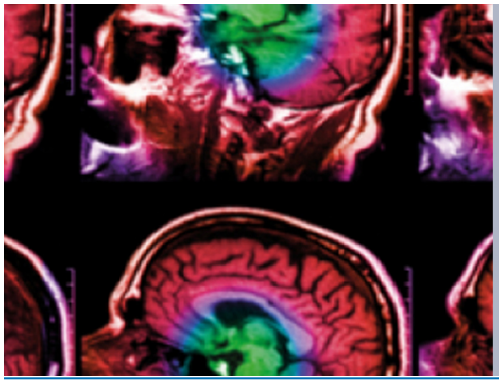
PAPER • OPEN ACCESS

Applications of a patient-specific whole-body CT-mesh hybrid computational phantom in second cancer risk prediction

To cite this article: Erika Kollitz *et al* 2022 *Phys. Med. Biol.* **67** 185011



IOP Publishing Phys. Med. Biol. 67 (2022) 185011 https://doi.org/10.1088/1361-6560/ac8851

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Applications of a patient-specific whole-body CT-mesh hybrid computational phantom in second cancer risk prediction

Erika Kollitz[1], Moritz Roew[1], Haegin Han[2], Marco Pinto[1], Florian Kamp[3,4,9], Chan Hyeong Kim[2], Marco Schwarz[5,6,9], Claus Belka[4], Wayne Newhauser[7,8], Katia Parodi[1,10] and George Dedes[1,10]

[1] Chair of Experimental Physics, Ludwig Maximilian University of Munich, Munich, Germany
[2] Department of Nuclear Engineering, Hanyang University, Seoul, Republic of Korea
[3] Department of Radiation Oncology and Cyberknife Center, University Hospital of Cologne, Cologne, Germany
[4] Clinic for Radiation Therapy and Radiation Oncology, Hospital of the Ludwig Maximilian University of Munich, Munich, Germany
[5] Azienda Provinciale per i Servizi Sanitari (APSS) Proton therapy Department, Trento, Italy
[6] Radiation Oncology Department, University of Washington, Seattle, Washington, United States of America
[7] Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana, United States of America
[8] Medical Physics, Mary Bird Perkins Cancer Center, Louisiana State University, Baton Rouge, Louisiana, United States of America
[9] Current Affiliation of Dr. Schwarz is the University of Washington, USA, and Current affiliation for Dr. Kamp is the University Hospital of Cologne, Germany
[10] These authors have contributed to this work equally.

E-mail: E.Kollitz@physik.uni-muenchen.de

Keywords: second cancer risk, proton therapy, computational phantoms

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
24 March 2022

REVISED
6 July 2022

ACCEPTED FOR PUBLICATION
9 August 2022

PUBLISHED
12 September 2022

Abstract

Objective. CT-mesh hybrid phantoms (or 'hybrid(s)') made from integrated patient CT data and mesh-type reference computational phantoms (MRCPs) can be beneficial for patient-specific whole-body dose evaluation, but this benefit has yet to be evaluated for second cancer risk prediction. The purpose of this study is to compare the hybrid's ability to predict risk throughout the body with a patient-scaled MRCP against ground truth whole-body CTs (WBCTs). *Approach.* Head and neck active scanning proton treatment plans were created for and simulated on seven hybrids and the corresponding scaled MRCPs and WBCTs. Equivalent dose throughout the body was calculated and input into five second cancer risk models for both excess absolute and excess relative risk (EAR and ERR). The hybrid phantom was evaluated by comparing equivalent dose and risk predictions against the WBCT. *Main results.* The hybrid most frequently provides whole-body second cancer risk predictions which are closer to the ground truth when compared to a scaled MRCP alone. The performance of the hybrid relative to the scaled MRCP was consistent across ERR, EAR, and all risk models. For all in-field organs, where the hybrid shares the WBCT anatomy, the hybrid was better than or equal to the scaled MRCP for both equivalent dose and risk prediction. For out-of-field organs across all patients, the hybrid's equivalent dose prediction was superior than the scaled MRCP in 48% of all comparisons, equivalent for 34%, and inferior for 18%. For risk assessment in the same organs, the hybrid's prediction was superior than the scaled MRCP in 51.8% of all comparisons, equivalent in 28.6%, and inferior in 19.6%. *Significance.* Whole-body risk predictions from the CT-mesh hybrid have shown to be more accurate than those from a reference phantom alone. These hybrids could aid in risk-optimized treatment planning and individual risk assessment to minimize second cancer incidence.

1. Introduction and background

One of the most devastating potential effects of radiotherapy in cancer treatment is the incidence of a radiogenic secondary malignancy. Secondary malignant neoplasms (SMNs) accounted for 17%–19% of all new cancer diagnoses in the US between 2005 and 2009 (Morton *et al* 2014), and it is estimated that 8% of all SMNs are

related to the patient's prior radiation treatment (Berrington de Gonzales *et al* 2011). Pediatric patients are at even greater risk, as children and young adults are more likely to live longer post-treatment, and because their smaller size places key organs at risk (OAR) closer to areas receiving therapeutic doses (ICRP 2007, Dracham *et al* 2018). Due to an overall increase in cancer survivor populations (estimated to reach 22.1 million in the US) (Miller *et al* 2019, Howlader *et al* 2021), there is increasing interest in improving patient quality of life post treatment by reducing long term adverse effects of radiation therapy (Oeffinger *et al* 2006, Newhauser *et al* 2016).

For exposures below 2.5 Sv (considered low dose in radiation oncology), the commonly accepted model for cancer induction risk due to radiation is the linear no-threshold (LNT) model described in the BEIR VII report (NRC 2006). This model relies heavily on data from the atomic bomb survivors, and the validity of this model is the subject of significant debate (Mossman 2012, Cardarelli and Ulsh 2018, Sykes 2020). Even accepting the LNT model as it is, current recommendations from the BEIR VII limit its validity to tissues receiving under 2.5 Sv. Alternate risk models have been proposed to complement the LNT model to extend into therapeutic exposures (Fontenot *et al* 2009, Rechner *et al* 2015, Stokkevåg *et al* 2016, Dasu and Toma-Dasu 2017).

Despite the uncertainties inherent in risk modelling, relative risk is important to consider when comparing treatment modalities and assessing individual patient risk. Furthermore, aside from simple risk assessment, there is interest in developing risk-optimized treatment planning methods to reduce the risk of SMNs while maintaining clinical tumor control standards (Rechner *et al* 2015, Zhang *et al* 2015, Rechner *et al* 2020). However, while this does begin to address the SMN risk inside the treatment field, studies have shown that secondary malignancies can appear throughout the body in tissues receiving relatively low dose. In particular, a study from Diallo *et al* (2009) found in a group of 115 pediatric patients who developed a solid SMN that 22% were more than 5 cm away and 13% were more than 10 cm away from the edge of the irradiated volume. Furthermore, the largest population of SMNs (31% of all reported) were observed in tissues receiving less than 2.5 Gy.

One of the main benefits of proton therapy is reduced dose to healthy tissue due to the dose deposition properties of the Bragg peak. In recent years a lot of research has been dedicated to investigating whether the tissue sparing from proton therapy actually results in reduced risk for developing an SMN compared to photon therapy (Chung *et al* 2013, Eaton *et al* 2015, DeLaney *et al* 2020, König *et al* 2020). One of the concerns regarding risk from proton therapy is the stray neutron dose generated in the treatment head or the patient (Schneider *et al* 2002, Jarlskog and Paganetti 2008, Hälg and Schneider 2020), particularly in passive scattering proton therapy, where additional neutrons are produced within the scatterers compared to active scanning proton therapy (Pérez-Andújar *et al* 2009, Newhauser *et al* 2016, AAPM TG 158 Kry *et al* 2017).

Physical measurements, Monte Carlo simulations, and analytical models have all been vital tools when investigating and modelling out-of-field dose. However, a key element needed to translate these out-of-field dose predictions into patient-specific organ doses and subsequently second cancer risk is knowledge of the patient's anatomy outside the treatment field (Newhauser and Durante 2011). In clinical cases, the extent of the patient CT images are limited and rarely extend across the whole-body, therefore this anatomical knowledge needs to be supplemented by some other means.

Advanced mesh-type reference computational phantoms (MRCPs) have been developed by international commission on radiological protection (ICRP) Task Group 103 under Committee 2 (Kim *et al* 2018) to replicate the ICRP 110 adult reference phantoms (ICRP 2009) in a 3D volumetric mesh format. These MRCPs provide multiple benefits over voxel-based phantoms, including more accurate modelling of fine structures and relative ease of modification without loss of information (Kim *et al* 2016). By scaling these MRCPs to patient-specific measurements, and then splicing them with known patient CT data, a hybrid whole-body phantom can be created which supplements the anatomy of the known in-field with a patient-specific approximation of the out-of-field anatomy. These hybrid phantoms retain the integrity of the in-field dose distribution, and most frequently provide the best estimate of out-of-field organ dose compared to an MRCP alone (Kollitz *et al* 2022).

The goal of this study is to assess if the ability of a patient-specific CT-mesh hybrid phantom to predict second cancer risk in organs throughout the body is impacted when considering multiple risk models and evaluate if the hybrid geometry could provide a benefit in risk assessment. In this specific case, focusing on the risk from the therapeutic beam and internally generated neutrons produced during active scanning proton therapy without the influence of the treatment room or beam nozzle.

2. Methods

Seven head and neck hybrid phantoms were created, each corresponding to a separate ground truth whole-body CT. For those hybrids, treatment plans were created using a research version of the RayStation 7 treatment planning system which were based on and adapted from real clinical plans delivered to proton therapy patients. These treatment plans were then simulated via an in-house Geant4 Monte Carlo code on the ground truth

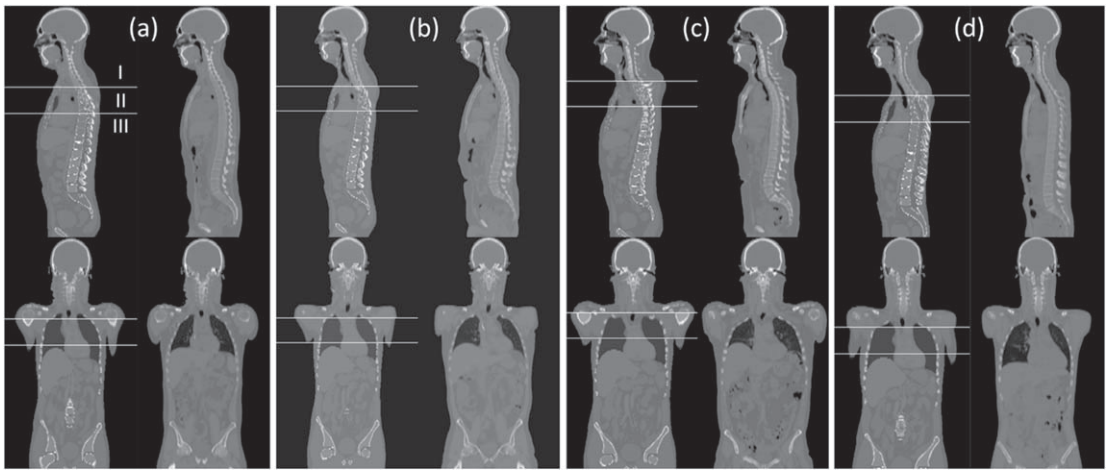


Figure 1. All of the male hybrid phantoms representing patients 1–4 side by side with their respective WBCTs. For each subfigure (a)–(d), the hybrid is on the left and the WBCT is on the right. The top row shows all images in sagittal view and the bottom row in coronal. The hybrid has each of the three segments (from top to bottom in each torso: in-field CT (I), transition (II), and out-of-field scaled MRCP (III)) demarcated by a white line and labelled in subfigure (a), where the shared anatomy is the most superior segment of the head and neck.

whole-body CT (WBCT), on the hybrid phantom, and on an MRCP scaled to patient measurements taken from the WBCT (in a clinical case the scaling could be adapted to use physical measurements). In the previous study of Kollitz *et al* (2022), it was found that the default MRCP was least frequently the best predictor of organ dose compared to a hybrid phantom or an MRCP scaled to patient measurements. For this reason, we chose not to include the default MRCP in our methods and analysis. The ability of the hybrid to predict equivalent dose to organs of interest compared to a mesh phantom alone was then analyzed with respect to the WBCT ground truth. The resulting organ equivalent doses were then input into five risk models (1 linear, 2 linear-plateau, 2 linear-exponential) for radiation-induced secondary cancer. Then, similarly to the equivalent dose, the hybrid was then analyzed in the context of risk predictive ability compared to the scaled MRCP alone with respect to the WBCT ground truth.

2.1. Hybrid phantom creation

The hybrid phantom design is composed of three segment types which form a continuous whole body computational phantom when combined:

- In-field—patient CT images of the treatment site
- Out-of-field—tetrahedral volumetric meshes cut from an MRCP scaled to patient measurements
- Transition—a blending region created by voxelizing the scaled MRCP, deformably registering the scaled MRCP to the patient CT, and extending and softening the deformation field to smoothly blend the out-of-field mesh anatomy into the patient CT anatomy.

Each of these seven hybrids was created following the methodology of Kollitz *et al* (2022). The MRCPs used in all hybrids in this study were developed to replace the ICRP 110 adult reference computational models (ICRP 2009, Kim *et al* 2018). The completed hybrids used in this study are shown in figures 1 and 2, where figure 1 shows the male hybrids with their respective WBCTs, and figure 2 shows the same with the female hybrids with all segments marked.

In its current state, many stages of hybrid creation are still manual, such as: selecting the point at which the scaled MRCP gets appended to the in-field CT, organizing image file data, and integrating the images into the Geant4 environment. A trained user following an instructional manual should be able to create a fully bespoke hybrid phantom from starting materials within a few hours, which includes processing time for scripts executing mesh voxelization, image deformation, and tetrahedral mesh excision. As work progresses on the hybrid phantoms, an increasing priority should be placed on the automation of hybrid production. However, at this early stage of development, it is important to note the not insignificant time investment needed to create a hybrid.

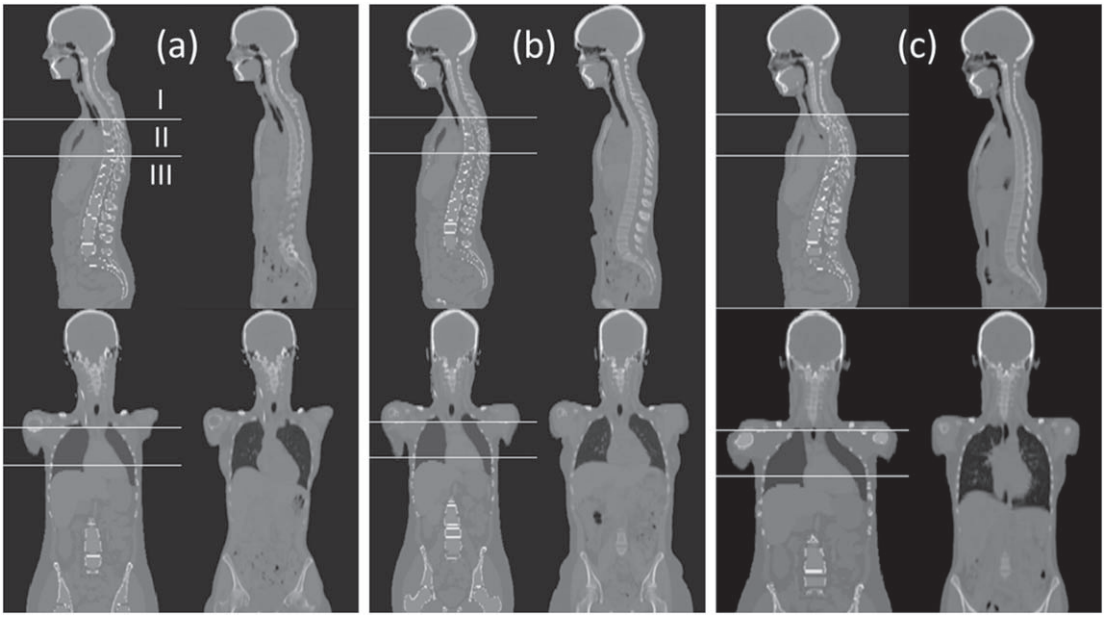


Figure 2. All of the female hybrid phantoms representing patients 5–7 side by side with their respective WBCTs. For each subfigure (a)–(c), the hybrid is on the left and the WBCT is on the right. The top row shows all images in sagittal view and the bottom row in coronal. The hybrid has each of the three segments (from top to bottom in each torso: in-field CT (I), transition (II), and out-of-field scaled MRCP (III)) demarcated by a white line and labelled in subfigure (a), where the shared anatomy is the most superior segment of the head and neck.

2.1.1. WBCT selection

Each of the seven hybrid phantoms were uniquely tailored to a different patient WBCT. These WBCTs were chosen from a selection of scans provided by the University Hospital of LMU Munich. Some criteria which guided WBCT selection were: relative lack of metallic artifacts (particularly from fillings in the mouth), mixed genders, and variety in patient size. For this study, 4 male WBCTs and 3 female WBCTs were used.

2.1.2. MRCP scaling

The adult MRCPs were provided by the ICRP Task Group 103, Committee 2 who developed them in two formats: a triangular surface mesh model with concentric hollow organs and a volumetric model with solid tetrahedral meshes. The surface mesh is easily manipulated in most commercial 3D modelling software, and here has been scaled to match the torso dimensions of the WBCT. The scaling was based on the previous work of Zvereva et al (2017) and Kollitz et al (2022), using bony anatomy of the CT image (spine height, pelvis width and depth), in a clinical situation the measurements would be taken directly from the patient (sitting height, hip circumference, or a more complex scaling factor from surface imaging). After scaling, the mesh was converted into a volumetric mesh before being cut and spliced into the hybrid and implemented into the Geant4 simulation geometry.

2.2. Treatment planning and simulation

A research version of RayStation 7 was used to contour and plan IMPT proton treatments for each of the seven treatment sites. Treatment plans were created on each treatment site's designated in-field CT. All plans centered on the head and neck region, however the treatment plans were greatly varied between patients, including different beam angle combinations, inclusion or exclusion of range shifters, and extent of tumor volumes.

Each of the treatment plans made for this study was based on and adapted from anonymized clinical proton treatment plans provided for reference by the Trento Proton Center. The beam setup and range uncertainties for these plans were adapted as closely as possible given the difference in available machines in our research version of RayStation 7. Likewise, dose constraints and objectives were initially implemented as is, and then adjusted to achieve dose distributions in the new anatomy which were as close as possible to the dose to targets and OAR in the original plans. When the resulting dose distribution differed significantly from the original plan, dose limit guides in the literature were consulted to confirm that the differing organs were still within clinical toxicity limits. The complete treatment plans as simulated in Geant4 are given in the supplementary materials (available online at stacks.iop.org/PMB/67/185011/mmedia).

2.2.1. Organ contouring

Two categories of organs were contoured, one for the purposes of treatment planning (local fine structure OAR, planning target volumes (PTVs), etc) and one for the purposes of post-simulation risk analysis. Some of these organs belonged to both groups depending on the treatment plan (left and right eyes, thyroid), however, some additional organs were contoured for the purposes of post-simulation risk analysis such as the whole brain and part of the lungs (when present in the CT).

All organs were contoured based on contouring atlases for the head and neck from the radiation therapy oncology group (RTOG) and NRG Oncology Center for Innovation in Radiation Oncology as well as by referencing real clinical active scanning proton treatment plans. Some of the structures contoured for treatment planning purposes in the CT in-field (such as the optic chiasm and brainstem) are likely to have some error in definition due to the low resolution ($1.034 \times 1.034 \times 5$ mm) and poor image contrast in the WBCT segments. The tumor volumes were challenging to place within the WBCT considering the WBCTs had no solid tumors and thus lacked the tissue displacement typical of images containing solid tumors. For example, if in the clinical case a tumor formed between two OARs, then the only options to replicate that volume in the WBCT were to either include some healthy tissue of the OAR in the contour or to reduce the tumor volume to exclude OAR tissue. For these cases, we took the second approach to keep the OAR volumes as consistent across the patient representations as possible. The choice to reduce excess overlap between the tumor volumes and the OARs was made to preserve the ability to use the clinical treatment plans' planned dose distributions as a guide for our own. If the tumor shape had been copied directly, it is possible that some overlap with OARs would have made it impossible to achieve the same dose distribution as in the reference clinical plans.

After contouring was complete, the smallest treated volume for a single plan was 59 cm^3 and the average volume of the largest PTV in each plan was 235 cm^3. An additional complication to contouring was patient positioning. The WBCTs were taken with a neutral head tilt, while the CTs of the reference treatment plans were taken with the head tilted back, making it more difficult to replicate the tumor volumes in the WBCT. With these considerations in mind, the tumor shape and volume from the reference plans were replicated as closely as possible.

The organs chosen for post-simulation analysis of secondary cancer risk were the brain, left and right eyes, thyroid, lungs, liver, kidneys, and bladder. These organs were chosen since they were relatively easy to contour, are spread throughout the torso, were present across both genders of patient, had equivalent meshes in the MRCPs, spanned several organs with specific organ risk coefficients using the BEIR VII model (NRC 2006), and/or were of particular interest when considering secondary cancer risk. While some of these organs were already contoured for treatment planning purposes (such as the eyes), organs not contained or partially contained in the treatment anatomy (lungs, liver, kidneys, and bladder) had to be contoured on the WBCT for dose analysis purposes.

2.2.2. Simulations and equivalent dose calculation

Each treatment plan was exported from RayStation using an in-house script which converted the beam angles and spot coordinates into a notation readable by Geant4 and which additionally converted the beam weighting from monitor units to particle number. The base implementation of Geant4 (Agostinelli *et al* 2003) was derived from that of Schmid et al (2015) and updated to be compatible with version 10.05.p01 and to construct tetrahedral mesh and hybrid geometry. The voxel-based segments were implemented in the form of DICOM images using the HU-based materials list from Schneider *et al* (2000) and the mesh-based segments used the organ materials from the original MRCPs (Kim *et al* 2018). The Monte Carlo physics models were configured as the default predefined QGSP_BIC_HP Geant4 physics list. For a given patient representation (WBCT, hybrid, scaled MRCP), 500 million particles were simulated for each treatment plan, separated into 10 batches for statistical analysis. This number of particles was chosen pragmatically as a compromise between ideal statistics (for example $<1\%$ uncertainty even for organs farthest from the treatment isocenter) and computational and temporal limitations. With this number of simulated particles the uncertainty in equivalent dose and risk quantities are around 1%–2% in the closer and larger out-of-field organs (lungs and liver) and around 10%–16% in the farther and smaller out-of-field organs (kidneys and bladder). We considered this to be an acceptable level of uncertainty given the practical limitations inherent in substantially increasing the number of simulated particles and given that the absolute dose difference in the far-field organs was small enough to be of little practical relevance especially when considering other uncertainties not accounted for in this study.

To calculate equivalent dose, the therapeutic proton component of dose was separated from the neutron component, so the appropriate weighting factors could be applied to each. Further, the neutron energy was needed to calculate the energy-specific weighting factor recommended by the ICRP 92 model (ICRP 2003). Using the scorers previously developed and tested (Kollitz *et al* 2022), the neutron dose scorer separated out the neutron component, first by tagging neutrons and all descendants of neutrons and then tallying all dose from the tagged particles as the neutron dose component. The other scorers developed in that work were used to track the

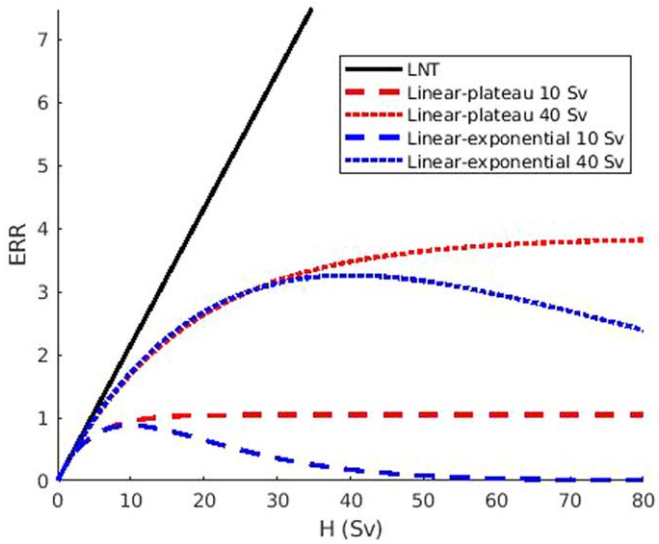


Figure 3. The shape of the 5 risk models used in this study: linear no-threshold (LNT), linear-plateau with an inflection at 10 Sv and at 40 Sv, and linear-exponential with an inflection at 10 Sv and at 40 Sv. This example uses the generic organ coefficient from the BEIR VII report for a patient with an age at exposure of 45 and an attained age of 65.

average neutron energy per voxel: one weighted by the neutron fluence through that voxel, one weighted by the amount of kinetic energy transferred from neutrons to charged particles within that voxel (KERMA). These two methods of obtaining neutron energy could potentially yield different weighting factors and therefore different estimates of equivalent dose from neutrons (Kollitz *et al* 2022). In this way, we could evaluate the potential impact of alternate neutron equivalent dose calculation methods on risk prediction when using Monte Carlo simulations.

2.3. Risk model implementation

This study examined both excess relative risk (ERR) and excess absolute risk (EAR). For each patient, the ERR for each organ was first calculated using the LNT risk model as:

$$ERR_T = H_T \times \mu_{T,ERR}, \tag{2.1}$$

where ERR_T is the ERR of a radiation-induced secondary malignancy developing in the organ of tissue type T, H_T is the equivalent dose delivered to the organ of interest, and $\mu_{T,ERR}$ is an organ specific risk coefficient. Similarly the EAR was first calculated as:

$$EAR_T = H_T \times \mu_{T,EAR}, \tag{2.2}$$

where EAR_T is the EAR of a radiation-induced secondary malignancy developing in the organ of tissue type T, H_T is the equivalent dose delivered to the organ of interest, and $\mu_{T,EAR}$ is an organ specific risk coefficient. The organ specific risk coefficient for both models was calculated according to the BEIR VII report depending on risk type (ERR or EAR), the age at exposure, attained age, and sex of the patient. For this study, since all of the treatment plans were fully hypothetical, the chosen age at exposure was 45 and attained age was 65. The LNT model of risk is commonly recommended only in tissues receiving less than 2.5 Sv by the BEIR VII report and the ICRP 99 Publication (ICRP 2005), therefore, some alternative nonlinear models to risk were also examined in this study. The linear-plateau and linear-exponential risk models with an inflection point at 10 and 40 Sv were chosen to represent a wide range of plausible dose-response relationships and to put this study in the context of previous literature (Sigurdson *et al* 2005, Ruben *et al* 2008, Fontenot *et al* 2009, Fontenot *et al* 2010, Rechner *et al* 2012, Rechner *et al* 2015). Figure 3 illustrates the shapes of the different risk models used in this study using the $\mu_{T,ERR}$ value for 'other' organs which do not have a unique coefficient defined by the BEIR VII.

Using the linear-exponential risk model, the EAR was calculated as:

$$EAR = H \times \mu_{T,EAR} \times e^{-\alpha H} \tag{2.3}$$

and the EAR using the linear-plateau risk model was calculated as:

$$EAR = \frac{\mu_{T,EAR}}{\alpha} \times (1 - e^{-\alpha H}). \tag{2.4}$$

Both using the same mathematical formalisms as Fontenot *et al* (2010) which were in turn based on the models suggested by Schneider *et al* (2006), where H is the equivalent dose, $\mu_{T,EAR}$ is the organ specific risk

Table 1. Values for curvature parameter α used with each risk model.

Model	Inflection point (Sv)	α, this work	α, Fontenot *et al* (2010)
LNT	N/A	N/A	N/A
Linear plateau	10	0.230 259	0.25
	40	0.057 565	0.068
Linear exponential	10	0.1	0.09
	40	0.025	0.025

coefficient for that organ for EAR, and α is a parameter to control the shape of the curve according to the respective inflection point (10 or 40 Sv). Equations for ERR were the same as the ones for EAR except the $\mu_{T,EAR}$ was replaced by $\mu_{T,ERR}$.

The inflection point in the linear-exponential model was defined as the equivalent dose point which yielded the maximum risk. As the linear-plateau model doesn't have a true mathematical inflection point, the inflection point was defined as the equivalent dose point which yielded 90% of the maximum risk of the model as equivalent dose approaches infinity. In addition to this, the model was defined such that the risk at 1 Sv was equivalent to that of the LNT model, adjusting the $\mu_{T,EAR/ERR}$ term if necessary. This was done to maintain the linear relationship at low doses. With these conditions, the α for each model and the μ_T for each model and each tissue were calculated. Table 1 gives the calculated α value for each model and inflection point as well as those used previously by Fontenot *et al* (2010). For all models which used an α value which differed from those used by Fontenot *et al* (2010), the organ-specific risk coefficients $\mu_{T,EAR/ERR}$ were calculated using both our and Fontenot's α values at the 1 Sv reference point mentioned previously. From this calculation, in all models there was a less than 1% difference between the modified $\mu_{T,EAR/ERR}$ risk coefficients.

2.4. Analysis methods

For both equivalent dose and risk, the better performing phantom for a given organ and patient was defined by whichever phantom gave the smallest percent difference in equivalent dose or risk from the WBCT ground truth. The phantoms were considered to be equal predictors of equivalent dose or risk for a given organ if the 25th and 75th percentiles of the organ quantity (equivalent dose or risk) overlapped. Additionally, in the case of one overestimating the ground truth and the other underestimating, the two phantoms were considered equal predictors if the absolute value of the difference of the 25th and 75th percentiles from the ground truth overlapped. The uncertainty percentiles were entirely based on the statistical variance between the 10 simulation batches.

For clarity, we define here a 'case' as an equivalent dose or risk evaluation in a specific organ for a given treatment plan.

For example, one case would be the KERMA-weighted equivalent dose in the left eye in Patient 3, which would then be used to evaluate the relative performance of the phantoms for that case. For risk, an example of a case would be the linear-plateau model with inflection at 40 Sv EAR prediction in the liver of Patient 4 based on the fluence-weighted equivalent dose calculation.

3. Results

This section addresses the hybrid performance in terms of equivalent dose and secondary cancer risk prediction. The different factors impacting the calculated risk (selected model, neutron energy calculation method, etc) are examined based on the effect on the hybrid phantom's predictive ability. The full quantitative data for equivalent dose and secondary cancer risk estimates are given in the supplementary materials.

3.1. Hybrid ability to predict equivalent dose

Regardless of whether the KERMA- or fluence-weighted neutron energy scoring was used for equivalent dose calculation, in all considered patients the hybrid provided better than or equal predictions compared to the scaled MRCP for all organs designated as in- or near-field. The in or near field organs were defined as the brain, the left and right eyes, and the thyroid. In all patients these organs are entirely contained in the in-field CT segment of the hybrid phantom shared by the WBCT. They are categorized as 'in or near field' since, due to the different treatment plans, some organs such as the thyroid may be in or out of the treatment field depending on whether the nodal volumes were irradiated for that patient. In 95% of the in-and-near-field cases the hybrid was better than the scaled MRCP alone, out of a total 56 cases (two equivalent dose scorers, four in-and-near-field organs, seven patients). In the remaining 3 (5%) cases, the scaled MRCP performed equivalently to the hybrid.

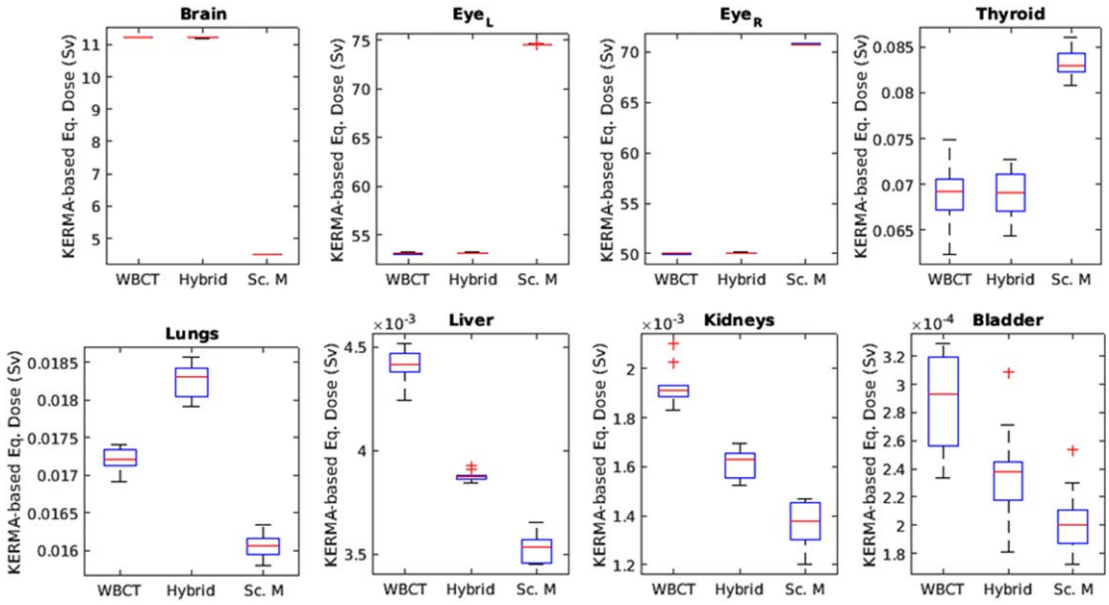


Figure 4. Boxplots showing the equivalent dose calculated using the KERMA-weighted neutron energy scorer for each of the selected organs in Patient 1 using the hybrid phantom, the scaled MRCP, and the ground truth WBCT. The central mark in each box is the median, the box edges are the 25th and 75th percentiles, the crosses are outliers (defined as points that are more than 1.5 times the interquartile range away from the 25th or 75th quartiles.), and the whiskers extend to the extreme data points not including outliers.

Table 2. Fraction of organs by phantom which gave the best equivalent dose prediction. In/near-field: brain, left and right eyes, thyroid. Out-of-field: lungs, liver, kidneys, bladder.

Organ location	Hybrid	Scaled MRCP	Equivalent performance
In/near-field	0.95	0.00	0.05
Out-of-field	0.48	0.18	0.34

This is consistent with what might be expected considering that by definition the hybrid and WBCT share the in-field anatomy while the scaled MRCP does not.

For the organs far outside the treatment field (defined here as the lungs, liver, kidneys, and bladder for all patients) the hybrid provided superior estimates than the scaled MRCP in 48% out of a total 56 cases and equivalent estimates in 34% of cases. In the remaining 18%, the scaled MRCP outperformed the hybrid phantom (table 2).

Figure 4 shows boxplots of the organ equivalent doses calculated using the KERMA-weighted neutron energy scorer for a single patient (Pat 1). The top row of boxplots show that for the in and near field organs, the hybrid accurately represents the equivalent dose received by the ground truth WBCT. Outside of the treatment field, in the transition and mesh regions of the hybrid phantom, typically the hybrid still exhibits an improvement over the scaled MRCP, but may differ from the ground truth WBCT value. Boxplots of the equivalent dose using the fluence-weighted scorer, and for the other patients, are given in the equivalent dose boxplots in the supplementary materials.

In addition to the number of cases in each rank, we can also look at the average magnitude of deviation from the WBCT for each rank category. For equivalent dose, this information is given in table 3. The average absolute value of the percent difference for superior hybrid cases in the in-field (where the hybrid shares the CT anatomy of the WBCT) is 0.18% and there are no inferior hybrid cases. In the out-of-field, the hybrid has similar average percent differences for both superior and inferior cases (9.3%–13.5%). This contrasts sharply with the scaled MRCP, where the average percent difference varied across superior and inferior cases from 2.4% to 25.5%. When the hybrid provides a superior out-of-field prediction, on average it provides an improvement on the percent difference by approximately 12% over the scaled MRCP. In cases where the scaled MRCP outperforms the hybrid, that the overall average percent differences lower, and also that the difference between the two phantoms is reduced to about 7%.

Table 3. Equivalent dose prediction average percent difference from WBCT of organs falling under the superior/inferior performance category for both phantoms. In/near-field: brain, left and right eyes, thyroid. Out-of-field: lungs, liver, kidneys, bladder.

	Hybrid		Scaled MRCP	
Organ location	Superior	Inferior	Superior	Inferior
In/near-field	0.18	N/A	N/A	30.62
Out-of-field	13.53	9.34	2.42	25.50

Table 4. Fraction of organs by phantom which gave the best risk prediction. In/near-field: brain, left and right eyes, thyroid. Out-of-field: lungs, liver, kidneys, bladder.

Organ location	Hybrid	Scaled MRCP	Equivalent performance
In/near-field	0.96	0.00	0.04
Out-of-field	0.518	0.196	0.286

3.2. Hybrid ability to predict risk

For each patient, the ERR and EAR was calculated for the five selected risk models and for both the fluence and KERMA-weighted neutron energy scorer. This resulted in 20 estimates of risk (10 each for ERR and EAR) for a given organ in a given patient representation (WBCT, hybrid, scaled MRCP) for each treatment plan. Overall, across all in-field organ risk estimates (brain, left and right eyes, and thyroid), the hybrid outperformed the scaled MRCP in 96% of cases and performed equally to scaled MRCP in 4%. Across the out-of-field organs (lungs, liver, kidneys, and bladder) the hybrid outperformed the scaled MRCP in 51.8% of cases, performed equivalently in 28.6% of cases, and performed worse than the scaled MRCP in 19.6% of cases (table 4).

Figure 5 shows the EAR per ten thousand person years (10k PY) predicted for all organs and all patients, using the KERMA-weighted neutron energy scorer for each of the 5 risk models (the corresponding figure for ERR is given in the supplementary materials). As all of the models share the linear model in low dose regions, it was possible to observe which cases choice of phantom had a higher impact than choice of risk model. It is also possible from this figure to see not only which phantom best predicted the ground truth for each EAR case, but also the magnitude of the difference with respect to the WBCT and to each other.

Similarly to the equivalent dose analysis, we can also look at the average absolute value deviation from each risk prediction rank category from both the hybrid and scaled MRCP cases. These values for the risk are given in table 5. While the numbers are slightly different, overall a similar pattern emerged as in the equivalent dose. In the in-field organs, the cases where the hybrid was superior had an average percent difference from the ground truth of 0.25%, and there were no inferior in-field hybrid cases. Out-of-field, the average percent difference between the superior and inferior hybrid cases ranged from 9.3% to 12.6%, while the same range for the scaled MRCP cases ranged from 3.2% to 25.9%.

For an out-of-field case where the hybrid gave the superior estimate, the difference between the hybrid and the scaled MRCP was 13.3%. Correspondingly, when the scaled MRCP gave the superior case, the average difference from the hybrid was 6.1%.

3.2.1. Impact of selected risk model

The five different risk models frequently had a significant impact on the magnitude of the risk prediction in organs receiving >1 Sv. However, the relative performance of the hybrid compared to the scaled MRCP was completely unaffected by choice of risk model. For a given organ subjected to a given treatment plan, if the hybrid outperformed the scaled MRCP using one risk model, it subsequently outperformed the scaled MRCP using every other risk model for both EAR and ERR.

Figure 6 shows the EAR for all patients for each of the five risk models (the corresponding figure for ERR is given in the supplementary materials). The organs receiving the highest dose have the greatest difference between the risk models. For organs receiving less than 2.5 Sv, where the linear-plateau and linear-exponential models are designed to match the LNT model, the difference in predicted risk between the models is minimal. Despite being matched to the LNT in the low dose-region, it is still possible to have some small deviations, particularly at extremely low dose. This is because the non-LNT models are inherently nonlinear, and the point of approximate linearity is matched at the 1 Sv equivalent dose point. In regions receiving close to zero dose, the

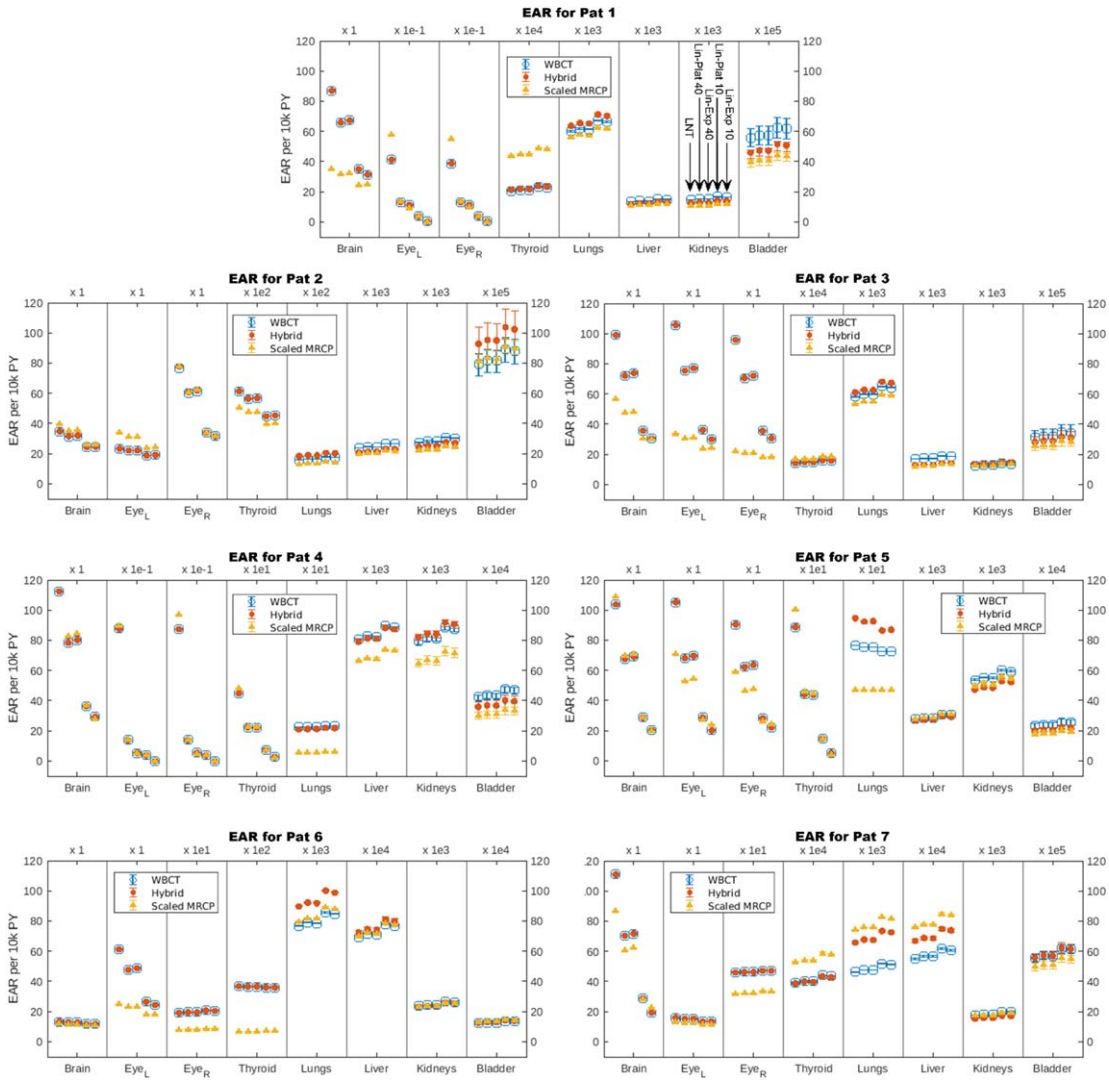


Figure 5. A comparison of the EAR predictions using the KERMA-weighted neutron energy scorer for equivalent dose for the 5 risk models across all 7 patients, all 8 selected organs, and all 3 patient representations. Each column corresponds to a particular organ where the WBCT, hybrid, and scaled MRCP are represented by a blue open circle, a red filled circle, and a yellow filled triangle respectively. Within a particular column, there are five points from left to right corresponding to each risk model. From left to right: LNT, linear-plateau with inflection at 40 Sv, linear-exponential with inflection at 40 Sv, linear-plateau with inflection at 10 Sv, and linear-exponential with inflection at 10 Sv. For visibility, the predicted risk and error for all models was multiplied by the specified power of 10 at the top of each of their respective columns. Error bars representing the 25th and 75th percentile are given for every measurement, but are only visible in some very far out-of-field organs.

Table 5. Risk prediction average percent difference from WBCT of organs falling under the superior/inferior performance category for both phantoms. In/near-field: brain, left and right eyes, thyroid. Out-of-field: lungs, liver, kidneys, bladder.

	Hybrid		Scaled MRCP	
Organ location	Superior	Inferior	Superior	Inferior
In/near-field	0.25	N/A	N/A	28.24
Out-of-field	12.56	9.29	3.15	25.92

models are going to deviate more than at the point of alignment. From this figure, across all risk models the scaled MRCP most visibly deviates from the hybrid and WBCT in tissues receiving between 0.1 and 11 Sv. This dose level is more likely to correspond to tissues in or near the treatment field and may include some organs which are entirely contained in the in-field CT, such as the brain. This region in turn corresponds to the part of the hybrid phantom which most heavily utilizes patient-specific anatomy.

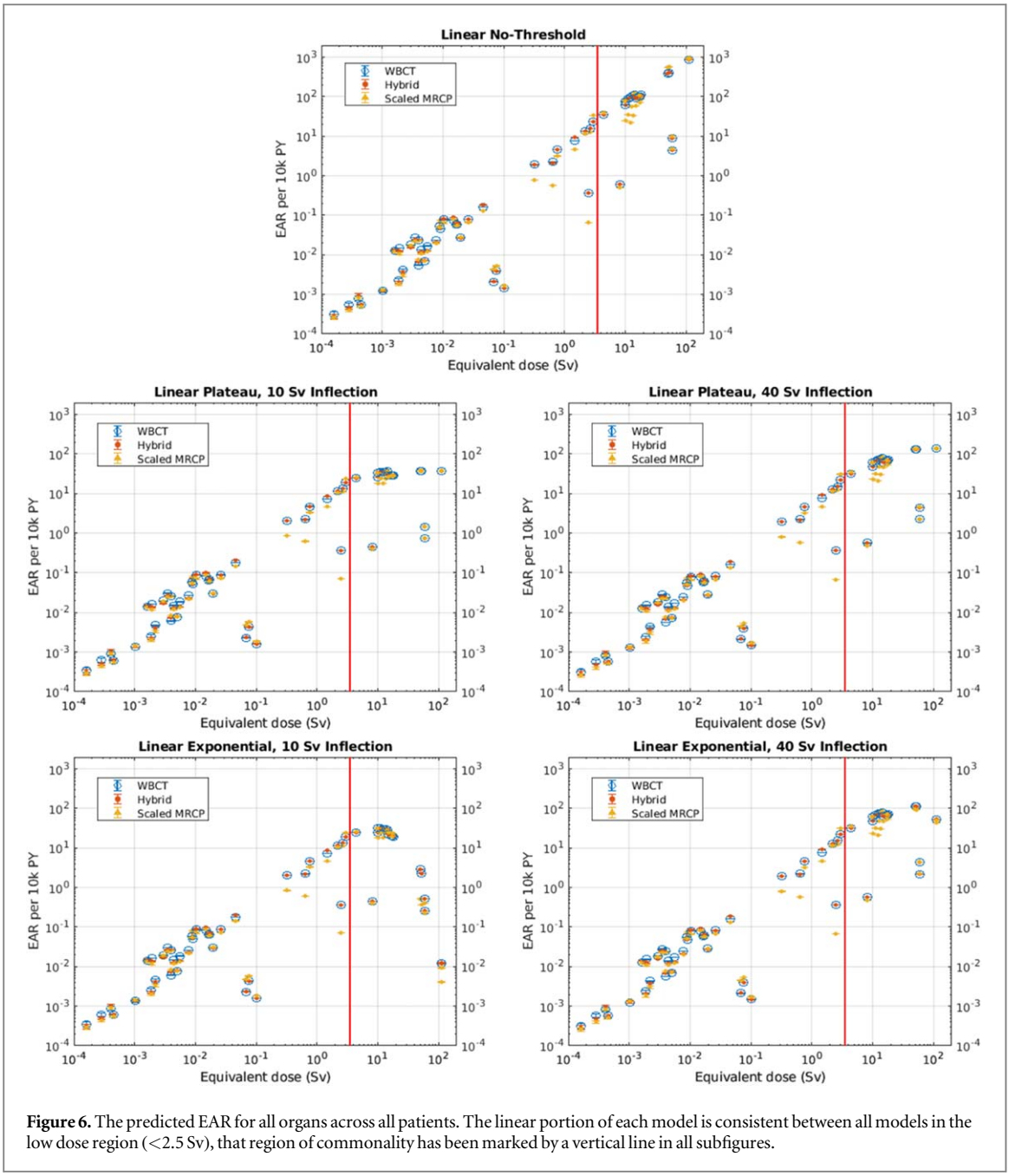


Figure 6. The predicted EAR for all organs across all patients. The linear portion of each model is consistent between all models in the low dose region (<2.5 Sv), that region of commonality has been marked by a vertical line in all subfigures.

3.2.2. Impact of neutron energy scoring type

The choice of KERMA- or fluence-weighted neutron energy scoring for equivalent dose calculation did not greatly impact the magnitude of predicted risk in any of the patient representations. Overall, using the KERMA-weighted scorer resulted in slightly higher absolute risk predictions compared to those using the fluence-weighted scorer. However, relative performance of the hybrid and scaled MRCP was not often affected. Across all the in-field organs (brain, left eye and right eyes, and thyroid), the choice of scorer did not impact relative hybrid performance at all. That is, for in-field organs if the hybrid provided a superior prediction of risk compared to the scaled MRCP using the KERMA-weighted scorer, then it likewise provided a superior prediction when using the fluence-weighted scorer.

However, outside of the treatment field, there were a few cases where changing the scorer also changed which phantom provided the closest estimate of risk relative to the ground truth WBCT. These changes were consistent for a given organ across ERR and EAR calculations and across all implemented risk models. For the 4 out-of-field organs (lungs, liver, kidneys, and bladder) across 7 patients, when switching from the fluence-weighted scorer to the KERMA-weighted scorer, in 64% of cases the switch did not impact relative performance. In 25% of cases, changing from the fluence-weighted scorer to the KERMA-weighted scorer resulted in the hybrid improving its relative rank compared to the scaled MRCP. In the remaining 11% of cases, changing from the fluence-weighted to the KERMA-weighted scorer resulted in the hybrid's relative rank worsening compared to the scaled MRCP.

Although in general the KERMA-weighted scorer yielded higher estimates of risk, quantitatively the difference between the two scorers is quite small in the out-of-field. For example, in the out-of-field organs the largest quantitative difference between the two scorers was observed in the lungs of the Pat 5 scaled MRCP using the LNT model, where the choice in scorer resulted in a change in the EAR from 4.53 to 4.68 per 10k PY and a change in the ERR from 1.10 to 1.14 for the fluence and KERMA-weighted scorers respectively. In all other out-of-field organs for all other patients, the difference in ERR was <0.01 and the difference in EAR per 10K PY was <0.09 between scorers.

For most cases, the deviation in risk estimation caused by the different scorers was less than the inherent systematic uncertainty due to the different risk models. For organs inside the field or receiving more than ~0.1 Sv, the risk model uncertainty was greater than the uncertainty introduced from the different scorer types. Similarly, for organs very far from the treatment field (typically receiving <2 mSv), the risk model systematic uncertainty outweighed the scorer choice uncertainty. Organs in between those two regions (for these plans typically the liver, kidneys, and sometimes lung) the scorer uncertainty was the same or larger than the risk model uncertainty. As an example of this, the EAR of the kidneys in the Pat 4 WBCT using the LNT model was 54.4 per 10k PY using the fluence-weighted scorer and 79.3 for the KERMA-weighted scorer. However, for the same organ and patient representation using the linear-plateau model with a 10 Sv inflection point, the EAR was 60.8 per 10k PY using the fluence-weighted scorer and 88.7 for the KERMA-weighted scorer. This means that, for this specific case, changing the risk model resulted in an estimated change of ~8 incidences of secondary cancer per 10k PY while changing the neutron energy scorer resulted in an estimated change of ~26 incidences per 10k PY.

4. Discussion

4.1. Conclusions from results

This study represents the first steps into exploring the applications of a patient-specific hybrid CT-mesh computational phantom. In addition to implementing and analyzing the hybrid's secondary cancer risk prediction, we aimed to reproduce and expand on the results of Kollitz *et al* (2022) (which used two WBCTs and four total hybrids) with respect to the hybrid's ability to predict equivalent dose to organs throughout the whole body. To this end, the equivalent dose results reflected the same pattern found in Kollitz *et al* (2022). The hybrid yielded as good as or better predictions of equivalent dose for all in-field organs across all patients. The selected out-of-field organs were most frequently best predicted by the hybrid (48% of all out-of-field organs), closely followed by equivalently predicted by both (34%). The least probable, though still likely, result was that the out-of-field organ was best predicted by the scaled MRCP, which occurred in 18% of the studied cases.

When expanding from the equivalent dose to the secondary cancer risk, a similar pattern emerged, but with a slightly stronger preference for the hybrid phantom in out-of-field organs. As with equivalent dose, the secondary cancer risk to the in-field organs was predicted by the hybrid as well as or better than the scaled MRCP in all cases. In the out-of-field cases, following the pattern of the equivalent dose predictions, the hybrid yielded a superior prediction of risk compared to the scaled MRCP in most out-of-field organs (51.8%) and an equivalent prediction (28.6%). The scaled MRCP was again the least likely to yield the closest prediction of risk at 19.6% of cases.

In both equivalent dose and risk, the hybrid exhibited a more consistent percent difference in the out-of-field compared to the scaled MRCP. Whether a hybrid case was superior (average difference of ~13%) or inferior (average difference of ~9%) to the scaled MRCP, the average percent difference was in the same range. For the scaled MRCP, however, the superior cases had an average difference of ~3% while the inferior cases had an average difference of ~26%. This indicates that the scaled MRCP has a much higher variance in predictive ability. Furthermore, when the hybrid is superior to the scaled MRCP, the gain in percent accuracy is on average approximately twice as high as when the scaled MRCP is superior to the hybrid.

One subtle aspect to note is that there are very rare cases in figures 5 and 6 where the scaled MRCP and hybrid appear to give equally good risk predictions for some organs in the in-field receiving higher levels of dose (for example the right eye in Pat 2 in figure 5). This could potentially have a few causes such as: (1) a small absolute difference in equivalent dose or risk which, despite being statistically separate, is difficult to visualize on a log–log scale or even a linear scale covering a large range of values or (2) regions where organs receive close to prescription levels of dose might be at least partially contained in the PTV contours (planned to receive high and uniform dose), leading to a more even distribution and similar predictions of dose at prescription levels despite the different anatomies and perturbed dose distribution between the MRCP and hybrid.

The hybrid phantom is able to retain the integrity of the in-field while providing an improvement over a scaled MRCP alone for whole-body equivalent dose and secondary cancer risk prediction. Furthermore, the hybrid's ability to predict risk compared to the scaled MRCP was not impacted by the choice of risk model or

type of risk. Therefore, despite the inherent uncertainties in calculating the absolute quantity of risk, the hybrid could be a useful tool for comparative ERR/EAR prediction when using any of the 5 models investigated.

The one factor that did change how the hybrid predicted risk relative to the scaled MRCP was the choice of neutron energy scorer. The scorer choice only changed relative performance in out-of-field organs, and even in that region the majority of the organ risks maintained their relative accuracy across both phantoms. However, when changing the scorer did effect the relative predictive performance, the most common change was the KERMA-weighted scorer favoring the hybrid phantom compared to the fluence-weighted scorer.

Both of the neutron energy scorers were designed to replicate as best as possible the conditions given by the ICRP 92 (ICRP 2003) such that the use of the energy-dependent neutron weighting factor model would be valid. However, due to the nature of the ICRP conditions (a mono-energetic field of neutrons impacting a target), an exact replication of these conditions for internally generated neutrons is impossible. The scorers as they stand represent a balance between the granularity of detail that is possible to access with Monte Carlo simulations and the generality of the neutron weighting factor model which accounts for probabalistic interactions while unable to know specifics.The relationship between the relative uncertainties of the scorer versus the risk model is likely due to a complicated interplay of the dominating factors at different dose levels, dose components, and distance from the field. At high doses (>2.5 Sv) the risk models will have the strongest influence over uncertainty due to different curve shapes after that equivalent dose point. Not only that, but at these levels the proton dose is dominant, and the change in neutron weighting factor from the different scorers is unlikely to affect the risk estimate as much. At lower doses (>0.1 Sv), the risk model influence is lessened, but in this dose region it is also likely that the dose is similarly dominated by the proton component of dose, so the change in neutron weighting factor has less of an influence. In regions far from the field, where both neutron dose dominates and where the neutron weighting factor differs significantly between the KERMA and fluence weighting, the choice of scorer is the stronger influencing factor. In the furthest extreme ranges from the treatment field, such as the brain in the prostate plan, the neutron weighting factor does not change significantly with KERMA versus fluence weighting and the risk model again becomes the strongest influence over uncertainty. Risk model choice can have a significant impact on the uncertainty, even in the range between 0 and 2.5 Sv where the models are supposed to approximate the LNT model, because of the inherent non-linearity of the other models mentioned in section 3.2.1.

In terms of individual organ performance trends, across all seven patients and all dose and risk quantities, the list of organs from most to least frequently superior to the scaled MRCP alone was the lungs, liver, kidneys, and bladder, with the lungs being superior over the scaled MRCP alone in at least 5 of the patients for all quantities and the liver in at least 4. However, we feel this does not necessarily showcase a particular benefit that the hybrid geometry brings to those organs, and rather is a factor of the particular conditions under which we designed the tests (namely that all the treatment plans were head and neck plans). This list of organs also correlates with distance of the organ from the isocenter, which was a key factor in whether or not a prediction in the out-of-field matched the ground truth WBCT (Kollitz *et al* 2022). Because of this, it is most likely the case that organs closest to the isocenter in the hybrid are the most likely to have accurate distances compared to the WBCT since they share the in-field anatomy (and the scaled MRCP does not). As the distance increases, the inter-patient anatomical differences begin to compound and result in varied distances to isocenter for organs farthest from the in-field region and therefore more varied performance. However, this does show that the organs closest to the in-field are the most likely to gain an advantage from using the hybrid geometry.

Despite the hybrid clearly showing a relative advantage in risk prediction over the scaled MRCP, it is still necessary to discuss the absolute advantage and whether that warrants the effort of constructing a hybrid phantom. Because the absolute risk predictions in this study are incomplete due to a lack of machine and room geometry in the simulation (besides range shifter blocks when required by the treatment plan), we will only be discussing the difference in predicted risk in terms of the internally generated neutrons.

Most organs designated as outside the treatment field (lungs, liver, kidneys, and bladder) received less than 1 Sv of equivalent dose. The only exception was the lungs in Pat 5, which received 1.46 or 1.48 Sv using the fluence-weighted and KERMA-weighted scorer respectively in the ground truth WBCT. These organs all fall under the LNT-matching portion of the investigated risk models, and therefore the absolute quantity of predicted risk for both ERR and EAR was relatively consistent across all the tested models. These are also the organs which yielded the lowest risk of secondary cancer according to those models. While the hybrid did most frequently provide a superior risk prediction over the scaled MRCP in this region, the risk predicted was quite small, on the scale of 10^{-2} EAR per 10k PY or 10^{-4} ERR for the kidneys and liver and 10^{-4} EAR per 10k PY or 10^{-5} ERR for the bladder. So while the hybrid might give a 15% improvement over the scaled MRCP (as it did for the EAR in the kidneys of Pat 4), that resulted in a difference between the hybrid and scaled MRCP of 0.0172 EAR per 10k PY. It is reasonable to ask whether unaccounted uncertainties might overtake any observed superiority of the hybrid, or whether this level of EAR is of interest when considering treatment planning optimization.

To address the first question, the hybrid phantom and the scaled MRCP are intrinsically similar in the out-of-field since the hybrid out-of-field is made from a segment taken directly from the scaled MRCP. If the only region of interest is contained completely in the mesh part of the hybrid phantom, then the effort needed to create a hybrid will likely mean it would be easier and equivalently useful to use a scaled MRCP alone considering that part of the anatomy would be the same in both. However, the real strength of the hybrid phantom lies in (1) the preservation of the in-field, (2) the inclusion of the deformed transition region, and (3) the ability to create whole body dose and risk distributions with a single simulation geometry. The preservation of the in-field region means that the dose distribution entering the region beyond the CT is as accurate as possible, in addition to having the deformable registration in the transition region working to match patient anatomy bordering the CT image. Furthermore, it is unlikely that there would be a clinical scenario where simultaneously the far out-of-field risk is of interest and the in and near field risk is not of interest, meaning that it is unlikely to encounter a scenario where the hybrid would not provide any benefit over the scaled MRCP.

To address the second question, the findings from Diallo *et al* (2009) reported that for their cohort of 115 pediatric patients experiencing a secondary cancer, 27% of all SMNs were in regions receiving <1 Gy, and a full 6% were observed in regions receiving <0.1 Gy. While this is not fully transferable to the results of this paper due to the use of absorbed dose in Diallo and equivalent dose here, as well as our lack of machine and room geometry, the key takeaway message is that tissues outside of the treatment field receiving very low doses are still potential host to a non-negligible number of secondary cancers. In addition to this, some of the most prevalent types of secondary cancer are sarcoma, thyroid, breast, and CNS (Jenkinson *et al* 2004, Diallo *et al* 2009, Berrington de Gonzalez *et al* 2013). Of those origin sites, only the thyroid was included in this study as the intent was to elucidate the hybrid's ability to predict the risk throughout the body, not make any statements on the magnitude or alleviation of risk. The inclusion of other sites in future studies could potentially show significant levels of risk in these sites outside of the treatment field.

Another factor to consider is the computational time necessary to use Monte Carlo simulations with the hybrid phantom. In this study, a single job of 1 million particles took between 3 and 26 h, depending on the treatment plan complexity and inclusion of range shifters. While this would be prohibitively lengthy in a clinical setting, it should be noted the patient geometry type did not have a large impact on the simulation time compared to the impact of the treatment plan itself (in a treatment plan taking 3 h to simulate there might be a difference of $\sim$30 min between geometry types, in a treatment plan taking 26 h it might be a difference of $\sim$3 h). So, the same temporal obstacles would be in place even if a full patient whole-body CT was available and is not limited to the hybrid phantom. Another hurdle is the requirement of a large number of particles to achieve better statistics far from the treatment field. Again, this is independent of the hybrid phantom, but remains a challenge when discussing whole-body Monte Carlo simulations. Some potential solutions include using the hybrid in combination with analytical models, so the benefit of patient-specific anatomy is maintained without the computational burden of a full Monte Carlo simulation. In addition to this, some recent software developments for GPU accelerated Monte Carlo simulations might make these kinds of whole-body simulations more feasible in a clinical environment (Qin *et al* 2017, Qin *et al* 2018, Adam *et al* 2020).

4.2. Impact and future directions

The hybrid phantom has demonstrated from the results that it possesses the capability to provide patient-specific estimates of equivalent dose and secondary cancer risk when out-of-field patient anatomy is unknown. The hybrid's predictive ability is unhindered by choice of risk type or model. This kind of hybrid could potentially be used for designing risk-optimized treatment plans, or reconstructing dose for a patient presenting for reirradiation.

However, there are quite a few things to consider for future studies. The first is the construction of a treatment machine and room in the simulation geometry. Particularly in the case of passive scattering plans, a significant component of stray radiation is produced inside the treatment head (Newhauser and Durante 2011), and the work of Englbrecht *et al* (2021) showed that there were some significant contributions to the neutron field from the gantry components and room structure in the 0.1–10 MeV energy range, coinciding with the peak of the ICRP 92 neutron weighting factor. Without this contribution, any estimate of risk cannot be accurate for a clinical scenario. For this study it was considered sufficient to only have the patient geometry constructed because (1) the absolute risk was of less importance in evaluating hybrid performance than relative accuracy and (2) only active scanning plans were simulated, and internally generated neutrons are the primary contributor to stray neutron dose in active scanning systems (Hälg and Schneider 2020). Despite the lack of this geometry, the scored equivalent dose (both total and neutron components) outside the of treatment field is within the range of expected values for scanned proton beams (Paganetti 2012, Hälg and Schneider 2020). On the whole, our values tend to underestimate equivalent dose compared to the range of studies examined in both Paganetti (2012) and

Hälg and Schneider (2020), possibly due to the reference studies' use of mono-energetic fields at a single angle versus a full treatment plan and our aforementioned lack of a full simulation geometry.

Other future developments to consider could be full DVH-based volumetric risk calculations, especially for organs in the transition region that are closer to the treatment field, for use in risk-optimized treatment planning. Furthermore, future studies could include not just ERR or EAR, but also include lifetime attributable risk (LAR). Not only this, but also some retrospective analysis could be performed where available clinical dose delivery data is compared against results obtained using hybrid phantoms.

Future work should also focus on cases where secondary cancer risk is of the most concern, such as in pediatric cases. A full library of MRCPs of varied body types are now available (Choi *et al* 2020), allowing a more refined initial selection of MRCP to match WBCT body type and dimensions. In addition to this, pediatric mesh phantoms and a corresponding library are currently under development (Choi *et al* 2021), enabling the creation and study of pediatric hybrid phantoms. Additionally, frequent sites of secondary cancer such as sarcomas, breast, CNS, or other OARs could be added to the risk calculations on a treatment site specific basis. Similarly, treatment plans targeting other primary cancer sites where secondary cancer is a concern for treatment planning could be investigated using the hybrid phantom methodology.

5. Conclusions

For the selected out-of-field organs, the hybrid phantom was most frequently the best predictor in terms of second cancer risk. The choice of a specific risk model did not alter the fact that the hybrid phantom is an overall better predictor compared to a fully generic mesh patient model, although it can affect the absolute level of the quantified risk. The accuracy of the prediction is not dependent on specific choice of organs, but rather on other parameters, predominantly the treatment plan itself and the relative location of the organs with respect to the treatment field.

The hybrid CT-mesh phantom provides a tool to supplement patient CT images with patient-specific out-of-field anatomy to provide more accurate predictions of equivalent dose and secondary cancer risk than generic anthropomorphic phantoms. However, further development is needed in automating the hybrid creation and in accounting for the contributions of the machine and room geometry to the dose components related to second cancer risk.

Acknowledgments

This work was funded through the Deutsche Forschungsgemeinschaft (DFG) as part of the Research Training Group Graduiertenkolleg (GRK) 2274 joint initiative for Advanced Medical Physics for Image-Guided Cancer Therapy. We would like to additionally thank and acknowledge the support from Erik Traneus and Rasmus Nilsson of RaySearch Laboratories AB (Stockholm, Sweden) for technical assistance and access to RayStation for research purposes. Finally, we would like to thank Dr Jonas Fontenot for consulting on the risk model coefficients and Dr Marco Riboldi for consulting on deformable image registration for the hybrid phantom.

ORCID iDs

Erika Kollitz https://orcid.org/0000-0002-6331-6171
Haegin Han https://orcid.org/0000-0002-7635-1611
Marco Pinto https://orcid.org/0000-0001-6835-2561
Marco Schwarz https://orcid.org/0000-0001-6964-7732
Katia Parodi https://orcid.org/0000-0001-7779-6690

References

Adam D P, Liu T, Caracappa P F, Bednarz B P and Xu X G 2020 New capabilities of the Monte Carlo dose engine ARCHER-RT: clinical validation of the varian TrueBeam machine for VMAT external beam radiotherapy *Med. Phys.* **47** 2537–49

Agostinelli S *et al* 2003 Geant4—a simulation toolkit *Nucl. Instrum. Methods Phys. Res.* A **506** 250–303

Berrington de Gonzalez A, Curtis R E, Kry S F, Gilbert E, Lamart S, Berg C D, Stovall M and Ron E 2011 Proportion of second cancers attributable to radiotherapy treatment in adults: a cohort study in the US SEER cancer registries *Lancet. Oncol.* **12** 353–60

Berrington de Gonzalez A, Gilbert E, Curtis R, Inskip P, Kleinerman R, Morton L, Rajaraman P and Little M P 2013 Second solid cancers after radiation therapy: a systematic review of the epidemiologic studies of the radiation dose-response relationship *Int. J. Radiat. Oncol. Biol. Phys.* **86** 224–33

Cardarelli J J 2nd and Ulsh B A 2018 It is time to move beyond the linear no-threshold theory for low-dose radiation protection *Dose Response* **16** 1–24

Choi C *et al* 2021 Development of skeletal systems for ICRP pediatric mesh-type reference computational phantoms *J. Radiol. Prot.* **41** 139–61
Choi C, Yeom Y S, Lee H, Han H, Shin B, Nguyen T T and Kim C H 2020 Body-size-dependent phantom library constructed from ICRP mesh-type reference computational phantoms *Phys. Med. Biol.* **65** 125014
Chung C S, Yock T I, Nelson K, Xu Y, Keating N L and Tarbell N J 2013 Incidence of second malignancies among patients treated with proton versus photon radiation *Int. J. Radiat. Oncol. Biol. Phys.* **87** 46–52
Dasu A and Toma-Dasu I 2017 Models for the risk of secondary cancers from radiation therapy *Phys Med.* **42** 232–8
DeLaney T, Yock T and Paganetti H 2020 Assessing second cancer risk after primary cancer treatment with photon or proton radiotherapy *Cancer* **126** 3397–9
Diallo I *et al* 2009 Frequency distribution of second solid cancer locations in relation to the irradiated volume among 115 patients treated for childhood cancer *Int. J. Radiat. Oncol. Biol. Phys.* **74** 876–83
Dracham C B, Shankar A and Madan R 2018 Radiation induced secondary malignancies: a review article *Radiat. Oncol J.* **36** 85–94
Eaton B R, MacDonald S M, Yock T I and Tarbell N J 2015 Secondary malignancy risk following proton radiation therapy *Front. Oncol.* **5** 261–7
Englbrecht F S, Trinkl S, Mares V, Rühm W, Wielunski M, Wilkens J J, Hillbrand M and Parodi K 2021 A comprehensive Monte Carlo study of out-of-field secondary neutron spectra in a scanned-beam proton therapy gantry room *Z. Med. Phys.* **31** 215–28
Fontenot J D, Lee A K and Newhauser W D 2009 Risk of secondary malignant neoplasms from proton therapy and intensity-modulated x-ray therapy for early-stage prostate cancer *Int. J. Radiat. Oncol. Biol. Phys.* **74** 616–22
Fontenot J D, Bloch C, Followill D, Titt U and Newhauser W D 2010 Estimate of the uncertainties in the relative risk of secondary malignant neoplasms following proton therapy and intensity-modulated photon therapy *Phys. Med. Biol.* **55** 6987–98
Hälg R A and Schneider U 2020 Neutron dose and its measurement in proton therapy-current State of Knowledge *Br. J. Radiol.* **93** 1107–20
Howlader N *et al* 2021 *SEER Cancer Statistics Review, 1975-2018* National Cancer Institute Bethesda, MD (https://seer.cancer.gov/csr/1975_2018/)
ICRP 2003 Relative biological effectiveness (RBE), quality factor (Q), and radiation weighting factor (wR). ICRP Publication 92 *Ann. ICRP* 1–121
ICRP 2005 Low-dose extrapolation of radiation-related cancer risk. ICRP Publication 99 *Ann. ICRP* **35** 1–153
ICRP 2007 The 2007 recommendations of the international commission on radiological protection. ICRP Publication 103 *Ann. ICRP* **37** 1–339
ICRP 2009 Adult reference computational phantoms. ICRP Publication 110 *Ann. ICRP* **39** 1–170
Jarlskog C Z and Paganetti H 2008 Risk of developing second cancer from neutron dose in proton therapy as function of field characteristics, organ, and patient age *Int. J. Radiat. Oncol. Biol. Phys.* **72** 228–35
Jenkinson H C, Hawkins M M, Stiller C A, Winter D L, Marsden H B and Stevens M C 2004 Long-term population-based risks of second malignant neoplasms after childhood cancer in Britain *Br. J. Cancer* **91** 1905–10
Kim C H *et al* 2016 The reference phantoms: voxel vs polygon *Ann ICRP. 2016* **45** 188–201
Kim C H *et al* 2018 New mesh-type phantoms and their dosimetric applications, including emergencies *Ann. ICRP* **47** 45–62
Kollitz E *et al* 2022 A patient-specific hybrid phantom for calculating radiation dose and equivalent dose to the whole body *Phys. Med. Biol.* **67** 035005
König L *et al* 2020 Secondary malignancy risk following proton versus x-ray treatment of mediastinal malignant lymphoma: a comparative modeling study of thoracic organ-specific cancer risk *Front. Oncol.* **10** 989–98
Kry S F, Bednarz B, Howell R M, Dauer L, Followill D, Klein E, Paganetti H, Wang B, Wuu C -S and George Xu X 2017 AAPM TG 158: measurement and calculation of doses outside the treated volume from external-beam radiation therapy *Med. Phys.* **44** e391–429
Miller K D, Nogueira L, Mariotto A B, Rowland J H, Yabroff K R, Alfano C M, Jemal A, Kramer J L and Siegel R L 2019 Cancer treatment and survivorship statistics *CA A Cancer J. Clin.* **69** 363–85
Morton L M, Onel K, Curtis R E, Hungate E A and Armstrong G T 2014 The rising incidence of second cancers: patterns of occurrence and identification of risk factors for children and adults *Am. Soc. Clin. Oncol. Educ. Book* **34** e57–67
Mossman K L 2012 The LNT debate in radiation protection: science versus policy *Dose-Response* **10** 1–13
National Research Council (NRC) 2006 *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2* (Washington, DC: The National Academies Press) (https://doi.org/10.17226/11340)
Newhauser W D and Durante M 2011 Assessing the risk of second malignancies after modern radiotherapy *Nat. Rev. Cancer* **11** 438–48
Newhauser W D, Berrington de Gonzalez A, Schulte R and Lee C 2016 A review of radiotherapy-induced late effects research after advanced technology treatments *Front. Oncol.* **6** 13–24
Oeffinger K *et al* 2006 Chronic health conditions in adult survivors of childhood cancer *N. Engl. J. Med.* **355** 1572–82
Paganetti H 2012 Assessment of the risk for developing a second malignancy from scattered and secondary radiation in radiation therapy *Health Phys.* **103** 652–61
Pérez-Andújar A, Newhauser W D and Deluca P M 2009 Neutron production from beam-modifying devices in a modern double scattering proton therapy beam delivery system *Phys. Med. Biol.* **54** 993–1008
Qin N, Pinto M, Tian Z, Dedes G, Pompos A, Jiang S B, Parodi K and Jia X 2017 Initial development of goCMC: a GPU-oriented fast cross-platform Monte Carlo engine for carbon ion therapy *Phys. Med. Biol.* **62** 3682–99
Qin N, Shen C, Tsai M Y, Pinto M, Tian Z, Dedes G, Pompos A, Jiang S B, Parodi K and Jia X 2018 Full Monte Carlo-based biologic treatment plan optimization system for intensity modulated carbon ion therapy on graphics processing unit *Int. J. Radiat. Oncol. Biol. Phys.* **100** 235–43
Rechner L A, Howell R M, Zhang R, Etzel C, Lee A K and Newhauser W D 2012 Risk of radiogenic second cancers following volumetric modulated arc therapy and proton arc therapy for prostate cancer *Phys. Med. Biol.* **57** 7117–32
Rechner L A, Eley J G, Howell R M, Zhang R, Mirkovic D and Newhauser W D 2015 Risk-optimized proton therapy to minimize radiogenic second cancers *Phys. Med. Biol.* **60** 3999–4013
Rechner L A, Modiri A, Stick L B, Maraldo M V, Aznar M C, Rice S R, Sawant A, Bentzen S M, Vogelius I R and Specht L 2020 Biological optimization for mediastinal lymphoma radiotherapy—a preliminary study *Acta Oncol.* **59** 879–87
Ruben J D, Davis S, Evans C, Jones P, Gagliardi F, Haynes M and Hunter A 2008 The effect of intensity-modulated radiotherapy on radiation-induced second malignancies *Int. J. Radiat. Oncol. Biol. Phys.* **70** 1530–6
Schmid S, Landry G, Thieke C, Verhaegen F, Ganswindt U, Belka C, Parodi K and Dedes G 2015 Monte Carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional changes in prostate cancer patients *Phys. Med. Biol.* **60** 9329–47

Schneider U, Agosteo S, Pedroni E and Besserer J 2002 Secondary neutron dose during proton therapy using spot scanning *Int. J. Radiat. Oncol. Biol. Phys.* **53** 244–51

Schneider U, Lomax A, Pemler P, Besserer J, Ross D, Lombriser N and Kaser-Hotz B 2006 The impact of IMRT and proton radiotherapy on secondary cancer incidence *Strahlenther Onkol.* **182** 647–52

Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions *Phys. Med. Biol.* **45** 459–78

Sigurdson A J *et al* 2005 Primary thyroid cancer after a first tumour in childhood (the childhood cancer survivor study): a nested case-control study *Lancet* **365** 2014–23

Stokkevåg C H, Fukahori M, Nomiya T, Matsufuji N, Engeseth G M, Hysing L B, Ytre-Hauge K S, Rørvik E, Szostak A and Muren L P 2016 Modelling of organ-specific radiation-induced secondary cancer risks following particle therapy *Radiother. Oncol.* **120** 300–6

Sykes P J 2020 Until there is a resolution of the Pro-LNT/Anti-LNT debate, we should head toward a more sensible graded approach for protection from low-dose ionizing radiation *Dose-Response* **18** 1–4

Zhang R, Mirkovic D and Newhauser W D 2015 Visualization of risk of radiogenic second cancer in the organs and tissues of the human body *Radiat Oncol.* **10** 107–16

Zvereva A, Schlattl H, Zankl M, Becker J, Petoussi-Henss N, Yeom Y S, Kim C H, Hoeschen C and Parodi K 2017 Feasibility of reducing differences in estimated doses in nuclear medicine between a patient-specific and a reference phantom *Phys. Med.: Eur. J. Med. Phys.* **39** 100–12