A Monte Carlo based radiation response modelling framework to assess variability of clinical RBE in proton therapy

J Eulitz⁵,⁶, B Lutz⁵, P Wohlfahrt¹,², A Dutz³, W Enghardt¹,², M Krause¹,²,³,⁵,⁶, C Karpowitz¹,²,³,⁶,⁷,⁸, E G C Troost¹,²,³,⁵,⁶,⁷, M Krause⁵,⁶, C Lühr⁴, and A Lütz⁹

1 Faculty of Medicine and University Hospital Carl Gustav Carus, OncoRay—National Center for Radiation Research in Oncology, Technische Universität Dresden, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
2 Faculty of Medicine and University Hospital Carl Gustav Carus, Department of Radiotherapy and Radiation Oncology, Technische Universität Dresden, Dresden, Germany
3 Helmholtz-Zentrum Dresden-Rossendorf, Institute of Radiooncology—OncoRay, Dresden, Germany
4 Helmholtz-Zentrum Dresden-Rossendorf, Institute for Radiation Physics, Dresden, Germany
5 National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany
6 German Cancer Research Center (DKFZ), Heidelberg, Germany
7 Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany
8 Helmholtz Association/Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Germany
9 German Cancer Consortium (DKTK), Partner Site Dresden, and German Cancer Research Center (DKFZ), Heidelberg, Germany

E-mail: Jan.Eulitz@uniklinikum-dresden.de

Keywords: proton therapy, clinical RBE, Monte Carlo, passive scattering

Abstract

The clinical implementation of a variable relative biological effectiveness (RBE) in proton therapy is currently controversially discussed. Initial clinical evidence indicates a variable proton RBE, which needs to be verified. In this study, a radiation response modelling framework for assessing clinical RBE variability is established. It was applied to four selected glioma patients (grade III) treated with adjuvant radio(chemo)therapy and who developed late morphological image changes on T1-weighted contrast-enhanced (T1w-CE) magnetic resonance (MR) images within approximately two years of recurrence-free follow-up. The image changes were correlated voxelwise with dose and linear energy transfer (LET) values using univariable and multivariable logistic regression analysis. The regression models were evaluated by the area-under-the-curve (AUC) method performing a leave-one-out cross validation. The tolerance dose TD50 at which 50% of patient voxels experienced toxicity was interpolated from the models. A Monte Carlo (MC) model was developed to simulate dose and LET distributions, which includes variance reduction (VR) techniques to decrease computation time. Its reliability and accuracy were evaluated based on dose calculations of the clinical treatment planning system (TPS) as well as absolute dose measurements performed in the patient specific quality assurance.

Morphological image changes were related to a combination of dose and LET. The multivariable models revealed cross-validated AUC values of up to 0.88. The interpolated TD₅₀ curves decreased with increasing LET indicating an increase in biological effectiveness. The MC model reliably predicted average TPS dose within the clinical target volume as well as absolute water phantom dose measurements within 2% accuracy using dedicated VR settings.

The observed correlation of dose and LET with late brain tissue damage suggests considering RBE variability for predicting chronic radiation-induced brain toxicities. The MC model simulates radiation fields in patients precisely and time-efficiently. Hence, this study encourages and enables in-depth patient evaluation to assess the variability of clinical proton RBE.
1. Introduction

Over the last few decades, the number of proton radiotherapy facilities has steeply increased (www.ptcog.ch). One of the reasons for this trend is the expectation of reduced radiation-induced side-effects compared to state-of-the-art photon treatments (Baumann et al 2016, Harrabi et al 2016) as the physical properties of protons offer the possibility to spare normal tissues surrounding the tumor from dose. Furthermore, protons are more effective at cell killing compared to photons, which is quantified in terms of the relative biological effectiveness (RBE). At present, a biological model considering a constant RBE of 1.1 is applied in proton beam treatment planning. However, a controversial debate about the necessity to review and improve this clinical RBE concept is currently ongoing (Mohan and Grosshans 2017, Lühr et al 2018b). A large number of in vitro (Chaudhary et al 2014, Pagano et al 2014b) and a few in vivo experiments (Saager et al 2018) suggest a variable RBE along the beam direction. Currently, it is unclear whether the variability of the RBE translates into clinically relevant differences in treatment outcome and whether treatment may need to be adapted (Lühr et al 2018a). Initial patient studies seeking clinical evidence of an increased RBE showed ambiguous results (Giantsoudi et al 2016, Peeler et al 2016, Underwood et al 2018).

One of the main determinants of in vitro proton RBE is the linear energy transfer (LET) besides physical dose and the values $\alpha$ and $\beta$ describing the radiobiological response to photon irradiation. The LET of protons increases while slowing down and therefore increases with penetration depth. In contrast, the LET of photons is essentially constant. A recent analysis of 31 pediatric ependymoma patients showed a correlation of image changes in magnetic resonance imaging (MRI) three months after proton therapy (PT) with dose and LET suggesting a variable proton RBE (Peeler et al 2016). However, radiation response models that assess the clinical relevance of variable RBE are lacking for chronic normal tissue complications of the central nervous system. An analysis of later time-points during follow-up is required, which may be more indicative of permanent adverse neurological defects.

The precise description of the radiation field in a patient is a prerequisite for modelling of biological radiation response. In particular, accurate dose and LET distributions in the patients are needed. Currently, clinical treatment planning systems (TPSs) cannot predict all relevant beam qualities for RBE prediction. Moreover, the accuracy of analytical pencil-beam (PB) dose calculation algorithms used in most TPSs was found to be limited, especially in heterogeneous patient geometries (Paganetti 2009, 2012, 2014a, Schuemann et al 2015). To model radiation response precisely, a method to predict the radiation field based on Monte Carlo (MC) simulations has been recommended (Paganetti et al 2008). MC treatment plan simulations can also be applied to verify clinical TPS dose and to predict absolute dose measurements in patient-specific quality assurance. However, to make MC-derived data clinically more accessible, methods that increase computational efficiency (Magni et al 2013, Méndez et al 2015) and an automated integration of the MC approach into the clinical workflow have been suggested (Verburg et al 2016).

In this study, a radiation response modelling framework has been established and retrospectively applied to four patients with glioma grade III who showed late morphological image changes in T1-weighted contrast-enhanced (T1w-CE) follow-up MR images. For this purpose, an MC simulation model was implemented to perform automated dose and LET simulations for individual patient treatment plans at the University Proton Therapy Dresden (UPTD). Robust variance reduction techniques were integrated to increase the simulation efficiency to enable RBE and dose verification studies with large patient cohorts.

2. Methods

2.1. Computational framework and efficiency

A UPTD-specific MC simulation model for automated simulation of passive scattering proton treatment plans has been developed (Eulitz 2017). It was implemented using TOPAS (Perl et al 2012), which is based on Geant4 (Agostinelli et al 2003) and dedicated to radiotherapy applications. The model contains an MC geometry of the universal nozzle of Ion Beam Applications (IBA, Louvain-La-Neuve, Belgium) in double-scattering (DS) mode (figure 1), which has been implemented and commissioned using water phantom measurements. In commissioning, 1 mm proton range and 2% dose accuracy for each of the eight double-scattering nozzle options are achieved.

We automated the simulation of proton treatment plans using the information included in the Digital Imaging and Communications in Medicine (DICOM) files provided by the TPS. The following parts of the simulation are automatically adjusted: patient-specific machine geometries (apertures and compensators), patient position including couch and gantry rotation, and patient anatomy and materials. We used the TOPAS built-in converter, which assigns materials based on computed tomography (CT) images representing the x-ray attenuation in Hounsfield Units (HU) (Schneider et al 2000), to synthesize the geometry of the patient. The CT-based assignment of material densities in simulation was adapted to reproduce the conversion of CT number to stopping
power as implemented in the clinical TPS XiO (Elekta, Stockholm, Sweden) used for DS treatment planning at UPTD (Wohlfahrt et al 2017). This allows for an unbiased comparison of MC dose simulations against TPS calculations.

For all simulations, we used TOPAS version 3.1 with default physics settings optimized for PT (Jarlskog and Paganetti 2008). All MC simulations were performed on the high-performance computation (HPC) cluster located at the Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany.

A variance reduction (VR) technique called geometrical splitting (GS) was implemented to decrease computation time. We applied GS settings similar to those that had earlier been customized for a comparable nozzle model (Méndez et al 2015) using the TOPAS built-in VR functionality (Perl et al 2012). In brief, two splitting planes (SPs) were inserted perpendicular to the beam axis within the nozzle (figure 1): one immediately upstream of the second scatterer (SP1) and one immediately upstream of the aperture farthest away from the isocenter (SP2) (figure 1). At each splitting plane, eight protons were generated (split) from one incident proton. If initial protons hit a surface parallel to the beam axis defined by a square volume in between both splitting planes (figure 1), Russian roulette (RR) was applied.

A region of interest (ROI) was defined as a disk with radius $R_{ROI}$ on a plane immediately downstream of the second aperture. The full particle phase space was scored on that plane. For phase space simulations, the particle production range cuts for electrons, positrons, and photons were increased from 0.05 mm to 50 mm to further save computation time (Paganetti 2002, Méndez et al 2015). For water phantom and CT simulations the TOPAS default range cut of 0.05 mm was applied for all particles. The phase space file was used as a beam source for water phantom and patient field simulations and was simulated four times, to decrease statistical uncertainty in a time-efficient way. Pre-calculated stopping-power ratios were used for protons and electrons during CT dose calculation (Perl et al 2012) with TOPAS default settings for proton (MeV) and electron (keV) energy binning.

Systematic errors can occur with the applied GS implementation, if a too-small ROI is chosen. Hence, robust $R_{ROI}$ values were iteratively derived by systematically comparing GS simulations against reference simulations without GS. The analysis was done for eight treatment fields, each representative of the eight nozzle options. Proton fluence, energy, and angular distributions as well as water phantom depth-dose curves were compared against reference simulations.

To derive planar fluence profiles the phase space was divided into concentric rings of equally sized areas of 3.14 cm$^2$. Proton fluences $\geq$ 50% of the fluence scored in the central ring were compared. The energy and angular spectra were evaluated in the region of the peak ($\geq$30% of maximum proton fluence) of both distributions using 4 MeV and 0.02° angular bins, respectively. The angular spectrum refers to the space angle $\Theta$, which is defined with respect to the beam axis. The GS simulations were normalized against reference simulations. For the planar fluence profiles, the fluence scored in the central concentric ring was used for normalization. The energy and angular spectra were normalized using the average fluence scored in the region of the peak.
The radius of the ROI, 
\[ R_{\text{ROI}} = \begin{cases} \alpha r_{\text{sq}}, & r < r_{\text{sq}} \\ \alpha r, & r \geq r_{\text{sq}} \end{cases} \]  
was approximated from iterative simulations. The scaling factor \( \alpha \) was determined to 4.5. The value \( r \) denotes the aperture radius and is defined as the maximum distance between the beam axis and the inner aperture opening. A square aperture (10 \( \times \) 10 cm\(^2\)) was used for the simulation of the eight treatment fields with an aperture radius \( r_{\text{sq}} = 71 \text{mm} \). In the rare case of larger aperture radii, \( R_{\text{ROI}} \) is scaled according to \( r \).

We evaluated our GS settings for a head, lung, and pelvis patient treatment field by comparing dose simulations with GS against reference dose simulations without GS. For this, the simulations were normalized by average CTV dose and in each spatial dimension, Gaussian filters were applied with a standard deviation of two voxels (2 mm) to reduce statistical MC noise and better resolve systematic dose deviations originating from geometrical splitting. The Gaussian filters were not applied in response modelling. For all simulations, approximately one million protons per square centimeter reached the phase space plane.

2.2. Dose verification
Due to the complexity of field formation in passive-scattering proton therapy, current clinical TPS do not provide accurate absolute dose calculations. Hence, the field-specific output factor (OF) needs to be measured to transform the relative dose distribution derived from the TPS to absolute doses. An OF for a treatment field is defined as the relationship between the number of monitor units (MU) and a characteristic field dose \( D \).

Due to the complexity of field formation in passive-scattering proton therapy, current clinical TPS do not provide accurate absolute dose calculations. Hence, the field-specific output factor (OF) needs to be measured to transform the relative dose distribution derived from the TPS to absolute doses. An OF for a treatment field is defined as the relationship between the number of monitor units (MU) and a characteristic field dose \( D \).

At UPTD, \( D_{\text{exp}} \) is determined as the average of four dose measurements in a water phantom at representative points within the plateau region of the spread-out-Bragg-peak.

For clinical dose verification, two different methods were implemented. In method 1 (M1), MC simulations were adapted to absolute patient dose measurements in a water phantom. In method 2 (M2), the dose of MC patient field simulations were normalized to dose calculations of the clinical TPS XiO (Elekta, Stockholm, Sweden) both performed in a water phantom. M2 allows for an unbiased comparison of TPS and MC dose calculations in the patient.

In method M1, we implemented an MC model of the ionization chamber inside the nozzle. The MU charge collection was assigned to a dose value, referred to as MU simulation (\( \text{MU}_{\text{sim}} \)), which was scored in a small pad geometry centered at the beam axis (Paganetti 2006). By comparing simulation and measurement of one clinical reference field in a water phantom (range: 16 cm, modulation: 10 cm, detector position: 11 cm) \( \text{MU}_{\text{sim}} \) was normalized once to the clinical MU at UPTD. We validated the OF prediction accuracy for seven other quality assurance reference fields and nine patient treatment fields. The relative dose \( D_{\text{rel}} \) simulated in the patient could be converted to absolute dose,

\[ D_{\text{abs}} = \frac{\text{MU}_{\text{exp}}}{\text{MU}_{\text{sim}}} \cdot D_{\text{rel}}, \]  
by applying the ratio \( \text{MU}_{\text{exp}}/\text{MU}_{\text{sim}} \), where the applied clinical MU \( \text{MU}_{\text{exp}} \) had been determined by patient-specific dose measurements in a water phantom for each treatment field prior to treatment.

For method M2, a separate water phantom simulation was performed for each patient field, in which the treatment head was configured as for the delivery of the field to the patient. The absolute dose \( D_{\text{abs}} \) can then be calculated as

\[ D_{\text{abs}} = \frac{D_{\text{W}}^\text{plan}}{D_{\text{W}}^\text{sim}} \cdot \frac{N_{\text{W}}}{N} \cdot D_{\text{rel}}, \]  
where \( D_{\text{W}}^\text{plan} \) and \( D_{\text{W}}^\text{sim} \) are the dose values in the water phantom of the treatment plan and the MC simulation, respectively. \( D_{\text{W}}^\text{plan} \) and \( D_{\text{W}}^\text{sim} \) are the average doses calculated from four characteristic spatial points in the field defined during clinical patient quality assurance. The different number of initial protons in the simulation of the water phantom, \( N_{\text{W}} \), and the patient, \( N \), are considered by the factor \( N_{\text{W}}/N \).

Method M2 was used to simulate the treatment plans of the four glioma patients (table 1) who were used in this study for radiation response modelling. Thus, the radiation response modelling framework remains independent from patient-specific measurements and an unbiased comparison between TPS and MC simulations is enabled. The MC dose simulations were compared with XiO dose calculations within the clinical TPS RayStation (RaySearch AB, Stockholm, Sweden). Dose volume histogram (DVH) statistics were evaluated for relevant structures such as the clinical target volume (CTV) and organs at risk (OAR).

2.3. RBE assessment
Treatment plans and MR scans of four glioma patients treated with (adjuvant) proton therapy (in combination with chemotherapy) at UPTD were considered (table 1). The patients were treated with conventionally
fractionated proton therapy (2 Gy(RBE) per fraction) to a total dose of 60 Gy(RBE). Plan information is provided in table 2. All patients received their first proton therapy fraction between September 2015 and June 2016 and developed post-therapeutic morphological MR image changes (IC), being contrast enhancements in T1w-CE MR sequences. These contrast enhancements result from damages to the blood-brain barrier and are clinically used to identify normal tissue complications (Mullins et al 2005). On average, follow-up MR scans were acquired every three-month after treatment.

Available patient data from biopsies and functional imaging were considered to distinguish image changes that may represent radiation-induced necrosis (radionecrosis) from image changes that may be signs of e.g. tumor progression, infection or gliosis. Patient PIII and PIV had a histologically confirmed radionecrosis. Patient PI and PII were highly suspected of having developed radionecrosis as no tumor constituents were found in follow-up 11C-methonine positron emission tomography (PET) imaging (Terakawa et al 2008) and biopsies.

For modelling, the latest available T1w-CE MR sequence within two years of recurrence-free follow-up was used. In the case of a partial resection of damaged tissue within this period, the last available MR sequence before resection was considered. The post-treatment MR images were rigidly registered to the planning CT scan and image changes were contoured within the Raystation TPS. All image change contours were validated by an experienced radiation oncologist (ET). Image voxels within the delineated region were labeled as 1. The remaining voxels were set to 0. Only brain tissue voxels that received more than 2% of the prescribed dose. Liquor-filled cavities, such as ventricles or surgical cavity, were also contoured on MR scans. Image voxels within these contours were excluded from analysis, because no image changes related to radionecrosis were expected in these regions.

For the four glioma patients, the MC framework was used to simulate, in addition to dose, track-averaged and dose-averaged LET distributions (LETT, LETD). Both the LETT and LETD represent two different methods for averaging the LET over a spectrum of protons and are frequently used as input parameters in RBE models for proton therapy. For LET calculation, the default TOPAS proton LET scorer was used, which gives the LET of primary and secondary protons including the energy deposited by associated secondary electrons (Perl et al 2012).

Two univariable ($M_1$, $M_2$) and two multivariable ($M_3$, $M_4$) logistic regression models (Table 3) were developed by a voxelwise correlation of image changes with either dose or LET as well as an interaction term of both parameters. The probability for an image change in a voxel

$$P_{IC} = \frac{1}{1 + e^{-s}},$$

Table 1. Characteristics of glioma patients with morphological image changes in T1-weighted contrast-enhanced (T1w-CE) MR images after (adjuvant) proton radio-(chemo)therapy.

| Patient | Diagnosis (WHO Grade) | VCTV (cm$^3$) | $\tilde{D}$ (Gy (RBE)) | Age at first radio-therapy (years) | Time to latest available image changes in T1w-CE MR (months after treatment) |
|---------|-----------------------|--------------|------------------------|----------------------------------|--------------------------------------------------|
| $P_I$   | Astrocytoma III       | 158.0        | 60                     | 41                               | 12                                               |
| $P_{II}$ | Astrocytoma III      | 196.5        | 60                     | 46                               | 23                                               |
| $P_{III}$ | Astrocytoma III     | 248.4        | 60                     | 41                               | 23                                               |
| $P_{IV}$ | Oligoastrocytoma III | 117.0        | 60                     | 51                               | 23                                               |

Abbreviations: VCTV, clinical target volume; $\tilde{D}$, prescribed dose, WHO, World Health Organization; RBE, relative biological effectiveness.

Table 2. Plan information for the four glioma patients treated with passive scattering proton therapy.

| Patient | Prescription dose (Gy/fx) | Number of fractions | R/cm | M/cm | Snout position/cm | Beam angles (gantry, couch) |
|---------|---------------------------|---------------------|------|------|-------------------|-----------------------------|
| $P_I$   | 2                         | 30                  | 11.0 | 8.0  | 18.5              | ($85^\circ$, $0^\circ$)     |
| $P_{II}$ | 2                         | 15                  | 9.4  | 7.7  | 17.2              | ($175^\circ$, $0^\circ$)    |
| $P_{III}$ | 2                         | 30                  | 9.2  | 7.7  | 15.3              | ($310^\circ$, $80^\circ$)   |
| $P_{IV}$ | 2                         | 30                  | 8.8  | 7.4  | 13.0              | ($70^\circ$, $300^\circ$)   |

Abbreviations: $P$, patient; fx, fraction; $R$, range; $M$, modulation.
was estimated based on a linear combination

\[ S = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \]

(5)
of the model predictors \( X_1 \) and \( X_2 \) representing the simulated beam qualities (dose, LET) with model coefficients \( \beta \). The model performance was assessed by the Area Under the Receiver-Operating-Characteristic (ROC) Curve (AUC).

A leave-one-out cross validation was performed to assess patient variability and model’s accuracy. The data set was split into four subsets, in which each subset contained data of three patients for training and the remaining patient for testing. For training, the logistic regression function of the scikit-learn python package (Pedregosa et al 2011) was used. For each model, the logistic regression coefficients obtained in the four cross-validation subsets were averaged. The averaged coefficients were used for the final prediction of image changes \( P_{IC} \).

### 3. Results

#### 3.1. Computational framework and efficiency

GS reduced the phase space computation time of the eight representative treatment fields by a factor of 5 to 15 for high (27.5 cm) to low (4.8 cm) field ranges, respectively, compared to reference simulations. The planar fluence had differences of up to 1.13% compared to reference simulations with an average difference of \(-0.03\%\) and a standard deviation of 0.12%. The statistical uncertainty of the MC simulation within each concentric ring was \(\leq 0.10\%\). The energy and angular spectra from the GS simulations differed at most by 0.46% and 0.61%, respectively, and the statistical uncertainties were below 0.10% within each energy and angular bin, respectively. Maximum differences for all depth-dose curves were below 1.80% in percentage of plateau dose (standard deviation of the difference of 0.60%).

The comparison of GS and reference simulation is illustrated in figure 2 for the treatment field representing nozzle option DS6. For visualization, the relative fluence differences (figures 2(E)–(G)) are shown in percentage of the averaged fluence in the beam center \(\phi_c\) of each reference simulation. The value \(\phi_c\) corresponds to the fluence scored in the central concentric ring for the planar profiles. For both spectra, \(\phi_c\) represents the average fluence in the central peak region (\(\geq 30\%\) of maximum fluence). The effect of increasing the range production cuts for electrons, positrons and gamma particles within the nozzle simulation on the proton fluence spectrum scored in the phase space scorer is negligible (see supplementary figure S1 (stacks.iop.org/PMB/64/225020/mmedia)). Small dose changes may occur for long-range fields (figure S1: T, X) at the water surface, but are still within the clinically acceptable tolerance range.

For the simulation of a head, lung, and pelvis treatment field, we obtained a maximum dose difference of 1.12% within the CTV when comparing simulations with and without GS (figure 3). The corresponding normalized CTV dose difference histograms are illustrated in figures 3(D)–(F) and table 4, respectively. With GS and increased particle production cuts the MC computation time of the head, lung and pelvis field was reduced by a factor of approximately 9.5, 3.5 and 2.5, respectively.

#### 3.2. Dose verification

Using geometrical splitting and approximately one million particles per square centimeter in the phase space plane, the maximum statistical uncertainty of the \(MU_{sim}\) calculation was 0.25%. The fourfold simulation of the phase space resulted in a statistical dose uncertainty in patients of about 0.5% in one CT voxel (1 \(\times\) 1 \(\times\) 2 mm\(^3\)). Using method M1 to determine the absolute dose, all clinical reference field and glioma patient field OF measurements were predicted within an accuracy of 2% (figure 4).

The simulated absolute dose distributions of the four glioma patients (table 1) were compared against XiO dose calculations using method M2 (figure 5). The simulated CTV mean dose was on average 0.58% below the...
planned CTV mean dose with a standard deviation of 0.88% and a maximum difference of 1.98%. Dose differences occurred in particular at distal field edges and in regions with high density gradients such as bone-air and bone-tissue interfaces. Dose statistics for CTV and relevant OARs are summarized in Table 5.

### 3.3. RBE assessment

For each of the four models $M_1$ to $M_4$, all logistic regression coefficients $\beta_i$ of the training cohorts were significantly different from zero with p-values less than 0.001. Table 6 shows the results of the leave-one-out cross-validation.
For the multivariable models $M_{III}$ and $M_{IV}$, dose response curves and the characteristic TD$_{50}$ parameter were calculated (see figures 6(B), (C), (E) and (F)). For $M_{III}$ a linear relation between TD$_{50}$ and the LET$_T$ was obtained with an intercept of 97.22 Gy and a slope of $-11.79$ Gy (keV/$\mu$m)$^{-1}$. For $M_{IV}$ the TD$_{50}$ was indirectly proportional to the LET$_T$. Using LET$_T$ instead of LET$_T$ for regression modelling, the models $M_{III}, M_{III}$, and $M_{IV}$ revealed very similar AUC values of 0.64, 0.86 and 0.85, respectively. In figure 6(D), the ROC curves for the two LET averaging methods are compared for models $M_{III}$ and $M_{IV}$. Calculated image change probabilities based on the averaged coefficients ($\beta$) and the mean AUC values derived by applying each trained model to the patient left out for training. For each model, image change probabilities were calculated using the $\beta$ applied to the full patient data and compared against the observed image changes. The AUC resulting for models $M_{I} - IV$, logit models; $\beta_{0-2}$, logistic regression coefficients; AUC, area under the curve; CI, confidence interval.

Table 5. Average differences of dose-volume histogram (DVH) parameters ($\Delta D_{\text{Ref}}, \Delta D_{\text{mean}}, \Delta D_{\text{max}}$) in percentage of prescribed dose for four simulated glioma patient proton plans versus treatment planning system calculations (XiO).

|                | $\Delta D_{\text{Ref}}$ in % | $\Delta D_{\text{mean}}$ in % | $\Delta D_{\text{max}}$ in % |
|----------------|-------------------------------|-------------------------------|-------------------------------|
|                | $\mu \pm \sigma$ max         | $\mu \pm \sigma$ max         | $\mu \pm \sigma$ max         |
| CTV           | $-1.19 \pm 0.50$ 2.05         | $-0.58 \pm 0.88$ 1.98         | $-0.54 \pm 1.39$ 2.57        |
| Brainstem     | $0.03 \pm 0.01$ 0.03          | $-0.79 \pm 1.03$ 2.42         | $-3.20 \pm 5.37$ 12.43       |
| Brain—CTV     | $0.00 \pm 0.01$ 0.02          | $-0.04 \pm 0.14$ 0.23         | $-0.87 \pm 1.16$ 2.57        |
| Optic nerve   | Left $0.15 \pm 0.88$ 1.42     | $0.61 \pm 1.83$ 3.33          | $-0.15 \pm 5.35$ 8.18        |
| Optic nerve   | Right $0.95 \pm 0.93$ 1.97    | $1.10 \pm 1.48$ 3.63          | $1.38 \pm 2.53$ 5.95         |

Abbreviations: CTV, clinical target volume; $D_{\text{Ref}}, D_{\text{mean}}, D_{\text{max}}$, difference in dose-volume histogram parameters between Monte Carlo simulation and the treatment planning system XiO; $\mu$, $\sigma$ and max, average, standard and maximum difference of four patients.

Table 6. Results of the leave-one out cross validation with dose and LET$_T$ for four glioma patients using four subsets each with three patients for model training and the remaining left-out patient for model testing.

|                | $\beta_0$ | $\beta_1$ | $\beta_2$ | AUC         |
|----------------|-----------|-----------|-----------|-------------|
|                | $\mu \pm \sigma$ | $\mu \pm \sigma$ | $\mu \pm \sigma$ | Mean 95% CI  |
| $M_I$          | $-6.36 \pm 0.12$ | $0.047 \pm 0.001$ | — | 0.67 0.53–0.82 |
| $M_{II}$       | $-4.51 \pm 0.17$ | $0.21 \pm 0.08$ | — | 0.64 0.49–0.80 |
| $M_{III}$      | $-17.66 \pm 2.51$ | $0.182 \pm 0.031$ | $2.15 \pm 0.51$ | 0.88 0.80–0.95 |
| $M_{IV}$       | $-9.45 \pm 0.67$ | $0.038 \pm 0.009$ | $0.038 \pm 0.007$ | 0.87 0.75–0.98 |

Abbreviations: $M_{I-IV}$, logit models; $\beta_{0-2}$, logistic regression coefficients; AUC, area under the curve; CI, confidence interval.
4. Discussion

Numerous in vitro and in vivo findings suggest a variable proton RBE. Therefore, there is a strong demand for studies investigating the clinical impact and verifying clinical evidence on a variable RBE. We implemented a full Monte Carlo patient simulation model that accurately predicts dose and LET distributions for patients treated with passive scattering proton therapy at UPTD. Based on that we established a radiation response modelling framework to assess the variability of clinical proton RBE and applied it to four glioma patients suspicious for (radio)necrosis. We found non-uniform spatial distributions of the necrotic lesions within the brain that were highly correlated with a combination of dose and LET with cross-validated AUC mean values of up to 0.88. This constitutes the first clinical indication for a variable clinical proton RBE for late brain toxicity.

The multivariable models with dose and LET performed remarkably better than the univariable models. The reason for that is the dedicated position of the image changes at the edge of the CTV. While model $M_{I}$ including only dose as predictor solely modelled the image changes within the high-dose region well, model $M_{III}$ using LET$_T$ better describes the image change voxels outside the CTV. However, both predictors perform poorly when applied to all image change voxels with cross-validated AUC values of 0.67 and 0.64, respectively. The full

Figure 5. Comparison of a simulated proton plan (A) to planned dose distributions derived from the treatment planning system XiO (B) for patient $P_{III}$. The relative dose difference in percent of the prescribed dose of 60 Gy (RBE) is illustrated in (C). The evaluation was performed within the treatment planning system RayStation. In the simulation, approximately $10^9$ primary protons reached the patient.

Figure 6. Results of the models predicting image changes on T1w-CE MR images. Receiver-operator-characteristic curves derived from a logistic regression analysis for different logit models (A) and linear energy transfer (LET) averaging methods (D). For the multivariable logistic regression models, projections for each of three LET$_T$ (1, 3, and 5 keV $\mu$m$^{-1}$) values are shown for model $M_{III}$ (B) and model $M_{IV}$ (E) together with the corresponding TD$_{50}$ lines (C) and (F). $M_{III}$ and $M_{IV}$ are hard to distinguish in (A) and LET$_T$ and LET$_D$ in (D).
image change distribution is only accurately captured when using combined functions of dose and LET. It has been shown that in vitro RBE models predict higher biological doses in particular at normal tissue structures at the edge of the CTV (Lühr et al. 2018b) when applied to clinical treatment fields. This effect is caused by a steep increase in LET at the distal edge of the proton beam. Hence, the spatial correlation of late brain toxicity with dose and LET indicates that proton RBE variability can play a role for clinical scenarios.

Both multivariable models discriminate responding brain tissue voxels from non-responding voxels similarly well. The predicted image change probabilities are comparable, too, as illustrated in figure 7. Hence, both LET treatment planning strategies seem appropriate and both models can support daily treatment planning by predicting risk regions for risk glioma patients. However, the incorporation of more patients is required to further validate the model. For that, the presented radiation response modelling framework provides a fast and automated solution. When applying the models to patients outside this cohort consideration of additional factors, such as diabetes, hypertension and chemotherapy, may be relevant.

Compared to model $M_{III}$, the mathematical form of model $M_{IV}$ is more similar to some common empirical in vitro based RBE models. However, when estimating an RBE from the T$_{50}$ distribution by assuming, e.g. an LET of 1 keV $\mu$m$^{-1}$ as reference radiation, model $M_{IV}$ would predict an RBE of about 5 for the high-LET range (5 keV $\mu$m$^{-1}$). This high RBE value seems unreasonable and illustrates the limitation of our approach in deriving absolute clinical RBE values from the clinical image data. One reason may be the simplified quantification of radiation damage due to contrast enhancements. Also, differences in the radio-sensitivity within the brain are ignored. Finally, the voxel-wise NTCP approach makes some assumptions itself, which may be too simplistic. For instance, all image voxels are treated to be independent in modelling. The late radiation response is interpreted as a local effect, which occurs within a voxel and is independent from the surrounding tissue. To better understand the effective size of the responding tissue unit and to quantify its damage the incorporation of functional imaging data may be helpful. This may potentially allow for deriving more reliable proton RBE values and for fitting $\alpha$ and $\beta$ values directly to patient outcome data.

A recent clinical RBE modelling study for brain (Peeler et al. 2016) used a model with dose and LET comparable to $M_{III}$ for radiation response modelling. When comparing their linear T$_{50}$ curve as a function of LET, with the one presented here, differences in slope and intercept are observed. These quantitative deviations most likely originate from the difference in the endpoint studied and patient cohort characteristics. Peeler et al. investigated early (three months post treatment) T2-FLAIR (fluid-attenuated inversion recovery) MR image changes for 31 pediatric patients treated for ependymoma. While early T2-FLAIR image changes are likely to be more closely correlated to the treatment time point, they are also more likely to indicate radiation edema. Radiation edema can disappear and do not necessarily indicate chronic radiation damage (Gutin et al. 1991). We considered late T1w-CE MR image changes as they may be correlated with radiological signs of radionecrosis (Dequesada et al. 2008, Walker et al. 2014).

There is a debate about the most accurate LET averaging method for proton RBE modelling. We used track-averaged LET primarily to facilitate comparability of our results with the other RBE modelling study for brain (Peeler et al. 2016). In our study, replacing LET$_{T}$ by LET$_{D}$ values resulted in very similar AUC values during modelling. Hence, the absolute difference between the LET$_{T}$ and LET$_{D}$ values had no major effect on the power of the models to differentiate voxels with or without image changes. The characteristic, though comparable, shape of the LET distributions for both averaging approaches, in particular the shallow rise within the CTV and sharp maximum at the distal field edge, is responsible for the high AUC values of the multivariable logistic models.

The foundation of the presented radiation response modelling framework is the MC simulation model that reliably predicts dose and LET distributions in patients. CTV mean doses of TPS proton treatment plans are predicted within 2% accuracy for four glioma patients. No systematic geometrical shifts are observed for dose differences between MC simulation and the TPS, which would indicate a patient misalignment in the simulation or an error in the generation of patient-specific beam-shaping components. Dose differences that occur at regions with high material density gradients, e.g. bone-tissue or tissue-air interfaces, could be referred to differences in proton scattering modelling of the underlying MC and PB algorithms (Sawakuchi et al. 2008). Hence, the MC simulation framework may be used as a reference for the verification of proton treatment plans in heterogenous patient geometries. Furthermore, it may be applied in clinical trials comparing proton therapy to conventional radiotherapy to avoid bias caused by systematic differences in calculated dose distributions (Schuemann et al. 2015, Zschack et al. 2016). As absolute dose measurements are predicted within the clinically acceptable tolerance of 2%, the framework could be applied to predict OF measurements in patient-specific quality assurance. This may help to reduce measurement effort for proton treatment using the DS technique.

The integration of VR methods with robust GS settings, which we adapted to the UPTD simulation environment, lead to a significant reduction in computation time without a clinically relevant loss of accuracy. The average CTV dose of reference MC simulations without VR are predicted within 0.3% accuracy for a head, lung, and prostate treatment field using VR. Hence, the framework is applicable to study larger patient cohorts. On the
HPC Hypnos5, the treatment field of a brain-tumor patient was simulated in about 2500 CPU hours using AMD Opteron 16-core.

These CPU times are required to reach a statistical dose uncertainty of 0.5% in a $1 \times 1 \times 2$ mm$^3$ CT voxel. The computation time can be further reduced if the convergence criterion is specified for larger voxel sizes. In routine proton treatment planning, dose grids up to $3 \times 3 \times 2$ mm$^3$ are used in particular for lung or prostate treatments. For such voxel extensions, the CPU time could be reduced by about a factor of 9 to reach a statistical dose uncertainty of 0.5% as it relates almost linearly to the particle number per voxel. For the verification of a typical clinical prostate field with a $3 \times 3 \times 2$ mm$^3$ planning grid, the CPU time would then be around 280 CPU hours. However, as the voxel-based analysis requires high regional precision, all MC calculations in this study were evaluated with the highest available spatial resolution defined by the CT Grid.

For retrospective studies, method M1 for absolute dose determination can further save computing time. Since absolute dose measurements are available for all irradiated patient fields in passive scattering at our institution (and probably most other centers), a separate water phantom simulation is not necessary. The reduction of the CPU time, especially for large fields (lung or prostate simulations) can be relevant. Method M1 conceptually
refers to the absolute dose measured prior to treatment for each patient field. Therefore, it directly adjusts the absolute dose to measured data circumventing the need for absolute dose model. However, method M1 introduces a dependency on patient-specific measurements to the radiation response modelling framework and a bias when comparing MC against TPS dose simulations. Since we want to provide a framework independent of individual patient measurements, which can also be directly used to verify TPS dose distributions, we used Method M2 in radiation response modelling.

In general, there is a probability that MR image changes considered as radionecrosis are caused by tumor progression (Walker et al. 2014). Albeit, based on the available histological and functional imaging data, in this study, the likelihood of tumor progression for the four selected patients is low.

Anatomical changes in the patient, especially after irradiation, remain a limitation as they complicate the assignment of the radiation field to the image change voxel. We used rigid co-registration because this is a standard procedure in our clinic for analyzing follow-up data. Also, differences between planned and delivered dose may occur. Inter-fractional patient positioning uncertainties may smear-out the dose and the LET distributions. However, for the proton therapy of head patients an inter-fractional setup uncertainty of about 1 mm can be achieved due to the usage of head fixation masks (Lowe et al. 2016). To roughly estimate the impact of image change detection uncertainties or patient setup uncertainties, a geometric robustness analysis was exemplarily performed for model $M_{III}$. The image change contours were shifted in eight spatial directions using vectors of $(\pm 1.0, \pm 1.0, \pm 1.0)$ mm. The value of 1.0 mm represented half the length of the largest voxel extension in follow-up MR images. For all shifts, the mean AUC values for model $M_{III}$ vary in the range of 0.84 to 0.89. Hence, the localization of the image changes influence the discrimination power of the model. However, for each shift the model remains accurate with an AUC value $\geq 0.80$.

Unlike image registration and contouring inaccuracies, uncertainties in dose and LET calculation can be more accurately quantified. An in-depth analysis of model uncertainties related to dose and LET calculation may help to improve the predictive power of the radiation response model. In particular, the prediction of the proton range could be improved by describing the patient geometry in the simulation using dual-energy CT data (Bär et al. 2018, Wohlfahrt et al. 2018) when it will be available for all patients.

The obtained correlation of image change with dose and increasing LET constitutes a clinical indication for a variable proton RBE for grade III glioma patients treated with adjuvant proton radio(chemo)therapy. Our implemented MC simulation model reliably characterized the radiation treatment fields in patients. It overcomes some limitations of the clinical TPS concerning dose calculation accuracy making the radiation response modelling more precise. The model’s ability to predict LET distributions for patients treated with passive scattering at UPTD enables systematic testing of the hypothesis of a variable local radiation response. With VR settings adapted to the UPTD simulation environment, dose and LET simulations can be performed in a reasonable time schedule. Hence, we encourage applying our response model to glioma patients outside this cohort to generate more clinical evidence and a better understanding of a variable proton RBE (Eulitz et al. 2019).

5. Conclusion

In summary, we set up a radiation-response modelling framework for analyzing clinical side effects to investigate RBE in patients receiving proton therapy at UPTD. The framework was applied to four glioma patients with morphological MR image changes within two years follow-up. We found a spatially variable dose response in the brain and a good prediction of the spatial distribution of the image changes when using both dose and LET as predictors. This may be indicative of a variable clinical RBE different from 1.1. Our work enables and encourages an in-depth patient study to generate evidence on the clinical RBE variability in proton therapy.

Acknowledgment

This research is funded by the European Social Fund and co-financed by tax funds based on the budget approved by the members of the Saxon State Parliament (Eulitz Jan, 100316833).

ORCID iDs

B Lutz https://orcid.org/0000-0003-3057-7831
A Lühr https://orcid.org/0000-0002-9450-6859
References

Agostinelli S et al 2003 Geant4—a simulation toolkit Nucl. Instrum. Methods Phys. Res. A 506 280–303
Bar E, Lalonde A, Zhang R, Lee K-W, Yang K, Sharp G, Liu B, Royle G, Bouchard H and Lu H-M 2018 Experimental validation of two dual-energy CT methods for proton therapy using heterogeneous tissue samples Med. Phys. 45 48–59
Baumann M, Krause M, Overgaard J, Debus J, Bentzen S M, Daart J, Richter C, Zips D and Bottfeld T 2016 Radiation oncology in the era of precision medicine Nat. Rev. Cancer 16 234–49
Chaudhary P et al 2014 Relative biological effectiveness variation along nonmonoenergetic and modulated Bragg peaks of a 62 MeV therapeutic proton beam: a preliminary assessment Int. J. Radiat. Oncol. Biol. Phys. 90 27–33
Dequesada I M, Quisling R G, Yachnis A and Friedman W A 2008 Can standard magnetic resonance imaging reliably distinguish recurrent tumor from radiation necrosis after radiosurgery for brain metastases? A radiographic-pathological study Neurosurgery 63 898–904
Eulitz J 2017 Commissioning of a model of the IBA universal nozzle in double scattering mode in the Monte–Carlo tool TOPAS at the University Proton Therapy Dresden Master Thesis (Dresden: Technical University Dresden)
Eulitz J et al 2019 Predicting late magnetic resonance image changes in glioma patients after proton therapy Acta Oncol. 58 1536–39
Giantosoudi D et al 2016 Incidence of CNS injury for a cohort of 111 patients treated with proton therapy for medulloblastoma: LET and RBE associations for areas of injury Int. J. Radiat. Oncol. Biol. Phys. 95 287–96
Gutin P H, Leibel S A and Sheline G E 1991 Radiation Injury to the Nervous System (New York: Raven Press)
Harrah S B, Bougoff N, Mohr A, Haberer T, Herfathet K, Combé S E, Debus J and Adebeg S 2016 Dosimetric advantages of proton therapy over conventional radiotherapy with photons in young patients and adults with low-grade glioma Strahlenther. Onkol. 192 759–69
Jarlskog C Z and Paganietti H 2008 Physics settings for using the Geant4 toolkit in proton therapy IEEE Trans. Nucl. Sci. 55 1018–25
Low E, Albertini F, Atkenhead A, Lomax A J and MacKay R I 2016 Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors Phys. Med. Biol. 61 413
Lühr A, von Neubeck C, Krause M and Troost E G C 2018a Relative biological effectiveness in proton beam therapy—current knowledge and future challenges Clin. Transl. Radiat. Oncol. 9 35–41
Lühr A et al 2018b ‘Radiobiology of proton therapy’: results of an international expert workshop Radiother. Oncol. 128 56–67
Magni S J R, Perl J, Facchini R, Schümann J and Paganietti H 2013 Geometrical splitting technique to improve the computational efficiency of Monte Carlo calculations for proton therapy Med. Phys. 40 041718
Méndez J R, Perl J, Schuemann J, Shin J, Paganietti H and Facchini R 2015 Improved efficiency in Monte Carlo simulation for passive-scattering proton therapy Phys. Med. Biol. 60 5019–35
Mohan R and Grosshans D 2017 Proton therapy—present and future Adv. Drug Deliv. Rev. 109 26–44
Mullins M E, Baret G D, Schaefer P W, Hochberg F H, Gonzalez R G and Lev M H 2005 Radiation necrosis versus glioma recurrence: conventional MR imaging clues to diagnosis Am. J. Neuroradiol. 26 1967–72
Pagani H 2009 Dose to water versus dose to medium in proton beam therapy Phys. Med. Biol. 54 4399
Pagani H 2006 Monte Carlo calculations for absolute dosimetry to determine machine outputs for proton therapy field Phys. Med. Biol. 51 2801–12
Pagani H 2014a Monte Carlo simulations will change the way we treat patients with proton beams today Br. J. Radiol. 87
Pagani H 2002 Nuclear interactions in proton therapy: dose and relative biological effect distributions originating from primary and secondary particles Phys. Med. Biol. 47 747–64
Pagani H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations Phys. Med. Biol. 57 899
Pagani H 2014b Relative biological effectiveness (RBE) values for proton beam therapy; Variations as a function of biological endpoint, dose, and linear energy transfer Phys. Med. Biol. 59 R419
Pagani H, Iang H, Parodi K, Slopesma R and Engelsman M 2008 Clinical implementation of full Monte Carlo dose calculation in proton beam therapy Phys. Med. Biol. 53 4825
Pedregosa F et al 2011 Scikit-learn: machine learning in python J. Mach. Learn. Res. 12 2825–30
Peleer C R, Mirkovic D, Titt U, Blanchard P, Gunther J R, Mahajan A, Mohan R and Grosshans D R 2016 Clinical evidence of variable proton biological effectiveness in pediatric patients treated for ependymoma Radiother. Oncol. 121 395–401
Perl J, Shin J, Schümann J, Facchini R and Paganietti H 2012 TOPAS: an innovative proton Monte Carlo platform for research and clinical applications Med. Phys. 39 6818–37
Saager M, Peschke P, Borns S, Debus J and Karger C P 2018 Determination of the proton RBE in the rat spinal cord: is there an increase towards the end of the spread-out Bragg peak? Radiother. Oncol. 128 115–20
Sawakuchi G O, Titt U, Mirkovic D and Mohan R 2008 Density heterogeneities and the influence of multiple Coulomb and nuclear scatterings on the Bragg peak distal edge of proton therapy beams Phys. Med. Biol. 53 4605
Schneider W, Bottfeld T and Schlegel W 2000 Correlation between CT number and tissue parameters needed for Monte Carlo simulations of clinical dose distributions Phys. Med. Biol. 45 459
Schuemann J, Giantosoudi D, Grassberger C, Motiebadd M, Min C H and Paganietti H 2015 Assessing the clinical impact of approximations in analytical dose calculations for proton therapy Int. J. Radiat. Oncol. Biol. Phys. 92 1157–64
Takami Y, Taiyama Y, Shibuya Y, Higashiya S, Takami T and Ohata K 2008 Diagnostic accuracy of [11C]-Methionine PET for differentiation of recurrent brain tumors from radiation necrosis after radiotherapy J. Nucl. Med. 49 694–5
Underwood T S A, Grassberger C, Bass R, MacDonald S M, Meyerson N M, Yeap B Y, Jenner B R and Paganietti H 2018 Asymptomatic late-phase radiographic changes among chest-wall patients are associated with a proton RBE exceeding 1.1 Int. J. Radiat. Oncol. 101 809–19
Verburg J M, Grassberger C, Dowell S, Schuemann J, Seco J and Paganietti H 2016 Automated Monte Carlo simulation of proton therapy treatment plans Technol. Cancer Res. Treat. 15 NP35–46
Walker A J, Ruzevick J, Malayeri A A, Rigamonti D, Lim M, Redmond K J and Kleinberg I 2014 Postradiation imaging changes in the CNS: how can we differentiate between treatment effect and disease progression? Future Oncol. 10 1277–97
Wohlhafert P, Mohler C, Hetschold V, Menkel S, Greilisch S, Krause M, Baumann M, Enghardt W and Richter C 2017 Clinical implementation of dual-energy CT for proton treatment planning on pseudo-nonmonoenergetic beams J. Radiat. Oncol. Biol. Phys. 97 427–34
Wohlhafert P, Mohler C, Richter C and Greilisch S 2018 Evaluation of stopping-power prediction by dual- and single-energy computed tomography in an anthropomorphic ground-truth phantom Int. J. Radiat. Oncol. 100 244–53
Zschaech S et al 2016 PRONTOX—proton therapy to reduce acute normal tissue toxicity in locally advanced non-small-cell lung carcinomas (NSCLC): study protocol for a randomised controlled trial Trials 17 543