

## Scientific Article

# Validation and Comparison of Radiograph-Based Organ Dose Reconstruction Approaches for Wilms Tumor Radiation Treatment Plans



Ziyuan Wang, PhD,<sup>a,\*</sup> Marco Virgolin, PhD,<sup>b</sup> Brian V. Balgobind, PhD,<sup>a</sup> Irma W.E.M. van Dijk, PhD,<sup>a</sup> Susan A. Smith, MPH,<sup>c</sup> Rebecca M. Howell, PhD,<sup>c</sup> Matthew M. Mille, PhD,<sup>d</sup> Choonsik Lee, PhD,<sup>d</sup> Choonik Lee, PhD,<sup>e</sup> Cécile M. Ronckers, PhD,<sup>f,g</sup> Peter A.N. Bosman, PhD,<sup>b</sup> Arjan Bel, PhD,<sup>a</sup> and Tanja Alderliesten, PhD<sup>a,h</sup>

<sup>a</sup>Department of Radiation Oncology, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands; <sup>b</sup>Life Sciences and Health Research Group, Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG, Amsterdam, the Netherlands; <sup>c</sup>Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas; <sup>d</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland; <sup>e</sup>Division of Radiation Oncology, University of Michigan, Ann Arbor, Michigan; <sup>f</sup>Princess Máxima Center for Paediatric Oncology, Utrecht, the Netherlands; <sup>g</sup>Institute of Biostatistics and Registry Research, Brandenburg Medical School, Neuruppin, Germany; <sup>h</sup>Department of Radiation Oncology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands

Received October 17, 2021; accepted June 22, 2022

## Abstract

**Purpose:** Our purpose was to validate and compare the performance of 4 organ dose reconstruction approaches for historical radiation treatment planning based on 2-dimensional radiographs.

**Methods and Materials:** We considered 10 patients with Wilms tumor with planning computed tomography images for whom we developed typical historic Wilms tumor radiation treatment plans, using anteroposterior and posteroanterior parallel-opposed 6 MV flank fields, normalized to 14.4 Gy. Two plans were created for each patient, with and without corner blocking. Regions of interest (lungs, heart, nipples, liver, spleen, contralateral kidney, and spinal cord) were delineated, and dose-volume metrics including organ mean and minimum dose ( $D_{\text{mean}}$  and  $D_{\text{min}}$ ) were computed as the reference baseline for comparison. Dosimetry for the 20 plans was then independently reconstructed using 4 different approaches. Three approaches involved surrogate anatomy, among which 2 used demographic-matching criteria for phantom selection/building, and 1 used machine learning. The fourth approach was also machine learning-based, but used no surrogate anatomies. Absolute differences in organ dose-volume metrics between the reconstructed and the reference values were calculated.

Sources of support: Financial support of this work was provided by Stichting Kinderen Kankervrij (KiKa; project no. 187). Dr Cécile Ronckers was supported by a grant from Dutch Cancer Society (#UVA2012-5517).

Disclosures: Dr Alderliesten, Dr Bel, and Prof. dr. Peter A.N. Bosman are involved in projects supported by Elekta. Dr Bel is involved in projects supported by Varian. KiKa, Elekta, and Varian were not involved in the study design, in the collation, analysis or interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

Data sharing statement: Research data are not available at this time.

\*Corresponding author.; E-mail: ziyuwang1991@gmail.com

<https://doi.org/10.1016/j.adro.2022.101015>

2452-1094/© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Results:** For  $D_{\text{mean}}$  and  $D_{\text{min}}$  (average and minimum point dose) all 4 dose reconstruction approaches performed within 10% of the prescribed dose ( $\leq 1.4$  Gy). The machine learning-based approaches showed a slight advantage for several of the considered regions of interest. For  $D_{\text{max}}$  (maximum point dose), the absolute differences were much higher, that is, exceeding 14% (2 Gy), with the poorest agreement observed for near-beam and out-of-beam organs for all approaches.

**Conclusions:** The studied approaches give comparable dose reconstruction results, and the choice of approach for cohort dosimetry for late effects studies should still be largely driven by the available resources (data, time, expertise, and funding).

© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

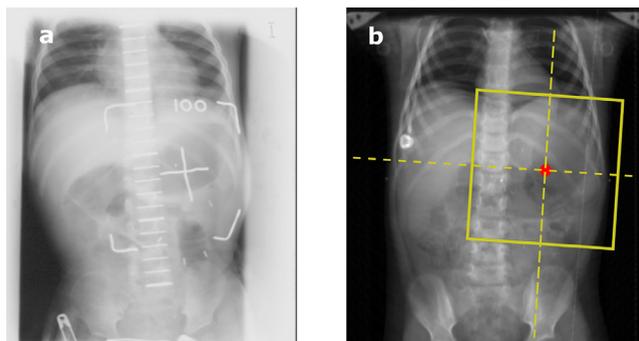
Childhood cancer survivors often experience treatment-related late adverse effects (LAEs), which have been linked to an increased risk of chronic morbidity and mortality.<sup>1-4</sup> In particular, radiation treatment (RT) is a component of approximately half of all cancer treatments<sup>5</sup> and is associated with a number of LAEs, such as second malignant neoplasms and cardiovascular disease.<sup>6-8</sup> Dose-response relationships can be translated into contemporary RT planning by providing dose constraints to organs-at-risk to help mitigate RT-related LAEs in future childhood cancer survivors.<sup>6</sup>

To date, most large retrospective cohorts of childhood cancer survivors include individuals treated in the precomputed tomography (CT) era of RT (before approximately 1999), for whom treatment was primarily based on conventional 2-dimensional (2D) radiographs.<sup>9,10</sup> For these cohorts, doses to organs of interest are not directly available so that methods must be applied to reconstruct the radiation dose; these methods often involve the use of physical or computational phantoms (3D models of human anatomy) or recent planning CTs of other patients as surrogate anatomies to overcome the lack of 3D anatomic imaging.<sup>9,11,12</sup> The mean organ dose and prescribed dose to individual body regions are the most commonly reported dose metrics.<sup>13-15</sup> As LAEs are found to be related to dose in specific organs (eg, second malignant neoplasms<sup>13</sup>) or

their subvolumes (eg, coronary artery<sup>16</sup>), there is a growing interest in using more refined dose and dose-volume metrics for dose-response modeling of LAEs.<sup>13,17</sup>

To provide more refined dose information, dose reconstruction is performed from the limited information available. Before dose reconstruction, patients' historic RT records must be abstracted for patient demographics (sex, age at RT, height, and weight) and specific treatment details, including beam energy, dose fractionation, field geometry, blocking details, field weighting, field location, and anatomic field borders. In some cases, field geometry superimposed on 2D radiographs is available. Figure 1a is an example of a historical 2D radiograph used for field placement.<sup>15</sup> The surrogate anatomy is chosen based on the available patient demographics, and the historical RT plan is then simulated on the surrogate anatomy using the available treatment details.<sup>12,14,18</sup> The resulting dose distribution can then be used to derive 3D organ dose-volume metrics for dose-response analysis.

There exist several options for the surrogate anatomy, each with its pros and cons. Stylized computational phantoms have a simplified geometric representation of human anatomy but can be easily scaled to different sizes and adapted to include additional organs or organ substructures. Stylized computational phantoms are commonly used for dosimetry in large retrospective childhood cancer survivor cohorts.<sup>9,11,18</sup> Voxel computational phantoms, on the other hand, are created from 3D medical images of real patients of specific ages and thus are more



**Fig. 1** A, An example of a historical 2-dimensional radiograph with field geometry (anteroposterior; AP) is shown. The white corners, made by 4 thin lead wires placed on the patient's body, depict the field boundaries. The white cross indicates the field isocenter, and the ruler indicates the field size. B, An example of a digitally reconstructed radiograph (derived from computed tomography) with field geometry (AP) is shown. The solid yellow lines depict the effective field boundary. The red point, which is the cross of the dashed yellow lines, indicates the field isocenter.

realistic but are rigid and cannot be flexibly scaled or repositioned.<sup>19</sup> Advanced boundary representation modeling methods offer a hybrid approach, which enables organ reshaping and repositioning while still allowing for realistic anatomic representation.<sup>20</sup> Such an approach has been used to develop a number of computational phantoms to represent average anatomies of the adult and pediatric populations<sup>20-22</sup> based on CT images and to represent different weight percentiles,<sup>23</sup> but these phantoms are typically only available for discrete ages.<sup>11,14,20,22,23</sup> Patient-specific CT images with organ delineations provide an alternative type of surrogate anatomy.<sup>12</sup> Due to the routine use of CT in RT planning, it is relatively easy to prepare and import a surrogate CT scan into a treatment planning system (TPS) for dose calculation compared with the computational phantoms. However, the planning CT scans typically only include anatomy near the site of treatment, whereas LAEs such as second solid tumors may occur throughout the entire body.<sup>24</sup>

There are different sources of uncertainty in RT dose reconstruction. One source of uncertainty comes from the difference between the surrogate anatomy and the historical patient's unknown anatomy during treatment.<sup>11,14</sup> The most commonly used patient-to-surrogate matching criteria are age and sex.<sup>9,12</sup> Different patient-to-surrogate matching criteria have also been investigated based on available patient demographics, such as height and weight percentiles<sup>10,23</sup> and water equivalent diameter (defined as average of the scanned range of the body).<sup>25</sup> When a 2D radiograph of the historical patient is available, more features can be used for dose reconstruction (eg, use 2D radiographs to guide 3D organ deformation<sup>26</sup>). Recently, machine learning (ML) has been leveraged in dose reconstruction approaches based on data sets of CT scans to predict 3D information (anatomy or dose) from the available 2D features<sup>38,42</sup>. Another source of uncertainty comes from the characterization of radiation beams used in the dose reconstruction calculations.<sup>14</sup> Different dose calculation algorithms, such as model-based algorithms are encountered in TPSs, Monte Carlo radiation transport simulation, and measurement-based algorithms are then used to estimate dose distributions in the surrogate anatomies<sup>9,10,14,27</sup>. However, studies have shown that the dose calculation algorithms in commercial TPSs underestimate out-of-field doses, while Monte Carlo simulations are more accurate.<sup>28,29</sup>

Because reconstructed doses are used for dose-response modeling,<sup>3,10,18</sup> the uncertainty in the reconstructed dose should be reported and incorporated into the models.<sup>14,30</sup> This is crucial for developing robust dose-response models that can be directly translated into contemporary RT planning, that is, used to define objectives for organs-at-risk. However, only a few study groups have reported a validation of their dose reconstruction approaches.<sup>12,18,26,31</sup>

In this study, we carried out a collaborative effort among several institutions to validate and compare 4 different dose reconstruction approaches. To simulate ground truth organ

doses for the validation, we used CT scans of recently treated patients with Wilms tumor and their clinical plans (with small adaptations) as reference data. For each of the contemporary patients, we extracted data analogous to what would be available in typical historic RT records<sup>15</sup> and asked the institutions to independently reconstruct the RT dose according to their previously published approaches.

## Methods and Materials

### Patient cohort, plan design, and organ delineations

We investigated Wilms tumor dose reconstructions because this type of kidney cancer is one of the most common types of childhood cancer in the abdominal region and the RT flank fields have not changed significantly over several decades.<sup>32-34</sup> We considered all pediatric patients (18 patients in total) treated for Wilms tumor between 2004 and 2016 in our hospital with a planning CT and complete treatment record. By creating an overview of a set of characteristics associated with these patients (ie, age, sex, height, weight, tumor laterality, treatment field) and considering the representativeness of these characteristics for the Wilms tumor patient population, we selected 10 patients with Wilms tumor. For these 10 patients, the age at the time of RT planning CT ranged from 2.5 to 5.5 years. A common longitudinal field-of-view shared by these CTs was between the 10th thoracic vertebra (T10) and the 1st sacral vertebra (S1). Detailed patient information can be found in [Table 1](#).

For each patient (P1-10) we created 2 typical Wilms tumor RT plans with 6 MV anteroposterior and posteroanterior parallel opposed flank fields.<sup>33,35</sup> The first plan involved open fields (P1-10<sub>R or L</sub>, where R or L refers to a right-sided or left-sided field) and the second (P1-10<sub>RB or LB</sub>) included small corner blocks (see block information in [Table 1](#)). The plans were developed by a pediatric radiation oncologist using the Oncentra TPS (version 4.3; Elekta AB, Stockholm, Sweden). We delineated regions of interest on the CT including lungs, heart, nipples, liver, spleen, contralateral kidney (ipsilateral kidneys surgically removed before RT), and a subvolume of the spinal cord from T10 to S1 using Velocity (version 3.2.0; Varian Medical Systems, Inc, Palo Alto, Calif, USA). For the lungs and heart, we delineated the portions of these organs that were imaged in the CT scans (for 8 and 5 out of 10 patients, the CT scans did not include complete lungs or heart, respectively).

### Reference dose calculation

For validation purposes, reference dose values were extracted from the 3D dose distributions of the designed

**Table 1 Characteristics of the 10 patients and the associated RT plans (2 plans per patient)**

Patient	Age (y)	Sex	Height (cm)	Weight (kg)	Tumor laterality	Field cranial/ caudal borders	Plan ID	Shielding (yes/no)	Shielded region border
P1*	3.0	M	92	13.0	Right	T9/S1	P1 <sub>R</sub>	No	
							P1 <sub>RB</sub>	Yes	Right body contour
P2	3.1	F	99	16.0	Right	T8/L4	P2 <sub>R</sub>	No	
							P2 <sub>RB</sub>	Yes	T8
P3	3.9	F	108	17.5	Right	T8/L4	P3 <sub>R</sub>	No	
							P3 <sub>RB</sub>	Yes	T8, T9, part of liver
P4	4.7	M	123	27.0	Right	T11/L4	P4 <sub>R</sub>	No	
							P3 <sub>RB</sub>	Yes	Right rib 9-10
P5	4.8	F	110	15.0	Right	T10/S1	P5 <sub>R</sub>	No	
							P5 <sub>RB</sub>	Yes	Right rib 9-10
P6*	5.1	F	122	22.0	Right	T10/S2	P6 <sub>R</sub>	No	
							P6 <sub>RB</sub>	Yes	Right rib 10
P7	2.5	F	93	14.0	Left	T9/S2	P7 <sub>L</sub>	No	
							P7 <sub>LB</sub>	Yes	S2
P8	4.2	M	106	15.5	Left	T11/L4	P8 <sub>L</sub>	No	
							P8 <sub>LB</sub>	Yes	Left rib 10
P9	4.2	F	115	20.0	Left	T10/L4	P9 <sub>L</sub>	No	
							P9 <sub>LB</sub>	Yes	Left rib 9-10
P10	5.5	M	116	18.0	Left	T10/S1	P10 <sub>L</sub>	No	
							P10 <sub>LB</sub>	Yes	Left rib 10

*Abbreviations:* F = female; L = left-sided plan; LB = left-sided plan with a block; M = male; R = right-sided plan; RB = right-sided plan with a block; RT = radiation treatment.

\*The prescribed dose of P1 and P6 was rescaled to 14.4 Gy in the dose reconstruction analysis.

T1 through T12 represent the 12 thoracic vertebrae; L1 through L5 represent the 5 lumbar vertebrae; S1 through S5 represent the 5 sacral vertebrae.

RT plans calculated on the CTs by the Oncentra TPS. All plans were designed assuming an Elekta LINAC treatment machine using 6 MV photons. A collapsed cone algorithm was used to calculate the dose, which was reported to achieve good in-field and near-field dose calculation accuracy.<sup>36</sup> The organ dose-volume metrics that were considered included mean dose ( $D_{\text{mean}}$ ), minimum dose ( $D_{\text{min}}$ ), maximum dose ( $D_{\text{max}}$ ), and the percentage of organ volume receiving at least 5 Gy and 10 Gy dose ( $V_5$  and  $V_{10}$ , respectively). Here, minimum dose and maximum dose refer to the minimum and maximum point dose to a region of interest in the reconstructed dose matrix. For the nipples, only  $D_{\text{mean}}$  was considered (as the nipple volume was small). For those patients where the volume of the heart or lungs was truncated (8/10 and 5/10 respectively), only  $D_{\text{max}}$  values representing the highest dose were used in the analysis.

### Preparation of input data for dose reconstruction

The following sections summarize the instructions shared among the participating institutes performing independent dose reconstructions according to their particular approach.

#### Digitally reconstructed radiographs with plotted field

Digitally reconstructed radiographs were generated from the CTs using the built-in module in the TPS for each beam's eye view (ie, anteroposterior and posteroanterior) with the field geometry plotted on top of it to simulate historical radiographs (Fig 1b). We selected an enhancement setting (min/max CT data threshold -300/3095, center 1500, width 3000, bone threshold 100, and bone enhancement factor 2.5) in Oncentra TPS that gave similar contrast (based on a visual check) as historical radiographs.

#### Data coding forms to describe patient and plan information

The RT record abstraction was prepared according to the data coding forms proposed by Stovall et al.<sup>9</sup> Patient information such as name and date of birth were anonymized. The RT details were abstracted and then checked by an experienced dosimetrist. Abstraction followed the methods described in Howell et al<sup>18</sup> and is briefly summarized here. In total, the data coding forms consisted of 3 pages. The first page was used to collect basic information such as the maximum target dose to each body region, which is defined as the sum of the prescribed dose from all overlapping fields, that is, the anteroposterior and posteroanterior fields. The second page included details on prescription(s) and treatment field parameters,

including dose, orientation, energy, field size, weighting, shielding, and anatomic borders. The third page was used to collect information about the proximity of organs of interest to the treatment fields, solely based on visually checking the prepared 2D radiographs. Proximity to the treatment field was specified as in-beam, at beam edge, near-beam, out-of-beam, or shielded.

### Dose reconstruction approaches

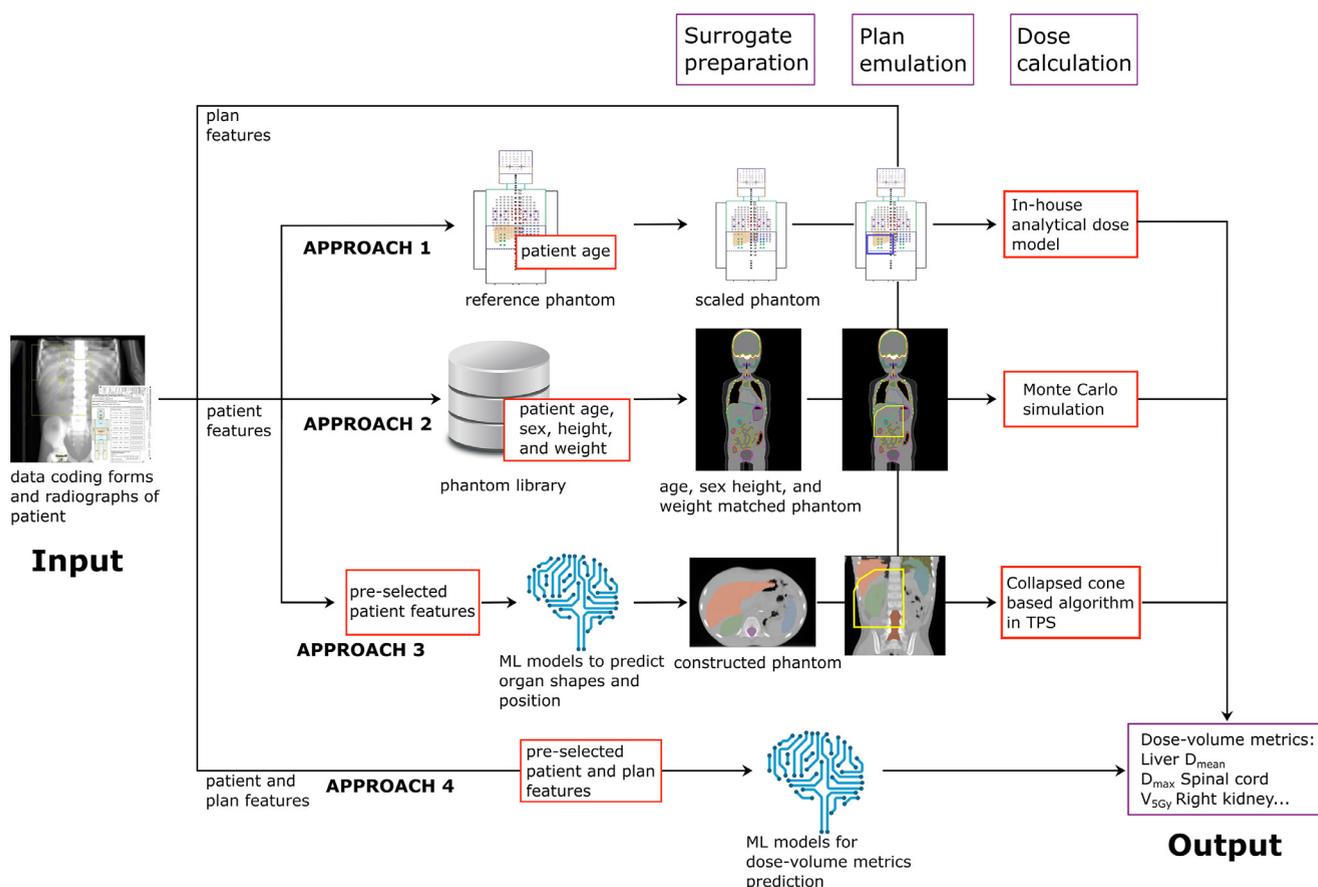
The 4 dose reconstruction approaches included in this study are listed, and the processes are summarized in Figure 2 and described in detail in the following paragraphs. The examples of the surrogate anatomies are illustrated for methods 1 to 3 in Figure 3.

1. Approach 1: An age-scaled stylized computational phantom-based approach<sup>9,18</sup>.
2. Approach 2: A multiple-feature matched computational phantom-based approach<sup>10,31</sup>.
3. Approach 3: A surrogate anatomy ML-based approach<sup>38</sup>.
4. Approach 4: A surrogate-free ML-based approach<sup>42,44</sup>.

#### Approach 1

The details of the age-scalable stylized computational phantom-based approach and its use in dose reconstructions for late effects studies are described in the literature<sup>9,18</sup>. The computational phantom consists of rectangular cuboids for the head, neck, trunk, arms, and legs; organs are specified by 3D grids of evenly spaced points.

For each of the 10 patients, the phantom was scaled to their age at RT by applying 3D scaling functions that account for nonuniform growth of different body regions<sup>45</sup>. Organs for each phantom were also scaled to age at RT according to the scaling functions that were applied to each of the respective body regions. RT plans were then reconstructed on the age-scaled phantoms based on the field parameters in the coding forms and a visual check of field placements compared with the radiographs<sup>9</sup>. Dose to all points in each organ were calculated using analytical dose models<sup>9</sup>, from which  $D_{\text{max}}$ ,  $D_{\text{mean}}$ , and  $D_{\text{min}}$  were reported. For the right and left nipple, doses were reported for a single point on each side. For the spinal cord, doses were the average of the central point in each vertebra (T10, T11, L1 to L5, and S1). For the spleen, no dose was reported as a dose grid that represents the spleen is not available in the computational phantom. For this study, only the RT plans with rectangular open-beam fields ( $n = 10$ ) were reconstructed using approach 1; however, in principle, blocking is also possible<sup>18</sup>.



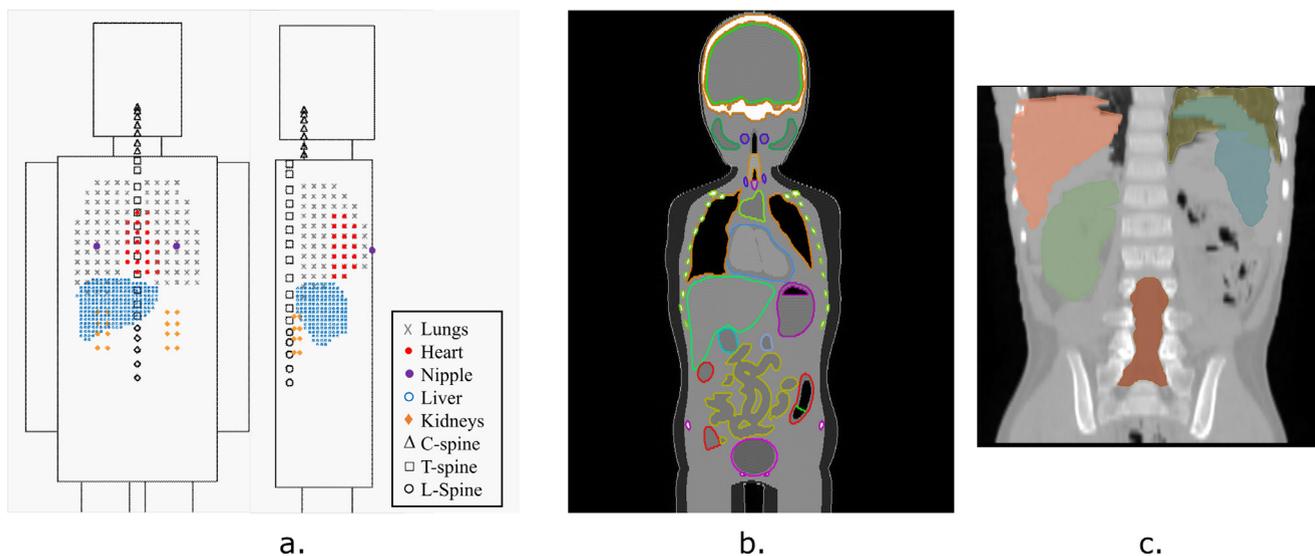
**Fig. 2** An illustration of steps taken by the 4 different dose reconstruction approaches given the same input data, that is, data coding forms and historical-like radiographs of the 2 beams.

### Approach 2

The multiple-feature matched computational phantom-based approach, based on a library ( $n = 351$ ) of whole-body computational phantoms covering a large portion of the population in the United States in terms of age, sex, height, and weight, was previously developed<sup>20,22</sup>. The approach considers multiple features (eg, age, sex, height, and weight) when selecting the surrogate phantom as available for a particular study. In this study, patient height, weight, and sex were provided in the data coding forms and were used to select the closest matched phantom from the phantom library as the surrogate anatomy. Next, the phantom (in the format of Digital Imaging and Communications in Medicine files<sup>43</sup>) was imported into a commercial TPS, Eclipse (Varian Medical Systems, Palo Alto, Calif, USA). RT plans were reconstructed based on the data coding forms and radiographs. Once the RT plans were created, the plan data were exported from the TPS for organ dose calculation using an RT-dedicated Monte Carlo transport code.<sup>10</sup> Additional details on this approach are available in the literature<sup>10,31</sup>. Right and left nipple dose were not reported as these structures were not explicitly defined in the phantom. Point doses in T10-S1 vertebrae were reported as a surrogate for the dose to the spinal cord.

### Approach 3

Approach 3 is the latest extension of the work<sup>38</sup>. The approach incorporates ML to automatically construct patient-specific phantoms. Among the several ML models we tested, we selected the model resulting from the gene-pool optimal mixing evolutionary algorithm for genetic programming (GP-GOMEA) for this study. GP-GOMEA is a state-of-the-art algorithm for learning interpretable ML models in the form of mathematical expressions,<sup>37</sup> and, in particular, GP-GOMEA was recently shown to have better prediction performance among several other models (eg, Least absolute shrinkage and selection operator and random forest) in the task of constructing individualized phantoms.<sup>38</sup> The training data set was similar to that used in<sup>38</sup>, with some enhancements. Specifically, the training data set included a larger database of 136 CTs of pediatric patients with cancer, in the age range of 1 to 8 years with more organs delineated. For each of the 136 CT scans, various features analogous to those available in historical radiographs were extracted from digitally reconstructed radiographs. Multiple ML models (1 per ROI) were trained to separately predict the most similar organs and body contours, and the most likely location of each organ's center of mass, based on the extracted features<sup>12,38</sup>. Next, the predicted best-matching organs



**Fig. 3** A, An illustration of the stylized computational phantom showing the organs (represented by 3-dimensional grids of points) used by approach 1. Note that since the time of this study, the heart model in this phantom has been updated<sup>18</sup>. B, A coronal view of an example of a computational phantom used by approach 2. The colored regions are representations of the segmented organs. C, A front sectional view of a patient-specific surrogate anatomy constructed by approach 3. The colored regions are representations of the “implanted” organs.

(which may belong to different surrogate CTs) were automatically “virtually implanted” at the predicted locations within the predicted body contour, forming a composite patient-specific phantom.<sup>38</sup>

Based on the input data, a list of the features of the 10 patients extracted from the coding forms and 2D radiographs was used as input for this approach. The result was 10 patient-specific phantoms, which were then imported into the Oncentra TPS for manual reconstruction of the RT plans on the phantom as described in<sup>12</sup>. Doses were calculated by the TPS based on the surrogate anatomy using the same collapsed cone algorithm as was used for the reference dose calculation.

#### Approach 4

Approach 4 is a dose prediction approach based on ML, which does not require any surrogate anatomy<sup>42</sup>. For this approach, we also used the ML algorithm GP-GOMEA to build the models. In the implementation of the approach for this study, ML models were generated to directly predict organ dose-volume metrics given a list of available 2D patient and plan features. The training data included 136 abdominal CTs of patients between 1 to 8 years old and 300 artificial Wilms tumor RT plans. The artificial RT plans were automatically generated by sampling within plan border ranges defined by an experienced clinical oncologist. Each plan was simulated on each CT, resulting in a total of 40,800 dose distributions<sup>44</sup>. The calculated organ dose-volume metrics were then used to train the ML models as response variables, whereas patients’ features available in historical records and detectable on digitally reconstructed radiographs, as well

as features of the RT plan, were used as explanatory variables. Separate ML models were then generated for each organ dose-volume metric. Based on the input data, the features of the 10 patients and 20 plans were put into the trained ML models from which organ dose-volume metrics were obtained.

#### Dose evaluation

To assess the level of agreement between the reference doses and the reconstructed doses obtained by the 4 approaches, we computed the absolute difference (subtracting the reconstructed value from the reference value and taking the absolute value) for organ mean, minimum, maximum dose,  $V_5$ , and  $V_{10}$  (denoted by  $DE_{mean}$ ,  $DE_{min}$ ,  $DE_{max}$ ,  $DE_{V5}$ , and  $DE_{V10}$ , respectively). To make results comparable, all the plans were normalized to a prescribed dose of 14.4 Gy.

In addition to providing the average and range of the differences for each of the organ dose-volume metrics, Wilcoxon rank-sum testing was performed to check whether differences between deviation distributions obtained by the various approaches were statistically significant ( $P < .05$ ).

#### Results

The average and range of the magnitude of the organ dose-volume metric differences obtained by the 4 approaches compared with the reference doses are

summarized in Table 2 (for  $DE_{\text{mean}}$ ,  $DE_{V_5}$ , and  $DE_{V_{10}}$ ) and Table 3 (for  $DE_{\text{min}}$  and  $DE_{\text{max}}$ ). The values of  $D_{\text{mean}}$  calculated by the 4 approaches along with the reference values for the 20 cases are presented in the Supplementary Materials. Most of the organs considered in this study are in-field or near-field organs (the contralateral kidney, spleen, liver, and spinal cord) for which the reference dose metrics calculated from the TPS can be considered to be ground truth for comparison purposes.

For  $DE_{\text{mean}}$ , an average deviation  $\leq 1.4$  Gy (10% of the prescribed dose) was found for most of the organs. Among the 7 organs in Table 2, approach 3 was found to have the smallest average  $DE_{\text{mean}}$  for 4 organs. However, for none of the organs considered were these differences found to be significantly smaller than the other 3 approaches ( $P > .05$ ).

The largest  $DE_{\text{mean}}$  values among organs for the 4 approaches were 5.1 Gy (right nipple) for approach 1, 3.3 Gy (liver) for approach 2, 4.7 Gy (spleen) for approach 3, and 3.6 Gy (right nipple and liver) for approach 4. For  $DE_{V_5}$  and  $DE_{V_{10}}$ , on average a deviation  $\leq 15\%$  of the volume was found for all the organs.

Except for the obtained values reported by approach 1 for the liver, an average of  $DE_{\text{min}} \leq 1.0$  Gy was found for all organs by all approaches. Approach 4 was found to achieve the smallest  $DE_{\text{min}}$  in both average and largest values for 4 out of 5 organs. For 4 out of 8 organs (both lungs, heart, and spleen),  $DE_{\text{max}}$  was found to be on average  $\geq 2.0$  Gy for all approaches. Among the 8 organs, approach 2 and approach 3 each had the smallest average  $DE_{\text{max}}$  for 3 organs, while approach 4 obtained the smallest average  $DE_{\text{max}}$  for 2 organs. The largest deviation (ie, worst case) reported for each organ's  $DE_{\text{min}}$  and  $DE_{\text{max}}$  over all approaches was 7.2 Gy and 14.3, respectively.

Approach 3 was found to have significantly smaller  $DE_{\text{min}}$  than the other 3 approaches for the spinal cord ( $P = .01$ ). No other distributions of  $DE_{\text{min}}$  or  $DE_{\text{max}}$  were found to be statistically different ( $P > .05$ ).

We observed that the reconstructed  $D_{\text{mean}}$  of the liver by all approaches had similar variations among the plans with the same laterality (see Supplementary Materials). The differences between  $D_{\text{mean}}$  for left-sided plans were smaller than for the right-sided plans. On average, the reconstructed  $D_{\text{mean}}$  values of the kidneys by all approaches had in general small differences ( $\leq 1.3$  Gy). The reconstructed  $D_{\text{mean}}$  of the spinal cord had similar values across all plans for any of the approaches. When the  $D_{\text{mean}}$  of an organ for plans with a corner block had a different value than for plans with a rectangular open field, approaches 2, 3, and 4 were found to be able to capture the trend (ie, decrease) of the differences but not always accurately (ie, the magnitude of decrease; eg, see  $P_{9L}$  vs  $P_{9LB}$  in the Supplementary Materials). Furthermore, no significant differences in  $DE_{\text{mean}}$  between plans with corner blocks applied and plans with open fields were found for these 3 approaches.

## Discussion

In this study, the performance of 4 organ dose reconstruction approaches were validated and compared for the same RT plans. Multiple institutes participated and independently performed the dose reconstruction using their own approach based on the same input data. A comprehensive analysis was performed to assess and compare the dosimetry results.

The results indicate that on average, the approaches achieved agreement within 10% of the prescribed dose for  $D_{\text{mean}}$  and  $D_{\text{min}}$  and within 15% of the organ volume for  $V_5$  and  $V_{10}$ . Lower agreement was observed for reconstructed  $D_{\text{max}}$  doses for all approaches for near-field and out-of-field organs (eg, kidneys and spleen for right-sided plans). For near-field organs, this is mainly attributed to the high dose gradient at the edge of the field. For out-of-field organs, additionally, the reference dose values were calculated by the collapsed cone algorithm in the Oncentra TPS. This algorithm is known to be subject to underestimation compared with Monte Carlo simulations,<sup>31</sup> as used in approach 2, and when compared with analytical models based on physical measurements, as applied in approach 1. Across all surrogate-based dose reconstruction approaches considered here, for in-beam/near-beam organs, the mismatch between anatomy of the surrogate phantom and the patient represents the main cause of reconstructed dose inaccuracies.<sup>11,14</sup>

For approach 1, the largest  $DE_{\text{min}}$  values for all organs were obtained, except for the spinal cord (largest for approach 4). This can potentially be due to the rough geometric modeling of the human anatomy in the approach 1 phantom (Fig 3).

Limitations of our study include the small number of patients and plans included, as well as the sole focus on Wilms tumor plans in the abdominal region, which provided a limited spread of anatomic patient and geometric plan variations. Second, in this study we did not consider the uncertainty introduced by using a different dose calculation algorithm in the reference case compared with the dose calculation algorithms used by approach 1 and approach 2.<sup>10,18</sup> Approach 3 and approach 4 (training stage) used the same dose calculation (collapsed cone algorithm) as was used for the reference dose calculation. Thus, a bias toward smaller dose differences for out-of-the-field organs may exist for the 2 ML-based approaches versus approach 1 and approach 2. Furthermore, some organ dose metrics were not reported for all approaches, such as the spleen dose of approach 1 and the right and left nipple dose of approach 2. Doses for these organs could be added in the future. Furthermore, for approach 1, dose reconstruction was not performed for plans involving corner blocks.

In general, compared with approach 1 and 2, approach 3 achieved slightly better average values for  $D_{\text{mean}}$  (up to

**Table 2 Average (range) of DE<sub>mean</sub> (in Gy), DE<sub>V5</sub>, and DE<sub>V10</sub> (in %) of organ dose reconstructions (for a subset of the organs) obtained for the 20 reconstruction cases**

Organ	DE <sub>mean</sub> (Gy)				DE <sub>V5</sub> (%)				DE <sub>V10</sub> (%)			
	Approach 1*	Approach 2	Approach 3	Approach 4	Approach 2	Approach 3	Approach 4	Approach 1	Approach 2	Approach 3	Approach 4	
R Nipple	1.0 (0.2-5.1)	<b>0.6 (0.0-2.8)</b>	<b>0.6 (0.0-2.8)</b>	0.6 (0.0-3.6)								
L Nipple	0.6 (0.0-1.5)	<b>0.3 (0.0-2.6)</b>	<b>0.3 (0.0-2.6)</b>	0.4 (0.0-1.5)								
Liver	1.4 (0.2-3.0)	1.4 (0.2-3.3)	<b>1.1 (0.1-2.3)</b>	1.4 (0.1-3.6)								
Spleen	<b>0.5 (0.0-1.3)</b>	0.9 (0.1-4.7)	0.9 (0.1-4.7)	0.9 (0.0-2.9)	15 (0-39)	7 (1-17)	10 (3-25)	13 (0-27)	8 (2-17)	10 (2-26)		
R Kidney	<b>0.6 (0.3-0.9)</b>	1.3 (0.6-2.2)	1.3 (0.1-2.5)	1.2 (0.5-2.3)	<b>3 (0-10)</b>	7 (0-34)	6 (0-27)	<b>3 (0-11)</b>	6 (0-35)	7 (0-23)		
L Kidney	0.8 (0.2-1.4)	0.7 (0.0-2.0)	0.7 ( <b>0.3-1.0</b> )	<b>0.4 (0.0-0.8)</b>	11 (4-20)	<b>11 (1-20)</b>	11 (5-22)	<b>6 (1-10)</b>	9 (1-18)	7 (2-14)		
Spinal cord	0.8 (0.2-1.4)	0.7 (0.0-2.0)	0.7 ( <b>0.3-1.0</b> )	<b>0.4 (0.0-0.8)</b>	6 (0-15)	6 (2-9)	<b>3 (1-6)</b>	4 (2-11)	4 (1-6)	<b>3 (0-4)</b>		

*Abbreviations:* DE<sub>mean</sub> = organ mean dose; DE<sub>V5</sub> = organ volume receiving at least 5 Gy dose; DE<sub>V10</sub> = organ volume receiving at least 10 Gy dose; L = left; R = right.  
 \* For approach 1, the reconstruction for plans with a corner block applied was not performed. The statistics of the results are based on reconstruction outcomes of plans with open fields only.  
 The doses for all plans were scaled to a prescribed dose of 14.4 Gy before comparison. For each dose-volume metric and for each organ, the smallest average of the deviation values among the approaches is indicated in bold. Similarly, the smallest range of the deviation values is formatted in bold in parentheses. The empty fields in the table indicate that the organ dose-volume metrics of the respective approach are not available for the respective approach.

0.3 Gy) and D<sub>max</sub> (up to 1.1 Gy) for in-field and near-field organs, indicating promising applications of leveraging ML for individualized phantom construction. However, this slight advantage does not apply for all organs and was not found to be statistically significant. Taking into account the relatively small amount of patient anatomies (136 CTs) used in the training stage, it is likely that the ML approach can perform better as more data become available to train the ML models; however, this remains to be seen.<sup>39,40</sup> A disadvantage of approach 3 is the possible unrealistic 3D anatomy (eg, overlapping organs), as organ shape and locations were predicted independently<sup>38</sup>. In this study, all 10 automatically assembled patient-specific phantoms had overlapping organ contours to some extent. In contrast, approach 2 uses more anatomically consistent phantoms containing a complete set of organs (we refer to<sup>31</sup> for available organs). Overall better results were obtained by approach 2 compared with approach 1 for D<sub>min</sub> and D<sub>max</sub>. This may indicate that more realistic phantoms (approach 2) are a better surrogate anatomy than stylized phantoms or that height and weight (used to select the representative phantom in approach 2) are important features to consider for patient-phantom matching (approach 1 only considers age). However, for a subset of the organs, approach 2 and approach 3 still had larger dose differences than approach 1, for example, DE<sub>mean</sub> for the right kidney was found to be 0.7 Gy smaller for approach 1 compared with approaches 2 and 3. So, even if approach 1 is arguably simpler than the others in that it uses stylized phantoms and relies on age alone, it remains competitive on some organs. This suggests that the position and shape of some organs (like the right kidney) remain very hard to predict well even with more complex matching criteria (approach 2) or ML (approach 3).

The predicted dose metric values of approach 4 were found to be comparable to the investigated surrogate-based approaches or even better (eg, for DE<sub>min</sub>). Furthermore, in general a smaller range (or largest value) of differences between the reconstructed organ dose-volume metrics and the reference metrics was observed for approach 4 compared with the other 3 approaches. This indicates that approach 4 has fewer outliers (ie, large deviations in dose reconstruction) and could be considered more robust. The downside of approach 4 is the absence of the entire 3D dose distribution as it can only predict dose metrics for which the ML models are trained. Nevertheless, the promising results of the 2 ML-based approaches indicate that leveraging ML on a data set where variability of patient anatomy is captured can benefit the accuracy, efficiency, and robustness of an ML-based dose reconstruction approach.

In epidemiologic studies of late effects, the level of detail of dosimetry that is required remains largely unknown.<sup>2,14</sup> The required dose accuracy for a study greatly depends on the type of tumor, organs of interest,

**Table 3** Average (range) of  $DE_{\min}$  and  $DE_{\max}$  values (in Gy) of dose reconstructions obtained for the 20 reconstruction cases.

Organ	$DE_{\min}$ (Gy)				$DE_{\max}$ (Gy)			
	Approach 1*	Approach 2	Approach 3	Approach 4	Approach 1*	Approach 2	Approach 3	Approach 4
R Lung					3.0 (0.2-13.2)	3.5 (0.1-12.3)	<b>2.9</b> (0.0-13.0)	3.5 ( <b>0.5-8.6</b> )
L Lung					3.4 ( <b>0.2-10.9</b> )	3.3 (0.0-11.0)	4.4 (0.1-11.9)	3.7 (0.9-14.3)
Heart					<b>2.0</b> (0.2-7.4)	3.6 (0.0-10.3)	2.4 (0.0-12.8)	3.2 ( <b>0.0-5.8</b> )
Liver	3.4 (0.2-14.2)	0.3 (0.0-1.0)	0.3 (0.0-0.5)	<b>0.1</b> ( <b>0.0-0.3</b> )	0.6 (0.0-1.8)	<b>0.3</b> ( <b>0.0-0.9</b> )	0.5 (0.0-1.6)	0.4 (0.0-1.2)
Spleen		0.8 (0.0-10.4)	0.8 (0.0-12.2)	<b>0.7</b> ( <b>0.0-6.5</b> )		6.1 (0.0-14.0)	3.1 (0.0-12.3)	<b>3.0</b> ( <b>0.1-10.1</b> )
R Kidney	0.9 (0.7-1.1)	0.2 (0.0-0.4)	<b>0.1</b> (0.0-0.2)	<b>0.1</b> ( <b>0.0-0.1</b> )	9.9 (9.3-10.6)	1.6 (0.2-6.5)	<b>0.5</b> ( <b>0.1-1.4</b> )	5.0 (2.7-6.1)
L Kidney	1.0 (0.4-1.5)	0.2 (0.0-0.4)	<b>0.1</b> ( <b>0.0-0.2</b> )	<b>0.1</b> ( <b>0.0-0.2</b> )	10.2 (8.7-12.2)	<b>1.1</b> ( <b>0.0-2.9</b> )	3.0 (0.2-7.5)	1.5 (0.2-3.6)
Spinal cord	0.7 (0.0-2.5)	0.4 (0.0-1.7)	<b>0.1</b> ( <b>0.0-1.3</b> )	0.9 (0.0-7.2)	0.5 (0.0-2.7)	0.6 (0.0-2.9)	<b>0.4</b> ( <b>0.0-1.8</b> )	0.5 (0.0-2.6)

The bolded numbers indicate the results with the smallest average deviation or smallest deviation range in brackets.  
Abbreviations:  $DE_{\max}$  = maximum dose;  $DE_{\min}$  = minimum dose; L = left; R = right.  
 $DE_{\min}$  is available for a subset of organs.  
\* For approach 1, the reconstruction for plans with a corner block applied was not performed. The statistics of the results are based on reconstruction outcomes of plans with open fields only.

available outcome data, and study design, such as cohort or case-control.<sup>26</sup> Some existing modeling studies use dose bins of 2 Gy, which indicates that the granularity of dose-effect relationships is limited to effects that can be distinguished at the level of 1 Gy dose difference.<sup>26</sup> One Gy dose bins are almost never used because there would be too few people in each category. However, in future studies, if enough data are available and better dose reconstruction accuracy is desirable, smaller dose bins should be used to obtain finer dose-effect relationship models, or the dose might be modeled as a continuous variable instead of a categorical variable. The question of what level of accuracy is needed can be more solidly answered after the models using finer dose bins or continuous variables are evaluated.

We chose several dose metrics for this study. The mean organ dose represents the average of the dose distributed in an organ and is commonly used to model organ-specific effects in published dose-effect studies.<sup>13,16</sup> The minimum organ dose is considered as an indication of whether the organ is located near the irradiated region. The maximum organ dose is considered as some organs (eg, the spinal cord) have a serial functionality, and thus an understanding of the maximum dose to the organ and the LAEs is required. We further considered  $DE_{V5}$  and  $DE_{V10}$ , as dose-volume histogram metrics are more commonly used as dose toxicity predictors and are used in current clinical practice of treatment planning for dose optimization and evaluation.<sup>17,41</sup>

Due to scarcity of data and relevant studies, it is currently difficult to provide clear clinical relevance of different dose metrics and to determine what level of inaccuracy in reconstruction is acceptable to still obtain a good dose-response modeling. In return, this is exactly why more research on the validation of dose

reconstruction approaches is needed, as well as studies on their use in dose-response modeling and the associated robustness of the dose-response modeling to deviations in the input (ie, the dose reconstruction values).

From a practical point of view, considering our results, we conclude that the selection of a dose reconstruction approach should be primarily based on the available historical data and the amount of time/effort and funding available for dose reconstruction. Approaches that require height and weight or 2D radiographs may not be appropriate when these data are not available, which is often the case. For the 4 approaches we compared, 2D radiographs are necessary for approach 3 and approach 4, but not for approach 1 and approach 2 (although 2D radiographs were used in this study by approach 1 and approach 2 to accurately position a plan on the phantom, which almost certainly benefitted the performance of these approaches). When height and weight information of the patient is not available, approach 2 will only use the age information and impute height and weight from standard growth tables (eg, using growth charts of the USA: <https://www.cdc.gov/growthcharts/index.htm>). When limited patient information is available, approaches 1 and 2 are more applicable compared with the ML-based, because age is the only patient feature needed to scale the phantoms. In terms of efficiency, approach 4 is a fully automatic pipeline to generate organ dose-volume metrics. Once the models are trained, the pipeline will automatically generate the required organ dose-volume metrics given the historical features of a patient and 2D radiographs in seconds. Approach 3 takes longer to run (minutes) as it handles 3D imaging to assemble a phantom according to ML predictions, but it is handy in that this is done automatically. The subsequent plan emulation and dose calculation step can also be carried out by an

automatic plan emulation pipeline<sup>44</sup>. In terms of time, if a cohort includes a large number of patients treated with similar types of RT plans and 2D radiographs are available, approach 4 is recommended as an efficient solution with competitive performance. However, if a study cohort includes a smaller amount of patient data associated with RT plans of large variability, the 3 surrogate-based approaches that provide individualized manual plan emulations are better choices.

## Conclusions

We compared the performance of 4 different dose reconstruction approaches for 2D radiograph-based organ dose reconstruction by using the same patient data set. On average all dose reconstruction approaches obtained  $D_{\text{mean}}$  with similar accuracy (deviation  $\leq 1.4$  Gy) for the investigated organs, which can provide reliable input for dose modeling using 2 Gy dose bins. Conversely, predictions for  $D_{\text{max}}$  were found to have much larger deviations, irrespective of the approach, suggesting that their use should be discouraged. A voxel phantom approach using multifeature matching (approach 2) provided the most realistic anatomy. An age-scaled phantom approach (approach 1) used the least amount of patient information while providing comparable dose reconstruction outcome for most of the investigated organs. Finally, ML-based individualized approaches (approach 3/4) achieved competitive results with the other 2 approaches, which are likely to further improve with more data, while employing automatic dose reconstruction procedures, which increases efficiency especially for larger cohorts.

## Acknowledgments

The authors thank Dr Petra S. Kroon and Dr Geert O. Janssens (Department of Radiation Oncology, University Medical Center Utrecht, Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands), Prof.dr. Marcel B. van Herk (Manchester Cancer Research Centre, Division of Cancer Sciences, University of Manchester, United Kingdom), Prof. David C. Hodgson (Department of Radiation Oncology, Princess Margaret Cancer Centre, Canada), and Dr Lorna Zadavec Zaletel (Department of Radiation Oncology, Institute of Oncology Ljubljana, Slovenia) for sharing (anonymized) data (CTs and patient features) of patients treated at their departments for use in the development of the 2 ML based approaches. The authors thank the Maurits en Anna de Kock Stichting for financing a high-performance computing system and Elekta for providing the research software ADMIRE for automatic segmentation for data preparation for the 2 ML-based approaches. Dr Matthew M. Mille and Dr Choonsik Lee performed their portions of this study using

the computational resources of the NIH high-performance computing Biowulf cluster (<http://hpc.nih.gov>).

## Supplementary materials

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.adro.2022.101015](https://doi.org/10.1016/j.adro.2022.101015).

## References

- Cheung YT, Brinkman TM, Li C, et al. Chronic health conditions and neurocognitive function in aging survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. *J Natl Cancer Inst.* 2017;110:411–419.
- Hoppe B, Howell R, Ladra M, et al. Spermatogenesis after testicular radiation exposure in children: Initial results from the Pediatric Normal Tissue Effects in the Clinic (PENTEC) initiative. *Int J Radiat Oncol Biol Phys.* 2019;105:E631–E632.
- Bowers DC, Liu Y, Leisenring W, et al. Late-occurring stroke among long-term survivors of childhood leukemia and brain tumors: A report from the Childhood Cancer Survivor Study. *J Clin Oncol.* 2006;24:5277–5282.
- Mulrooney DA, Armstrong GT, Huang S, et al. Cardiac outcomes in adult survivors of childhood cancer exposed to cardiotoxic therapy: A cross-sectional study from the St. Jude lifetime cohort. *Ann Intern Med.* 2016;164:93–101.
- Jairam V, Roberts KB, James BY. Historical trends in the use of radiation therapy for pediatric cancers: 1973–2008. *Int J Radiat Oncol Biol Phys.* 2013;85:e151–e155.
- Constine L, Ronckers C, Hua CH, et al. Pediatric Normal Tissue Effects in the Clinic (PENTEC): An international collaboration to analyse normal tissue radiation dose–volume response relationships for paediatric cancer patients. *Clin Oncol.* 2019;31:199–207.
- Turcotte LM, Liu Q, Yasui Y, et al. Temporal trends in treatment and subsequent neoplasm risk among 5-year survivors of childhood cancer, 1970–2015. *JAMA.* 2017;317:814–824.
- Bates JE, Howell RM, Liu Q, et al. Therapy-related cardiac risk in childhood cancer survivors: An analysis of the Childhood Cancer Survivor Study. *J Clin Oncol.* 2019;37:1090–1101.
- 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.
- 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.
- 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–R302.
- Wang Z, van Dijk IW, Wiersma J, et al. Are age and gender suitable matching criteria in organ dose reconstruction using surrogate childhood cancer patients' CT scans? *Med Phys.* 2018;45:2628–2638.
- de Gonzalez AB, Gilbert E, Curtis R, et al. Second solid cancers after radiation therapy: A systematic review of the epidemiologic studies of the radiation dose–response relationship. *Int J Radiat Oncol Biol Phys.* 2013;86:224–233.
- Bezin JV, Allodji RS, Mège JP, et al. A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose–response relationships. *J Radiol Prot.* 2017;37:R1–18.
- van Dijk IW, MC Cardous-Ubbink, 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.
16. Hahn E, Jiang H, Ng A, et al. Late cardiac toxicity after mediastinal radiation therapy for Hodgkin lymphoma: Contributions of coronary artery and whole heart dose-volume variables to risk prediction. *Int J Radiat Oncol Biol Phys.* 2017;98:1116–1123.
  17. Gagliardi G, Constine LS, Moiseenko V, et al. Radiation dose—volume effects in the heart. *Int J Radiat Oncol Biol Phys.* 2010;76:S77–S85.
  18. Howell RM, Smith SA, Weathers RE, et al. Adaptations to a generalized radiation dose reconstruction methodology for use in epidemiologic studies: An update from the MD Anderson Late Effect Group. *Radiat Res.* 2019;192:169–188.
  19. Xu XG, Eckerman KF. *Handbook of Anatomical Models for Radiation Dosimetry.* Boca Raton: CRC Press; 2009.
  20. Geyer AM, O'Reilly S, Lee C, et al. 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.
  21. 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;55:133–162.
  22. Lee C, Lodwick D, Hurtado J, et al. The UF family of reference hybrid phantoms for computational radiation dosimetry. *Phys Med Biol.* 2010;55:339–363.
  23. Cassola V, Milian F, Kramer R, et al. 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.
  24. Diallo I, Haddy N, Adjadj E, et al. Frequency distribution of second solid cancer locations in relation to the irradiated volume among 115 patients treated for childhood cancer. *Int J Radiat Oncol Biol Phys.* 2009;74:876–883.
  25. 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.
  26. Ng A, Brock KK, Sharpe MB, et al. 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.
  27. Wang Z, Balgobind B, Virgolin M, et al. How do patient characteristics and anatomical features correlate to accuracy of organ dose reconstruction for Wilms' tumor radiation treatment plans when using a surrogate patient's CT scan? *J Radiol Prot.* 2019;39:598–619.
  28. Howell RM, Scarboro SB, Kry SF, et al. Accuracy of out-of-field dose calculations by a commercial treatment planning system. *Phys Med Biol.* 2010;55:6999–7008.
  29. Chen W-Z, Xiao Y, Li J. Impact of dose calculation algorithm on radiation therapy. *World J Radiol.* 2014;6:874–880.
  30. Ntentas G, Darby SC, Aznar MC, et al. Dose-response relationships for radiation-related heart disease: Impact of uncertainties in cardiac dose reconstruction. *Radiation Oncol.* 2020;153:155–162.
  31. Mille MM, Jung JW, Lee C, et al. Comparison of normal tissue dose calculation methods for epidemiological studies of radiotherapy patients. *J Radiol Prot.* 2018;38:775–792.
  32. van Dijk IW, Oldenburger F, Cardous-Ubbink MC, et al. Evaluation of late adverse events in long-term Wilms' tumor survivors. *Int J Radiat Oncol Biol Phys.* 2010;78:370–378.
  33. Jereb B, Burgers JMV, Tournade MF, et al. Radiotherapy in the SIOP (International Society of Pediatric Oncology) nephroblastoma studies: A review. *Med Pediatr Oncol.* 1994;22:221–227.
  34. Termuhlen AM, Tersak JM, Liu Q, et al. Twenty-five year follow-up of childhood Wilms tumor: A report from the Childhood Cancer Survivor Study. *Pediatr Blood Cancer.* 2011;57:1210–1216.
  35. GJ D'angio, Tefft M, Breslow N, et al. Radiation therapy of Wilms' tumor: Results according to dose, field, post-operative timing and histology. *Int J Radiat Oncol Biol Phys.* 1978;4:769–780.
  36. Ahnesjö AJM. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Med Phys.* 1989;16:577–592.
  37. Virgolin M, Alderliesten T, Witteveen C, Bosman PAN. Improving model-based genetic programming for symbolic regression of small expressions. *Evol Comput.* 2021;29:211–237.
  38. Virgolin M, Wang Z, Alderliesten T, Bosman PA. Machine learning for automatic construction of pediatric abdominal phantoms for radiation dose reconstruction. *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications. International Society for Optics and Photonics.* Bellingham, WA: SPIE; 2020.
  39. Wei Z, Wang W, Bradfield J, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am J Hum Genet.* 2013;92:1008–1012.
  40. Zhou L, Pan S, Wang J, Vasilakos AVJN. Machine learning on big data: Opportunities and challenges. *Neurocomput.* 2017;237:350–361.
  41. Kirkpatrick JP, van der Kogel AJ, Schultheiss TE. Radiation dose—volume effects in the spinal cord. *Int J Radiat Oncol Biol Phys.* 2010;76:S42–S49.

## Further readings

42. Virgolin Marco, Wang Ziyuan, Balgobind Brian, et al. Surrogate-free machine learning-based organ dose reconstruction for pediatric abdominal radiotherapy. *Physics in Medicine & Biology.* 2020;65(24):245021.
43. Griffin Keith T, et al. Conversion of computational human phantoms into DICOM-RT for normal tissue dose assessment in radiotherapy patients. *Physics in Medicine & Biology.* 2019;64(13):13NT02.. In press.
44. Wang Ziyuan, Virgolin Marco, Bosman A Peter, et al. Automatic radiotherapy plan emulation for 3D dose reconstruction to enable big data analysis for historically treated patients. *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications.* 2019. *International Society for Optics and Photonics.* 2019.
45. Gupta Aashish Chandra, Shrestha Suman, Owens Constance, Smith A Susan. Development of an age-scalable 3D computational phantom in DICOM standard for late effects studies of childhood cancer survivors. *Biomedical Physics & Engineering Express.* 2020;6(6). In press.