NTCP modelling for high-grade temporal radionecrosis in a large cohort of patients receiving pencil beam scanning proton therapy for skull base and head and neck tumors

C Schröder MD , A Köthe MSc , C De Angelis MSc , L Basler MD , G Fattori PhD , S Safai PhD , D Leiser MD , AJ Lomax PhD , DC Weber MD

PII: S0360-3016(22)00093-1
DOI: https://doi.org/10.1016/j.ijrobp.2022.01.047
Reference: ROB 27453

To appear in: *International Journal of Radiation Oncology, Biology, Physics*

Received date: 19 May 2021
Revised date: 4 January 2022
Accepted date: 26 January 2022

Please cite this article as: C Schröder MD , A Köthe MSc , C De Angelis MSc , L Basler MD , G Fattori PhD , S Safai PhD , D Leiser MD , AJ Lomax PhD , DC Weber MD , NTCP modelling for high-grade temporal radionecrosis in a large cohort of patients receiving pencil beam scanning proton therapy for skull base and head and neck tumors, *International Journal of Radiation Oncology, Biology, Physics* (2022), doi: https://doi.org/10.1016/j.ijrobp.2022.01.047

NTCP modelling for high-grade temporal radionecrosis in a large cohort of patients receiving pencil beam scanning proton therapy for skull base and head and neck tumors

Schröder C, MD[1,2,*,$]christina.schroeder@ksw.ch christina.schroeder@ksw.ch; Köthe A[1,3*], MSc; De Angelis C[1], MSc; Basler L[1], MD; Fattori G[1], PhD; Safai S[1], PhD; Leiser D[1], MD; Lomax AJ[1,3], PhD; Weber DC, MD [1,4,5]

[1]Center for Proton Therapy, Paul Scherrer Institute, Villigen, Switzerland
[2]Institute for Radiation Oncology, Cantonal Hospital Winterthur (KSW), Winterthur, Switzerland
[3]ETH, Department of Physics, Zürich, Switzerland
[4]University Hospital Zürich, Zürich, Switzerland
[5]University Hospital of Bern, Inselspital, University of Bern, Bern, Switzerland

[*]These authors contributed equally to this analysis.

[$]Corresponding author**:** Damien C. Weber, MD, Center for Proton Therapy, Paul Scherrer Institute, 5232 Villigen PSI West, Switzerland
Christina Schröder, MD, Institute for Radiation Oncology, Cantonal Hospital Winterthur Winterthur, Switzerland

Abstract

Purpose/Objectives: To develop a normal tissue complication probability (NTCP) model including clinical and dosimetric parameters for high-grade temporal lobe radionecroses (TRN) after pencil beam scanning (PBS) proton therapy (PT).

Materials/Methods: Data of 299 patients with skull base and Head and Neck tumors treated with PBS PT with a total dose of ≥60 Gy_{RBE} from 05/2004–11/2018 were included. Patients with a ≥ grade (G) 2 TRN (CTCAE v5.0 criteria) were considered as having a high-grade TRN. Nine clinical and 27 dosimetric parameters were considered for structure-wise modelling. After elimination of strongly cross-correlated variables, logistic regression models were generated using penalized LASSO regression. Bootstrapping was performed to assess parameter selection robustness. Model performance was evaluated via cross-correlation by assessing the area under the curve of receiver operating characteristic curves (AUC-ROC) and calibration with a Hosmer-Lemeshow test statistic.

Results: After a median radiological follow-up of 51.5 months (range, 4–190), 27 (9%) patients developed a ≥ G2 TRN. Eleven patients had bitemporal necrosis, resulting in 38 events in 598 temporal lobes for structure-wise analysis. During Bootstrapping analysis, the highest selection frequency was found for prescription dose (PD), followed by Age, V_{40Gy}[%], Hypertension (HBP) and D_{1cc}[Gy]. During cross validation Age*PD* D_{1cc}[Gy]*HBP was superior in all described test statistics. Full cohort structure wise and patient wise models were built with a maximum AUC-ROC of 0.79 (structure-wise) and 0.76 (patient-wise).

Conclusion: While developing a logistic regression NTCP model to predict ≥ G2 TRN, the best fit was found for the model containing Age, PD, D_{1cc}[Gy] and HBP as risk factors. External validation will be the next step to improve generalizability and potential introduction into clinical routine.

Key words

brain radiation necrosis, proton therapy, pencil beam scanning, Normal Tissue Complication Probability, Logistic regression modelling, skull base tumors, head and neck tumors.

1 Introduction

Radionecroses (RN) are well-documented side effects after high-dose radiation to the brain that can occur in a considerable number of patients, especially for those with radio-resistant tumors in vicinity of critical structures, such as skull-base chordoma, chondrosarcoma or cancers of the head and neck region (1-17). For these tumors, the temporal lobes are especially at risk given the close anatomical proximity of the target volumes. The incidence of temporal RN (TRN) reported in more recent series including both photon and proton treatment of skull base or otorhinolaryngology (ORL) cancers ranges from 2.3 – 17.1% (1-14, 17). For proton specific data the incidence of TRN ranges from 9.7 – 17.1% (1, 3, 6, 7, 17). For Carbon Ions, an incidence of up to 65 % has been observed, although the total number of patients in these analyses was low (8, 15, 16). Most of the reported data refer to RN of any grade with most authors using the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) criteria. Grade 1 (G1) toxicities, defined as asymptomatic lesions not requiring interventions, are the most common type of RN. Those lesions may be followed-up and are usually completely regressive over time (18). Radiation necrosis of a higher grade needing at least corticosteroid treatment appear in up to 6% of patients (3, 6, 9). RN usually appear after a median of 17 – 34 months but may occur for at least up to 5 years after radiotherapy (1-16).

Whereas G1 RN are often without consequences for the patients, higher grade RN can severely affect a patient's well-being and quality of life (QoL). In tumors requiring a high radiation dose, it may not always be possible to avoid risking small, asymptomatic RN to ensure tumor control. However, high-grade RN may permanently impair the QoL of the patient. Most existing analysis regarding risk factors for RN have focused on RN of any grade with proton specific data being generally scarce.

In this analysis, we therefore aim to develop a logistic regression model combining clinical and dosimetrical factors to predict high-grade RN in a large cohort of patients treated with proton therapy (PT) and assess parameters associated with an increased risk of developing high-grade RN.

2 Methods and Material

2.1 Patient and treatment characteristics

The institutional database was queried to identify patients with skull-base or ORL tumors treated with pencil beam scanning (PBS) proton therapy (PT) with a minimal total prescription dose of > 60 Gy_{RBE} in conventional fractionation and curative intent at initial presentation. The analysis was approved by the local ethics committee of "XXX, Anonymized for Review". All patients had a follow-up period of at least 4 months. Patients with mixed photon/proton treatment plans, re-irradiated patients and patients with insufficient follow-up were excluded from this analysis. Two hundred and ninety-nine patients treated between 05/2004 – 11/2018 were identified.

Out of the 299 patients, the majority presented with skull-base chordoma (n = 184; 61.5 %), and skull-base chondrosarcoma (n = 73; 24.4 %). The ORL cohort consisted of adenoid cystic carcinoma (n = 25, 8.4 %), nasopharyngeal carcinoma (NPC) (n = 6, 2 %), carcinomas of the nasal cavity/paranasal sinuses (n = 6, 2 %) and other head and neck primaries (n = 4, 1.3 %). Patients treated for chordoma received a median total treatment dose of 74 Gy_{RBE} (range, 70 – 75.6 Gy_{RBE}), those treated for chondrosarcoma 70 Gy_{RBE} (range, 67.4 – 76 Gy_{RBE}) and patients with head and neck tumors 70 Gy_{RBE} (range, 66 – 77.4 Gy_{RBE}). All patients underwent normofractionated treatments (1.8-2.2 Gy_{RBE}/fraction). The majority of patients with ORL tumors (n = 23, 60.5 %) received a total treatment dose of 66 – 70 Gy. In 15 patients (39.5 %), a dose escalated treatment was done with 9 of those patients (23.7 %)

receiving a total treatment of up to 74 Gy and 6 patients (15.8 %) receiving a total treatment dose of up to 77.4 Gy. The mean volume of the irradiated high-dose volume was 97.1 cc (standard deviation (SD) 102.5 cc). Patient age ranged from 2 to 84 years (median, 45) with 24 patients being under the age of 18. Additional patient and treatment characteristics are detailed in Table 1.

2.2 Proton therapy planning and delivery

For treatment planning, patients received a planning CT (2 mm slice thickness) with customized immobilization (bite block or thermoplastic mask). If indicated, patients received an additional CT with i.v. contrast agents with immobilization along with the planning CT (without i.v. contrast agents). Additionally, MRI (T1, T1 with contrast, T2, FLAIR and diffusion weighted for more recent treatments) and/or PET-CT scans were added for target definition if applicable. Gross tumor volume (GTV, if applicable) and/or clinical target volume (CTV) were contoured according to the tumor specific guidelines. A planning target volume (PTV) margin was added to account for setup uncertainties (depending on treatment region and immobilization technique, cranial/head and neck 4 – 5 mm). Organs at risk were delineated on CT and MRI according to standard delineation atlases (20). For delineation of the temporal lobes, the planning CT was registered with the MRI through a case-specific region of interest, followed by visual assessment and manual refinement. The anatomy of the temporal lobes has been well described and there are comprehensive contouring guideline for organs at risk in this anatomical region (19, 20). Treatment planning was done using the in-house developed treatment planning system ("XXX, Anonymized for Review") with a dose calculation grid size of 4-5 mm. For the anatomical sites of patients treated in this cohort, the typically assumed 3% range uncertainty for protons is also adequately compensated using our defined PTV margins. A relative biological effectiveness (RBE) compared to megavoltage photon therapy of 1.1 was used for treatment planning. PT was delivered using the pencil

beam scanning (PBS) technique for all patients. Daily kV/kV imaging and if necessary additional 3D positioning using the remote positioning system (until 2018) (21) or in-room CT (since 2013) was used for patient setup verification. Weekly control CTs were performed to monitor possible changes in the sinuses. If necessary, the treatment plan was adapted accordingly.

2.3 Scoring of toxicity and calculation of outcome parameters

RN were defined as contrast enhanced lesions on T1- and/or edema on T2-weighted sequences (6, 22, 23). Figure 1 shows an example of axial T1 MRI slices with contrast for three different patients. RN were scored according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) v5.0 criteria. Grade 2 (G2) RN was defined as causing moderate symptoms, indication for corticosteroids and G3 as causing severe symptoms, indication for medical interventions (e.g. surgery). G4 necrosis is defined as life-threatening with an indication for urgent intervention. The time to the first appearance of the RN was calculated from the start of the proton therapy.

2.4 Clinical and dosimetric parameters used logistic regression modelling

In total, 11 clinical and 26 dosimetric parameters were initially considered for this analysis. The clinical parameters were location of primary tumor (skull base/ORL), prescription dose, volume of the temporal lobe (cc), volume of the PTV (cc), diabetes mellitus (DM), hypertensive blood pressure (HBP), dyslipidemia (DL), smoking, sex, age at the start of proton therapy and use of any chemotherapy. Location of primary tumor, DM, HBP, DL smoking, sex and the use of chemotherapy were binary parameters, whereas prescription dose, volume of the temporal lobe and the age at the start of proton therapy were treated as continuous variables.

The dosimetric parameters (for temporal lobe) used were minimum dose received by X cc of the volume in Gy_{RBE} and percent (D0.5cc, D1cc and D2cc) as well as the minimum (Dmin), mean (Dmean) and maximum (Dmax) dose in the volume in Gy_{RBE} and percent. Furthermore, the volume receiving at least a dose of X Gy_{RBE} (VXGy) in percent and cc was analysed in 10 Gy bins ($V10Gy_{RBE}$ – $V70Gy_{RBE}$). A summary of these dosimetric values in the cohort can be found in table S2 in the supplement.

2.5 Statistical analysis: Logistic regression modelling

Structure-wise logistic regression models for ≥ G2 TRN were developed on a structure-wise basis. In a first step, all possible model parameters were evaluated against every other based on the Spearman's ranked correlation coefficient (SRCC) (24). Parameters with a SRCC of > 0.8 were considered strongly cross-correlated. To avoid bias from strongly correlated parameters in the following bootstrapping selection procedure, for each group of strongly correlated parameters, one "representative" was selected for further analysis based on correlation to outcome (TRN) and redundant variables removed. This process is visualized in more detail in figure S1 in the supplemental material.

To assess each parameter's selection robustness, 2000 bootstrap samples of the dataset were created and logistic regression models for ≥ G2 TRN were generated based on least absolute shrinkage and selection operator (LASSO). Models were fit on all patients with complete data (complete-case analysis). The selected regularization parameter for each bootstrap sample was the minimal parameter that led to a 4-parameter model. Based on the parameter selection frequency, logistic models were built including 1 to 5 of the most relevant parameters from the bootstrap analysis.

Additionally, due to the imbalanced nature of the data the Akaike information criterion (AIC) (26) and the Bayes information criterion (BIC) (27) were calculated for all those logistic regression models to allow the selection of a reasonable number of model parameters.

For model performance evaluation, structure wise cross validation (leave-one-out (LOO), 15-fold cross validation) was performed. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were calculated to elaborate on the classification ability. For evaluation of the model calibration and goodness of fit, a Hosmer-Lemeshow (HL) test statistic and calibration curve were calculated (28). For this, patients were grouped into 10 bins containing an equal number of patients. Additionally, binary cross-entropy (CE = -Loglikelihood/number of samples) was calculated to compare the likelihood of our data fitting the different models. Additionally, full cohort, structure wise and patient wise predictions were calculated for the models including the mentioned parameters. NTCP was calculated using the following formula:

$$NTCP_{TRN} = (1 + e^{-z})^{-1} \text{ with}$$

$$z = \text{constant} + \beta_1 * x_1 + \beta_2 * x_2 + \ldots + \beta_n * x_n$$

The patient wise prediction was calculated as (29):

$$NTCP_{TRN/patient} = 1 - (1 - NTCP_{TRN/TLright}) (1 - NTCP_{TRN/TL\ left})$$

After selection of the most robust set of parameters from the bootstrapping and cross-validation procedure including performance assessment, the final model was fit by penalized regression (LASSO) to reduce the risk for overfitting. The optimal regularization parameter was selected by 10-fold cross-validation and deviance minimization. Confidence intervals were estimated with un-penalized logistic regression on bootstrap samples. For statistical

analyses, IBM SPSS statistics 25 (Statistical Package for Social Sciences, SPSS, IBM Corp., Armonk, NY) and MATLAB® R2020a (Mathworks Inc, Natick, MA, USA) were used.

3 Results

3.1 Incidence and timing of temporal radionecrosis

The median clinical follow-up of all patients was 60.5 months (range, 11 – 190) with a median radiological follow-up of 51.5 months (range, 4 – 190).

A total of 75 patients developed a RN of any grade of the brain tissue after proton therapy. Twelve patients (4 %) developed a non-temporal necrosis of any grade of the distant brain only, whereas 63 patients (21.1 %) developed a TRN of any grade. Of those, 27 (9 %) patients developed a ≥ G2 RN with 17 patients (5.7 %) having a G2 and 10 patients (3.3 %) a G3 TRN. There were no G4 or G5 TRN. The median time to any grade TRN was 23 months (range 6 – 74 months) with a median time to ≥ G2 TRN of 20 months. The 1-, 3- and 5-year rates of ≥ G2 TRN were 1.3 %, 8.0 % and 9.0 %.

Of the 27 patients with ≥ G2 TRN, 11 patients (40 %) developed a bitemporal RN, 10 (37 %) patients a RN of the left temporal lobe and 6 (23 %) patients of the right temporal lobe. For the present analysis, the temporal lobes of each patient were analysed separately, therefore resulting in 38 ≥ G2 TRN in 598 temporal lobes (6.4 %) that were used for the analysis below.

3.2 Modelling

3.2.1 Multicollinearity parameter removal

For clinical parameters, there were no cross-correlated parameters with either other clinical or dosimetrical parameters with a SRCC > 0.8, so all clinical parameters were included in the logistic regression modelling. The strongest correlation with the incidence of ≥ G2 TRN were found for Age (R = 0.160), HBP (R = 0.137) and prescription dose (PD, R = 0.118).

As for dosimetric parameters, the strongest correlation with the incidence of ≥ G2 TRN was seen for D1cc (Gy) (R = 0.199), however, there were numerous strong cross-correlations between the dosimetric parameters that consequently had to be removed from the analysis. A total of 6 dosimetric parameters were finally included in the logistic regression modelling. These parameters are D1cc (Gy), Dmax (%), V40Gy (%), V20Gy (cc), Dmin (Gy) and V10Gy (%). This parameter selection process is visualized in figures S1 and S2 in the supplement.

3.2.2 Bootstrapping and parameter selection

During the bootstrapping analysis the highest selection frequency (SF) was found for prescription dose (SF 96.9 %), followed by Age (SF 89.3 %), V_{40Gy}[%] (SF 50.3 %), HBP (SF 48.7 %) and D_{1cc} [Gy] (SF 47.4 %). Figure S3 in the supplement shows the selection frequency in the 2000 bootstrapped samples.

Both the AIC and BIC continued to decrease with an increasing amount of parameters used in the model, but changes in BIC were negligible for more than 4 parameters, indicating that models with two to four parameters were appropriate given the statistical power of our dataset and to avoid overfitting. Based on that, models with two to five parameters were chosen for further evaluation, including the parameters prescription dose, age, V_{40Gy}[%], HBP and D_{1cc}[Gy].

3.2.3 Performance evaluation

Cross validation results on all model combinations showed that the model including Age*PD*D1cc*HBP was superior in all considered test statistics. It showed an AUC-ROC of 0.76, a cross entropy of 0.22 and a not significant HL p-value (0.18) suggesting that calibration as well as classification ability of a model built on these parameters are promising on unknown data. A more detailed overview over the results of the structure wise 15-fold cross validation process are detailed in table S1 in the supplement.

Full cohort models were built based on this set of model parameters which resulted in good classification ability for the patient-wise (AUC-ROC 0.79) and the structure-wise model (0.76). This is visualized in Figure 2. Furthermore, we observed good calibration on our dataset confirming that model predictions reflect the probability of TRN incidence appropriately compared to the actually observed incidence rate. While the structure-wise model showed excellent calibration (slope 1.13, HL p-value 0.66), the patient-wise model showed a slight overestimation of events (slope 0.8) which could be explained by the model being fit at a structure-level and then expanded to a patient-based approach, likely introducing some uncertainty. Calibration curves are visualized in Figure 3. Final model parameters, associated 90% confidence intervals as well as the performance evaluation on the final models are summarized in table 2.

4 Discussion

TRN is an adverse event that may occur in a substantial number of patients treated with radiotherapy which is observed in 2-17% of patients treated for ORL and skull base tumors (1-14, 17). Data for proton therapy shows a higher overall incidence of TRN, ranging from 9.7% – 17.1% (1, 8, 13, 17) compared to photon therapy with 2.3% - 15%) (2, 5, 11, 12, 14). In our analysis, the incidence of any-grade TRN of 20.7 % and 9 % for ≥ G2 TRN is slightly higher than published data for other proton therapy series (1, 3, 6, 7, 17). These observed

discrepant numbers may have several explanations. First, the median treatment dose in our cohort was 74 Gy_{RBE} and therefore exceeding that of most cited papers for both proton and photon therapy (1-5, 7, 8, 10-14). Only Mc Donald et al reported a higher median treatment dose of 75.6 Gy_{RBE} (1, 3, 6, 7). Additionally, the median radiological follow-up in this cohort was 51.5 months, therefore exceeding that of most published data for proton therapy with 20 – 38 months (1, 6-8, 17).

The majority of the reported TRN are CTCAE asymptomatic grade 1 TRN, not requiring further medical intervention. To our knowledge, the majority of existing data on TRN risk factors are based on TRN of any grade (2, 4-6, 8, 10-12, 14). However, skull base tumors like chordoma and chondrosarcomas are radio-resistant and require a high total treatment dose. According to Basler et al., a treatment dose of below 66 Gy to the GTV for skull base chordoma and chondrosarcoma results in a significant increase in local failure (30). Strongly constraining the dose to the temporal lobes in order to avoid asymptomatic TRN might therefore result in an impaired tumor control probability for these tumors. However, ≥ G2 RN can severely impair the QoL of a patient, either due to the necrosis itself or required interventions. Because of this clinical relevance together with a potential bias from undetected G1 TRN events, we decided to develop a logistic regression NTCP model that focused on ≥ G2 TRN. We therefore developed a logistic regression NTCP model that focused on ≥ G2 TRN.

Given the imbalanced nature of this cohort and the low percentage of patients with an event, a large focus during the modelling process was on the robustness of the final model. Therefore, this comprehensive statistical algorithm was used to ensure a robust parameter selection and model. The model providing the best fit in this analysis includes (Fig. 1-3) both clinical (Age, HBP) and dosimetric risk factors (prescription dose, D1cc (Gy) of the temporal

lobes). Numerous clinical and dosimetric risk factors have been identified for any grade TRN in published analyses. Notably, data on proton therapy-only cohorts are scarce. Kitpanit et al. performed an analysis including 234 patients with head and neck tumors in which they did not find a correlation between clinical parameters and TRN. However, this group has found a number of dosimetric parameters that were associated with a higher risk of TRN with the highest AUC for absolute volume V50Gy and D2cc, suggesting a role of high radiation dose delivered to the temporal lobes (3). However, their cutoff for D2cc for any grade RN was ≤62 Gy_{RBE}, which will be difficult to achieve in clinical practice if a cohort includes very radio-resistant tumors like chordoma. Factors associated with high dose areas like Dmax or D1cc were also identified in several photon-specific analysis, suggesting the high dose to a small volume being at least a contributing factor (2, 12), as well as in a joint analysis of protons and carbon ion therapy (17). Other dosimetric factors include Dmean (Gy) (4, 5) and dose-volume parameters of the temporal lobes like V30 – V70Gy (6, 11, 14). As for clinical parameters, age was identified as a risk factor in an analysis by Lee et al in head and neck cancer patients treated with 3D photon radiation therapy (4). Other clinical risk factors include concurrent systemic therapy (4, 12, 14), male sex and race (10, 13).

We would also like to discuss why we opted for a structure-wise modelling strategy, which assumes an independence of the temporal lobes. While clinical parameters, such as patient age, hypertension or sex apply to both temporal lobes equally, there is a difference in dosimetry and thus also in risk for RN which is different for each structure. While we acknowledge the possibility that there is an interplay between the temporal lobes, a lack of data in the literature in favor of that theory combined with the majority of papers correlating dose to necrosis on a structure-wise basis led us to perform the modeling at a structure level (3, 6, 8, 14).

The limitations of this analysis lie in its retrospective nature. Although patients with insufficient follow-up were excluded from this analysis, a minimal underestimation of events cannot be excluded. However, follow-up data was comprehensively collected by the department's study office and regularly discussed in a multidisciplinary team. Also, the median follow-up is exceeding that of most published data. Also, only patients without primary brain tumors and a total treatment dose of more than 60 Gy_{RBE} were included in this cohort. Therefore, the risk is estimated to be minimal. Another limitation lies in the lack of external validation to date. External validation will be necessary to confirm the results in a different cohort and therefore enable general application of this model as well as introduction into clinical routine.

Finally, we would like to propose suggestions, how this model could be integrated into a clinical workflow to support decision making for treatment planning or patient counseling. While the selection of risk factors already gives a hint whether a patient would need to be considered at high-risk for TRN, the estimation of the concrete patient risk could be further used for comprehensive patient counseling. Quantification of risk for a side effect is a concept easier to understand for a patient compared to any dose metric. In a next step, the model predictions – in combination with a dedicated threshold – could be used to assign patients to a low- or high-risk group. High risk group patients could then have a further planning optimization step, whose goal for example could be the reduction of D_1 to the temporal lobes or even LET-optimized planning. While based on our data we would suggest a threshold of 7.5 % for a structure-wise estimation (specificity (70%) and sensitivity (79%)), the selection of this threshold could be performed at the institutional level if further emphasis on either classification sensitivity or specificity is desired.

Conclusion

While developing logistic regression NTCP models to predict ≥ G2 TRN, both clinical and dosimetric parameters proved to be risk factors. The best fit was found for the model containing Age, PD, D1cc (Gy) and HBP as risk factors. External validation would be the next step to improve generalizability and potential introduction into clinical routine, potentially allowing for patient-specific planning to minimize high grade TRN for high-risk patients.

Funding Statement

This work was partially supported by the grant #2018-223 of the Strategic Focal Area "Personalized Health and Related Technologies (PHRT) of the ETH Domain

Data availability statement

Research data are not available at this time due to sensible patient information.

Conflicts of Interest

None

Acknowledgements

We would like to thank Prof. Daniel R. Zwahlen (KSW) for reviewing the manuscript.

References

1. Engeseth GM, Mohamed AS, Stieb S, Fuller CD, Garden AS, Rosenthal DI, et al. Radiation Associated Brain Necrosis following Proton Therapy for Head and Neck Skull Base and Intracranial Tumors. International Journal of Radiation Oncology Biology Physics. 2019;105(1):S5-S6.

2. Huang J, Kong FF, Oei RW, Zhai RP, Hu CS, Ying HM. Dosimetric predictors of temporal lobe injury after intensity-modulated radiotherapy for T4 nasopharyngeal carcinoma: a competing risk study. Radiat Oncol. 2019;14(1):31.

3. Kitpanit S, Lee A, Pitter KL, Fan D, Chow JCH, Neal B, et al. Temporal Lobe Necrosis in Head and Neck Cancer Patients after Proton Therapy to the Skull Base. Int J Part Ther. 2020;6(4):17-28.

4. Lee AW, Ng WT, Hung WM, Choi CW, Tung R, Ling YH, et al. Major late toxicities after conformal radiotherapy for nasopharyngeal carcinoma-patient- and treatment-related risk factors. Int J Radiat Oncol Biol Phys. 2009;73(4):1121-8.

5. Lu L, Sheng Y, Zhang G, Li Y, OuYang PY, Ge Y, et al. Temporal lobe injury patterns following intensity modulated radiotherapy in a large cohort of nasopharyngeal carcinoma patients. Oral Oncol. 2018;85:8-14.

6. McDonald MW, Linton OR, Calley CS. Dose-volume relationships associated with temporal lobe radiation necrosis after skull base proton beam therapy. Int J Radiat Oncol Biol Phys. 2015;91(2):261-7.

7. Merchant TE, Hua CH, Sabin ND, Ezell SE, Madey MA, Wu S, et al. Necrosis, Vasculopathy, and Neurological Complications After Proton Therapy for Childhood Craniopharyngioma: Results From a Prospective Trial and a Photon Cohort Comparison. International Journal of Radiation Oncology Biology Physics. 2016;96(2):S120-S1.

8. Miyawaki D, Murakami M, Demizu Y, Sasaki R, Niwa Y, Terashima K, et al. Brain injury after proton therapy or carbon ion therapy for head-and-neck cancer and skull base tumors. Int J Radiat Oncol Biol Phys. 2009;75(2):378-84.

9. Pehlivan B, Ares C, Lomax AJ, Stadelmann O, Goitein G, Timmermann B, et al. Temporal lobe toxicity analysis after proton radiation therapy for skull base tumors. Int J Radiat Oncol Biol Phys. 2012;83(5):1432-40.

10. Santoni R, Liebsch N, Finkelstein DM, Hug E, Hanssens P, Goitein M, et al. Temporal lobe (TL) damage following surgery and high-dose photon and proton irradiation in 96 patients affected by chordomas and chondrosarcomas of the base of the skull. Int J Radiat Oncol Biol Phys. 1998;41(1):59-68.

11. Su SF, Huang SM, Han F, Huang Y, Chen CY, Xiao WW, et al. Analysis of dosimetric factors associated with temporal lobe necrosis (TLN) in patients with nasopharyngeal carcinoma (NPC) after intensity modulated radiotherapy. Radiat Oncol. 2013;8:17.

12. Zeng L, Tian YM, Sun XM, Chen CY, Han F, Xiao WW, et al. Late toxicities after intensity-modulated radiotherapy for nasopharyngeal carcinoma: patient and treatment-related risk factors. Br J Cancer. 2014;110(1):49-54.

13. Zhang Y, Huo W, Adams JA, Sanford NN, Lam MB, Lu Y, et al. Temporal Lobe Necrosis After Proton for Nasopharyngeal Carcinoma: Predictive Factors and Clinical RBE Estimation. International Journal of Radiation Oncology Biology Physics. 2017;99(2):E386-E.

14. Zhou X, Ou X, Xu T, Wang X, Shen C, Ding J, et al. Effect of dosimetric factors on occurrence and volume of temporal lobe necrosis following intensity modulated radiation therapy for nasopharyngeal carcinoma: a case-control study. Int J Radiat Oncol Biol Phys. 2014;90(2):261-9.

15. Hagiwara Y, Koto M, Bhattacharyya T, Hayashi K, Ikawa H, Nemoto K, et al. Long-term outcomes and toxicities of carbon-ion radiotherapy in malignant tumors of the sphenoid sinus. Head Neck. 2020;42(1):50-8.

16. Kishimoto R, Mizoe JE, Komatsu S, Kandatsu S, Obata T, Tsujii H. MR imaging of brain injury induced by carbon ion radiotherapy for head and neck tumors. Magn Reson Med Sci. 2005;4(4):159-64.

17. Gillmann C, Lomax AJ, Weber DC, Jakel O, Karger CP. Dose-response curves for MRI-detected radiation-induced temporal lobe reactions in patients after proton and carbon ion therapy: Does the same RBE-weighted dose lead to the same biological effect? Radiother Oncol. 2018;128(1):109-14.

18. Rahmathulla G, Marko NF, Weil RJ. Cerebral radiation necrosis: a review of the pathobiology, diagnosis and management considerations. J Clin Neurosci. 2013;20(4):485-502.

19. Kiernan JA. Anatomy of the temporal lobe. Epilepsy Res Treat. 2012;2012:176157.

20. Sun Y, Yu XL, Luo W, Lee AW, Wee JT, Lee N, et al. Recommendation for a contouring method and atlas of organs at risk in nasopharyngeal carcinoma patients receiving intensity-modulated radiotherapy. Radiother Oncol. 2014;110(3):390-7.

21. XXX, Anonymized for Review

22. Shah R, Vattoth S, Jacob R, Manzil FF, O'Malley JP, Borghei P, et al. Radiation necrosis in the brain: imaging features and differentiation from tumor recurrence. Radiographics. 2012;32(5):1343-59.

23. Wang YX, King AD, Zhou H, Leung SF, Abrigo J, Chan YL, et al. Evolution of radiation-induced brain injury: MR imaging-based study. Radiology. 2010;254(1):210-8.

24. Spearman C. The proof and measurement of association between two things. American Journal of Psychology. 1904;15:72-101.

25. Efron B. 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife. Annals of Statistics. 1979;7(1):1-26.

26. Akaike H. Citation Classic - a New Look at the Statistical-Model Identification. Current Contents/Engineering Technology & Applied Sciences. 1981(51):22-.

27. Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978;6(2):461-4.

28. Hosmer DW, Lemeshow S. Goodness of Fit Tests for the Multiple Logistic Regression-Model. Communications in Statistics Part a-Theory and Methods. 1980;9(10):1043-69.

29. Kothe A, van Luijk P, Safai S, Kountouri M, Lomax AJ, Weber DC, et al. Combining Clinical and Dosimetric Features in a PBS Proton Therapy Cohort to Develop a NTCP Model for Radiation-Induced Optic Neuropathy. Int J Radiat Oncol Biol Phys. 2021.

30. Basler L, Poel R, Schroder C, Bolsi A, Lomax A, Tanadini-Lang S, et al. Dosimetric analysis of local failures in skull-base chordoma and chondrosarcoma following pencil beam scanning proton therapy. Radiat Oncol. 2020;15(1):266.

Figure and Table captions

Tables:

Table 1 Patient and treatment characteristics

Table 2 AUC-ROC, HL p-value and CE values for full cohort models (structure wise and patient wise) and model parameters with regression coefficient of the best performing model A*PD*D1cc*HBP

Figures:

Figure 1 Axial T1 with contrast MRI of three different patients treated with proton therapy showing uni- and bilateral radiation necrosis (red arrows) of the temporal lobe

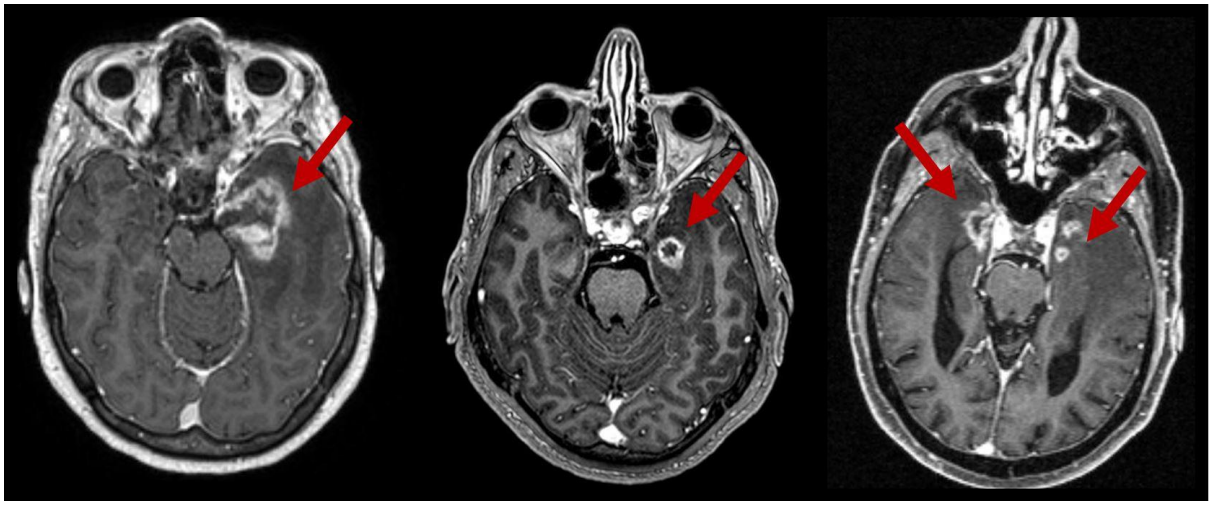
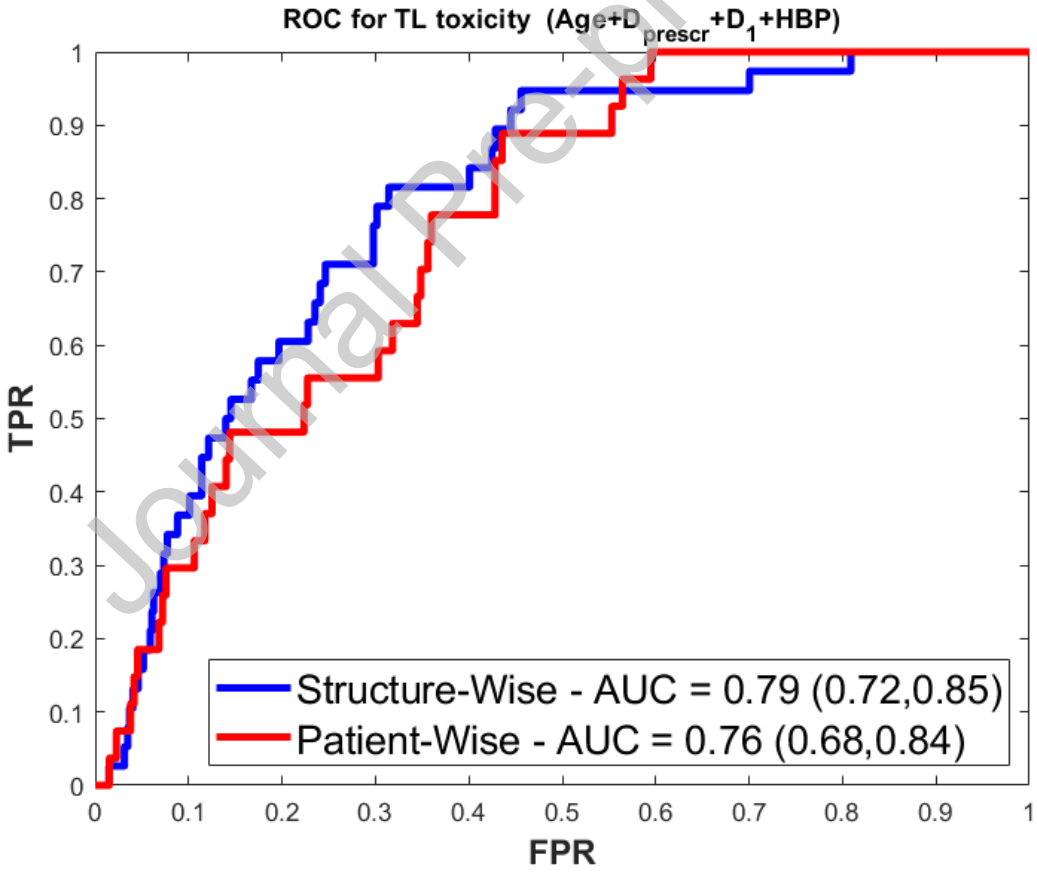


Figure 2 Structure wise and patient wise ROC curves for the model containing Age*PD*D1cc*HBP



Figures 3 a) and b) Structure wise and patient wise calibration curves of the logistic regression model including Age*PD*D1cc*HBP

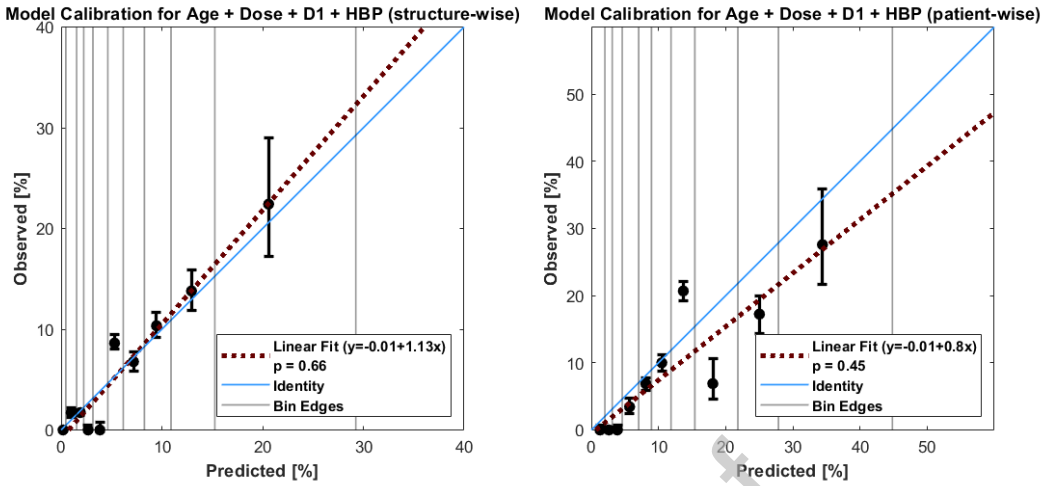


Table 1 Patient and treatment characteristics

		n	%
Sex	female	144	48
	male	155	52
Localization	Skull base	261	87
	ORL	38	13
RT at initial presentation	yes	242	81
	no	57	19
Diabetes Mellitus	present	15	5
	not present	278	93
	missing	6	2
Hypertensive Blood Pressure	present	61	20
	not present	230	77
	missing	8	3
Dyslipidemia	present	20	7
	not present	268	90
	missing	11	4
Smoking	ever	64	21
	never	141	47
	missing	94	31
Use of any Chemotherapy	yes	15	5
	no	283	95
	unknown	1	0

Total		299	100
Dose (Gy_{RBE})	median (range)		74 (66 - 77.4)
Age at RT (years)	median (range)		45 (2 - 84)

Table 2 The area under the curve of the receiver operating characteristic (AUC-ROC) with 95% confidence intervals, Hosmer-Lemeshow test statistic p-value and cross-entropy (CE) values for full cohort models (structure wise and patient wise) and model parameters with regression coefficient (90% confidence intervals) of the best performing model Age*PD*D1cc*HBP

Model parameters				
constant	Age (β_1)	PD (β_2)	D1cc (β_3)	HBP (β_4)
-23.25 (-47.5,-14.7)	0.031 (0.01,0.05)	0.204 (0.07,0.49)	0.06 (0.03,0.17)	0.584 (-0.03,1.36)

Performance evaluation					
structure wise			patient wise		
AUC-ROC	HL p-value	CE	AUC-ROC	HL p-value	CE
0.79 (0.72,0.85)	0.66	0.21	0.76 (0.68,0.84)	0.45	0.28