Contents lists available at ScienceDirect

Physica Medica

Image-based data mining applies to data collected from children

Lydia J. Wilson [a,*], Abigail Bryce-Atkinson [b], Andrew Green [b], Yimei Li [c], Thomas E. Merchant [a], Marcel van Herk [b], Eliana Vasquez Osorio [b], Austin M. Faught [a], Marianne C. Aznar [b]

[a] *St. Jude Children's Research Hospital, Department of Radiation Oncology, Memphis, TN, USA*
[b] *Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK*
[c] *St. Jude Children's Research Hospital, Department of Biostatistics, Memphis, TN, USA*

ABSTRACT

Purpose: Image-based data mining (IBDM) is a novel voxel-based method for analyzing radiation dose responses that has been successfully applied in adult data. Because anatomic variability and side effects of interest differ for children compared to adults, we investigated the feasibility of IBDM for pediatric analyses.
Methods: We tested IBDM with CT images and dose distributions collected from 167 children (aged 10 months to 20 years) who received proton radiotherapy for primary brain tumors. We used data from four reference patients to assess IBDM sensitivity to reference selection. We quantified spatial-normalization accuracy via contour distances and deviations of the centers-of-mass of brain substructures. We performed dose comparisons with simplified and modified clinical dose distributions with a simulated effect, assessing their accuracy via sensitivity, positive predictive value (PPV) and Dice similarity coefficient (DSC).
Results: Spatial normalizations and dose comparisons were insensitive to reference selection. Normalization discrepancies were small (average contour distance < 2.5 mm, average center-of-mass deviation < 6 mm). Dose comparisons identified differences (p < 0.01) in 81% of simplified and all modified clinical dose distributions. The DSCs for simplified doses were high (peak frequency magnitudes of 0.9–1.0). However, the PPVs and DSCs were low (maximum 0.3 and 0.4, respectively) in the modified clinical tests.
Conclusions: IBDM is feasible for childhood late-effects research. Our findings may inform cohort selection in future studies of pediatric radiotherapy dose responses and facilitate treatment planning to reduce treatment-related toxicities and improve quality of life among childhood cancer survivors.

1. Introduction

Thanks to major advancements in cancer care, 85% of pediatric oncology patients now survive 5 years or more, and many of them become long-term survivors [1]. Among the cancer treatment options available today, radiotherapy remains one of the most commonly used tools, involved in the treatment of approximately half of all childhood cancer cases [2]. Radiation, however, is associated with toxicities that increase the risk of severe chronic health conditions, including cardiac, endocrine, and neurologic conditions, in long-term survivors of childhood cancer [3]. Therefore, radiotherapy research increasingly focuses on mitigating or avoiding its long-term sequalae. Such research either directly [4–6] or indirectly [7] relies on our understanding of the relation between the physical dose of radiation delivered to a specific organ or structure and its biologic effects, known as the dose–response relationship. Unfortunately, the mathematical form and precise anatomic location relevant to this relation is often uncertain or unknown [8].

Conventional methods of elucidating dose–response relationships involve dose-volume analyses. These methods summarize heterogeneous 3D dose distributions into descriptive metrics for whole tissues and organs. This limits analysis, however, to an *a priori* selection of a critical organ, sub-organ, or tissue. Recent developments in radiation biology have revealed that this selection is not as straight forward as previously thought. For example, some biologic effects are associated with radiation doses deposited at distant anatomic locations [9], whereas others appear to be associated with the radiation doses in only a portion or sub-volume of an organ [10,11]. Furthermore, the relevance of such summary statistics is unclear in the age of highly conformal radiation treatments with complex dose distributions [12]. The widespread availability of such highly conformal radiation treatments also increases the importance of precisely understanding radiation sensitivity at a sub-organ level because it is now often possible to optimize

treatments to avoid specific sensitive volumes, if they are known.

Voxel-based analysis methods, also known as image-based data mining (IBDM), have gained popularity in recent years for their ability to maintain the spatial complexity of entire 3D dose distributions, their freedom from *a priori* anatomic assumptions, and their independence from structure delineations [13]. These methods comprise two primary steps: spatial normalization to a reference anatomy and voxelized dose comparison. Several studies have demonstrated the potential of IBDM to reveal details of the dose–response relation that had eluded traditional dose-volume methods [14–18]. For example, one study revealed that radiation dose in a small region across the base of the heart strongly correlates with the overall survival of patients with lung cancer [16]. Another study of radiation-induced trismus (difficulty in mouth opening) revealed a significant correlation between the severity and the dose in and near the ipsilateral masseter [19]. Correlations with the IBDM-identified volume were stronger than that in *a priori*-identified structures. All of these studies, however, focused on data acquired from and health effects relevant to adults.

Dose-response relationships for childhood cancer treatments represent a niche field with unique challenges. Childhood cancers are histologically distinct from adult cancers and are exceedingly rare, accounting for only 1% of all cancer cases [20]. Therefore, pediatric-specific study cohorts accumulate slowly. Furthermore, not only are children's young bodies inherently more vulnerable to radiation damage than adults', but they also face several decades of survivorship longer than is expected of most adult survivors. As a result, the specific side effects of interest among childhood cancer survivors often differ from those for survivors of adult cancer. For example, heart disease [21] and subsequent malignant neoplasms [22] are of particular concern to childhood cancer survivors due to their long latency of a decade or more. Both of these effects are also associated with low doses of radiation [23–25]. This differs from adult dose–response studies, which tend to focus on high-dose volumes [8].

Many of these unique challenges also complicate the applicability of IBDM to data from children. Although Palma et al. [26] recently demonstrated the utility of IBDM for adult multi-center studies, the relatively larger and age-related anatomic variability characteristic of childhood, which may result in larger uncertainties associated with the spatial normalization needed for IBDM, remained an outstanding question. A recent study investigated this spatial normalization in CT images from a small cohort of children and suggested that satisfactory normalization is feasible [27].

Many childhood cancers occur in the brain, where MRI is often used because of its superior soft-tissue contrast. However, CT-based analyses remain useful thanks to their direct relevance to radiotherapy treatment planning data. In addition, because large pediatric cohorts are rare and require accrual over many years, the heterogeneity of MR protocols and image quality across time or between centers in multi-institution studies, as are needed to increase small pediatric cohort sizes, present other challenges to the IBDM workflow. Hence, a logical next step will be to develop IBDM using MRI to improve soft-tissue spatial normalization in the brain [28,29] but the feasibility of the full IBDM pipeline in data from children must first be established, and CT offers valuable information to this effect. For example, we must first determine whether residual uncertainties (due to large anatomic variations) from the normalization process affect the ability of IBDM to identify a small "sensitive" sub-volume in a large population of children. The specific consequences of the characteristics of the dose distribution (e.g., dose gradient) are also unknown. Quantifying those uncertainties is important because they will inform sample-size and power calculations in future studies of the late effects of radiotherapy in children.

Here, we present a comprehensive end-to-end test of IBDM in CT data obtained from a large population of children for a simulated effect. We tested the performance of a published binary IBDM approach and quantified the role of three key steps in IBDM: selection of reference anatomy, sensitivity to the dose-distribution characteristics (using both simplified and modified clinical dose distributions), and robustness to the incidence rate (i.e., relative frequency) of the effect of interest in the population by using incidence rates representative of side effects associated with cranial irradiation in children.

2. Methods

We begin this section by defining our patient population and hypotheses (Section 2.1) before, for the reader's convenience, providing a brief description of the published IBDM method that we implemented in this study (Section 2.2). The sections that follow (2.3 to 2.6) describe our comprehensive end-to-end test in detail.

2.1. Population

We tested the applicability of IBDM to pediatric studies by using data from a cohort of 167 children. These children previously received proton radiotherapy for primary craniopharyngioma brain tumors at St. Jude Children's Research Hospital. Craniopharyngioma is a centrally located tumor arising in the suprasellar region of the intracranial compartment. We selected this cohort for its wide age range (10 months to 20 years), even sex distribution, and well-defined treatment protocol (Table 1). The patient data used in this study included their treatment-planning CT images, structure sets approved by a board-certified radiation oncologist, and clinically delivered radiotherapy dose distributions exported from a commercial treatment-planning system (Eclipse v13.7, Varian Medical Systems, Palo Alto, CA). All patients received focal proton therapy comprising two (lateral opposed) or three (lateral opposed plus apex) beams. All treatment-planning CT images were acquired in the head-first supine orientation with pixels ranging in size from 0.78 to 1.17 mm. All but one image (1.5 mm thickness) had a slice thickness of 1 mm. Structure sets for all patients included delineations of the target volumes and five organs at risk: brainstem, left cochlea, right cochlea, left optic nerve and right optic nerve. All structures were contoured according to the RT2CR (NCT01419067) or RT3CR (NCT02792582) clinical trial protocols. This study was approved by the St. Jude Institutional Review Board.

We used these data to test the hypothesis that it is feasible to use IBDM with data obtained from children to identify a sensitive sub-volume in which a systematic difference in dose deposition is associated with a binary effect of interest.

2.2. IBDM methodology

We performed the two primary steps of IBDM analyses by using the framework and software developed at the University of Manchester [14,16,19,30,31].

2.2.1. Spatial normalization

Spatial normalization in IBDM aims to map individual patient anatomy and the related treatment information (e.g., dose distributions) to a common reference anatomy. IBDM generally performs this normalization through a two-step process: rigid, or affine image registration to grossly align each patient to the reference anatomy, followed

Table 1
Demographic and treatment details of children included in this study.

Characteristic	Value
No. patients	167
No. males (%)	83 (49.7)
Mean age $\pm$ 1σ [y]	9.6 $\pm$ 4.6
Mean target volume $\pm$ 1σ [mm^3]	26.3 $\pm$ 19.1
Prescribed dose [Gy$_{RBE}$]	54
Prescribed no. fractions	30
No. who received chemotherapy (%)	0 (0)
No. who received surgery (%)	167 (100)

by deformable image registration (DIR) to fine-tune the normalization. We implemented three iterations of the rigid registration, allowing scaling only, then translation only, and then translation and rotation. This preliminary alignment only considers the anatomy within a region of interest that is automatically generated according to clinically delineated structures. We defined our region of interest as the bounding box encompassing the clinically delineated brain contour expanded by a 5 mm margin. Next, the IBDM DIR uses a B-spline DIR algorithm in the Nifty Registration package (NiftyReg, KCL, UK) [32] with the default registration parameters and regularization via a bending-energy penalty term. This step considers the entire patient image (i.e., anatomy both within and outside the region of interest).

The deformation matrix that results from the anatomic spatial normalization is applied to the associated radiotherapy treatment data (e.g., dose distributions and structure delineations) to obtain versions of each individual's radiotherapy treatment data propagated to the reference anatomy. IBDM aims to take the uncertainties associated with the spatial normalization into account during the voxelized dose comparison by applying a 3D Gaussian blurring kernel to each deformed dose distribution prior to performing the dose comparison. IBDM defines this blurring kernel by using the standard deviation (SD) in the center-of-mass locations of structures that are routinely delineated and within the registration region of interest. Specifically, it averages the SD in each of the superior-inferior, anteroposterior, and left–right directions across all delineated structures. This SD is calculated relative to the center of mass of each corresponding structure in the reference anatomy. In our study, the clinically delineated structures within the registration region of interest that contributed to this 3D Gaussian blurring kernel were the brainstem, left and right cochlea, and left and right optic nerves.

2.2.2. Voxelized dose comparison

In binary voxel-wise dose comparison, patients are grouped according to a binary effect measure, namely those whose data report an effect of interest in one group (E for "effect") and those without the effect in the other group (NE for "no effect"). The Manchester Tool Kit (MTK) for binary analysis [16,19] calculates a Student *t*-test statistic in each voxel of the reference anatomy, resulting in a 3D distribution of *t*-test statistic magnitudes, also referred to as a *t*-map. The MTK then performs a permutation test to identify the *t*-values associated with random discrepancies between the radiation dose distributions of the two groups of patients. The permutation test is a nonparametric test [33] that conservatively corrects for multiple comparisons to protect against erroneously significant results, which are likely given the large sample size [30,34–36]. Permutation testing randomizes the binary effect labels among the patients in the cohort several times (1,000 times in our study). For each permutation, MTK calculates a new *t*-map and records the most extreme positive and negative *t*-test statistics [14,16], considering all voxels within the reference anatomy skin contour (i.e., including anatomy within and outside the registration region of interest). This procedure tests the null hypothesis that there is no difference between the average dose distributions of the two groups ($H_0 : \overline{D_E} = \overline{D_{NE}}; H_1 : \overline{D_E} \neq \overline{D_{NE}}$). Ranking the recorded extreme *t*-statistics for all permutations enables one to assess the significance of the observed *t*-statistic (i.e., that obtained from the true effect labels, before random permutation). We determined statistical significance at the $\alpha = 0.01$ significance level, which is common in permutation tests to control for the family-wise error rate [36]. Practically, this entails comparing the 99th percentile in the ranked extreme *t*-statistics with the observed *t*-statistic. Additionally, the MTK indicates volumes in which the dose relates to the investigated effect by using the ranked extreme *t*-statistics for various p-values. The volumes are found by segmenting iso-*t* levels in the observed *t*-map. The MTK functionality is described in detail elsewhere [30].

2.3. Assessing the impact of the reference anatomy

We selected the reference anatomy from among the treatment planning CTs of the children in the cohort. To assess the sensitivity of the IBDM analysis to the selection of the reference anatomy, we considered three selection methods: the most representative among the ages, brain volumes, and target volumes. We determined representativeness as the median value because it is less susceptible to outliers than is the mean. Specifically, we selected images from the female and male patients closest to the median age, the patient with the median brain volume, and the patient with the median target volume.

2.4. Evaluation of spatial normalization

We visually assessed the performance of the DIR by calculating the mean CT of all registered images for each reference image. Additionally, we evaluated the performance of the DIR via surface-distance and center-of-mass analyses for all delineated structures. For surface-distance analysis, we calculated two distance metrics: the 95% Hausdorff ($DT_{H,95}$) [37] and average (DT_{avg}) distances as,

$$DT_{H,95} = \max\{DT_{H,95:\ t,\ s\to t}, DT_{H,95:\ s\to t,\ t}\} \tag{1}$$

$$DT_{avg} = avg\{DT_{avg:\ t,\ s\to t}, DT_{avg:\ s\to t,\ t}\} \tag{2}$$

respectively, where t represents the reference-anatomy contour and $s\to t$ represents the individual subject's contour propagated to the reference anatomy [27]. For center-of-mass analysis, we calculated the Euclidian norm of the center-of-mass displacement (ΔCOM), or.

$$\Delta COM = \|COM_{s\to t} - COM_t\| \tag{3}$$

where t and $s\to t$ are defined as above. We calculated the ΔCOM value for each patient and then averaged it for all patients (ΔCOM_{avg}).

2.5. Simplified dose distributions

To assess the strengths and limitations of IBDM analyses of the pediatric data, we used simple simulated 3D radiation dose distributions inspired by the methods of Chen et al. [14]. By using simulated dose distributions, we could choose the location and size of the sensitive volume of interest and maintained full control of the dosimetric characteristics within and outside the volume of interest. This enabled us to systematically test and quantify the sensitivity of IBDM to the various dose distribution characteristics it may encounter in clinical assessments.

We generated the simplified simulated dose distributions for each individual patient's anatomy prior to spatial normalization with an in-house code written in MATLAB (vR2019a, MathWorks, Natick, MA) and tools included in the Computational Environment for Radiotherapy Research (version 5.2) [38]. We defined these dose distributions in relative dose units and designed them to simulate analyses of out-of-field dose, as is highly relevant to pediatric late effects, such as second cancers. The test and control dose distributions comprised a homogeneous background dose of 100% assigned everywhere within each patient's skin contour. The test dose distributions additionally contained a cubic test volume of side length *S* centered in each patient's brain contour. The dose in this test volume was higher than the background dose level by an amount ΔD. This test volume mimics a sensitive sub-volume in which the dose is associated with an effect of interest. Out-of-field dose is not explicitly calculated by commercial treatment planning systems and so it is often estimated for research purposes using Monte Carlo methods [39]. However, these simulations suffer from statistical noise, especially outside of the therapeutically irradiated volume. Therefore, each test and control dose distribution included a unique additive random noise component (Fig. 1). We determined the random noise on a voxel-by-voxel basis by randomly sampling from a Gaussian distribution ($\mu = 0$;
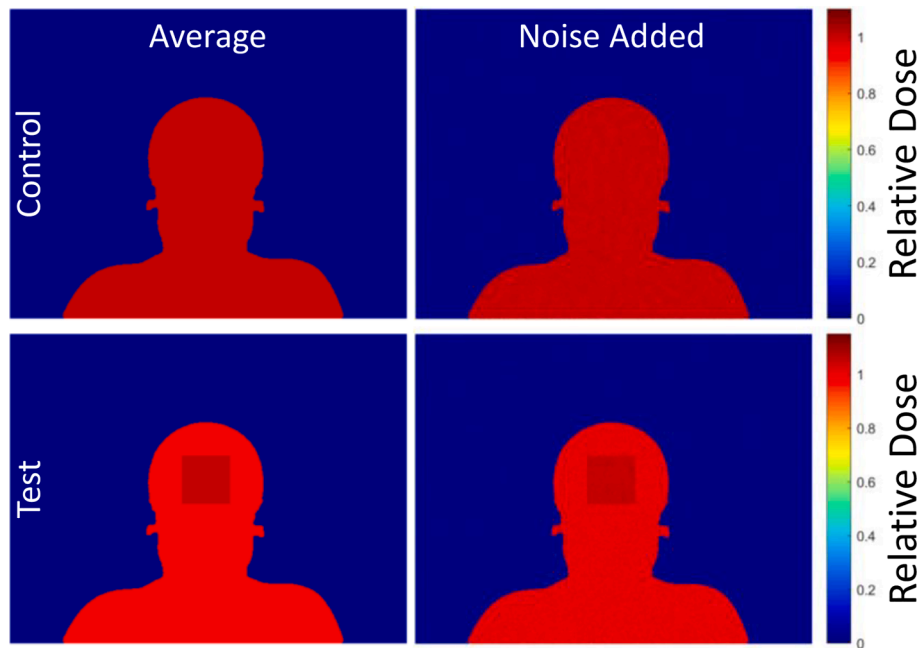
Fig. 1. Simple Simulated Dose Distributions. Representative examples of the control (top row) and test (bottom row) simple simulated dose distributions before (left column) and after (right column) random noise addition. Examples are shown for the 5% random noise level. Test examples are for a test volume with a side length of 6 cm and a dose discrepancy of 5% above the background dose level.

$\sigma = \sigma_{noise}$).

To assess the limits of IBDM functionality for pediatric analyses, we varied these simple simulated dose distributions based on three distinct parameters. Specifically, we generated test dose distributions for all combinations of three noise levels (σ_{noise}), three test-volume side lengths (S), and three dose deviations between the test volume and background dose level (ΔD). This produced 27 test conditions. We additionally generated control dose distributions for each of the three σ_{noise} levels, resulting in a total of 30 simulated dose distributions for each patient.

The three noise levels were defined by 1.96 σ_{noise} = 2%, 5%, and 7%

(Fig. 2). We selected these levels to span the noise and uncertainty levels representative of out-of-field dose calculations from Monte Carlo simulations [40] to analytical algorithms [41,42]. The three test-volume side lengths were $S = 0.5$ cm, 2.0 cm, and 6.0 cm (Fig. 2). These side lengths were selected to span potential critical-structure volumes on the order of the cochlea to the individual lobes of the brain. The three dose deviations were $\Delta D = 0.1\%$, 1%, and 5% (Fig. 2). We selected these relative magnitudes to span from stray-dose magnitudes [41] to clinically acceptable therapeutic dose deviations [43].
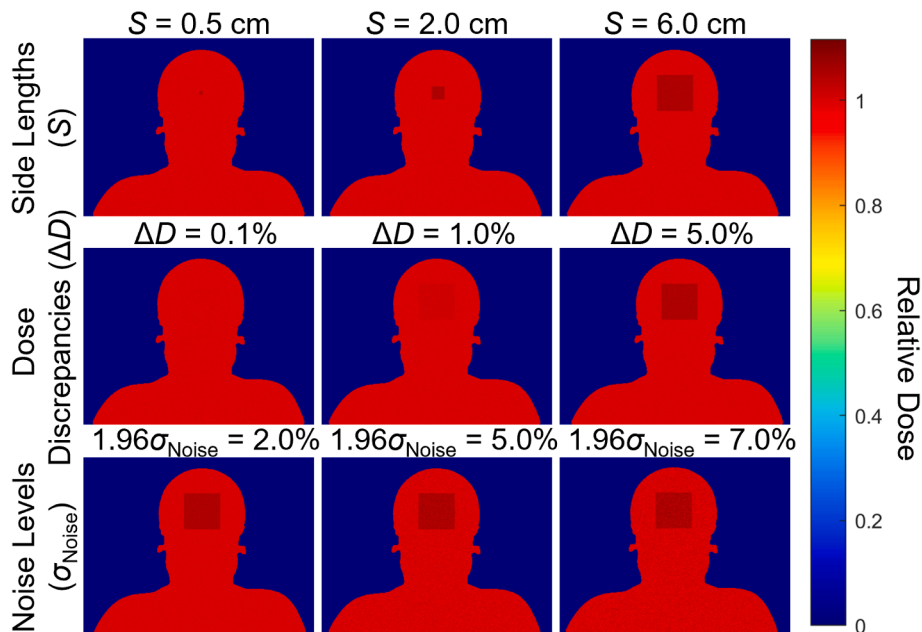

Fig. 2. Simple Simulated Dose Distribution Parameters. Representative examples of the parameters that we varied in generating the simple simulated dose distributions. Examples for the various side lengths (S) are shown with a dose discrepancy (ΔD) of 5% and noise level of 1.96 σ_{noise} = 2%. Examples for varying ΔD are shown for $S = 6$ cm and 1.96 σ_{noise} = 2%. Examples for varying σ_{noise} are shown for $S = 6$ cm and $\Delta D = 5\%$.

2.5.1. Simulated effects

Binary IBDM analysis requires that patient data be associated with a binary effect label. For this computational test, we assigned each patient's data a label of Effect (E) or No Effect (NE). Patient data assigned the E label included a test dose distribution (with a higher-dose cube of a certain size, S), whereas data assigned the NE label included the control dose distribution (without cube). We only considered E–NE groupings sharing the same level of noise, i.e., both E and NE dose distributions were generated with the same σ_{noise} such that the only systematic difference between dose distributions in the E and NE groups was the presence (or absence) of the test volume.

We assessed the sensitivity of the pediatric IBDM analyses to two characteristics of the binary patient labeling: the relative frequency of the simulated effect (i.e., the relative proportion of E labels among the cohort) and the individual patients who experienced the simulated effect (i.e., the specific patient data assigned the E label). We considered three relative frequencies of E, representing three different incidence rates of the simulated effect of interest: 50%, 20%, and 10%. We considered the relative frequency of 50% because this provides the highest level of statistical power. We selected the 20% and 10% frequencies to approximate the rates of subsequent neoplasms [44] and vasculopathies [45,46], respectively, which are two effects of interest among survivors of childhood brain cancer who received proton therapy.

We randomly assigned the E or NE labels to the patient data by using the random number generator in MATLAB vR2019a. To generate the 50% frequency labeling, we randomly assigned the E or NE label to the patient data with equal probability, resulting in 84/167 assigned the NE label and 83/167 assigned the E label. Similarly, to generate the 10% grouping, we assigned the E label to 10% of the cohort at random, and the remainder received the NE label. To define the labels for the 20% relative frequency and enable a test for sensitivity to individual patient status (i.e., E or NE labeling), we followed methods similar to those of k-fold cross validation [47]. First, we randomly partitioned the patient data into five subsections. We then generated five unique binary label combinations by assigning patient data in one subsection the E label and assigning the remaining patients the NE label. In this way, the patient data in each of the five subsections received the E label in one label combination. For each grouping strategy, we ensured that age and sex did not significantly differ by performing a Student t-test for age and a Pearson χ^2 test for sex.

2.5.2. Evaluation of dose comparison

The goal of the IBDM dose comparison was to identify the presence and location of the dose discrepancy between test and control dose distributions that is related to a simulated effect (i.e., identify and locate the test volume representing a sensitive sub-volume). To assess the sensitivity to the reference anatomy selection, we performed IBDM analyses for all four reference anatomies with the effect labels characterized by a 50% relative frequency of E and NE labels. To test the sensitivity to the relative frequency of simulated effects, we performed additional IBDM analyses with the best-performing reference anatomy for the 10% and 20% relative frequency label configurations. To test the sensitivity to the individual patient status (i.e., E or NE labeling), we performed IBDM analyses with the best-performing reference anatomy for all five folds of the 20% relative frequency label configuration. Finally, to test the sensitivity to dosimetric characteristics, we repeated all above-described tests for all 27 test configurations (Fig. 3). This resulted in a total of 270 IBDM analyses.

We quantified the performance of the IBDM dose comparison in each of these 270 analyses by calculating the Dice similarity coefficient (DSC) [48] of the reference patient's test volume (V_T) compared to the 99th percentile iso–t-map volume (V_{99th}), calculated as.

$$DSC = \frac{2 \times (V_T \cap V_{99th})}{V_T + V_{99th}} \tag{4}$$

2.5.3. Statistical sensitivity tests

We tested whether the results of the IBDM dose comparisons are associated with several components of the analysis method via multifactor ANOVA models with post-hoc Bonferroni (Dunn) t-tests. We fitted three separate multifactor ANOVA models (Fig. 3). The first model included all analyses performed with the 50% relative frequency and assessed whether the results differed based on the reference-anatomy selection and the three defining features of the test dose distributions (i.e., S, ΔD, and σ_{noise}). The second model included all IBDM analyses performed with the 20% relative frequency and assessed whether the results differed based on adjustments in the individual patient effect status (i.e., the folds). The final ANOVA model included the IBDM analyses performed for all three relative frequencies with the best-performing reference anatomy and assessed differences in the results based on adjustments to the relative frequency (i.e., incidence rate of the simulated effect). All statistical tests were performed with SAS version
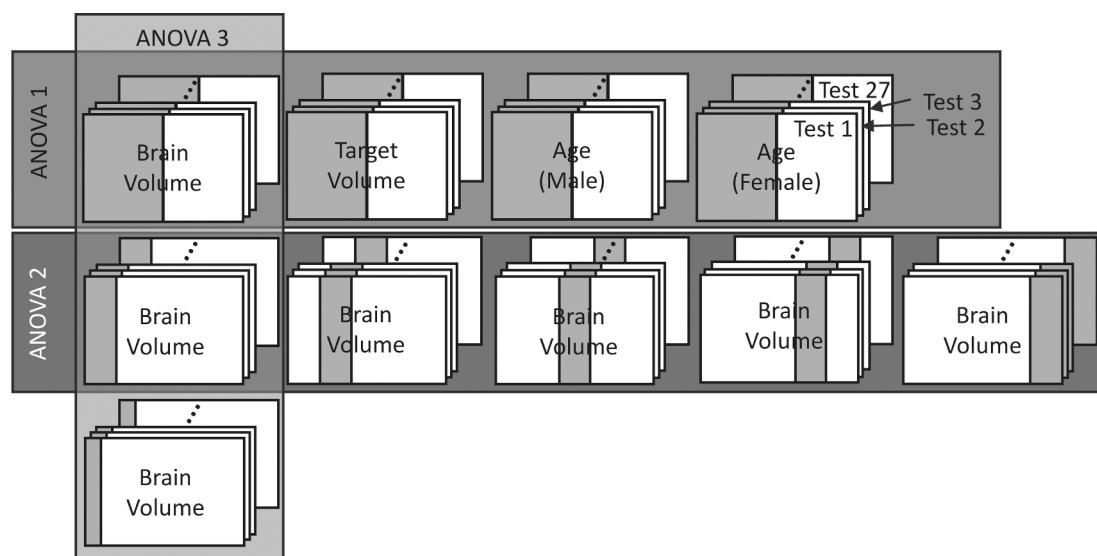

Fig. 3. Image Based Data Mining Analyses and Statistical Tests. Visual representation of the 270 analyses (each gray/white box) and three robust multifactor ANOVA tests performed. Each box represents the cohort of patients for a single test configuration (see Section 2.5) with the text referring to the method of reference-patient selection whereas the gray-shaded region represents the effect group and white the no-effect group. The width of the shaded region indicates the relative frequency of pediatric data assigned to the effect group (i.e., 50%, 20%, or 10%) and its location the k-fold grouping (see Section 2.5.1).

9.4 (SAS Institute, Cary, NC).

2.6. Modified clinical dose distributions

Although the simple simulated dose distributions enabled us to interrogate specific features of the dose distributions and their effects on IBDM performance when analyzing pediatric data, they served a primarily computational purpose by mimicking only analyses far from the therapeutic treatment field, which are of relevance to late effects like second cancer [25] and heart disease [23,24]. Other effects of interest in childhood survivorship, like cognitive deficits, appear to be related to exposures to tissues near the targeted volume [49]. In these cases, IBDM analyses must contend with spatially correlated noise (i.e., entrance doses) to identify anatomic volumes in which a threshold dose level separates the dose distributions of patients who experienced an effect of interest from those who did not. To more accurately mimic this situation, we performed a test based on the planned dose distributions from the radiotherapy treatments of the 167 children in the cohort.

Quantifying and assessing the performance of IBDM requires dose distributions that contain a systematic dose difference in a predefined anatomic volume, which correlates with the simulated effect labels (i.e., a known dose difference in a known volume from which the ability of IBDM to identify that volume can be quantified and judged), as we achieved by using the simple simulated dose distributions. To meet these requirements while maintaining clinical realism, we modified the clinical dose distributions by replacing the dose inside each patient's brainstem contour with the average magnitude of the dose distribution in that patient's brainstem (Fig. 4). We chose the brainstem as the volume of interest for this experiment because of its close proximity to the target volume, making it a particularly challenging case with which to assess the IBDM method's sensitivity to spatially correlated noise. Finally, we smoothed the resulting dose distribution with a Gaussian smoothing kernel of $\sigma = 1.0$ mm to achieve clinically realistic dose gradients (for proton therapy) around the edges of the brainstem. By these methods, we obtained dose distributions with clinically realistic therapeutic exposures and consistent dose discrepancies in a predefined volume (i.e., the brainstem).

We separated the patient data into new Effect (E) and No Effect (NE) groups for this analysis according to a threshold level among the average brainstem dose magnitudes. We tested IBDM performance with modified clinical dose distributions at the same relative frequencies of E labels as we used for the simplified dose distributions: 50%, 20%, and 10%. We assigned the effect labels by determining the three corresponding threshold dose levels: the median ($\tilde{D}$), 80th percentile ($D_{80\text{th}}$), and 90th percentile ($D_{90\text{th}}$) doses (Fig. 5). We assigned all patient data whose average brainstem doses were equal to or below the threshold to the NE group and all whose average brainstem doses were above the threshold to the E group.

This portion of the study aimed to assess the ability of IBDM to recognize and locate a dose discrepancy according to a threshold level in the presence of clinically realistic, spatially correlated noise (i.e., therapeutic and entrance doses). Therefore, we quantified the performance of IBDM by comparing the reference patient's brainstem contour to the 99th percentile iso–t-map volume via three metrics: sensitivity, or the fraction of the reference brainstem covered by the 99th percentile iso–t-map volume; positive predictive value (PPV), or the fraction of the 99th percentile iso–t-map located within the reference brainstem, and the DSC (see Equation (4)).
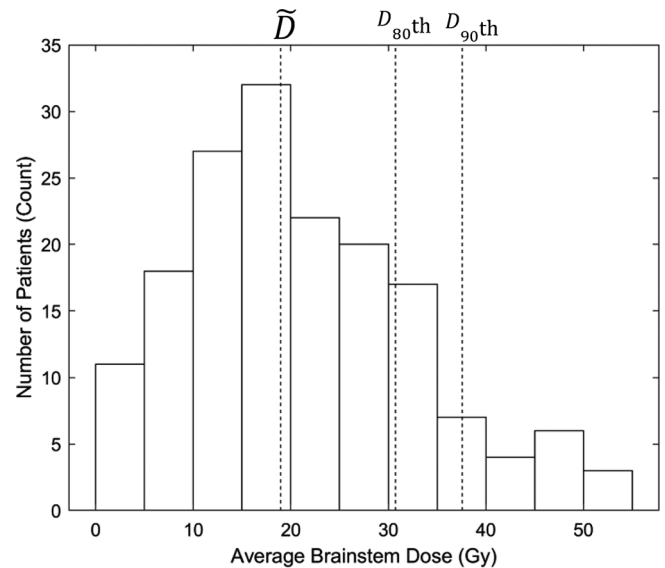

Fig. 5. Distribution of Average Brainstem Doses. Histogram showing the distribution of average doses within the brainstem contours of patients for the tests in modified clinical dose distributions. Dashed vertical lines indicate the median ($\tilde{D}$), 80th- ($D_{80\text{th}}$) and 90th-percentile ($D_{90\text{th}}$) dose levels, which served as the threshold levels for dividing the pediatric data into effect and no effect groups.
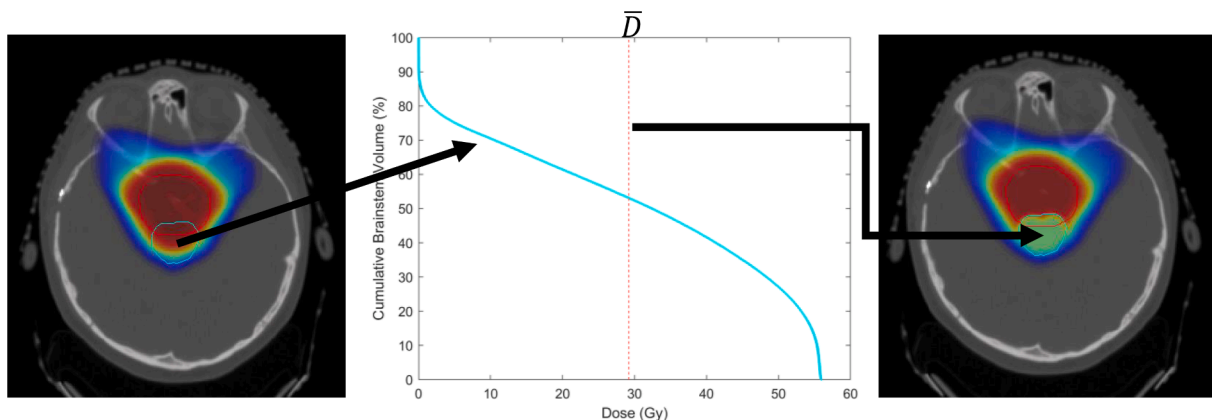

Fig. 4. Method for Modifying Clinical Dose Distributions. The left image shows a representative example of a planned dose distribution (color wash) overlaid on the corresponding treatment planning CT image (grayscale) with the brainstem (cyan) and target volume (red) contours. The center image shows the dose-volume histogram for the brainstem (cyan), with the mean brainstem dose ($\overline{D}$) indicated by the red dashed line. The right image shows the final modified clinical dose distribution, with the average dose magnitude assigned everywhere inside the brainstem contour. (color wash scale, green is 54%). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Evaluation of spatial normalization

Fig. 6 shows the original reference CT images and average treatment-planning CT images at two distinct points in the IBDM workflow: before and after DIR to the best-performing reference anatomy. Qualitatively, the improved clarity after DIR, especially in the regions shown with the soft-tissue window-level depicting the finer brain structure, indicated lower intensity variability among the mapped images and an overall high accuracy of the DIR, despite large age-related variations in anatomy. The corresponding images for all four reference patients (see Section 2.3) are supplied in Appendix A. Quantitatively, Table 2 shows the metrics computed in surface-distance and center-of-mass analyses. These metrics revealed that the discrepancies between the individual subject anatomy propagated to the reference coordinate system and the reference anatomy were small (i.e., $DT_{H,95} < 9$ mm, $DT_{avg} < 3$ mm, and $\Delta COM < 6$ mm), with a slight improvement for the reference anatomy selected on the basis of the median brain volume. Relatedly, Table 3 shows the dimensions of the Gaussian blurring kernel used to take the effects of DIR uncertainties into account in the voxelized dose comparison for all four reference anatomies. These dimensions were also small (i.e., SD < 4 mm).

3.2. Evaluation of dose comparison

The various strategies of assigning E and NE labels to patient data did not change the distribution of age and sex (p > 0.05 for all). IBDM

analyses among the 270 test cases were largely successful. Of the 270 tests, 219 (81%) identified a family-wise significant difference between the dose distributions of the E and NE groups at the $\alpha = 0.01$ significance level. All comparisons that failed the family-wise test (i.e., 51 of 270) involved either the smallest test volume (31 of 51, 60%) (i.e., the volume simulating a small sensitive sub-volume of interest; $S = 0.5$ cm) or the smallest introduced dose difference (49 of 51, 96%) (i.e., $\Delta D = 0.1\%$), or both (29 of 51, 57%). DSC analyses from voxel-wise comparisons revealed a peak in frequency at DSC magnitudes between 0.9 and 1. Fig. 7 shows representative examples of the iso–t-maps for IBDM comparisons in simple dose distributions, along with the corresponding DSC magnitudes. These simple dose distributions were characterized by a dichotomous dose distribution with a constant average dose difference in the test volume and no average dose difference elsewhere. Therefore, the iso–t-maps are dichotomized and lack a smooth transition through significance levels. Appendix B contains the corresponding images for all 270 tests.

3.3. Statistical sensitivity tests

Fig. 8 shows violin plots of DSC magnitudes from all IBDM analyses (i.e., including those in which the family-wise comparison failed to recognize a difference) that were partitioned by the various parameters considered in the sensitivity tests. Fig. 8a-d show the results of the sensitivity tests included in ANOVA 1 (shown in Fig. 3). The IBDM results, as quantified by DSC, were insensitive to the method of reference-anatomy selection (Fig. 8a). This suggests that these methods are robust to the method of reference-anatomy selection despite large variations in
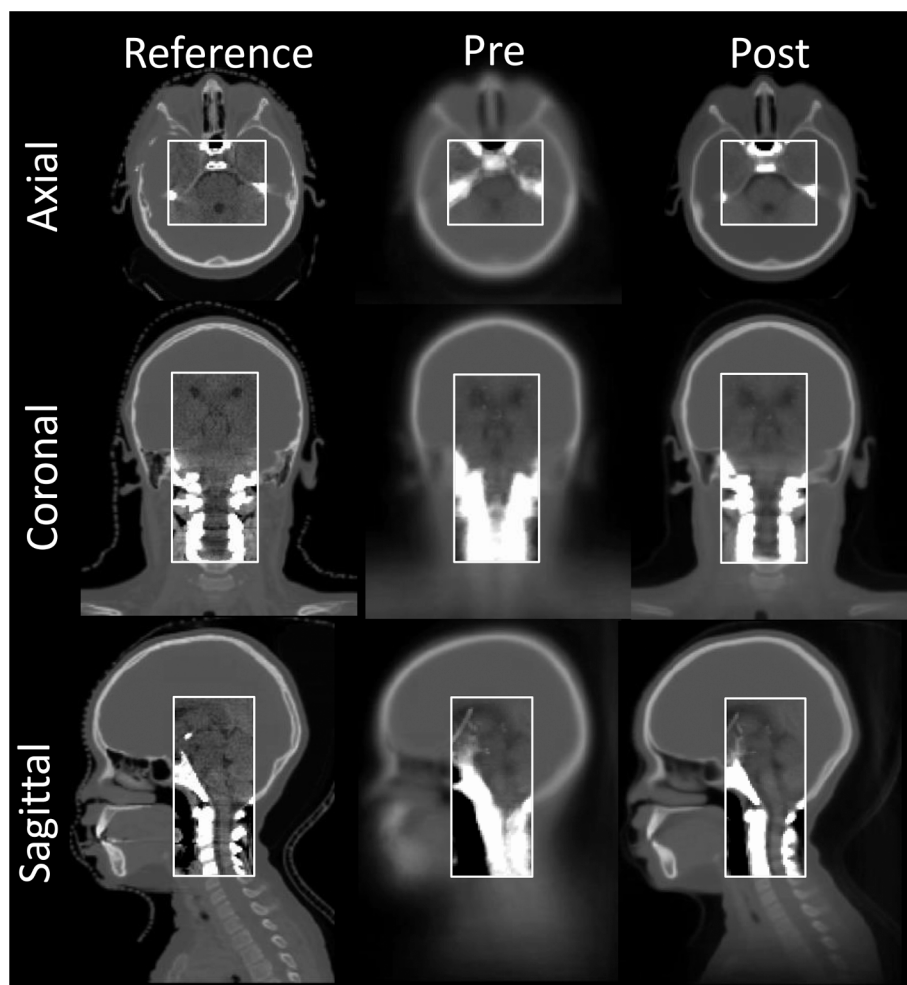

Fig. 6. Average CT Images. Treatment planning CT images shown for the reference anatomy (Reference) and those averaged over all 167 children before (Pre) and after (Post) the deformable image registration to the reference. The anatomy outside the white squares is shown in a bone window-level while the squares show a soft tissue window-level. Note that some brain structures are visible in the soft-tissue level, even with the known limitations of soft-tissue visualization of CT in the brain. This example is shown for the best-performing reference anatomy.

Table 2

Quantitative Evaluation of Deformable Image Registration. Measures of the deformable-image-registration accuracy include the 95% Hausdorff ($DT_{\text{H},95}$) and average (DT_{avg}) distances between surfaces (see Eqs. (1) and (2)) and the Euclidian norm of the center of mass displacement (ΔCOM_{avg}) (see Equation (3)). All metrics are reported as the value ± one standard deviation. Bold numbers show the lowest metric (i.e., best performance) for each row.

Volume of Interest	Metric	Reference Anatomy Selection Method			
		Age (Male)	Age (Female)	Brain Volume	Target Volume
Brainstem	$DT_{\text{H},95}$ mm	5.44 ± 2.51	6.95 ± 1.52	**4.78 ± 1.68**	6.99 ± 2.74
	DT_{avg} mm	2.15 ± 1.06	2.79 ± 0.60	**1.90 ± 0.61**	2.44 ± 0.95
	ΔCOM mm	2.60 ± 2.39	4.42 ± 1.04	**2.51 ± 1.54**	3.19 ± 1.66
Left Cochlea	$DT_{\text{H},95}$ mm	2.10 ± 3.16	2.57 ± 2.18	1.59 ± 0.69	**1.54 ± 1.02**
	DT_{avg} mm	1.05 ± 2.83	0.99 ± 1.90	**0.68 ± 0.46**	0.71 ± 0.66
	ΔCOM mm	1.44 ± 3.16	1.58 ± 2.23	**0.96 ± 0.80**	1.00 ± 1.16
Right Cochlea	$DT_{\text{H},95}$ mm	1.99 ± 3.12	1.89 ± 1.23	**1.80 ± 1.05**	1.98 ± 2.69
	DT_{avg} mm	1.02 ± 2.90	**0.72 ± 0.89**	0.76 ± 0.65	1.05 ± 2.34
	ΔCOM mm	1.26 ± 3.22	1.09 ± 1.21	**0.89 ± 0.96**	1.30 ± 2.80
Left Optic Nerve	$DT_{\text{H},95}$ mm	8.06 ± 4.84	6.50 ± 3.07	5.03 ± 3.14	**4.80 ± 2.30**
	DT_{avg} mm	2.19 ± 2.86	2.17 ± 2.28	**1.57 ± 1.36**	1.72 ± 1.50
	ΔCOM mm	3.58 ± 3.57	5.52 ± 3.31	**2.97 ± 2.06**	2.70 ± 1.89
Right Optic Nerve	$DT_{\text{H},95}$ mm	8.12 ± 4.84	5.63 ± 2.61	5.22 ± 3.58	**4.86 ± 2.35**
	DT_{avg} mm	2.30 ± 2.53	1.83 ± 1.76	**1.55 ± 1.30**	1.80 ± 1.51
	ΔCOM mm	3.53 ± 3.29	4.60 ± 3.00	2.99 ± 2.22	**2.87 ± 2.07**

Table 3

Dimensions of the Gaussian Blurring Kernel. Dimensions used to take the effects of uncertainties associated with mapping individual patient anatomies to a reference anatomy into account for each of the four reference anatomies considered. Bold values represent the smallest among the four reference patients.

Reference Anatomies	Anteroposterior (mm)	Left-Right (mm)	Superior-Inferior (mm)
Median Age (F)	2.67	2.63	2.11
Median Age (M)	3.25	1.68	2.02
Median Brain Volume	1.85	**1.47**	**1.34**
Median CTV	**1.66**	1.93	1.90
Abbreviations: F – female; M – male; CTV – clinical target volume			

age-related anatomy within the cohort. The DSC magnitudes were also sensitive to the test-volume side length (S) (Fig. 8b), insensitive to the magnitude of the dose discrepancy for ΔD greater than or equal to 1% (Fig. 8c) and significantly improved for the lowest noise level of 2% (Fig. 8d). Fig. 8e shows the results for ANOVA 2 (shown in Fig. 3). This plot reveals that the IBDM results were insensitive to the individual patient labels. This suggests that the IBDM method reliably assesses dosimetric discrepancies and is robust to inter-individual anatomic variations in children. Finally, Fig. 8f shows the results for ANOVA 3 (shown in Fig. 3). This plot indicates that the IBDM results were significantly worse with the 10% relative frequency of the E labels but were there was no difference between 20% and 50% relative frequencies. This suggests that a lower bound exists for the incidence rate of effects of interest that are suitable for this type of analysis, i.e., 36

events for 20%.

3.4. Modified clinical dose distributions

Table 4 lists the calculated metrics comparing the 99th percentile iso–t-volume to the reference brainstem from all three tests in modified clinical dose distributions. These values show that IBDM successfully identified systematic differences between the dose distributions of the E and NE groups with a family-wise $p \leq 0.001$ in all cases. The calculated sensitivities reveal that IBDM consistently recognized the dose differences present in the brainstem, even in the test with the lowest power (i. e., relative frequency of 10%). The PPV and DSC, however, were relatively lower, only reaching a maximum of 0.25 and 0.40, respectively. Fig. 9 shows the iso–t-maps for the three tests in modified clinical dose distributions. These images show that the 99% iso–t-volume covered the brainstem (i.e., the volume in which we introduced a systematic dose difference between the E and NE groups) despite the presence of clinically realistic, spatially correlated noise from the therapeutic dose distribution. This significant volume, however, also encompassed additional voxels outside of the brainstem due to correlations of the brainstem dose with systematic differences in the target coverage and beam arrangement for the two treatment protocols included in this study. Nevertheless, the most critical point is that the IBDM method must identify the entire sensitive sub-volume and not miss part of the volume of interest because of unavoidable noise. Interestingly, the DSC goes up with lower event rate, showing the potential of IBDM to overestimate sensitive regions because of correlations in the data.

4. Discussion

We tested the applicability of IBDM to pediatric data by using the imaging and treatment data from a cohort of 167 children who previously received proton therapy for primary brain tumors at St. Jude Children's Research Hospital. We assessed the strengths and limitations of IBDM with simulated effects and simple and modified clinical dose distributions. Our findings revealed that IBDM, for a structure that is central in the brain, is applicable to pediatric data despite large age-related anatomic variation and the small patient numbers and incidence rates characteristic of childhood late-effects research. Our findings also show that IBDM reliably identifies systematic dose differences as small as 1%, regardless of the selection of reference anatomy. The performance, however, depends on the magnitude of random noise, the relative frequency of the effect of interest in the cohort, and the size of the sensitive sub-volume (i.e., the side length, S, of the test volume in this study). It should be noted, however, that the DSC itself is sensitive to the volume being compared with lower DSC magnitudes more likely for smaller volumes. Therefore, although our IBDM results were sensitive to S, the extent to which the apparent sensitivity is attributable to the IBDM method rather than the inherent nature of DSC is unclear.

Although our study is the first to evaluate the feasibility of the complete IBDM workflow in pediatric CT data, the spatial normalization step can be compared to previous CT-based studies implementing IBDM in adult cohorts of lung cancer patients [16,30,31], head and neck cancer patients [19] and a recent study evaluating spatial normalization in data from children [27]. Beasley et al. found a DIR uncertainty (SD of center of mass) of 4.0, 0.34 and 0.30 mm in the left–right, anteroposterior and superior-inferior directions for the muscles of mastication [19]. We found similar DIR uncertainty values for our selected brain structures, with slightly less deviation, most likely because of the rigidity of structures contained within the skull. Veiga et al. found DT_{avg} and ΔCOM_{avg} values of 0.7 ± 1.0 mm and 2.9 ± 5.7 mm, respectively, for DIR of the central nervous system in children [27]. Although our ΔCOM_{avg} was consistent with that of Veiga et al., we found a slightly higher DT_{avg} (Table 1). It should be noted, however, that we calculated DT_{avg} as an unsigned average and it is unknown whether the metric
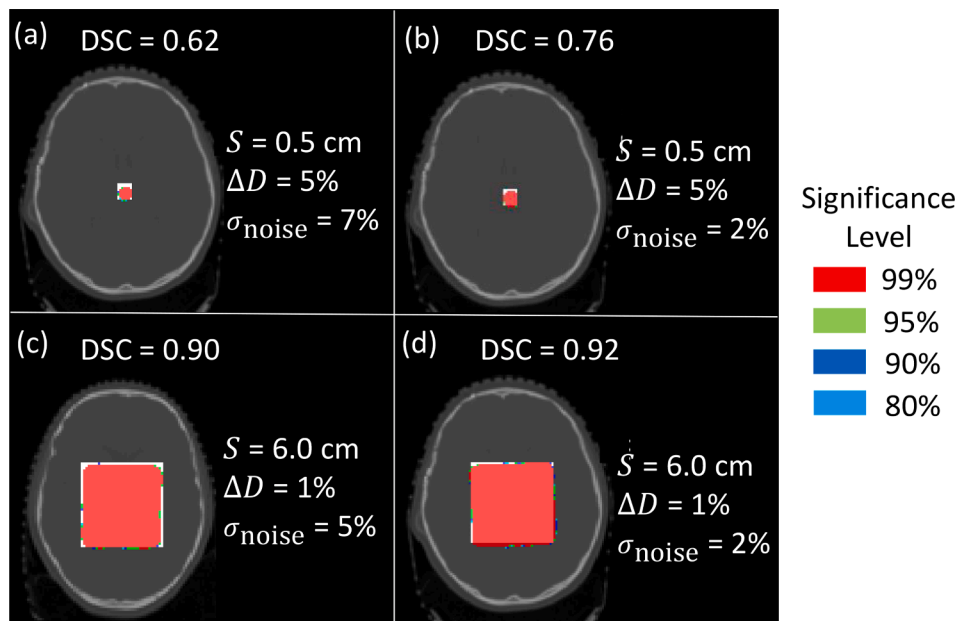
Fig. 7. Representative Iso–*t*-maps. Representative examples of the iso–*t*-maps from image-based data mining tests in the simple dose distributions, along with the corresponding Dice similarity coefficients (DSC, see Eq. (4)). Grayscale images show the reference anatomy CT with the test volume burned into the image (i.e. white squares). The color wash shows the voxels that were significant at various significance levels. (A) The iso–*t*-map for the test with the DSC closest to average among all tests. (B) The test with the median DSC magnitude. (C) An example with the most abundant DSC magnitude when binned in steps of 0.1. (D) An example with the maximum DSC among all tests. All images are shown for the slice at the center of the corresponding test volume and the white text indicates the DSC magnitude and the corresponding characteristics of the test dose distribution: side length of the test volume (S), dose discrepancy (ΔD), and noise level (σ_{noise}).

reported by Veiga et al. is signed or unsigned. Furthermore, we assessed DIR accuracy by using contours, which are widely known to depend on inter-observer variation [50]. The effect of this inter-observer variability is somewhat less severe in ΔCOM_{avg} assessment but has a strong effect on DT_{avg} and may also contribute to the differing results from Veiga et al. and our study.

Our study has several notable strengths. First, it represents the first end-to-end test of voxel-based analysis methods in a pediatric cohort. Our study also used well-characterized data from a large cohort of children with cancer. The data were collected under two successive clinical trials, thereby minimizing potential confounding factors related to imaging protocols, structure contours, and treatment variations. We also minimized confounding factors by using simulated dose distributions. This yielded complete control of the location, volume, and magnitude of dose deviations analyzed, which is a necessary step to objectively quantify the performance of the IBDM method and assess its sensitivity to various dosimetric characteristics.

Our study also had some limitations. First, our use of default DIR algorithm parameters may have limited the achievable DIR accuracy. Our metrics assessing DIR accuracy, however, were comparable to those in other IBDM studies. Future studies are needed to identify the optimal DIR parameters for pediatric IBDM analyses. Additionally, the ability of the registration algorithm to accurately match structures within the brain may be further limited by the lack of soft tissue contrast in CT imaging. This lack of contrast may also account for the small number of outliers (4/167 patients) that we observed with large registration discrepancies in which no obvious reason for registration failure was apparent. We precisely assessed the DIR accuracy, however, by propagating the spatial normalization to structure contours that spanned the brain volume, thereby minimizing the likelihood of overlooking gross deformation errors and still found a reasonable registration accuracy (< 6 mm in the best-performing reference), despite the low soft-tissue contrast of CT. Furthermore, if a few cases within a large cohort lead to registration failures, those cases can be discarded without affecting the statistical power of the analysis. Incorporating MR or other multi-modal imaging techniques would increase the information available to the registration algorithm and would enable a more comprehensive assessment of the registered brain structures, and likely a higher registration accuracy, as has been suggested by two methodological studies [28,29] using smaller cohorts (i.e., 50 adults and 6 children, respectively). However, the additional required image registration step (i.e.,

MR to CT) would also add additional uncertainty. Despite this limitation of CT-based IBDM, our results confirm the feasibility of applying IBDM in data from children, for the first time, and provide a reference point for future studies evaluating IBDM using MR images. We are currently generalizing the methods reported here for MR-based spatial normalization to perform further IBDM studies in children with brain tumors. Another limitation of this study is its computational nature, which only addresses the feasibility of IBDM in pediatric cohorts without providing new clinical results. Thorough benchmarking, however, is required to understand future clinical results and allow them to guide clinical practice. Having achieved this proof of principle, future studies will be able to apply these IBDM methods to analyze the dose response for some of the most common side effects of radiotherapy treatments among childhood cancer survivors with various central nervous system disorders.

Because childhood cancer is rare, large cohorts such as those used in previous adult IBDM studies may not be achievable. Our study shows that the smaller cohort size typical of pediatric cancers (i.e., often at best only a few hundred patients) provides a large enough sample size to successfully implement the IBDM methods for incidence rates representative of common effects of interest among childhood cancer survivors (i.e., 20% and 10%). Our findings suggest that cohorts of similar size and power are feasible for successful implementation of IBDM and pave the way for future studies to establish this. The specific sample size required will depend on several factors, such as the relative incidence rate of the effect of interest among the cohort, the success of the spatial normalization and the dynamic range represented in the dose distributions.

Our findings in modified clinical dose distributions highlight a challenge common to all data-mining endeavors in that the results rarely reveal the desired information cleanly. Instead, data mining results require expert interpretation to realize the valuable information. Our data were derived from children who were treated in one of two clinical protocols, with approximately half of the children in each protocol. The children treated in one protocol tended to have slightly higher target doses delivered by three treatment beams – two lateral beams and one apex beam. In contrast, the children treated in the other protocol tended to have relatively lower target doses delivered by two lateral treatment beams alone. The higher target doses and apex beam were also associated with higher brainstem doses; therefore, the grouping for a 50% relative frequency of E labels generally divided the cohort of children
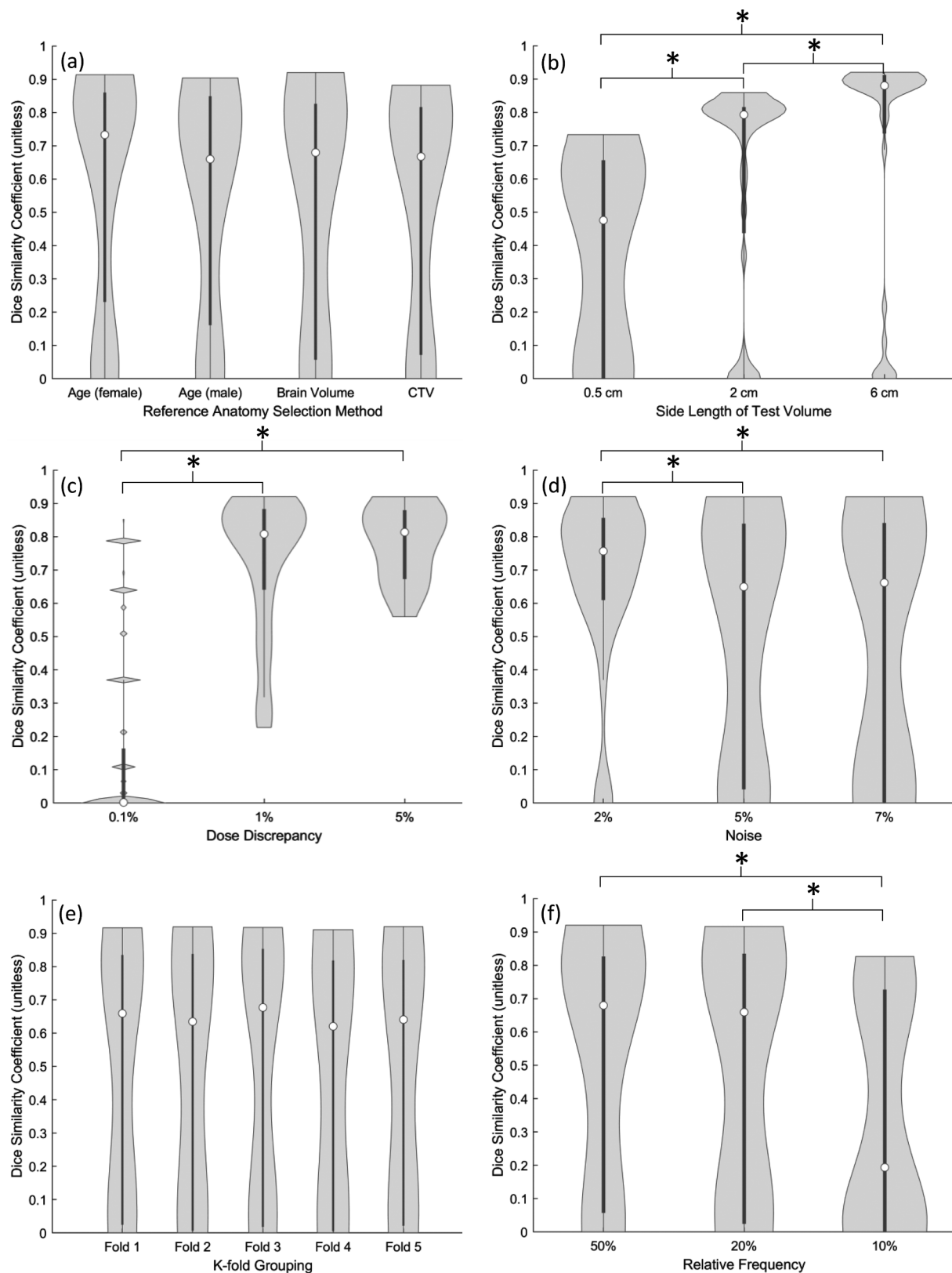
Fig. 8. Sensitivity Test Results. Violin plots of the calculated Dice similarity coefficients comparing the *t*-map volume significant at the 99% confidence level to the test volume. Data include all analyses performed with simple dose distributions. A-D) Data for comparisons with the 50% relative frequency grouping partitioned by (A) method of reference-anatomy selection, (B) test-volume side length, (C) dose discrepancy between test volume and background dose level, and (D) random noise level. E) Data for comparisons with the 20% relative frequency separated by the patient labeling combination. F) Data for the test for sensitivity to the relative frequency of the patients assigned to the Effect group. * Post-hoc Dunn tests significant at the α = 0.05 significance level. The width of each violin plot represents the probability density of each Dice similarity coefficient magnitude. White circles indicate the median of each Dice distribution; thick vertical bars show the inter-quartile range; and thin bars represent 1.5 × the interquartile range.

Table 4

Quantitative Evaluation in Modified Clinical Dose Distributions. Table of metrics quantifying the performance of the image-based data mining analysis in modified clinical dose distributions. Metrics include the family-wise *p*-value, sensitivity, positive predictive value (PPV), and dice similarity coefficient (DSC, see Eq. (4)) comparing the 99th percentile *iso-t*-volume to the brainstem of the reference anatomy. Higher magnitudes of sensitivity, PPV, and DSC indicate improved performance with 1.00 being the maximum attainable magnitude for all three metrics.

Relative Frequency (%)	p	Sensitivity	PPV	DSC
50	<0.001	1.00	0.04	0.07
20	<0.001	1.00	0.10	0.18
10	<0.001	0.98	0.25	0.40

according to their treatment protocol. This understanding helps explain the lower PPV we observed in the 50% relative frequency test (Table 4). Although we expected significant results outside of the brainstem in the modified clinical dose distribution experiment, which involved the spatially correlated therapeutic doses, the spatial distribution of these significant results for the 50% relative frequency test (Fig. 9) clearly demonstrated that the additional significant voxels outside of the brainstem were largely located in the target volume and along the beam

paths (i.e., where the protocol-related systematic dose differences existed). The 20% and 10% relative-frequency groupings, however, more evenly mixed the data obtained from the children in the two protocols in the NE group. Therefore, these analyses were less susceptible to the protocol-related covariates. This highlights the critical importance of having a deep understanding of the dataset under consideration. Relatedly, Fig. 9 also reveals a trend of improving DSC with increasing significance level. Therefore, another potential avenue of excluding erroneously identified volumes is by assessing significance at higher levels. A practical limitation, however, is that evaluating higher significance requires a large number of permutations, and therefore prohibitively larger computation time and data-storage requirements. Future work could explore leveraging methods to extrapolate to smaller p-values without the need for additional permutations [51].

A further consideration is that the target volume for all children in this cohort was located in the same, well-defined region of the brain. Hence, the high-dose volume occurred in the same location for all patients after spatial normalization. This contrasts with other tumor sites in which IBDM has been previously applied, where tumor location can be highly varied within the anatomy of interest (e.g., lung) and spatial correlations between the dose distribution and heterogeneous anatomic locations may impact IBDM analyses [26,52]. Further work to assess the
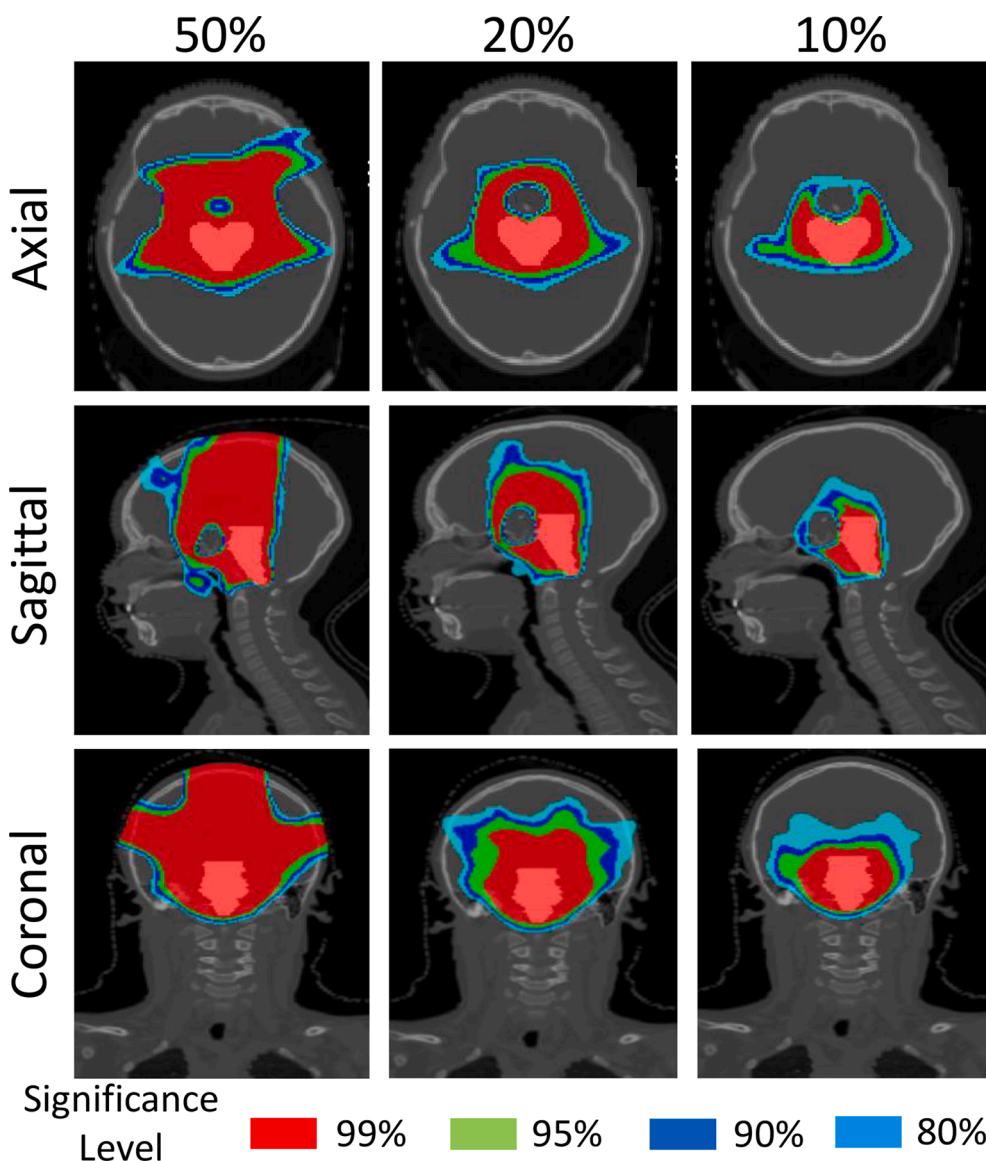

Fig. 9. Iso–*t*-maps from Modified Clinical Dose Distributions. Iso–*t*-maps resulting from image-based data mining comparisons in modified clinical dose distributions at three relative frequencies of the simulated effect of interest: 50%, 20%, 10%. The grayscale image shows the reference anatomy with the brainstem (i.e. the volume to which we introduced a systematic dose difference) burned into the image (shown in white). The color wash indicates the significance level of the iso–*t*-map. All images show the slice in the geometric center of the reference brainstem.

effect of tumor location consistency on IBDM success is required.

This raises an important consideration in choosing a dataset that is well suited to IBDM analyses: possible covariates and confounders. This consideration will guide the decision of whether the effects of confounders on the results can be interpreted and taken into account manually (as was done here) or explicitly considered with methods like per-voxel cox regression [30] or whether a different dataset containing confounders that are only weakly associated with the effect of interest should be sought. Other potential considerations in determining the suitability of a cohort include the presence and effects of healthy tissue deformations (e.g., from tumor growth or surgery) on DIR accuracy.

Finally, it should also be noted that IBDM results indicate correlation and should not be confused with causation. The resulting iso–*t*-maps provide an assessment of differences between dose distributions that are correlated with differences in effect status. Whether the indicated correlation occurs as a direct association between the dose and the effect of interest or is due to a confounding factor must be interpreted carefully. Even if the results suggest that the correlation is a direct association between the dose and the effect of interest, further studies other than IBDM are required to establish a causal relation between the absorbed dose and the biologic effect.

5. Conclusion

This study is the first to comprehensively test the applicability, strengths, and limitations of voxel-based analysis methods, i.e., IBDM, with pediatric data. Our findings demonstrated that such voxel-based analysis methods apply to data from pediatric cohorts, despite large age-related anatomic variations. The results of this study can guide future studies leveraging IBDM to improve understanding of radiation dose–response relationships in children by guiding their selection of an ideal cohort. Looking forward, such studies have the potential to improve radiotherapy by informing treatment planning to reduce treatment-related toxicities and improve quality of life among the survivors of childhood cancer.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data statement

The data that support the findings of this study are available from the corresponding author, LJW, upon reasonable request

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2022.05.003.

References

[1] Howlader N, Noone A, Krapcho M, Miller D, Brest A, Yu M, et al. SEER Cancer Statistics Review, 1975–2016. Bethesda, MD: National Cancer Institute; 2018.

[2] Hudson MS. Jude LIFE Study Population. Department of Epidemiology: St. Jude Children's Research Hospital; 2019.

[3] Suh E, Stratton KL, Leisenring WM, Nathan PC, Ford JS, Freyer DR, et al. Late mortality and chronic health conditions in long-term survivors of early-adolescent and young adult cancers: a retrospective cohort analysis from the Childhood Cancer Survivor Study. Lancet Oncol 2020;21:421–35.

[4] Wilson LJ, Newhauser WD. Generalized approach for radiotherapy treatment planning by optimizing projected health outcome: preliminary results for prostate radiotherapy patients. Phys Med Biol 2021;66.

[5] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013;107:267–73.

[6] Rechner LA, Eley JG, Howell RM, Zhang R, Mirkovic D, Newhauser WD. Risk-optimized proton therapy to minimize radiogenic second cancers. Phys Med Biol 2015;60:3999–4013.

[7] Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system. Med Phys 2000;27:701–11.

[8] ASTRO. Quantitative analyses of normal tissue effects in the clinic (QUANTEC). New York, NY: American Society for Radiation Oncology; 2010.

[9] Siva S, MacManus MP, Martin RF, Martin OA. Abscopal effects of radiation therapy: A clinical review for the radiobiologist. Cancer Lett 2015;356:82–90.

[10] Schwartz DL, Hutcheson K, Barringer D, Tucker SL, Kies M, Holsinger FC, et al. Candidate Dosimetric Predictors of Long-Term Swallowing Dysfunction After Oropharyngeal Intensity-Modulated Radiotherapy. Int J Radiat Oncol 2010;78:1356–65.

[11] Toussaint L, Indelicato DJ, Stokkevåg CH, Lassen-Ramshad Y, Pedro C, Mikkelsen R, et al. Radiation doses to brain substructures associated with cognition in radiotherapy of pediatric brain tumors. Acta Oncol 2019;58:1457–62.

[12] Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. Brit J Radiol 2019;92:20190270.

[13] Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: A methodological cookbook. Physica Med 2020;69:192–204.

[14] Chen C, Witte M, Heemsbergen W, van Herk M. Multiple comparisons permutation test for image based data mining in radiotherapy. Radiat Oncol 2013;8:293.

[15] Acosta O, Drean G, Ospina JD, Simon A, Haigron P, Lafond C, et al. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. Phys Med Biol 2013;58:2581–95.

[16] McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. Eur J Cancer 2017;85:106–13.

[17] Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello MC, et al. A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage. Int J Radiat Oncol Biol Phys 2016;96:127–33.

[18] Witte MG, Heemsbergen WD, Bohoslavsky R, Pos FJ, Al-Mamgani A, Lebesque JV, et al. Relating Dose Outside the Prostate With Freedom From Failure in the Dutch Trial 68 Gy vs. 78 Gy. Int J Radiat Oncol. 2010;77:131-8.

[19] Beasley W, Thor M, McWilliam A, Green A, Mackay R, Slevin N, et al. Image-based Data Mining to Probe Dosimetric Correlates of Radiation-induced Trismus. Int J Radiat Oncol Biol Phys 2018;102:1330–8.

[20] Howlader N, Noone A, Krapcho M, Miller D, Bishop K, Kosary C, et al. SEER Cancer Statistics Review, 1975–2014. Bethesda, MD: National Cancer Institute; 2017.

[21] Mulrooney DA, Armstrong GT, Huang S, Ness KK, Ehrhardt MJ, Joshi VM, et al. Cardiac Outcomes in Adult Survivors of Childhood Cancer Exposed to Cardiotoxic Therapy: A Cross-sectional Study. Ann Intern Med 2016;164:93–101.

[22] Facoetti A, Barcellini A, Valvo F, Pullia M. The Role of Particle Therapy in the Risk of Radio-induced Second Tumors: A Review of the Literature. Anticancer Res 2019;39:4613–7.

[23] Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of Ischemic Heart Disease in Women after Radiotherapy for Breast Cancer. N Engl J Med 2013;368:987–98.

[24] Bates JE, Shrestha S, Liu Q, Mulrooney DA, Smith SA, Leisenring WM, et al. Low-dose radiation to cardiac substructures and late-onset cardiac disease: A report from the Childhood Cancer Survivor Study (CCSS). J Clin Oncol. 2021;39:10027.

[25] Diallo I, Haddy N, Adjadj E, Samand A, Quiniou E, Chavaudra J, 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 2009;74:876–83.

[26] Palma G, Monti S, Pacelli R, Liao Z, Deasy JO, Mohan R, et al. Radiation Pneumonitis in Thoracic Cancer Patients: Multi-Center Voxel-Based Analysis. Cancers 2021;13:3553.

[27] Veiga C, Lim P, Anaya VM, Chandy E, Ahmad R, D'Souza D, et al. Atlas construction and spatial normalisation to facilitate radiation-induced late effects research in childhood cancer. Phys Med Biol 2021;66:105005.

[28] Monti S, Paganelli C, Buizza G, Preda L, Valvo F, Baroni G, et al. A novel framework for spatial normalization of dose distributions in voxel-based analyses of brain irradiation outcomes. Phys Med 2020;69:164–9.

[29] Skaarup M, Lundemann MJ, Darkner S, Jørgensen M, Marner L, Mirkovic D, et al. A framework for voxel-based assessment of biological effect after proton radiotherapy in pediatric brain cancer patients using multi-modal imaging. Med Phys 2021;48:4110–21.

[30] Green A, Vasquez Osorio E, Aznar MC, McWilliam A, van Herk M. Image Based Data Mining Using Per-voxel Cox Regression. Front Oncol 2020;10.

[31] McWilliam A, Dootson C, Graham L, Banfill K, Abravan A, van Herk M. Dose surface maps of the heart can identify regions associated with worse survival for lung cancer patients treated with radiotherapy. Phys Imag Radiat Oncol 2020;15: 46–51.

[32] Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, et al. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed 2010;98:278–84.

[33] Camargo A, Azuaje F, Wang H, Zheng H. Permutation - based statistical tests for multiple hypotheses. Source Code Biol Med 2008;3:15.

[34] Dudoit S, Shaffer JP, Boldrick JC. Multiple Hypothesis Testing in Microarray Experiments. Statist Sci 2003;18:71–103.

[35] Belmonte M, Yurgelun-Todd D. Permutation testing made practical for functional magnetic resonance image analysis. IEEE Trans Med Imaging 2001;20:243–8.

[36] Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology 2011;48:1711–25.

[37] Kim HS, Park SB, Lo SS, Monroe JI, Sohn JW. Bidirectional local distance measure for comparing segmentations. Med Phys 2012;39:6779–90.

[38] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. Med Phys 2003;30:979–85.

[39] Newhauser WD, Schneider C, Wilson L, Shrestha S, Donahue W. A review of analytical models of stray radiation exposures from photon- and proton-beam radiotherapies. Radiat Prot Dosim 2018;180:245–51.

[40] Wilson LJ, Pirlepesov F, Moskvin V, Li Z, Guo Y, Li Y, et al. Proton therapy delivery method affects dose-averaged linear energy transfer in patients. Phys Med Biol 2021;66:074003.

[41] Wilson LJ, Newhauser WD, Schneider CW, Kamp F, Reiner M, Martins JC, et al. Method to quickly and accurately calculate absorbed dose from therapeutic and stray photon exposures throughout the entire body in individual patients. Med Phys 2020;47:2254–66.

[42] Hauri P, Hälg RA, Besserer J, Schneider U. A general model for stray dose calculation of static and intensity-modulated photon radiation. Med Phys 2016;43: 1955–68.

[43] Moyers MF, Toth TL, Sadagopan R, Chvetsov AV, Unkelbach J, Mohan R, et al. 202 - Physical Uncertainties in the Planning and Delivery of Light Ion Beam Treatments. Am Assoc Physicists Med 2020.

[44] Friedman DL, Whitton J, Leisenring W, Mertens AC, Hammond S, Stovall M, et al. Subsequent Neoplasms in 5-Year Survivors of Childhood Cancer: The Childhood Cancer Survivor Study. JNCI 2010;102:1083–95.

[45] Kralik SF, Watson GA, Shih CS, Ho CY, Finke W, Buchsbaum J. Radiation-Induced Large Vessel Cerebral Vasculopathy in Pediatric Patients With Brain Tumors Treated With Proton Radiation Therapy. Int J Radiat Oncol Biol Phys 2017;99: 817–24.

[46] Lucas J, LeVine D, Ismael Y, Hsu C-Y, Darrow K, Faught A, et al. Rare-12. Vasculopathy in pediatric craniopharyngioma patients treated with surgery and radiotherapy. Neuro-Oncol. 2020;22. iii444-iii.

[47] Anguita D, Ghio A, Ridella S, Sterpi D. K-fold cross validation error rate estimate in support vector machines. Proc Int Conf Data Mining 2009.

[48] Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology 1945;26:297–302.

[49] Acharya S, Guo Y, Patni T, Li Y, Wang C, Gargone M, et al. Association Between Brain Substructure Dose and Cognitive Outcomes in Children With Medulloblastoma Treated on SJMB03: A Step Toward Substructure-Informed Planning. J Clin Oncol 2022;40:83–95.

[50] Wright JL, Yom SS, Awan MJ, Dawes S, Fischer-Valuck B, Kudner R, et al. Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO Consensus Paper. Pract Radiat Oncol 2019;9:65–72.

[51] Knijnenburg TA, Wessels LFA, Reinders MJT, Shmulevich I. Fewer permutations, more accurate P-values. Bioinformatics 2009;25:i161–8.

[52] Cella L, Monti S, Xu T, Liuzzi R, Stanzione A, Durante M, et al. Probing thoracic dose patterns associated to pericardial effusion and mortality in patients treated with photons and protons for locally advanced non-small-cell lung cancer. Radiother Oncol 2021;160:148–58.