


# 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?

Ziyuan Wang<sup>1,5</sup> , Brian V Balgobind<sup>1</sup>, Marco Virgolin<sup>2</sup>, Irma W E M van Dijk<sup>1</sup>, Jan Wiersma<sup>1</sup>, Cécile M Ronckers<sup>3,4</sup>, Peter A N Bosman<sup>2</sup>, Arjan Bel<sup>1</sup> and Tanja Alderliesten<sup>1</sup>

<sup>1</sup> Department of Radiation Oncology, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands

<sup>2</sup> Life Sciences and Health group, Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG, Amsterdam, the Netherlands

<sup>3</sup> Department of Pediatric Oncology, Amsterdam UMC, University of Amsterdam, Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands

<sup>4</sup> Princess Máxima Center for Pediatric Oncology, Heidelberglaan 25, 3584 CS, Utrecht, the Netherlands

E-mail: [z.wang@amc.uva.nl](mailto:z.wang@amc.uva.nl)

Received 17 January 2019, revised 25 March 2019

Accepted for publication 9 April 2019

Published 7 May 2019



CrossMark

## Abstract

In retrospective radiation treatment (RT) dosimetry, a surrogate anatomy is often used for patients without 3D CT. To gain insight in what the crucial aspects in a surrogate anatomy are to enable accurate dose reconstruction, we investigated the relation of patient characteristics and internal anatomical features with deviations in reconstructed organ dose using surrogate patient's CT scans. Abdominal CT scans of 35 childhood cancer patients (age: 2.1–5.6 yr; 17 boys, 18 girls) undergoing RT during 2004–2016 were included. Based on whether an intact right or left kidney is present in the CT scan, two groups were formed each containing 24 patients. From each group, four CTs associated with Wilms' tumor RT plans with an anterior-posterior—posterior-anterior field setup were selected as references. For each reference, a 2D digitally reconstructed radiograph was computed from the reference CT to simulate a 2D radiographic image and dose reconstruction was performed on

<sup>5</sup> Author to whom any correspondence should be addressed.

the other CTs in the respective group. Deviations in organ mean dose ( $DE_{\text{mean}}$ ) of the reconstructions versus the references were calculated, as were deviations in patient characteristics (i.e. age, height, weight) and in anatomical features including organ volume, location (in 3D), and spatial overlaps. Per reference, the Pearson's correlation coefficient between deviations in  $DE_{\text{mean}}$  and patient characteristics/features were studied. Deviation in organ locations and  $DE_{\text{mean}}$  for the liver, spleen, and right kidney were moderately correlated ( $R^2 > 0.5$ ) for 8/8, 5/8, and 3/4 reference plans, respectively. Deviations in organ volume or spatial overlap and  $DE_{\text{mean}}$  for the right and left kidney were weakly correlated ( $0.3 < R^2 < 0.5$ ) in 4/4 and 1/4 reference plans. No correlations ( $R^2 < 0.3$ ) were found between deviations in age or height and  $DE_{\text{mean}}$ . Therefore, the performance of organ dose reconstruction using surrogate patients' CT scans is primarily related to deviation in organ location, followed by volume and spatial overlap. Further, results were plan dependent.

Keywords: dose reconstruction, dose comparison, anatomical variation, childhood cancer, abdominal tumors

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Childhood cancer patients are cured about 80% of the times nowadays. Compared to adult cancer patients, they have a longer life expectancy, but a lower tolerance to ionizing radiation [1, 2]. Thus, childhood cancer survivors that received radiation treatment (RT) are more prone to late adverse effects in their lifetime [1, 3]. Having precise knowledge of dose to organs at risk (OARs) during treatment could enable understanding of the radiation dose and late adverse effects relationship, and ultimately improve the design of future treatments. However, obtaining dose to OAR information is hampered by the fact that the majority of patients with long-term (>20 years) follow-up were treated in the era when 2D radiographic images taken by a radiotherapy simulator were used for RT field localisation and dose calculation (e.g. pre-1990s in the Netherlands) [4, 5]. 3D dose distributions are therefore absent.

To address the absence of 3D images for these patients, various dose reconstruction methods have been developed. Phantom-based methods provide either physical measurements or dose computation, using physical or computational phantoms that simulate the human anatomy, respectively [5]. Compared to physical phantoms with fixed geometries, computational phantoms have advantages in versatility, efficiency, and safety. The recently introduced 'hybrid' computational phantoms combine voxel-based images with advanced boundary representation methods that allow organ reshaping and repositioning, achieving both realistic and flexible anatomy [6–9].

Phantoms were originally used in prospective dosimetry studies for radiological protection purposes [9]. So-called reference phantoms were designed to simulate the representative anatomy of a specific age and gender in the population matching the published reference human anthropomorphic values [6, 10, 11]. However, anatomical variations exist in the population both in the outer body and in internal anatomy [12]. Also, in RT dosimetry, rather than a low, homogeneous dose over the scanned body volume as in CT exams, the dose distribution is mostly heterogeneous in the region of interest as high dose is delivered to the tumor site and surrounding normal tissues are blocked as much as possible. Therefore, for

retrospective *individualised* RT dosimetry studies, the accuracy of a dose reconstruction largely depends on how similar the geometry of the phantom was to the specific geometry of the individual patient [9, 13]. While phantom libraries were developed to provide more variability in anatomies [14–18], it is still not clear how to best select/deform a phantom for a patient for whom 3D image data is absent. A recent study on CT dosimetry matched phantoms by age and gender, height and weight, and water equivalent diameter (defined as average of the scanned range of the body), and found the average difference in organ mean dose to be 24%, 14% and 12%, respectively [19]. In a study on RT dosimetry, it was shown that lungs and heart shape, visible in 2D radiographs, could guide the deformation of a 3D organ model to provide satisfactory organ dose reconstruction results ( $\leq 0.3$  Gy mean organ dose deviation in lungs, heart, and breasts) [20].

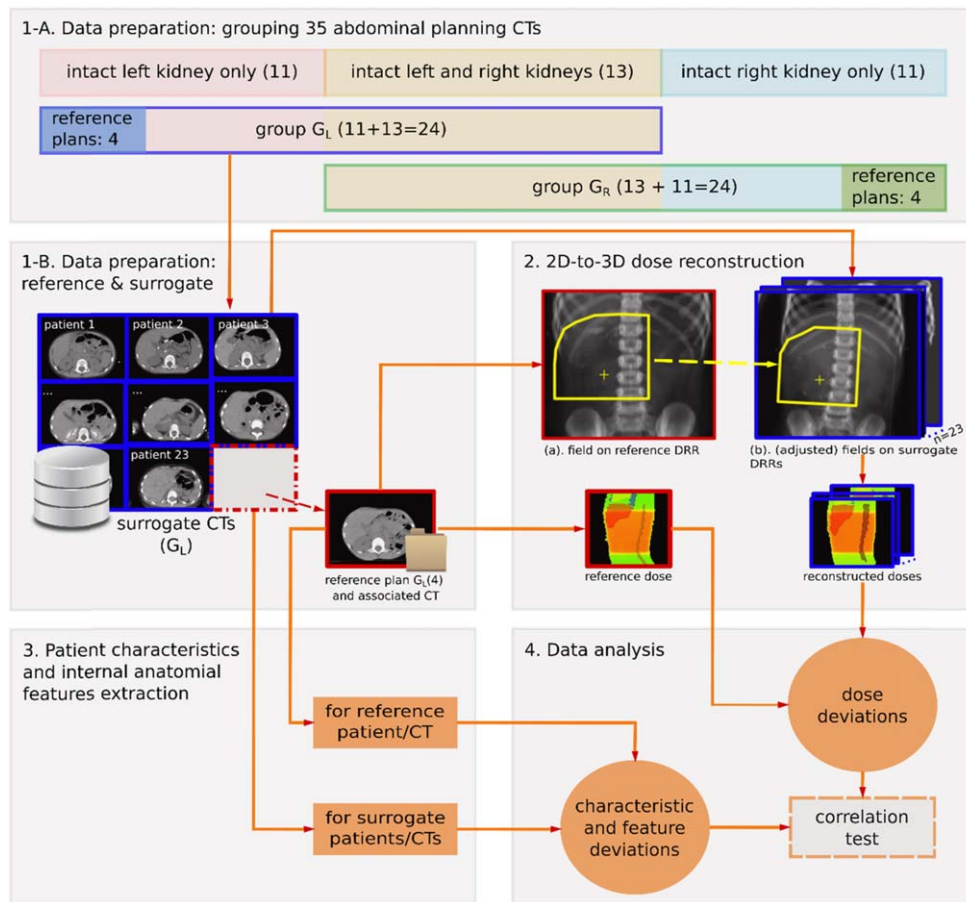
Studies using pediatric phantom libraries are rather scarce, and limited in variability in internal anatomy [11, 17, 18]. The availability of CT scans, acquired for RT planning purposes of recently-treated childhood cancer patients, provides possibilities for creation of a library of CT scans that represents the anatomical variations among childhood cancer patients. The idea is that one of the CT scans from the library can be selected that closely resembles the historically-treated patient's anatomy to perform 3D organ dose reconstruction. However, the available data of the historically-treated patients is limited, and it is yet unknown how to best select the corresponding surrogate CT scan.

In a previous study, we assessed a 2D-to-3D dose reconstruction approach (3D dose reconstruction for patients with only 2D imaging data available) using age- and gender-matched surrogate CTs, which is the literature standard of matching patients for organ dose reconstruction purposes [21]. We focused on abdominal CTs of childhood cancer patients and Wilms' tumor (the most common type of kidney cancer in children) RT plans, which in the 2D-era were planned on abdominal anterior-posterior—posterior-anterior (AP-PA) simulator films [22]. Although for a subset of reconstructions the obtained dose reconstruction errors were small, the use of an age- and gender-matched surrogate CT scan did not guarantee sufficient accuracy [21]. Since geometrical differences in anatomy were considered to be the source of dose reconstruction uncertainty [9, 13], it is expected that a better understanding of patient characteristics (e.g. height) and internal anatomical features (e.g. organ location) in relation to organ dose reconstruction accuracy will benefit the selection of the best-matched surrogate CT scan(s). Moreover, ongoing research is assessing the feasibility of modeling internal anatomy based on parameters that are available in the 2D setting (e.g. height, anatomy visible in 2D radiographs) [23, 24].

In this study, we investigated the correlation of a selection of patient characteristics and internal anatomical features with organ dose reconstruction accuracy. Again, we focused on recently used Wilms' tumor RT plans in our medical center, as for these an AP-PA field setup similar to historical abdominal RT planning was still applied. Different from the previous study, we did not select surrogate CTs based on pre-assumed matching criteria such as age and gender, but used all available CTs as surrogates. Furthermore, we quantified deviations in patient characteristics and anatomical features and investigated their relation to dose reconstruction accuracy.

## 2. Materials and methods

A flowchart illustrating our approach is presented in figure 1. The key idea is that reference data (each reference plan with associated reference CT) are used to simulate a patient for whom only 2D data is available. Therefore, a digitally-reconstructed-radiograph (DRR) is



**Figure 1.** Pipeline of the methodology. From the 35 planning CTs, two groups of 24 CTs each are formed:  $G_L$  (with left kidney) and  $G_R$  (with right kidney) (block 1-A). In block 1-B, a reference plan with associated CT is selected from the  $G_L$  group and the remaining 23 CTs are used as surrogate CTs. In block 2, (a) illustrates the DRR (AP view) derived from the reference CT with the effective field borders of the reference plan (in solid yellow lines), and (b) illustrates the DRRs derived from the surrogate CTs with the adjusted field borders of the reference plan (only one case fully visible). The yellow cross points indicate the field isocenters. Further, patient characteristic/feature deviations are extracted and correlations with dose deviations are investigated (blocks 3 and 4).

created from the reference CT to simulate the historical radiograph and the reference RT fields are projected on this DRR. However, for the internal anatomical feature extraction and reference dose calculation, the actual reference CT is used, to enable comparison with the surrogate CTs and the reconstructed doses. This will be explained in more detail in the following sections.

### 2.1. Data preparation

We retrospectively included a cohort of 35 childhood cancer patients, who received RT at the Academic Medical Center/Emma Children’s Hospital (AMC) ( $n = 33$ ) or at the University

Medical Center Utrecht/Princess Máxima Center for Pediatric Oncology (UMC Utrecht) ( $n = 2$ ) from 2004 to 2016, for whom CT scans of the abdomen were available. Data were anonymised prior to analysis. The requirement of ethical approval of this study involving human subjects has been waived by the local medical ethics review committee.

The CT scans from AMC (voxel size:  $0.9375 \times 0.9375 \times 3 \text{ mm}^3$ ) were acquired on a GE Lightspeed RT 16 CT machine and the CT scans from UMC Utrecht (voxel size  $0.9805 \times 0.9805 \times 3 \text{ mm}^3$ ) were acquired on a Philips Brilliance Big Bore 16 CT machine. Patients' ages at the time of scanning ranged from 2.1 to 5.6 years, as Wilms' tumor mostly occurs within this age group [22]. The patients were diagnosed with different types of childhood tumors with Wilms' tumor as the largest disease category (21 out of 35).

In Wilms' tumor patients, mostly, only one kidney is affected, making the other kidney an OAR during RT. Consequently, CT scans of patients who underwent nephrectomy at one side cannot be used as a surrogate CT for RT plans irradiating the kidney bed at the other side. Two groups of patients were thus formed (figure 1(1-A)):  $G_L$  consisting of patients with an intact left kidney and  $G_R$  consisting of patients with an intact right kidney present in the CT. Further, 13 patients with both intact kidneys were included in both groups. In total, each group consisted of 24 patients. We selected four right-sided Wilms' tumor plans for  $G_L$  and four left-sided Wilms' tumor plans for  $G_R$  as reference RT plans. The associated reference CTs are from patients treated at various ages: two  $\sim 3$ -year-old patients, four  $\sim 4$ -year-old patients, and two  $\sim 5$ -year-old patients. Since the RT treatment protocol for Wilms' tumor remained virtually the same from the pre-1990 era until now [25, 26], the RT plans used in this study and the corresponding results are a solid basis for applying our approach to patients treated in the 2D-era with historical plans.

Detailed information of patient characteristics and patient grouping can be found in table 1.

OARs including the liver, spleen, kidney(s), and spinal cord [22], were manually delineated (Velocity, version 3.2.0, Varian Medical Systems, Inc., Palo Alto, CA, US) by two experienced RT technologists and checked by a radiation oncologist. In order to compare the organ volume and organ dose in the spinal cord, we delineated a sub volume of the spinal cord (which is present in all included CT scans) using the vertebrae as a reference. This sub volume starts at the 10th thoracic vertebra and ends at the 1st sacrum vertebra. From this point on, the term *spinal cord* refers to this specific sub volume.

## 2.2. Plan reconstruction and dose calculation

We selected reference plans of recently-treated patients that were found to be representative of historical Wilms' tumor plans. These plans used an AP-PA field set up, with a prescribed dose of 14.4 Gy applied in eight fractions. According to the clinical SIOP WT 2001 protocol [27], the treatment field of a typical right-sided Wilms' tumor 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 Wilms' tumor covers the major part of the left flank with the vertebral column, the left part of the liver, and the spleen. RT plan field characteristics of the selected reference plans are provided in table 1.

To simulate the historical treatment based on 2D simulator films, the AP-PA field shapes of each reference RT plan were visualised on DRRs that were derived from the corresponding reference CT Rscan. This was achieved by applying a standard rendering method from Oncentra treatment planning system (TPS) (version 4.3, Elekta, Stockholm, Sweden). To mimic the contrast in historical simulator films, we selected an enhancement setting (min/max CT data threshold  $-300/3095$ , center 1500, width 3000, bone threshold 100, and bone

**Table 1.** Overview of patient characteristics, grouping, usage as reference, and the field features of the reference RT plans.

Patient ID	Group	Age (years)	Gender	Height (cm)	Weight (kg)	Tumor site	Used as reference (Plan ID)	Field cranial/caudal borders	Blocked region border
1	G <sub>L</sub> /G <sub>R</sub>	2.2	M	90	15	Right suprarenal gland	N		
2	G <sub>L</sub>	2.3	F	90	12	Left suprarenal gland	N		
3	G <sub>L</sub>	2.4	M	89	13	Right kidney	N		
4	G <sub>R</sub>	2.5	F	93	14	Left kidney	G <sub>R</sub> (1)	Th8/S1	S1
5	G <sub>L</sub> /G <sub>R</sub>	2.6	F	92	13	Left suprarenal gland	N		
6	G <sub>L</sub> /G <sub>R</sub>	2.8	F	105	12	Left suprarenal gland	N		
7 <sup>a</sup>	G <sub>R</sub>	2.8	M	98	18	Left kidney	N		
8	G <sub>R</sub>	2.9	F	98	13	Right kidney	N		
9	G <sub>L</sub>	3.0	M	92	13	Right kidney	N		
10	G <sub>R</sub>	3.1	M	92	10	Left kidney	N		
11	G <sub>L</sub>	3.2	F	99	16	Right kidney	G <sub>L</sub> (1)	Th8/L4	Th8
12 <sup>a</sup>	G <sub>R</sub>	3.2	F	95	13	Left kidney	N		
13	G <sub>L</sub>	3.3	M	96	15	Right kidney	N		
14	G <sub>L</sub> /G <sub>R</sub>	3.3	M	106	17	Ductus choledochus	N		
15	G <sub>R</sub>	3.5	F	98	15	Bilateral kidneys	N		
16	G <sub>L</sub>	3.6	F	98	14	Right kidney	N		
17	G <sub>L</sub> /G <sub>R</sub>	3.7	M	101	14	Right lower abdomen	N		
18	G <sub>L</sub>	3.8	M	102	15	Right kidney	G <sub>L</sub> (2)	Th12/L4	Right rib 11
19	G <sub>R</sub>	3.8	M	104	18	Left suprarenal gland	N		
20	G <sub>L</sub>	3.9	F	108	18	Right kidney	G <sub>L</sub> (3)	Th8/L4	Right rib 8-Th9
21	G <sub>L</sub> /G <sub>R</sub>	3.9	M	100	17	Retroperitoneal	N		
22	G <sub>R</sub>	4.2	F	115	20	Left kidney	G <sub>R</sub> (2)	Th10/L4	Left rib 10
23	G <sub>R</sub>	4.2	F	101	16	Left kidney	N		
24	G <sub>R</sub>	4.2	M	106	16	Left kidney	G <sub>R</sub> (3)	Th11/L4	Left rib 10
25	G <sub>L</sub> /G <sub>R</sub>	4.7	M	119	23	Left suprarenal gland	N		
26	G <sub>L</sub>	4.8	F	110	15	Right kidney	N		
27	G <sub>L</sub> /G <sub>R</sub>	4.8	F	103	17	Pelvic region	N		

**Table 1.** (Continued.)

Patient ID	Group	Age (years)	Gender	Height (cm)	Weight (kg)	Tumor site	Used as reference (Plan ID)	Field cranial/caudal borders	Blocked region border
28	G <sub>L</sub> /G <sub>R</sub>	4.9	M	105	18	4th ventricle	N		
29	G <sub>L</sub>	4.9	M	123	28	Right kidney	G <sub>L</sub> (4)	Th11/L4	Right rib 10
30	G <sub>L</sub> /G <sub>R</sub>	4.9	F	110	22	Para-aortic metastasis of WT left kidney	N		
31	G <sub>L</sub> /G <sub>R</sub>	5.1	M	109	17	4th ventricle	N		
32	G <sub>R</sub>	5.2	M	116	18	Left kidney	G <sub>R</sub> (4)	Th11/L5	Left rib 10
33	G <sub>L</sub> /G <sub>R</sub>	5.2	F	115	24	Left suprarenal gland	N		
34	G <sub>R</sub>	5.3	F	117	22	Left kidney	N		
35	G <sub>L</sub> /G <sub>R</sub>	5.6	F	113	20	4th ventricle	N		

Abbreviations and notations: F = female; M = male; Th1 through Th12 represent the 12 thoracic vertebrae; L1 through L5 represent the five lumbar vertebrae; S1 through S5 represent the sacral vertebrae.

<sup>a</sup> Data from UMC Utrecht.

enhancement factor 2.5) that gives similar contrast (based on a visual check) as in historical simulator films.

Each reference RT plan was reconstructed on the other 23 patients' CT scans in the respective group. The reference patient's CT scan on which the plan was originally applied and the associated calculated dose are referred to as  $CT_{ref}$  and *reference* dose. The surrogate patients' CT scans and the calculated doses in the respective group are referred to as  $CT_{sur}$  scans and *reconstructed* doses for *that reference plan*. Since for each group there are four reference patients, a  $CT_{ref}$  is at the same time a  $CT_{sur}$  for the other three reference plans.

A reference plan was reconstructed on  $CT_{sur}$  scans in Oncentra TPS as described in our previous study [21]. First, we duplicated the reference plan and positioned it on  $CT_{sur}$  with necessary adjustments to correct for anatomical differences as visible between the DRR derived from the  $CT_{sur}$  and the reference DRR. (figure 1(2)). We then scaled the monitor units to make the dose at the field isocenter the same as in the reference plan [28]. The adjusted plans were checked by an experienced pediatric radiation oncologist (BVB).

We chose a treatment unit in our current TPS (Elekta Linac with a multi-leaf collimator beam limiting device, energy: 6 MV) that is similar to the historical ones, and calculated all the doses for both  $CT_{ref}$  and  $CT_{sur}$  scans to enable a consistent comparison. For reference RT plans that included the use of blocks that are no longer available in the current TPS, we used a multi-leaf collimator (leaf width 0.5 cm) as a substitute to shape the block contour. For all plans, a collapsed cone algorithm was used to calculate the dose, which is considered accurate at in-field and near-field dose [29].

### 2.3. Deviations in patient characteristics and internal anatomical features

Independent from the organ dose reconstruction, the deviations in patient characteristics and internal anatomical features (organ volume, organ location, and organ spatial overlap) between the surrogate and reference patients/CTs were measured (figure 1(3)).

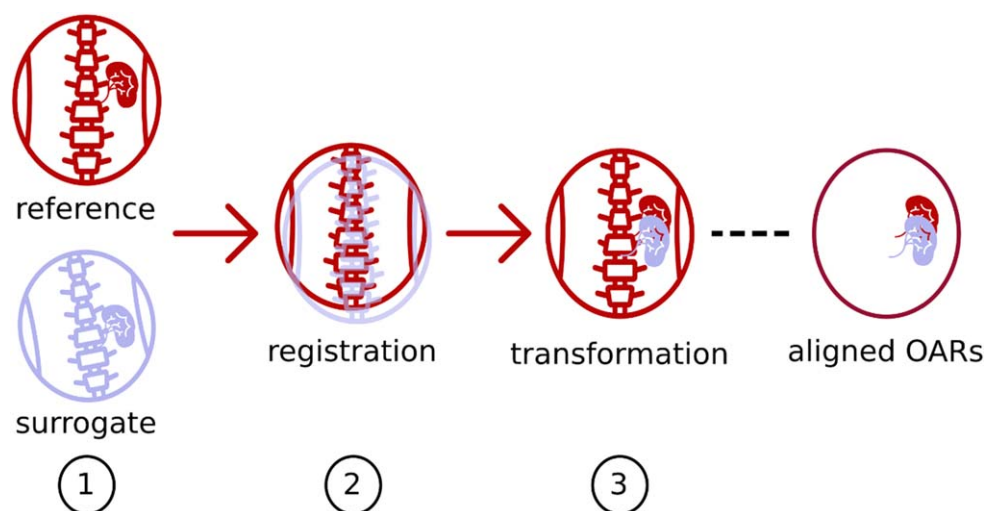
**2.3.1. Deviations in patient characteristics.** We calculated the signed difference in patient age, height, and weight between surrogate patient and reference patient based on table 1.

#### 2.3.2. Deviations in patient internal anatomy

**2.3.2.1. Pre-processing: transform surrogate OARs to reference anatomy.** For each reference plan, we applied a preprocessing step to transform the OAR contours associated with the  $CT_{sur}$  scans to the coordinate system of the  $CT_{ref}$  (see figure 2). This is needed to be able to compute deviations in anatomical features correctly, i.e. discounting for positioning differences. To achieve this, we use image registration techniques. Specifically, since the localisation and adaptation of the treatment plan used the vertebrae and the body shape as reference anatomy visible in DRRs, the registrations were based on the spinal cord, an OAR fixed inside the vertebrae, and the body contours. Two different types of registration were performed: without and with scaling of the contours. The first type of registration was solely performed on the spinal cord, to ensure alignment of all the patients. The latter was performed on the combination of the spinal cord and the body contour, with the possibility to additionally shrink/expand the  $CT_{sur}$ . This enables us to consider only the residual anatomical deviations after the field adjustment was performed in the dose reconstruction, because that partly corrects for anatomical deviations in bony anatomy and body size.

Given a reference plan, the steps to transform OAR contours were:





**Figure 2.** Illustration of transforming an OAR (in this example the left kidney) from a surrogate anatomy to the coordinate system of a reference anatomy based on registration of the spinal cord and outer body contour. Note: in the image, the shape of the vertebral column instead of the spinal cord is used for illustration purpose.

- (1) Convert OARs and body contours of each CT scan to binary images using the Visualisation Toolkit [30]. The binary images of the contours allow the registration methods (in step 2) to solely focus on the shape of the contours.
- (2) Perform automatic Euler (no scaling) or affine (allow scaling) registration using Elastix software [31]. The binary images of the spinal cord, and of the spinal cord combined with the body contour, for non-scaled and scaled registration, respectively, of each  $CT_{sur}$  were registered to that of the  $CT_{ref}$ .
- (3) Apply the deformation vector fields obtained from the registrations to the remaining binary images of OARs of  $CT_{sur}$ , to transform surrogate OARs to the coordinate system of the  $CT_{ref}$ .

Now all the OARs (non-scaled or scaled) are aligned in the coordinate system of  $CT_{ref}$  (see figure 2).

The anatomical deviations between the reference and surrogate CTs were measured based on the non-scaled and scaled contours. In the remainder, the notation O refers to the original non-scaled surrogate OARs, and the notation S refers to the scaled version of surrogate OARs.

**2.3.2.2. Volume.** The organ volume of an OAR was quantified based on the respective OAR contour. Organ volume deviation for an OAR was computed as the (signed) difference between the volume of the reference OAR and the one of the surrogate OAR, for both O and S.

**2.3.2.3. Location.** Organ locations were defined by the geometrical center of the OARs after alignment of the two contour sets based on the registration of the spinal cord (in the case for scaled contours, combined with the body contour), as described above. The deviation in organ

location between the surrogate OARs and reference OARs was measured as signed difference along: right-left (RL), AP, cranial-caudal (CC) as well as the Euclidean distance.

**2.3.2.4. Spatial overlap.** *Spatial overlap after alignment of geometrical centers of OARs: DSC<sub>1</sub>.* The first type of spatial overlap of OARs was assessed by calculating the Dice similarity coefficient (DSC) of the reference OAR contour and the surrogate OAR contour, after alignment of the geometrical centers of the two contours (no rotation performed). This measurement is commonly done to assess agreements in organ shape [32].

$$DSC_1(OAR_{ref}, OAR_{sur}) = \frac{2|V(OAR_{ref} \cap OAR_{sur})|}{|V(OAR_{ref})| + |V(OAR_{sur})|},$$

$$OAR = \{liver, spleen, kidneys, spinal\ cord\}$$

where  $OAR_{ref}$  refers to the reference OAR and  $OAR_{sur}$  refers to the surrogate OAR.  $V$  denotes the volume of the corresponding region.  $DSC_1$  was computed both for O and S.

*Spatial overlap after alignment of spinal cord (and body contour) of the respective CTs: DSC<sub>2</sub>.* In a dose reconstruction case, the anatomical deviations in OAR volume, shape, and location will influence the reconstruction quality in a joint fashion. Thus, we additionally considered a second type of spatial overlap measurement by directly calculating the DSC of the  $OAR_{sur}$  and  $OAR_{ref}$ , after they were pre-processed as described above, without aligning them on the center of mass. To distinguish from the previously introduced  $DSC_1$  after alignment of the geometrical centers, we denote this DSC as  $DSC_2$ . Similarly,  $DSC_2$  was computed both for O and S.

#### 2.4. Data analyses

The organ mean dose was calculated in the reconstructed and the reference plans. To assess the level of agreement between the reconstructed and the reference dose, we computed absolute deviations in organ mean dose ( $DE_{mean}$ , in Gy) for each OAR by subtracting the reconstructed mean dose from the reference mean dose for that organ. In addition, the relative  $DE_{mean}$  was computed as  $DE_{mean}$  divided by the reference organ-specific mean dose (reported in %). The average deviation magnitude was also reported (i.e.  $|DE_{mean}|$  in Gy and in %).

The anatomical deviations in organ volume, location, and spatial overlap based on the non-scaled OARs and scaled OARs between each reference patient and all associated surrogate patients were computed. As it is reasonable to expect the deviations based on scaled OARs to be smaller (in the case of DSC larger values indicate smaller deviations) in absolute values than the ones based on non-scaled ones, a paired one-sided  $t$ -test was performed to test this hypothesis, using statistical significance set at  $p < 0.05$ .

So far, deviations between the reference and surrogate anatomies of three patient characteristics (age, height, and weight) and seven internal anatomical features (organ volume, organ location in RL, AP, and CC directions and Euclidean distance, OAR spatial overlap  $DSC_1$  and  $DSC_2$ ) were introduced. Deviations in patient characteristics were measured based on the original data. Deviations in the anatomical features were measured based on both non-scaled and scaled OARs. Therefore, in total 17 types of deviations were generated. We calculated the Pearson's correlation coefficients of the 17 types of characteristic/feature deviation with  $DE_{mean}$ . For deviations that can only be positive (i.e. Euclidean distances, DSCs), their correlations with  $|DE_{mean}|$  were tested.

### 3. Results

#### 3.1. Organ dose reconstruction results

Boxplots of the reconstructed organ mean doses together with the reference organ mean dose for each reference plan are presented in figure 3. The average  $|DE_{\text{mean}}|$  in Gy and relative  $|DE_{\text{mean}}|$  in % are summarised in table 2.

For  $G_R$  plans (i.e. plans among patients with an intact right-sided kidney, i.e. primary left-sided Wilms' tumor), the reference mean liver doses vary from 2 to 6 Gy and the average liver  $|DE_{\text{mean}}|$  values for the reconstructions vary from 0.7 to 1.2 Gy (relative  $|DE_{\text{mean}}|$  14%–32%). The reference mean spleen doses vary from 7 to 14 Gy for  $G_R$  plans and large deviations were found in the reconstructions, with the average  $|DE_{\text{mean}}|$  values from 0.2 to 3.7 Gy (relative  $|DE_{\text{mean}}|$  1%–34%). For  $G_L$  plans, the reference mean liver doses vary from 4 to 13 Gy, and the average  $|DE_{\text{mean}}|$  values for the reconstructions vary from 1.6 to 3.2 Gy (relative  $|DE_{\text{mean}}|$  16%–41%). The reference mean spleen doses are low (<1 Gy) for  $G_L$  plans and average  $|DE_{\text{mean}}|$  values <0.3 Gy were found in the reconstructions (relative  $|DE_{\text{mean}}|$  22%–78%). For both  $G_R$  and  $G_L$  plans, the contralateral kidneys have a relatively low reference mean dose (<3 Gy), and small average  $|DE_{\text{mean}}|$  values (<1 Gy) were obtained for the kidneys in the reconstructions, while the relative  $|DE_{\text{mean}}|$  values (>18%) were comparable to the liver and the spleen. For both  $G_L$  and  $G_R$  plans, the reference mean spinal cord doses are relatively high (>8 Gy), meanwhile, small average  $|DE_{\text{mean}}|$  values were obtained in the spinal cord in the reconstructions (<1 Gy, relative  $|DE_{\text{mean}}|$  <10%).

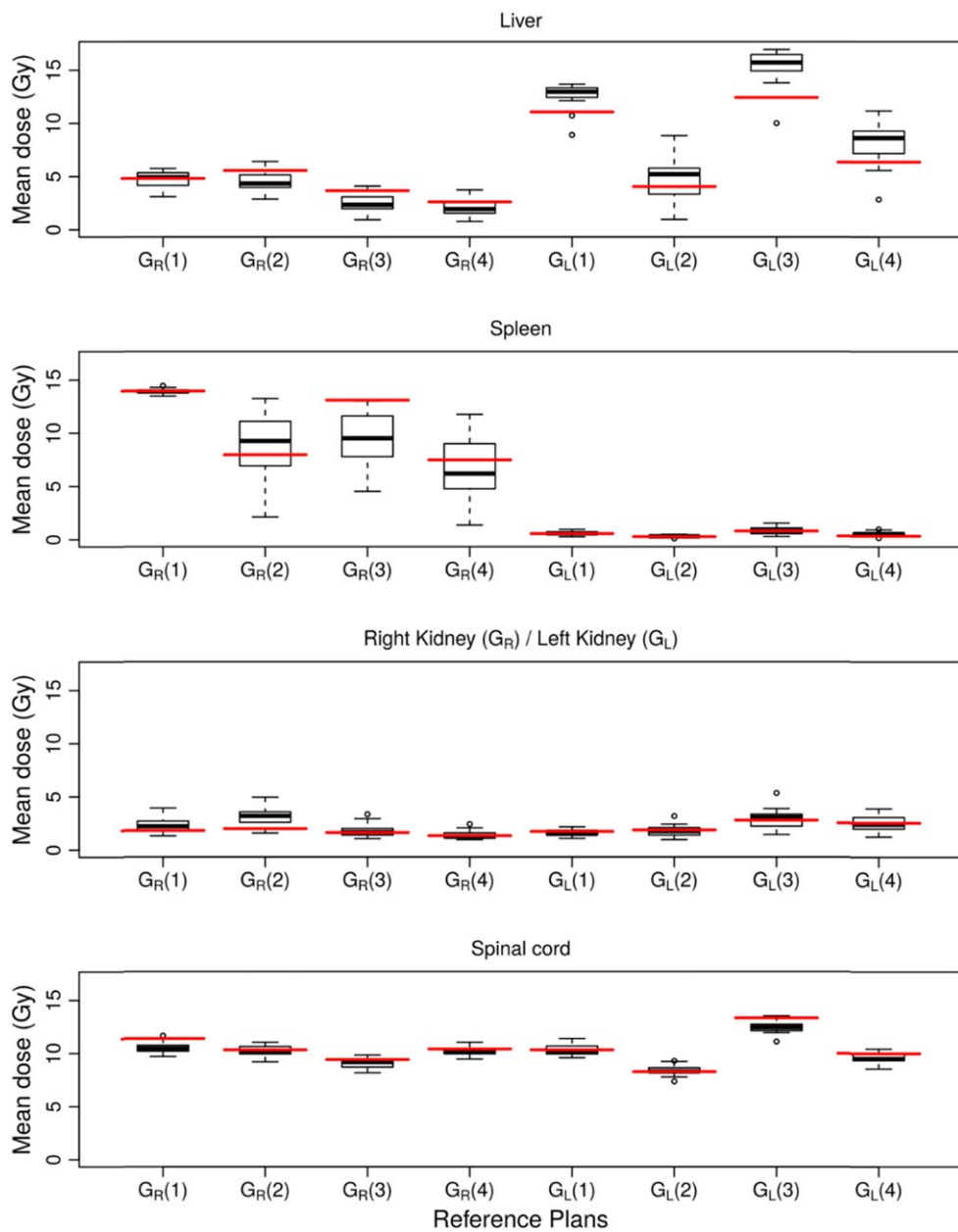
#### 3.2. Internal anatomical deviations

The distribution of the geometrical centers of all the OARs, when aligned by the geometrical center of the spinal cord of each CT (without scaling), is presented in figure 4. The average anatomical deviations and the  $p$ -values of the one-sided  $t$ -test comparing the deviations based on the non-scaled and scaled OARs are summarised in table 3.

We observed that for the liver, spleen, and right kidney 4/7 of the feature deviations were significantly reduced using scaled OARs compared to non-scaled OARs. For the left kidney, no significant differences were found ( $p > 0.05$ ). Of note, no comparison of deviations in anatomical features for the spinal cord is provided as the spinal cord was used as reference to align or scale the patient anatomies.

#### 3.3. Correlations

The top three highest correlations between the deviation in characteristics/features and  $DE_{\text{mean}}$  (or  $|DE_{\text{mean}}|$ ) (Pearson's R-squared) are summarised per OAR for each  $G_L$  and  $G_R$  plans in figure 5. For the liver, in general, the features that correlate most with  $DE_{\text{mean}}$  are organ location in AP direction for  $G_R$  plans, and organ location in CC direction for  $G_L$  plans. For the spleen, in general, organ location in CC direction correlates most with  $DE_{\text{mean}}$  for  $G_R$  plans and organ location in RL correlates most with  $DE_{\text{mean}}$  for  $G_L$  plans. For the right kidney, only organ location in RL direction has a moderate correlation with  $DE_{\text{mean}}$  for all the reference  $G_R$  plans. For the left kidney and the spinal cord, none of the features had an  $R^2 > 0.5$  correlation with  $DE_{\text{mean}}$  (or  $|DE_{\text{mean}}|$ ). There are differences in the most three correlated features among the same type ( $G_R$  or  $G_L$ ) of plans. For instance only for reference plan  $G_R(3)$ , a moderate to strong correlation ( $R^2 > 0.5$ ) between  $DSC_2$  and  $|DE_{\text{mean}}|$  for the liver and the spleen, and between Euclidean distance and the  $|DE_{\text{mean}}|$  for the spleen can be observed. Also, only for reference plan  $G_R(1)$ , a moderate correlation ( $R^2 > 0.5$ ) is found



**Figure 3.** Boxplot of the reconstructed organ mean dose per reference plan. The whiskers extend to the 10th and 90th percentiles, the boxes extend from the 25th to the 75th percentiles, the thick horizontal line inside the box indicates the 50th percentile (i.e. the median), and the circles indicate outliers. The red line indicates the reference organ mean dose. We combined the boxplots for the right kidneys in G<sub>R</sub> and the left kidneys in G<sub>L</sub> reference plans in one row.

**Table 2.** Average organ  $|DE_{\text{mean}}|$  (in Gy and in relative values %) of dose reconstructions obtained for each reference plan.

Reference plans	Liver	Spleen	Left Kidney	Right Kidney	Spinal cord
G <sub>R</sub> (1)	0.7 Gy, 14%	0.2 Gy, 1%		0.6 Gy, 36%	0.9 Gy, 8%
G <sub>R</sub> (2)	1.2 Gy, 22%	2.5 Gy, 32%		1.2 Gy, 59%	0.4 Gy, 4%
G <sub>R</sub> (3)	1.2 Gy, 32%	3.7 Gy, 29%		0.4 Gy, 25%	0.4 Gy, 5%
G <sub>R</sub> (4)	0.8 Gy, 31%	2.5 Gy, 34%		0.3 Gy, 21%	0.3 Gy, 3%
G <sub>L</sub> (1)	1.8 Gy, 16%	0.1 Gy, 22%	0.3 Gy, 18%		0.5 Gy, 4%
G <sub>L</sub> (2)	1.6 Gy, 41%	0.1 Gy, 25%	0.4 Gy, 24%		0.3 Gy, 4%
G <sub>L</sub> (3)	3.2 Gy, 25%	0.3 Gy, 33%	0.7 Gy, 25%		0.9 Gy, 7%
G <sub>L</sub> (4)	2.2 Gy, 35%	0.3 Gy, 78%	0.6 Gy, 24%		0.5 Gy, 5%
<b>Average</b>	<b>1.6 Gy, 27%</b>	<b>1.2 Gy, 32%</b>	<b>0.5 Gy, 23%</b>	<b>0.6 Gy, 35%</b>	<b>0.5 Gy, 5%</b>

Note: the reported values were rounded to 0.1 Gy and 1%.

between  $DSC_2$  and  $|DE_{\text{mean}}|$  for the right kidney. The results further indicate no or only a very small correlation ( $R^2 < 0.5$ ) between deviation in patient age, height, or weight and  $DE_{\text{mean}}$ .

Regarding the difference between scaled and non-scaled OARs, of the 21 times that deviations in anatomical features were found to be correlated with  $DE_{\text{mean}}$  (or  $|DE_{\text{mean}}|$ ) with Pearson's R-squared  $> 0.5$ , 16 times a stronger correlation was found for feature deviations based on scaled OARs compared to non-scaled OARs and, only 5 times it was the other way around.

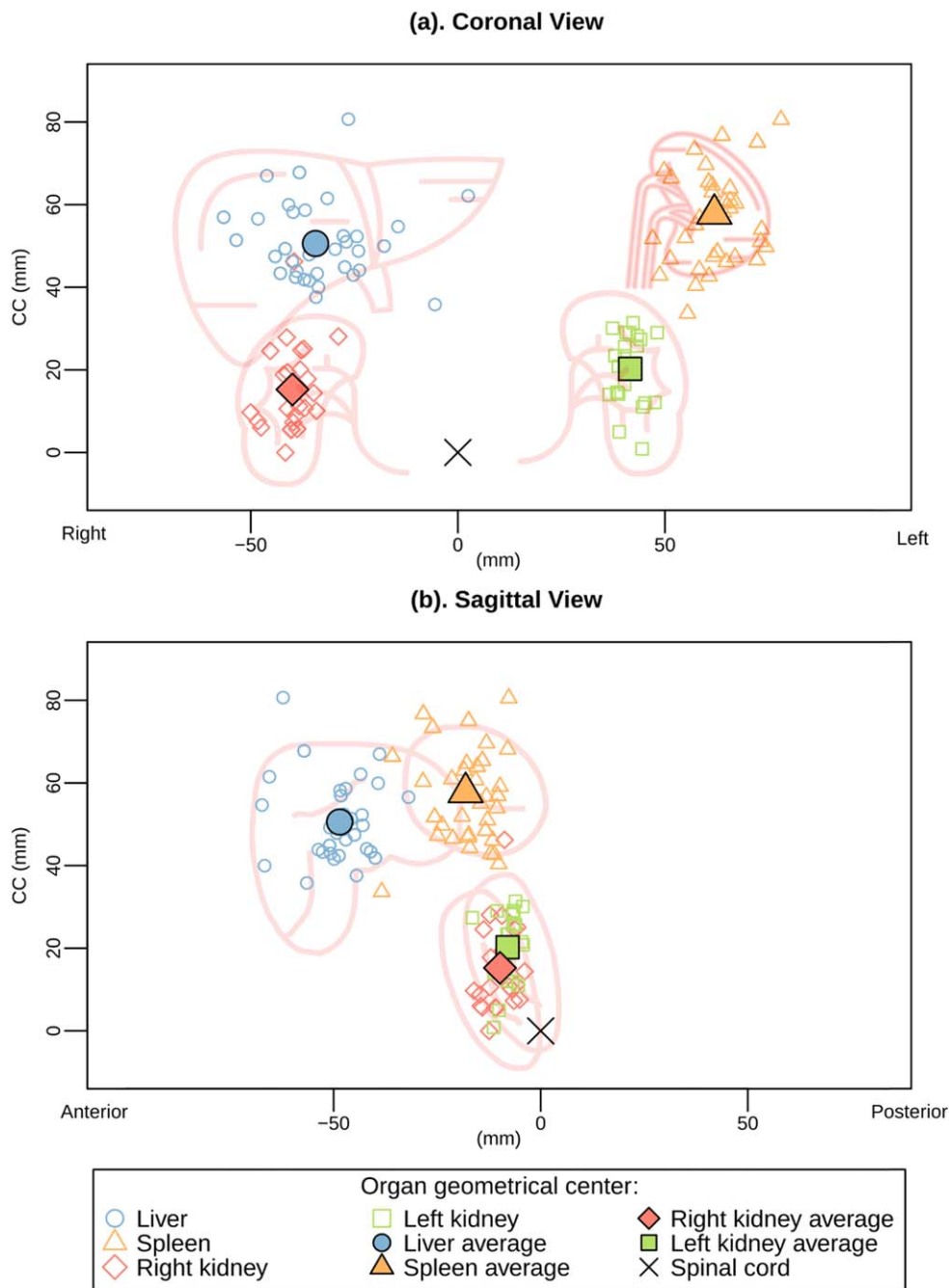
For each type of the plan (G<sub>R</sub>/G<sub>L</sub>), we selected the anatomical feature that on average had the largest correlation among the four reference plans. Scatter point plots of the deviation in the anatomical feature versus  $DE_{\text{mean}}$  (or  $|DE_{\text{mean}}|$ ) for four OARs are presented in figure 6. Plots for the spinal cord were not included because no correlations ( $R^2 < 0.4$ ) between feature deviations and  $DE_{\text{mean}}$  were found. When the mean organ dose is small (e.g. for the right kidney and the spleen for G<sub>L</sub> plans), even in case of a strong correlation between deviation in the feature and deviation in  $DE_{\text{mean}}$ , the scatter plots still illustrate small slopes.

#### 4. Discussion

In this study, for the first time to our knowledge, the relation of deviations in organ dose reconstruction using surrogate CT scans with deviations in patient characteristics such as age, height, and weight, and internal anatomical features such as organ location was comprehensively investigated. The focus was on Wilms' tumor RT plans from childhood cancer patients with an age range of 2–6 years during treatment.

Our analysis indicates that the performance of organ dose reconstruction using surrogate childhood cancer patients' CT scans is not consistently highly correlated to any of the tested features. In general, it is primarily related to organ location, organ volume, and spatial overlap of organs, and not related to patient age, height and weight. Further, it is dependent on the particular RT plan. These findings will guide the selection/construction of a surrogate anatomy to ensure accurate dose reconstruction.

A large variety in the  $DE_{\text{mean}}$  values was observed for the liver and the spleen for the investigated reference RT plans. For the liver and the spleen, organ location in all three directions were found to be the features that correlated most with organ  $DE_{\text{mean}}$ . This can be explained by the OAR location relative to the reference RT field: the liver and the spleen (for G<sub>R</sub> plans) were mostly located on the edge of the field (i.e.  $\sim 1/3$ – $2/3$  organ volume inside

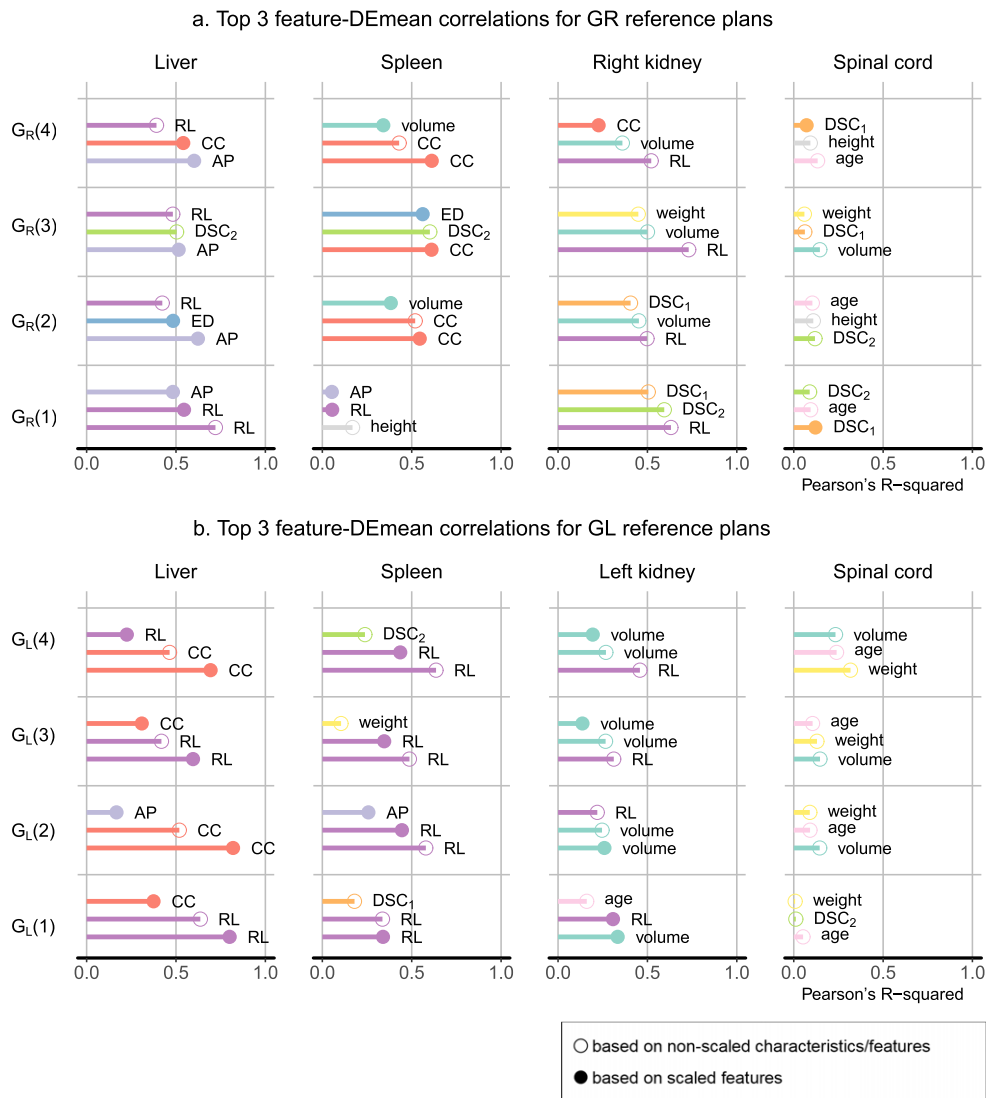


**Figure 4.** OAR geometrical center distributions in coronal view (a) and sagittal view (b), when the OARs (based on delineations) of each CT are aligned by the geometrical center of the spinal cord (black cross). Each non-filled symbol represents the geometrical center of an OAR associated with a CT included in this study. The filled symbols of each type represent the average locations of the OARs. The shapes of the OARs were added for illustration purpose.

**Table 3.** Average deviations in OAR volume, OAR location (in three directions and in terms of Euclidean distance), and organ spatial overlap (DSC<sub>1</sub> and DSC<sub>2</sub>) between reference and surrogate OARs, based on non-scaled and scaled OARs; and the *p*-value of the one-sided paired *t*-test based on non-scaled and scaled OARs (significance threshold: 0.05).

OAR	Liver			Spleen			Right kidney			Left kidney		
	O	S	<i>p</i>	O	S	<i>p</i>	O	S	<i>p</i>	O	S	<i>p</i>
Feature												
Volume (cm <sup>3</sup> )	133	109	0.01	35	39	NS	29	27	NS	24	31	NS
RL (mm)	14	12	<0.001	9	5	<0.001	5	4	0.03	3	4	NS
AP (mm)	9	10	NS	8	8	NS	4	3	NS	3	4	NS
CC (mm)	11	13	NS	14	12	0.02	11	11	NS	11	11	NS
Euclidean distance (mm)	23	22	NS	20	18	<0.001	14	13	0.01	13	14	NS
DSC <sub>1</sub>	0.70	0.72	<0.001	0.61	0.61	NS	0.74	0.75	0.02	0.76	0.75	NS
DSC <sub>2</sub>	0.60	0.68	<0.001	0.39	0.49	<0.001	0.57	0.59	0.01	0.64	0.61	NS

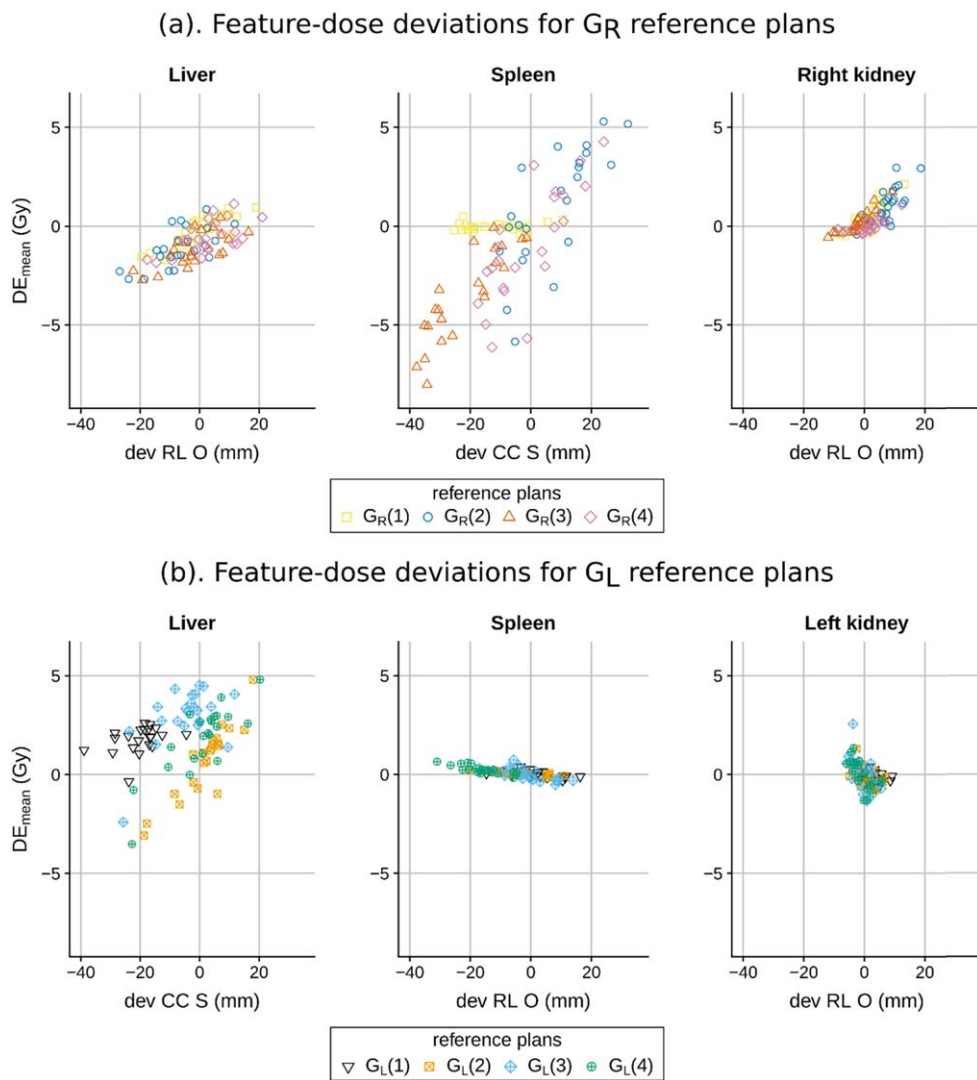
Abbreviations: O = Original (non-scaled) contours; S = Scaled contours; *p* = *p*-value; NS = not significant (when *p*-values >0.05).



**Figure 5.** The three highest squared Pearson’s correlation coefficients (i.e. R-squared) between deviations in patient characteristics/anatomical features and  $DE_{mean}$  (or  $|DE_{mean}|$ , if the deviation characteristic/feature has a scalar quantity) for OARs are presented as lollipop plots for each reference plan in  $G_R$  (a) and  $G_L$  (b). Text following each lollipop indicates the characteristic/feature. The solid or open circle of a lollipop indicates whether the characteristic/feature is scaled or not.

the field), where a steep dose gradient is present. Moreover, the liver and the spleen were found to have a larger organ location variation compared to the kidneys (figure 4). The contralateral kidneys, with only a small part of the volume in-field for the reference RT plans, received a low mean dose and were found to have smaller  $|DE_{mean}|$  but similar relative  $|DE_{mean}|$  compared with the liver and the spleen. At the same time, there are also differences between the two kidneys. The right kidney, which is located more caudally than the left kidney to accommodate the liver, was found to have larger  $|DE_{mean}|$  and smaller  $DSC_2$  values





**Figure 6.** Scatter plots of the feature deviations versus  $DE_{mean}$  for which the largest correlations were found for reference plans in  $G_R$  (a) and  $G_L$  (b). Each point represents one reconstruction case, with color and shape referring to the reference plan.

than the left kidney. While moderate correlations between anatomical features deviations (e.g. location in CC direction,  $DSC_2$ , and organ volume) and  $DE_{mean}$  (or  $|DE_{mean}|$ ) were found for the right kidney, no clear correlations were found for the left kidney. Similar to the left kidney, small  $DE_{mean}$  values were obtained for the spinal cord (average  $|DE_{mean}| < 1$  Gy, relative  $< 1\%$ ) and no correlations were found with any of the feature deviations. This could be explained by the following facts: (1) the spinal cord is fixed in the vertebrae while the positioning and adaptation of the RT fields is mainly based on bony anatomy of the vertebrae, and (2) the spinal cord is an inevitable in-field OAR according to field definitions due to its anatomic vicinity to the affected kidney. The spinal cord thus received a highly homogeneous

dose that is insensitive to small shape/size variations. Generally speaking, when the  $DE_{\text{mean}}$  is small, it is hard to find its correlation with deviation in the considered characteristics/features. To summarise, the quality of organ dose reconstruction is not only affected by the anatomical deviations respective to the reference anatomy, but also affected by anatomical variations, in particular organ location, respective to the field borders (e.g. in-field, on the edge of the field, or near-field).

Since there is a steep dose gradient on the field boundary, the field site (left/right) and shape of the Wilms' tumor reference plan influenced the correlations between feature deviations and  $DE_{\text{mean}}$ . For example, the field of reference plan  $G_R(1)$ , with 8th thoracic vertebra and 1st sacral vertebra as the cranial and caudal field borders, respectively, fully covered the spleen in  $CT_{\text{ref}}$  and  $CT_{\text{sur}}$  scans. Therefore, reconstructed mean doses for the spleen were virtually the same; accordingly, no strong correlation was found between the deviation in anatomical features and the dose deviation in the spleen. For plans where a larger part of the liver/spleen is intersected by the cranial field border than by the left/right field border,  $DE_{\text{mean}}$  for liver/spleen is more sensitive to organ location deviations along CC direction than along RL direction (see figure 6). Also, organ location variations were larger along the CC direction than along the AP and RL directions. This can be partly explained by the difference in stages of respiration captured in the CT scans, as respiration-induced organ motion mainly occurs along the CC direction [33].

In addition to variation in organ location, variations in organ volume and spatial overlap after alignment of the geometrical organ centers (i.e.  $DSC_1$ ) represent OAR anatomical variations that were expected to additionally contribute to the reconstructed dose deviations. However, the results of the correlation tests showed that deviation in these features does not generally correspond to deviation in mean dose. It indicated that these features alone do not dominate the organ mean dose accuracy, especially when the geometrical center of the OARs have large variations (in the case for the liver and the spleen). Furthermore, Euclidean distance and  $DSC_2$  are measurements that combine organ location distances in three directions, and organ volume, location, and spatial overlap differences, respectively. There were also no (moderate) correlations found between deviation in these features and  $|DE_{\text{mean}}|$ . Considering the definition of the measurements, they were isotropic in measuring dissimilarity in anatomy: a deviation in organ location to the right and the left could result in the same value for both the Euclidean distance and  $DSC_2$ . However, the relative location of the RT field to an OAR is anisotropic: displacement in different directions could lead to a totally different dosimetric change.

The low correlations found for deviations in age, height, and weight with  $DE_{\text{mean}}$  suggest that these features are insufficiently specific indicators of a surrogate CT to ensure accurate organ dose reconstruction. Thereby, this study confirms and extends the findings from our previous work that indicated that using only age and gender as matching criteria in organ dose reconstruction using surrogate childhood cancer patients' CT scans is not sufficient [21].

The patient characteristics age, height, and weight were analyzed because they are commonly investigated in phantom-based dose reconstruction studies for matching a patient with a phantom [14, 34]. Organ volume was included as a commonly used measurement for organ variation between patients. Organ location deviations were quantified not only in terms of Euclidean distance, but also in three orthogonal directions separately, to provide insight on displacements with certain directions. Two types of spatial overlap were included to provide overall assessments on OAR similarity combining similarity in organ volume and shape with orientation ( $DSC_1$ ) and with organ location ( $DSC_2$ ). In a previous feasibility study from our group that used machine learning techniques to predict anatomical similarity from available

features of historically-treated patients [23], several notions of patient similarity were proposed. However, these notions of patient similarity were not validated in terms of dose reconstruction accuracy, which is ultimately the goal of dose reconstruction. Some similarity notions are not included in this article such as the Hausdorff distance between two aligned OAR contours, so-called overall organ constellation, and a modification of the organ volume histogram overlap (OVH), an isotropic measurement of organ overlap that is similar to the concept of  $DSC_2$  [35]. In preliminary studies, we found that the Hausdorff distance was highly correlated with the DSC and resulted in similar correlations with dose deviations. For the OVH we did not find better correlations with deviations in dose than for  $DSC_2$ . As we believe that  $DSC_2$  represents a more direct and easy-to-understand overlap measurement, we presented  $DSC_2$  instead of OVH. Furthermore, no correlation between the overall organ constellation (sum of square of pairwise Euclidean organ distances) and the deviation in organ mean dose was found in our preliminary study.

In this study, anatomical deviations were measured based on both the original non-scaled and scaled OARs. The option where the OARs of the surrogate CT were scaled with respect to the spinal cord and body contour of the reference anatomy was investigated because the reference plans were also ‘rescaled’ based on the visible anatomy in a 2D image (described in section 2). The comparison of the results showed that anatomical deviations between surrogate and reference OAR were typically reduced when based on the scaled OAR except for the left kidney. Also, stronger correlations between deviations in features and deviations in organ mean dose were found based on scaled OARs than non-scaled OARs. It might, therefore, be of interest for future studies concerning the selection of similar patients/anatomies, to consider using scaled anatomies (based on 2D measurements), instead of original anatomies. On the other hand, for example, for organ location along the AP direction for all OARs, no significant reduction in deviation was found based on scaled contours compared to non-scaled contours. Even for the significantly reduced anatomical deviations, they were not eliminated: at most, scaling OARs reduced deviations for up to 50% of the deviations based on non-scaled contours. These observations again indicate that internal organ anatomies were not determined by the patient’s spinal cord and body contour.

A limitation of our study is the small number of patient data included, which provided a limited spread of anatomical variations in the 2–6 years old age range. It is expected that more statistically convincing, and probably stronger correlations will be found, once we include more patients. Secondly, our study was focused on Wilms’ tumor plans and OARs in the abdominal region. Nevertheless, as flank irradiation with AP-PA field set-up is typical in historical RT planning for other types of abdominal cancers [1], we expect that the general conclusions on characteristic/feature deviations in relation to dose deviations may also hold for other type of abdominal RT plans. Furthermore, in this study we limited ourselves to assessing the organ mean dose, which lacks the spatial information of a 3D dose distribution, while 3D dose distributions would provide the potential to develop more detailed models between treatment and late adverse effects. The anatomical variations of OARs between patients precluded direct voxel-to-voxel dose comparisons between the reference and surrogate CT. A possible solution is to employ deformable image registration techniques to allow for voxel-to-voxel dose comparison from two different anatomies. However, the use of deformable image registration would introduce new forms of uncertainty which remain to be quantified and, subsequently, need to be assessed in terms of their impact on dose reconstruction applications [36].

## 5. Conclusion

Internal anatomical deviations in organ location were found to primarily dominate dose reconstruction outcomes in mean dose for OARs, however, the linear correlations were not consistently high for all the investigated reference plans. Deviation in organ volume and spatial overlap contributed to a lesser extent to the mean dose deviations in OARs. Patient characteristics, such as age, height and weight were not linearly correlated with deviations in  $DE_{\text{mean}}$ . It is thus suggested that for organ dose reconstruction, efforts should be made to predict these relevant anatomical features from available (2D) information, and to estimate the extent of OAR coverage by the RT field based on RT plan features.

## Acknowledgments

Financial support of this work was provided by Stichting Kinderen Kankervrij (KiKa; project no. 187). Dr Ronckers is supported by the Dutch Cancer Society (KWF; grant no. UVA20212-5517). The authors thank Koen F Crama and Rianne (M.)A J de Jong for their help with organ delineation, and Raquel Dávila Fajardo (department of Radiation Oncology, UMC Utrecht Cancer Center, Utrecht, the Netherlands) for sharing the data of two patients treated at the UMC Utrecht/Princess Máxima Center for Pediatric Oncology for inclusion in this study. We further thank Dr Astrid van der Horst for manuscript proofreading.

## Conflicts of interest

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

## ORCID iDs

Ziyuan Wang  <https://orcid.org/0000-0001-6336-0300>

## References

- [1] Olch A J 2013 *Pediatric Radiotherapy Planning and Treatment* (Boca Raton, FL: CRC Press) (<https://doi.org/10.1201/b14554>)
- [2] Merchant T E *et al* 2013 Children's oncology Group's 2013 blueprint for research: radiation oncology *Pediatr. Blood Cancer* **60** 1037–43
- [3] Bhakta N *et al* 2017 The cumulative burden of surviving childhood cancer: an initial report from the St Jude lifetime cohort study (SJLIFE) *The Lancet* **390** 2569–82
- [4] Thwaites D I and Tuohy J B 2006 Back to the future: the history and development of the clinical linear accelerator *Phys. Med. Biol.* **51** R343
- [5] Stovall M *et al* 2006 Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies *Radiat. Res.* **166** 141–57
- [6] Cassola V *et al* 2009 FASH and MASH: female and male adult human phantoms based on polygon mesh surfaces: I. Development of the anatomy *Phys. Med. Biol.* **55** 133
- [7] Bolch W *et al* 2010 Hybrid computational phantoms for medical dose reconstruction *Radiat. Environ. Biophys.* **49** 155–68

- [8] Na Y H *et al* 2010 Deformable adult human phantoms for radiation protection dosimetry: anthropometric data representing size distributions of adult worker populations and software algorithms *Phys. Med. Biol.* **55** 3789
- [9] Xu X G 2014 An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history *Phys. Med. Biol.* **59** R233
- [10] ICRP 2002 Basic anatomical and physiological data for use in radiological protection: reference values: ICRP Publication 89 *Ann. ICRP* **32** 1–277
- [11] Stabin M *et al* 2011 Realistic reference adult and paediatric phantom series for internal and external dosimetry *Radiat. Prot. Dosimetry* **149** 56–9
- [12] Ackerman M J 1998 The visible human project *Proc. IEEE* **86** 504–11
- [13] Bezin J V *et al* 2017 A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose–response relationships *J. Radiol. Prot.* **37** R1
- [14] Cassola V *et al* 2011 Standing adult human phantoms based on 10th, 50th and 90th mass and height percentiles of male and female caucasian populations *Phys. Med. Biol.* **56** 3749
- [15] Segars W *et al* 2013 Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization *Med. Phys.* **40** 043701
- [16] Broggio D *et al* 2011 Construction of an extended library of adult male 3D models: rationale and results *Phys. Med. Biol.* **56** 7659
- [17] Tward D *et al* 2011 Generating patient-specific dosimetry phantoms with whole-body diffeomorphic image registration 2011 *IEEE 37th Annual Northeast Bioengineering Conf. (NEBEC)* (New York: IEEE) pp 1–2
- [18] Geyer A M *et al* 2014 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.* **59** 5225
- [19] Stepusin E J *et al* 2017 Assessment of different patient-to-phantom matching criteria applied in Monte-Carlo based computed tomography dosimetry *Med. Phys.* **44** 5498–508
- [20] Ng A *et al* 2012 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.* **84** e557–63 2012-1-1
- [21] Wang Z *et al* 2018 Are age and gender suitable matching criteria in organ dose reconstruction using surrogate childhood cancer patients' CT scans? *Med. Phys.* **45** 2628–38
- [22] van Dijk I W *et al* 2010 Evaluation of late adverse events in long-term wilms' tumor survivors *Int. J. Radiat. Oncol. Biol. Phys.* **78** 370–8
- [23] Virgolin M *et al* 2018 On the feasibility of automatically selecting similar patients in highly individualized radiotherapy dose reconstruction for historic data of pediatric cancer survivors *Med. Phys.* **45** 1504–17
- [24] Ng A *et al* 2012 Navigator channel adaptation to reconstruct three dimensional heart volumes from two dimensional radiotherapy planning data *BMC Med. Phys.* **12** 1
- [25] D'angio G J *et al* 1978 Radiation therapy of Wilms' tumor: Results according to dose, field, post-operative timing and histology *Int. J. Radiat. Oncol. Biol. Phys.* **4** 769–80
- [26] Jereb B *et al* 1994 Radiotherapy in the SIOP (International Society of Pediatric Oncology) nephroblastoma studies: a review *Med. Pediatr. Oncol.* **22** 221–7
- [27] Furtwängler R *et al* 2013 Clear cell sarcomas of the kidney registered on International Society of Pediatric Oncology (SIOP) 93-01 and SIOP 2001 protocols: a report of the SIOP renal tumour study group *Eur. J. Cancer* **49** 3497–506
- [28] Khan F M and Gibbons J P 2014 *Khan's the Physics of Radiation Therapy* 5th edn (Philadelphia: Lippincott Williams & Wilkins)
- [29] Ahnesjö A 1989 Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media *Med. Phys.* **16** 577–92
- [30] Schroeder W M, Martin K and Lorensen B 2004 *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics* 4th edn (New York: Kitware)
- [31] Klein S *et al* 2010 Elastix: a toolbox for intensity-based medical image registration **29** 196–205
- [32] Zou K H *et al* 2004 Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports *Acad. Radiol.* **11** 178–89
- [33] Huijskens S C *et al* 2015 Quantification of renal and diaphragmatic interfractional motion in pediatric image-guided radiation therapy: a multicenter study *Radiother. Oncol.* **117** 425–31
- [34] Lee C *et al* 2015 Reconstruction of organ dose for external radiotherapy patients in retrospective epidemiologic studies *Phys. Med. Biol.* **60** 2309–24

- 
- [35] Kazhdan M *et al* 2009 A shape relationship descriptor for radiation therapy planning *Med. Image. Comput. Comput. Assist. Interv.* **12** 100–8
- [36] Sotiras A, Davatzikos C and Paragios N 2013 Deformable medical image registration: a survey *IEEE Trans. Med. Imaging* **32** 1153–90