian F, Kramer R, de Oliveira Lira C, Khoury H. 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.
24. Segars W, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys.* 2010;37:4902–4915.
25. Segars W, Norris H, Sturgeon GM, et al. The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization. *Med Phys.* 2015;42:4719–4726.
26. Stepusin EJ, Long DJ, Marshall EL, Bolch WE. Assessment of different patient-to-phantom matching criteria applied in Monte-Carlo based computed tomography dosimetry. *Med Phys.* 2017;44:5498–5508.
27. Ng A, Brock KK, Sharpe MB, Moseley JL, Craig T, Hodgson DC. Individualized 3D reconstruction of normal tissue dose for patients with long-term follow-up: a step toward understanding dose risk for late toxicity. *Int J Radiat Oncol Biol Phys.* 2012;84:e557–e563.
28. Lee C, Jung JW, Pelletier C, et al. Reconstruction of organ dose for external radiotherapy patients in retrospective epidemiologic studies. *Phys Med Biol.* 2015;60:2309–2324.
29. Pritchard-Jones K, Bergeron C, de Camargo B, et al. Omission of doxorubicin from the treatment of stage II–III, intermediate-risk Wilms' tumour (SIOP WT 2001): an open-label, non-inferiority, randomised controlled trial. *The Lancet.* 2015;386:1156–1164.
30. Ahnesjö A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Med Phys.* 1989;16:577–592.
31. Khan FM, Gibbons JP. *Khan's the Physics of Radiation Therapy*. Philadelphia: Lippincott Williams & Wilkins; 2014.
32. R Core Team. *R: A Language and Environment for Statistical Computing [computer program]. Version 3.2.3.* Vienna, Austria: R Foundation for Statistical Computing; 2015.
33. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys.* 2006;33:3874–3900.
34. Krieger T, Sauer OA. Monte Carlo-versus pencil-beam/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom. *Phys Med Biol.* 2005;50:859.
35. Johansen S, Olsen DR, Danielsen T, Malinen E. Contralateral breast doses following radiotherapy of the breast and regional lymph nodes: measurements and treatment planning calculations. *Radiother Oncol.* 2007;82:332–336.
36. Ebert M, Haworth A, Kearvell R, et al. Comparison of DVH data from multiple radiotherapy treatment planning systems. *Phys Med Biol.* 2010;55:N337.
37. Markelj P, Tomažević D, Likar B, Pernuš F. A review of 3D/2D registration methods for image-guided interventions. *Med Image Anal.* 2012;16:642–661.