Radiotherapy and Oncology 130 (2019) 164–171

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Development and validation of NTCP models for acute side-effects resulting from proton beam therapy of brain tumours

Almut Dutz [a,b,c,*], Armin Lühr [a,b,c], Linda Agolli [a,d], Esther G.C. Troost [a,b,c,d,e], Mechthild Krause [a,b,c,d,e], Michael Baumann [a,b,c,d,e,f], Xavier Vermeren [g], Dirk Geismar [g,h,i], Emily F. Schapira [j], Meghan Bussière [j], Jillian E. Daly [j], Marc R. Bussière [j], Beate Timmermann [g,h,i,k,1], Helen A. Shih [j,1], Steffen Löck [a,c,d,e,1]

[a] *OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden – Rossendorf;* [b] *Helmholtz-Zentrum Dresden – Rossendorf, Institute of Radiooncology – OncoRay;* [c] *German Cancer Consortium (DKTK), partner site Dresden, and German Cancer Research Center (DKFZ), Heidelberg, Germany;* [d] *Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden;* [e] *National Center for Tumor Diseases (NCT), Partner Site Dresden; German Cancer Research Center (DKFZ), Heidelberg; Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden; Helmholtz Association / Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Dresden, Germany;* [f] *German Cancer Research Center (DKFZ), Heidelberg, Germany;* [g] *West German Proton Therapy Center Essen (WPE);* [h] *Department of Particle Therapy;* [i] *West German Cancer Center (WTZ), University Hospital Essen, Germany;* [j] *Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, USA;* [k] *German Cancer Consortium (DKTK), partner site Essen, Germany*

ARTICLE INFO

Article history:
Received 6 April 2018
Received in revised form 19 June 2018
Accepted 22 June 2018
Available online 20 July 2018

Keywords:
NTCP models
Brain tumours
Acute side-effects
Proton beam therapy

ABSTRACT

Background: The limited availability of proton beam therapy (PBT) requires individual treatment selection strategies, such as based on normal tissue complication probability (NTCP). We developed and externally validated NTCP models for common acute side-effects following PBT in brain tumour patients in effort to provide guidance on optimising patient quality of life.

Methods: An exploration cohort including 113 adult brain tumour patients who underwent PBT was investigated for the following endpoints: alopecia, scalp erythema, headache, fatigue and nausea. Dose–volume parameters of associated normal tissues were used for logistic regression modelling. Statistically significant parameters showing high area under the receiver operating characteristic curve (AUC) values in internal cross-validation were externally validated on two cohorts of 71 and 96 patients, respectively.

Results: Statistically significant correlations of dose–volume parameters of the skin for erythema and alopecia were found. In internal cross-validation, the following prognostic parameters were selected: V35Gy (absolute volume receiving 35 Gy) for erythema grade $\geq$1, D2% (dose to 2% of the volume) for alopecia grade $\geq$1 and D5% for alopecia grade $\geq$2. Validation was successful for both cohorts with AUC >0.75. A bivariable model for fatigue grade $\geq$1 could not be validated externally. No correlations of dose–volume parameters of the brain were seen for headache or nausea.

Conclusion: We developed and successfully validated NTCP models for scalp erythema and alopecia in primary brain tumour patients treated with PBT.

Proton beam therapy (PBT) is a promising irradiation technique because it achieves high dose conformity while simultaneously reducing dose to normal tissues surrounding the target volume [1–4]. Evidence suggests that PBT leads to advantageous patient-reported outcome measures compared to photon radiotherapy (RT) [5]. Although the number of operating PBT centres increases steadily (77 facilities worldwide, January 2018), the availability of PBT is still limited [6]. Hence, the selection of patients who are likely to benefit from PBT by reduction of side-effects is crucial to provide personalised radiotherapy. More than 100 ongoing clinical trials are being conducted to demonstrate the potential benefit of PBT, but only 5 of those randomise patients between photon and proton irradiation [7]. Underlying reasons include physician and patient preference, that many PBT centres are non-academic stand-alone institutions, i.e. not situated at photon RT departments or academic institutions and the cost coverage for PBT is heterogeneous between insurances [8]. New trial concepts clustering photon centres with a regional proton centre to facilitate randomisation studies are being implemented. Despite this, alternatives to randomised trials can be beneficial [9]. One treatment selection strategy may be the so-called "model-based approach" [10–12]. It

* Corresponding author at: OncoRay – National Center for Radiation Research in Oncology, Fetscherstr. 74, PF 41, 01307 Dresden, Germany.
E-mail address: almut.dutz@oncoray.de (A. Dutz).
[1] Shared last authorship.

https://doi.org/10.1016/j.radonc.2018.06.036
0167-8140/

is based on the prediction and comparison of normal tissue complication probabilities (NTCP) for proton and photon treatment plans. Various NTCP models based on photon therapy data are available for a multitude of side-effects [13–24]. Due to differences in dose distributions and relative biological effectiveness (RBE), however, NTCP models derived from photon treatment may not be reliable for PBT. Blanchard et al. [25] externally validated photon-derived NTCP models for head-and-neck cancer patients treated with PBT. The models for four of five investigated endpoints still performed well. Nevertheless, there is an urgent need for externally validated PBT-based NTCP models for other endpoints.

NTCP models for different endpoints following brain irradiation with photons have been developed from dosimetric and clinical outcome data [17–24]. Only one publication is available for combined proton and photon treatment [26]. Modelled late side-effects within the mentioned datasets include radiation necrosis [17,18], cognitive deterioration [18,19], cranial neuropathies [20], endocrine dysfunction [21,26], hearing loss, otitis and tinnitus [17,22,26] and visual impairment [17,23,24]. Acute side-effects during cranial photon irradiation have been investigated [27–30], but were rarely modelled. Mahadevan et al. [30] showed that dose–volume parameters of follicle hair bearing scalp were correlated to acute alopecia in a small patient cohort treated with intensity-modulated radiotherapy (IMRT). Sung et al. [31] correlated moderate to severe radiation dermatitis with dose–volume parameters of the skin in 101 patients having undergone IMRT. To the best of our knowledge there is no literature on PBT-based NTCP models for side-effects in primary brain tumour patients.

Thus, as a first step, we investigated common acute side-effects following PBT that often adversely affect patients' quality of life [32–41]. We developed and externally validated PBT-based NTCP models for alopecia, erythema and fatigue which in combination with NTCP models on late side effects may be used for future treatment selection [10].

Patients and methods

Patient data

The study design is presented in Fig. 1 and the patient cohorts are characterised in Table 1. All patients included in the study were over 18 years of age, had a diagnosis of a primary brain tumour, received normo-fractionated PBT and were treated consecutively within a certain time period at each institution. The exploration cohort consisted of 113 patients treated within clinical studies (DRKS00007670, DRKS00008569 and NCT02824731) between 12/2014 and 11/2016 at the Department of Radiotherapy and Radiation Oncology of the University Hospital Carl Gustav Carus Dresden. The validation cohort 1 comprised 71 patients from the West German Proton Therapy Centre Essen (treated between 07/2013 and 08/2016 within a clinical registry study (DRKS00004384)) and validation cohort 2 consisted of 96 patients treated at Massachusetts General Hospital in Boston (treated between 11/2013 and 12/2015). This retrospective trial was approved by the local Ethics Committee (EK219062016) and the institutional review board at the external institutions.

Radiotherapy planning and treatment delivery

Treatment plans were calculated using a double scattering PBT technique (XiO, Elekta AB, Stockholm, Sweden) for the exploration cohort and for validation cohort 2 (except for two cases) and an active scanning technique (XiO, [2013–2014]; RayStation, RaySearch AB, Stockholm, Sweden [2015–2016]) for validation cohort 1. For delineation and treatment planning, a computed tomography (CT) scan with slice thicknesses of 2 mm (exploration cohort), 1 mm (validation cohort 1) or 1.25 mm (validation cohort 2) was acquired for each patient and rigidly registered with (post-operative) magnetic resonance imaging (MRI) scans. Clinical target volumes (CTVs) were created based on the gross tumour volume or resection cavity taking into account information from MRI and the original tumour histology as well as anatomical boundaries. The target volume for the passive scattering treatment plans was the CTV with an in-beam margin of 3.5% (distal or proximal CTV depth) + 3 mm in the exploration cohort. In validation cohort 2, the distal range expansion to the CTV was 3.5% + 1 mm with an additional modulation expansion of twice the distal expansion. For validation cohort 1, a planning target volume (PTV) was constructed by adding a 5 mm margin to the CTV. For treatment planning, one to four beams in a patient individual beam setting were applied according to clinical practice in the respective centres; the RBE was considered to be constant at 1.1. The dose–volume

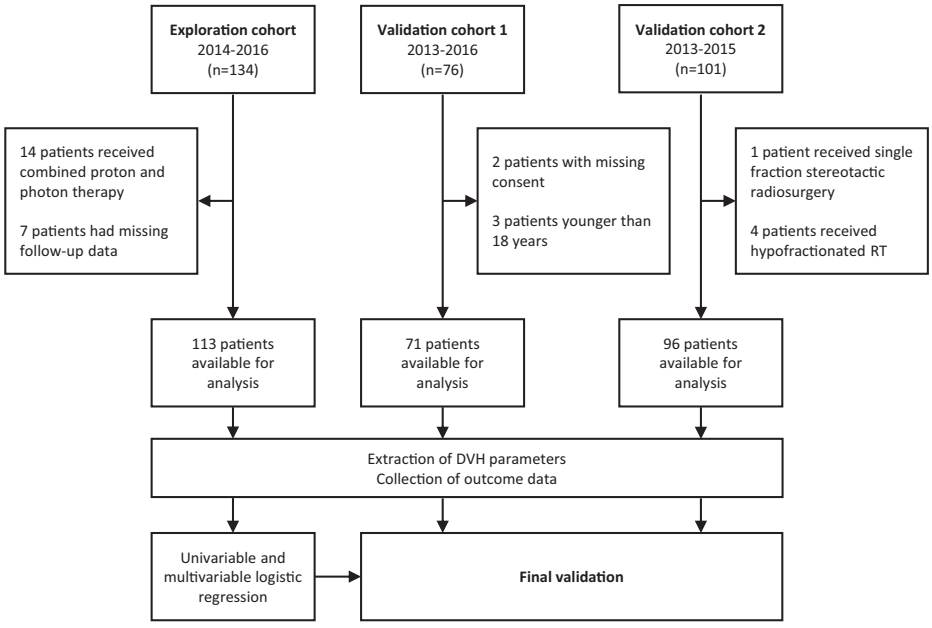

Fig. 1. Study design. RT, radiotherapy; DVH, dose–volume histogram.

Table 1
Comparison of patient and tumour characteristics between the exploration cohort and the two validation cohorts. Statistically significant *p*-values are marked in bold.

Characteristics	Exploration cohort (12/2014–11/2016) Median (range)	Validation cohort 1 (07/2013–08/2016) Median (range)	*p*-value	Validation cohort 2 (11/2013–12/2015) Median (range)	*p*-value
Age (years)	49.3 (21.2–79.9)	45.1 (18.1–79.6)	0.085	45.9 (18.5–89.3)	0.094
Total dose (Gy)	60.0 (30.0–74.0)	54.0 (42.5–60.0)	**<0.001**	54.0 (36.0–70.2)	**<0.001**
Dose per fraction (Gy)	2.0 (2.0–2.0)	2.0 (1.8–2.5)	**<0.001**	1.8 (1.7–1.8)	**<0.001**
Clinical target volume (cm^3)	153.9 (1.2–499.1)	50.9 (2.1–368.1)	**<0.001**	28.1 (1.2–267.2)	**<0.001**
	No of pts (%)	**No of pts (%)**		No of pts (%)	
Gender			0.094		**0.003**
Male	68 (60)	33 (46)		38 (40)	
Female	45 (40)	38 (54)		58 (60)	
Diagnosis			**<0.001**		**<0.001**
Pituitary adenoma	10 (9)	1 (1)		11 (11)	
Meningeoma	15(13)	26 (37)		38 (40)	
Vestibular schwannoma	0 (0)	0 (0)		9 (9)	
Craniopharyngeoma	1 (1)	3 (4)		2 (2)	
Astrocytoma (I)*	1 (1)	2 (3)		1 (1)	
(Oligo)astrocytoma and Oligodendroglioma (II)*	10 (9)	13 (18)		14 (15)	
(Oligo)astrocytoma and Oligodendroglioma (III)*	24 (21)	8 (11)		11 (11)	
Glioblastoma (IV)*	30 (27)	8 (11)		0 (0)	
Other	22 (19)	10 (14)		10 (10)	
Location of the tumour			0.11		**<0.001**
Brain	77 (68)	39 (55)		37 (39)	
Skull base	35 (31)	30 (42)		57 (59)	
Other	1 (1)	2 (3)		2 (2)	
Surgery			**0.011**		0.067
Yes/No	99/14 (88/12)	51/20 (72/28)		75/22 (78/22)	
Simultanous CTx			0.49		0.45
Yes/No	31/82 (27/73)	16/55 (23/77)		22/74 (23/77)	

Abbreviations: No, number; pts, patients; CTx, chemotherapy.
*Tumour classification according to WHO Classification of Tumours of the Central Nervous System, Fourth Edition.

constraints differed slightly between the exploration cohort and validation cohort 1 (Supplementary Table S1). Patient individual planning objectives were used in validation cohort 2. During PBT, patient positioning was daily verified by manual registration of orthogonally paired X-ray images with digitally reconstructed radiographs of the planning CT. In validation cohort 2, most patients were immobilised using a stereotactic bite block system.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2018.06.036.

Endpoint definition and extraction of dose–volume parameters

Treatment-related side-effects were prospectively scored before treatment (baseline), weekly during RT, at the end of RT and 3 months after RT using the Common Terminology Criteria for Adverse Events (CTCAE) version 4 scoring system. Alopecia, erythema, headache (pain), fatigue and nausea were investigated. The maximum severity scores of toxicity during treatment and follow-up were analysed. For patients with a non-zero baseline value and an increase in severity during treatment, the maximum score was used regardless the pre-treatment value. All endpoints were dichotomised at grade $\geq$1 (grade 0 vs remaining) and grade $\geq$2 (grade 0 and 1 vs remaining). Especially for young patients even relatively mild forms of side-effects (grade 1) may have an impact on their quality of life, while side effects grade $\geq$2 are more severe and therefore clinically more relevant for the general patient population.

All dose distributions were retrospectively exported using RayStation v6.0 scripts. Absolute volume parameters V*x*Gy (volume of an organ at risk (OAR) in cm^3 receiving *x* Gy) ranging from 5 Gy to 60 Gy in increments of 5 Gy and dose parameters D*x*% (dose in *x*% of the OAR volume in Gy) ranging from 5% to 95% in increments of 10% as well as D2% representing near maximum doses were analysed. These 23 dose–volume parameters were extracted for brain excluding the target volume (brain-CTV), brainstem and skin (body contour cropped by 3 mm).

Statistical analyses

The impact of both dose–volume parameters and clinical variables on the endpoints was evaluated. Uni- and multivariable logistic regression were carried out on the exploration cohort,

$$NTCP = \frac{1}{1 + e^{-(\beta_0 + x\beta)}}, \quad (1)$$

with the fit parameter β_0 and the parameter vector $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_n)$ for n variables $x = (x_1,\ldots,x_n)$. The clinical parameters age, prescribed total dose, CTV, concomitant chemotherapy (CTx), tumour location (brain or skull base), gender and surgery were investigated. Continuous values were included as continuous variables. To determine the final prognostic parameters, the area under the receiver operating characteristic curve (AUC) was calculated. This value equals 0.5 for non-prognostic models and 1 for perfect models. For development of NTCP models the following steps were performed: (i) A 3-fold internal cross-validation was conducted 333 times on the exploration cohort to identify prognostic dose–volume parameters. (ii) Dose–volume parameters showing a significant association ($p < 0.05$) to the investigated endpoint in univariable logistic regression and the largest AUC value of the internal validation folds were preselected. (iii) Clinical parameters showing a significant association to the endpoint in logistic regression were tested for correlation with the dose–volume parameters selected during step (ii) using the Spearman correlation coefficient ρ. Multivariable logistic regression models containing the independent clinical parameters ($|\rho| < 0.5$) and one dose–volume parameter were built as described

in step (i). Uni- or multivariable models with the largest AUC value in internal cross-validation were selected as final NTCP models. (iv) The models derived from the exploration cohort were applied without changes to both validation cohorts, i.e. the models were evaluated using the dose–volume parameters of the validation cohorts and the model coefficients derived from the exploration cohort. The resulting probabilities were compared to the observed outcome by calculating the AUC value. The 2.5th and 97.5th percentiles of 1000 bootstrap samples were used to estimate the 95% confidence intervals (CI) of the AUC values. Prognostic performance was assessed by AUC values and calibration plots. Binary variables were compared between the exploration and each of the validation cohorts using exact Fisher's tests. Differences in categorical variables were evaluated by chi-squared tests and in continuous variables by Mann–Whitney's U tests. Correlations between dose–volume parameters were assessed using the Spearman correlation coefficient ρ. For all statistical analyses, two-sided tests were performed using IBM SPSS Statistics 23 (IBM Corporation, Armonk, NY) and in-house written Python (version 2.7.10) programmes using the module scipy (statsmodels). *P*-values <0.05 were considered statistically significant.

Results

Glioblastoma was the most common diagnosis in the exploration cohort, whereas meningioma was most frequent in the validation cohorts leading to significant differences regarding CTV ($p < 0.001$), prescribed total dose ($p < 0.001$) and surgery ($p = 0.011$ for validation cohort 1). For all cohorts, the incidence of the investigated acute side-effects after PBT is given in Table 2. Overall, less alopecia and erythema grade 1–2 occurred in the validation cohorts compared to the exploration cohort ($p < 0.001$). No statistically significant differences in the frequencies of fatigue and nausea were found between the exploration and validation cohorts. Headches were less common in validation cohort 2 compared to the exploration cohort ($p < 0.001$). The low incidence rates of nausea and somnolence grade ≥2 in all cohorts did not allow for NTCP modelling.

Side-effects were tested for correlation with clinical cofactors in the exploration cohort (Supplementary Table S2). Erythema as well as alopecia occurred significantly more often for larger CTV ($p < 0.024$), prescribed total dose ($p < 0.037$) and performed surgery (erythema grade ≥1: $p < 0.001$, alopecia grade ≥2: $p = 0.030$). Fatigue grade ≥1 showed a higher prevalence in women compared to men ($p = 0.005$).

Dose–volume parameters of the skin in the high dose region were significantly correlated to erythema and alopecia. For erythema, the parameter revealing the largest AUC values in cross-validation was V35Gy (grade ≥1: AUC = 0.75, grade ≥2: AUC = 0.77). For alopecia grade ≥1 D2% (AUC = 0.88) and for alopecia grade ≥2 D5% (AUC = 0.82) were selected. As the clinical cofactors associated to erythema and alopecia were significantly correlated with each other and with the selected dose–volume parameters of the skin (Supplementary Table S3), univariable NTCP models including the above mentioned dose–volume parameters were finally developed (Table 3).

The high-dose parameter D2% of brain-CTV was significantly associated with fatigue grade ≥1 and revealed an AUC value of 0.60. As the cofactor gender was not correlated to this dose–volume parameter ($\rho = 0.15$), a bivariable model was built that showed an improved performance in cross-validation compared to the univariable model (AUC = 0.68), Table 3. None of the dose–volume parameters of brain-CTV showed significant correlation to fatigue grade ≥2, pain and nausea.

The final NTCP models developed on the exploration cohort were applied without changes to the validation cohorts showing high AUC values for erythema and alopecia (Table 3). Exemplarily, for erythema grade ≥1 and V35Gy of the skin, the validation AUC was 0.87, 95%CI: [0.77–0.95] in validation cohort 1 and 0.80 [0.71–0.89] in validation cohort 2. Similar results were obtained for alopecia. Thus, both validation cohorts confirmed the results of the exploration cohort. The regression curves and calibration plots for alopecia grade ≥2 and D5% as well as erythema grade ≥1 and V35Gy of the skin are shown in Fig. 2. The calibration plots for both endpoints showed a right shift for both validation cohorts. External validation of the bivariable NTCP model for fatigue grade ≥1 revealed low AUC values in both cohorts (validation cohort 1: 0.45 [0.31–0.61], validation cohort 2: 0.52 [0.40–0.64]). Since only one case of erythema grade ≥2 was reported in the validation cohorts, the NTCP models for this endpoint could not be validated.

Table 2
Comparison of the baseline-corrected acute side-effects (CTCAE v4.0) between exploration (E) and validation cohorts (V1, V2). The number N of available datasets is given. *P*-values represent results of the Pearson chi-squared test. Statistically significant *p*-values are marked in bold.

Toxicity Cohort	Grade 0 No of pts (%)	Grade 1 No of pts (%)	Grade 2 No of pts (%)	Grade 3 No of pts (%)	*p*-value
Alopecia					
E (*N* = 111)	15 (14)	26 (23)	70 (63)	0 (0)	
V1 (*N* = 71)	22 (31)	35 (49)	14 (20)	0 (0)	**<0.001**
V2 (*N* = 96)	59 (61)	16 (17)	21 (22)	0 (0)	**<0.001**
Erythema					
E (*N* = 113)	15 (13)	57 (51)	40 (35)	1 (1)	
V1 (*N* = 71)	32 (45)	39 (55)	0 (0)	0 (0)	**<0.001**
V2 (*N* = 96)	63 (66)	32 (33)	1 (1)	0 (0)	**<0.001**
Fatigue					
E (*N* = 112)	35 (31)	53 (47)	21 (19)	3 (3)	
V1 (*N* = 71)	18 (25)	45 (63)	6 (9)	2 (3)	0.12
V2 (*N* = 96)	40 (42)	47 (49)	9 (9)	0 (0)	0.063
Nausea					
E (*N* = 98)	82 (84)	13 (13)	3 (3)	0 (0)	
V1 (*N* = 71)	62 (87)	9 (13)	0 (0)	0 (0)	0.32
V2 (*N* = 96)	87 (91)	5 (5)	4 (4)	0 (0)	0.15
Pain					
E (*N* = 113)	57 (51)	33 (29)	18 (16)	5 (4)	
V1 (*N* = 71)	49 (69)	15 (21)	6 (8)	1 (2)	0.079
V2 (*N* = 96)	75 (78)	19 (20)	2 (2)	0 (0)	**<0.001**

Abbreviations: No, number; pts, patients.

Table 3
Logistic modelling results: Dose–volume parameters of different organs at risk and clinical cofactors for acute erythema, alopecia, and fatigue. Mean AUC values for 3-fold cross-validation (333 repetitions) and external validation are given. *P*-values were calculated on the exploration cohort, confidence intervals (CI) were obtained using 1000 bootstrap samples. Fitting parameters β_i as defined in Eq. (1). Statistically significant *p*-values are marked in bold.

Model	AUC	(95% CI)	β_0	(95% CI)	β_1	(95% CI)	*p*-Value	β_2	(95% CI)	*p*-Value
Erythema grade ≥1										
Model parameter					V35Gy Skin (cm^3)					
Exploration	0.75	(0.54–0.90)	1.00	(0.32–1.69)	0.09	(0.02–0.15)	**0.008**			
Validation 1	0.87	(0.77–0.95)								
Validation 2	0.80	(0.71–0.89)								
Erythema grade ≥2										
Model parameter					V35Gy Skin (cm^3)					
Exploration	0.77	(0.64–0.89)	−1.54	(−2.20–−0.88)	0.06	(0.03–0.08)	**<0.001**			
Validation 1	*									
Validation 2	0.84	(0.77–0.91)								
Alopecia grade ≥1										
Model parameter					D2% Skin (Gy)					
Exploration	0.88	(0.73–0.99)	−0.94	(−2.14–0.27)	0.10	(0.05–0.15)	**<0.001**			
Validation 1	0.82	(0.70–0.92)								
Validation 2	0.84	(0.75–0.92)								
Alopecia grade ≥2										
Model parameter					D5% Skin (Gy)					
Exploration	0.82	(0.69–0.95)	−1.33	(−2.91–−0.47)	0.08	(0.05–0.11)	**<0.001**			
Validation 1	0.77	(0.63–0.89)								
Validation 2	0.85	(0.76–0.92)								
Fatigue grade ≥1										
Model parameter					D2% Brain-CTV (Gy)			Gender		
Exploration	0.68	(0.50–0.84)	−0.90	(−2.30–0.51)	0.03	(0.00–0.06)	0.067	1.28	(0.33–2.23)	**0.009**
Validation 1	0.45	(0.31–0.61)								
Validation 2	0.52	(0.40–0.64)								

Abbreviations: AUC, area under the receiver operating characteristic curve; V*x*Gy, absolute volume receiving *x* Gy; D*x*% dose in *x*% of the volume of the organ at risk; CTV, clinical target volume.
* External validation not applicable due to zero incidence.

Discussion

Our study on adult primary brain tumour patients receiving PBT investigated the relation of acute side-effects and dose to associated OARs. Overall, PBT was well tolerated with very low incidences of side-effects' grade ≥3. Dose–volume parameters in the high-dose region were prognostic for erythema and alopecia. Fatigue was found to be associated with gender and high doses to the brain-CTV. The NTCP models for erythema as well as alopecia showed a similar or even improved performance for the validation cohorts compared to the exploration cohort.

Several relatively small studies documented acute side-effects in brain tumour patients receiving proton therapy according to CTCAE v3 or v4 [32–35,42]. The incidence rate of alopecia grade2 in the exploration cohort (63%) is comparable to the results of Hauswald et al. [34] (68%), even though the 19 glioma I-II patients were treated with an active scanning technique and a median dose of 54 Gy. Weber et al. [35] investigated 39 skull base meningioma patients treated with spot-scanning to a total dose of 56 Gy and Shih *et al.* [32] conducted a study on 20 glioma patients treated with a passive scattering technique to a total dose of 54 Gy. Both studies did not report any cases of alopecia grade 2, but an incidence rate for alopecia grade 1 of 60% and 85%, respectively. In two studies by Grosshans *et al.* [33] and Brown et al. [42] investigating 15 skull base chordoma and chondrosarcoma patients treated with spot-scanning to doses of 69.8 Gy and 68.4 Gy and 19 medulloblastoma patients receiving a total dose of 54 Gy, respectively, no erythema grade 2 was documented. This is comparable to validation cohort 1 in our study. The incidence rate of fatigue grade ≥1 both for the exploration (69%) and validation cohort (75%) is in line with the findings of Grosshans et al. [33] (66%). For nausea grade ≥1, Shih et al. [32] reported similar results with 20%, compared to the exploration cohort with 16%.

Our findings regarding the association of skin dose–volume parameters with alopecia and erythema are in line with other publications, in which photon doses to the skin were associated with these side-effects. Reduction in the irradiated volume V24Gy and V30Gy of follicle hair bearing scalp prevented alopecia in an 11-field IMRT [30]. In our study, the skin dose–volume parameters V25Gy showed also a significant association with alopecia. Nevertheless, the prognostic performance was slightly superior for D2% (alopecia grade ≥1) as well as for D5% (alopecia grade ≥2) in internal cross-validation. Another study investigated side-effects after conventional photon RT in 61 cranially irradiated patients. More than 50% of the patients receiving a follicle dose >43 Gy developed permanent moderate to severe alopecia [43]. The higher dose value of this prognostic parameter compared to our findings may be due to the endpoint of permanent manifestation of alopecia which was investigated in this study. Sung et al. [31] found the dose–volume parameter V35Gy as most significant predictor for radiation dermatitis grade ≥2 after hybrid IMRT for 101 breast cancer patients. As we found the same dose–volume parameter as the most predictive parameter for erythema, skin reaction in thoracic regions may be similar to those after cranial irradiation. Mendelsohn et al. [44] reported early dry desquamation (erythema grade ≥1) if the total skin dose during conventional radiotherapy does not exceed 30 Gy and acute moist desquamation (erythema grade ≥2 and 3) as an effect of total skin dose >40 Gy. Our results are in line with these findings as we predict that reduction of V35Gy lowers the risk of erythema.

The dose–volume parameters of the skin were highly correlated so that other parameters may be relevant on other cohorts (Supplementary Fig. S1). A principal component analysis was conducted (Supplementary Tables S5–S7, Figs. S2 and S3) to create independent parameters, the so-called principal components, representing different dose regions. NTCP models based on these

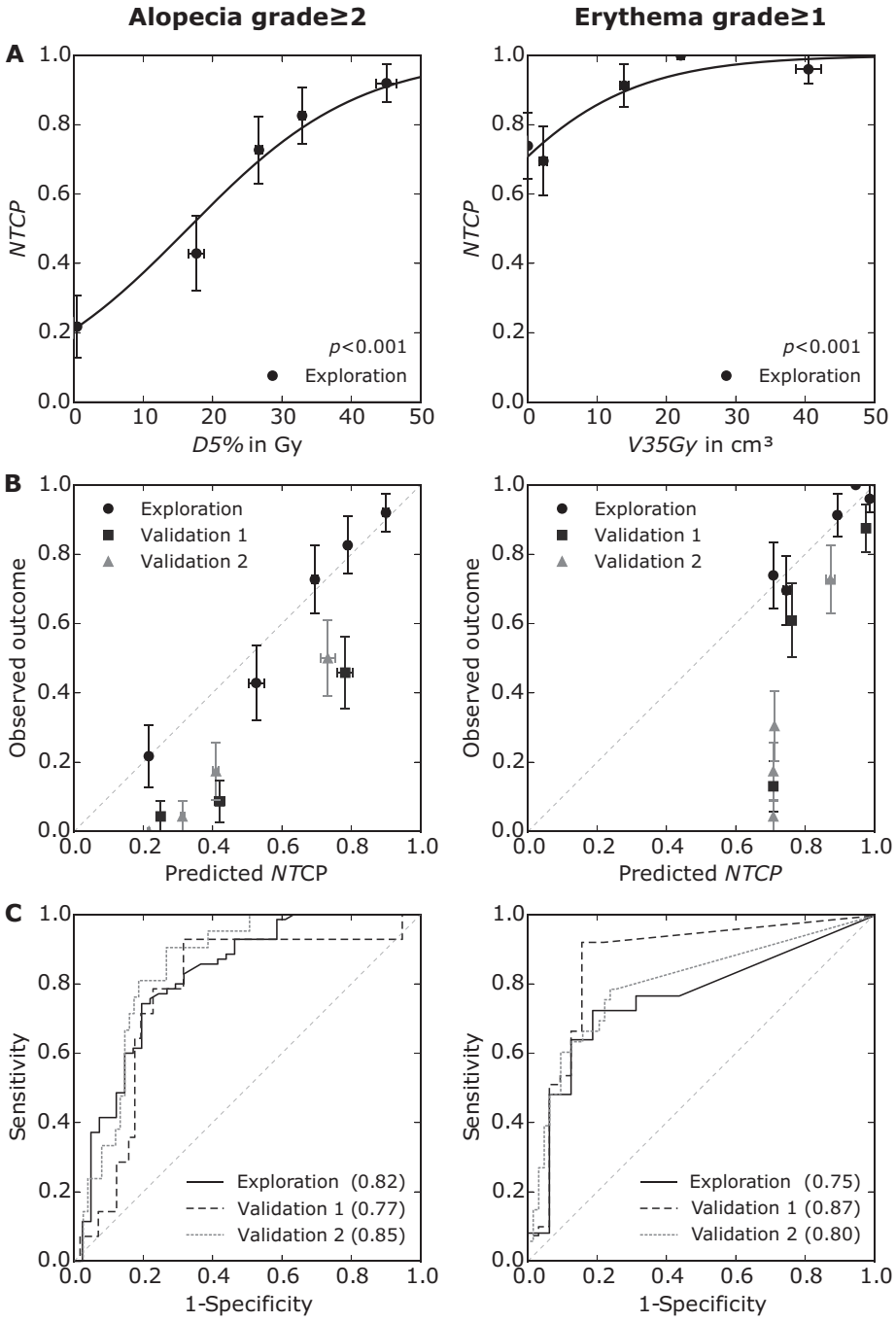

Fig. 2. Modelling results for alopecia grade ≥2 (left) and erythema grade ≥1 (right). **A** Logistic regression curves based on the dose–volume parameters D5% and V35Gy of the skin, respectively. **B** Calibration plots for these models. A perfectly calibrated model would lead to the grey line. Patients were sorted according to the parameter value and grouped into equally sized groups. The mean values are represented by data points, with error bars showing the standard deviation. **C** Receiver operating characteristic curves. Areas under the curve are given in parentheses. Abbreviations: NTCP, normal tissue complication probability.

components showed similar AUC values in internal and external validation, but they are more difficult to interpret. A shift in the calibration curves for alopecia and erythema was observed in both validation cohorts for the presented models as well as the models based on principal components.

Different rates of erythema and alopecia could be caused by differences in several patient and tumour characteristics, such as tumour type and volume (Table 1). Nevertheless, other considerable factors such as malnutrition, age and tobacco abuse history may be more important [45]. Different treatment techniques in exploration (passive scattering) and validation cohort 1 (active scanning) may also lead to differences in acute side-effects. A study investigating skin dose differences between spot scanning and passively scattered PBT in prostate cancer patients found a lower skin dose in patients receiving actively formed PBT compared to patients treated with passive PBT [46]. In many brain tumour patients, however, the target volume is located directly under the scalp, so that skin doses differ less between both techniques. A comparison of dose–volume parameters of the skin between the exploration and validation cohort 1 showed significant differences only in the low dose range (Supplementary Table S4). These findings and the shift in calibration for validation cohort 2 (passively

scattered PBT), support the hypothesis, that the treatment techniques are probably not the reason for different toxicity incidences. Hence, the differences may most likely be due to different toxicity scoring by the physicians, even though the same grading system was used. Di Maio et al. [47] detected frequent under-reporting of subjective side-effects, such as alopecia and fatigue, by physicians at three different centres to a variable extent compared to patients' self-assessment. The incorporation of Patient-Reported Outcome versions of the CTCAE scale (PRO-CTCAE™) in clinical trials may reduce such centre-specific differences in toxicity assessment [48].

One possibility to account for this potential variability in toxicity assessment may be a centre-specific recalibration of the developed models. We successfully tested this approach for both validation cohorts by adapting the constant coefficient of the logistic models only (intercept recalibration). Thus, after recalibration, e.g. using retrospective data, the models may be used by other proton therapy centres.

For alopecia and erythema, univariable NTCP models have been developed. The clinical cofactors surgery, prescribed dose and CTV were highly correlated to both acute side-effects and dose–volume parameters of the skin. As diagnosis determines the dose prescription, tumour volume and margin size, patients with large CTV had significantly higher prescribed total doses ($p < 0.001$), leading to larger dose–volume parameter values. Due to these high correlations, the clinical cofactors were not included in potential multivariable NTCP models, since including redundant information would not improve their prognostic ability.

For acute fatigue, we developed a bivariable model including the dose–volume parameter D2% of brain-CTV and the clinical cofactor gender. This model could not be validated. Gulliford et al. [29] investigated the association of dose to cranial structures and acute fatigue in 67 photon therapy patients and found significantly higher mean and maximum dose values for posterior fossa, brainstem and cerebellum for patients suffering from acute fatigue grade ≥ 2 ($p < 0.01$). Ferris et al. [28] found a significant correlation between patient-reported fatigue based on Multidimensional Fatigue Inventory scores and maximum dose to the brainstem as well as medulla in 124 photon therapy patients ($p < 0.05$). We did not find any statistically significant dose relationship between dose to the brainstem and acute fatigue for patients treated with passively scattered proton beams; instead, we found a significant correlation between fatigue grade ≥ 1 and female gender. One of the few studies investigating cancer-related fatigue according to the EORTC QLQ-C30 scale also found female gender associated with greater fatigue severity [49]. However, as there are only a few studies on gender differences in cancer-related fatigue, there is still concern about the prevalence and severity of this side-effect [50].

Finally, no association between dose to the brain and acute nausea and pain was found in our study. Radiation dose to substructures of the brain or concomitant medication such as analgesics may play a role. To the best of our knowledge, no NTCP models based on photon therapy data regarding these endpoints are available in the literature.

In general, NTCP models can be used for patient individual treatment selection [10]. The models developed in this study describe acute side-effects occurring after irradiation of the skin. Development of NTCP models for severe long-term side-effects and neurocognitive changes associated with dose to hippocampus, frontal and temporal lobe is needed additionally. This can be done using the implemented methodology as soon as data on long-term side-effects are available. If centre-specific differences in toxicity assessment are considered carefully, a combination of late toxicity models and the presented models for acute side-effects may support clinicians with finding the most beneficial treatment for brain tumour patients.

We developed and externally validated NTCP models relating dose–volume parameters of the skin to acute erythema and alopecia in cranial irradiated PBT patients. Validation confirmed that the presented dose–volume parameters are strongly associated with these endpoints for all investigated cohorts, but the calibration differed. After the development of NTCP models for late side-effects, a combination of NTCP models may have potential to serve as a reliable tool for future treatment selection [10].

Acknowledgements

The excellent work of the clinical trials offices under the direction of Monique Falk in Dresden and Stefanie Schulze Schleithoff in Essen as well as the clinical staff is highly acknowledged, including the work of Sabine Frisch for data collection and transfer.

Conflicts of interest

None.

Disclaimers

None.

References

[1] Baumann M, Krause M, Overgaard J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer 2016;16:234.
[2] Harrabi SB, Bougatf N, Mohr A, et al. Dosimetric advantages of proton therapy over conventional radiotherapy with photons in young patients and adults with low-grade glioma. Strahlenther Onkol 2016;192:759–69.
[3] Kristensen I, Nilsson K, Nilsson P. Comparative proton and photon treatment planning in pediatric patients with various diagnoses. IJPT 2015;2:367–75.
[4] Palm Å, Johansson KA. A review of the impact of photon and proton external beam radiotherapy treatment modalities on the dose distribution in field and out-of-field; implications for the long-term morbidity of cancer survivors. Acta Oncol 2007;46:462–73.
[5] Verma V, Simone II CB, Mishra MV. Quality of life and patient-reported outcomes following proton radiation therapy: A systematic review. J Natl Cancer Inst 2018;110. djx208–djx208.
[6] Particle Therapy Co-Operative Group. Particle therapy facilities in operation. https://www.ptcog.ch/index.php/facilities-in-operation (accessed on 31.01.2018); 2018.
[7] Mishra MV, Aggarwal S, Bentzen SM, Knight N, Mehta MP, Regine WF. Establishing evidence-based indications for proton therapy: an overview of current clinical trials. Int J Radiat Oncol Biol Phys 2017;97:228–35.
[8] Lühr A, von Neubeck C, Krause M, Troost EGC. Relative biological effectiveness in proton beam therapy – current knowledge and future challenges. Clin Transl Radiat Oncol 2018;9:35–41.
[9] Lühr A, von Neubeck C, Pawelke J, et al. Radiobiology of proton therapy: results of an international expert workshop. Radiother Oncol 2018 [Epub ahead of print].
[10] Langendijk JA, Lambin P, Ruysscher DD, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.
[11] Quik E, Feenstra T, Postmus D, et al. Individual patient information to select patients for different radiation techniques. Eur J Cancer 2016;62:18–27.
[12] Widder J, van der Schaaf A, Lambin P, et al. The quest for evidence for proton therapy: model-based approach and precision medicine. Int J Radiat Oncol Biol Phys 2016;95:30–6.
[13] Christianen ME, Schilstra C, Beetz I, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. Radiother Oncol 2012;105:107–14.
[14] Wopken K, Bijl HP, van der Schaaf A, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. Radiother Oncol 2014;113:95–101.
[15] Beetz I, Schilstra C, van der Schaaf A, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:101–6.
[16] Dijkema T, Raaijmakers CP, Haken RKT, et al. Parotid gland function after radiotherapy: the combined Michigan and Utrecht experience. Int J Radiat Oncol Biol Phys 2010;78:449–53.
[17] Burman C, Kutcher G, Emami B, Goitein M. Fitting of normal tissue tolerance data to an analytic function. Int J Radiat Oncol Biol Phys 1991;21:123–35.
[18] Lawrence YR, Li XA, el Naqa I, et al. Radiation dose-volume effects in the brain. Int J Radiat Oncol Biol Phys 2010;76:S20–7.

[19] Gondi V, Hermann BP, Mehta MP, Tomé WA. Hippocampal dosimetry predicts neurocognitive function impairment after fractionated stereotactic radiotherapy for benign or low-grade adult brain tumors. Int J Radiat Oncol Biol Phys 2012;83:e487–93.
[20] Meeks SL, Buatti JM, Foote KD, Friedman WA, Bova FJ. Calculation of cranial nerve complication probability for acoustic neuroma radiosurgery. Int J Radiat Oncol Biol Phys 2000;47:597–602.
[21] Pai HH, Thornton A, Katznelson L, et al. Hypothalamic/pituitary function following high-dose conformal radiotherapy to the base of skull: demonstration of a dose-effect relationship using dose-volume histogram analysis. Int J Radiat Oncol Biol Phys 2001;49:1079–92.
[22] Lee TF, Yeh SA, Chao PJ, et al. Normal tissue complication probability modeling for cochlea constraints to avoid causing tinnitus after head-and-neck intensity-modulated radiation therapy. Radiat Oncol 2015;10:194.
[23] Brizel DM, Light K, Zhou SM, Marks LB. Conformal radiation therapy treatment planning reduces the dose to the optic structures for patients with tumors of the paranasal sinuses. Radiother Oncol 1999;51:215–8.
[24] Martel MK, Sandler HM, Cornblath WT, et al. Dose-volume complication analysis for visual pathway structures of patients with advanced paranasal sinus tumors. Int J Radiat Oncol Biol Phys 1997;38:273–84.
[25] Blanchard P, Wong AJ, Gunn GB, et al. Toward a model-based patient selection strategy for proton therapy: external validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort. Radiother Oncol 2016;121:381–6.
[26] Marzi LD, Feuvret L, Boulé T, et al. Use of gEUD for predicting ear and pituitary gland damage following proton and photon radiation therapy. Br J Radiol 2015;88:20140413.
[27] Lee VH, Ng SC, Leung T, Au GK, Kwong DL. Dosimetric predictors of radiation-induced acute nausea and vomiting in IMRT for nasopharyngeal cancer. Int J Radiat Oncol Biol Phys 2012;84:176–82.
[28] Ferris MJ, Zhong J, Switchenko JM, et al. Brainstem dose is associated with patient-reported acute fatigue in head and neck cancer radiation therapy. Radiother Oncol 2018;126:100–6.
[29] Gulliford SL, Miah AB, Brennan S, et al. Dosimetric explanations of fatigue in head and neck radiotherapy: an analysis from the PARSPORT Phase III trial. Radiother Oncol 2012;104:205–12.
[30] Mahadevan A, Sampson C, LaRosa S, et al. Dosimetric analysis of the alopecia preventing effect of hippocampus sparing whole brain radiation therapy. Radiat Oncol 2015;10:245.
[31] Sung K, Ting H, Chao P, et al. Predicting grade 2 acute radiation-induced dermatitis after hybrid intensity modulation radiotherapy for breast cancer using a logistic regression normal tissue complication probability model. Eur J Cancer 2016;60:e4.
[32] Shih HA, Sherman JC, Nachtigall LB, et al. Proton therapy for low-grade gliomas: results from a prospective trial. Cancer 2015;121:1712–9.
[33] Grosshans DR, Zhu XR, Melancon A, et al. Spot scanning proton therapy for malignancies of the base of skull: treatment planning, acute toxicities, and preliminary clinical outcomes. Int J Radiat Oncol Biol Phys 2014;90:540–6.
[34] Hauswald H, Rieken S, Ecker S, et al. First experiences in treatment of low-grade glioma grade I and II with proton therapy. Radiat Oncol 2012;7:189.
[35] Weber DC, Schneider R, Goitein G, et al. Spot scanning-based proton therapy for intracranial meningioma: long-term results from the Paul Scherrer Institute. Int J Radiat Oncol Biol Phys 2012;83:865–71.
[36] Combs SE, Kessel K, Habermehl D, Haberer T, Jäkel O, Debus J. Proton and carbon ion radiotherapy for primary brain tumors and tumors of the skull base. Acta Oncol 2013;52:1504–9.
[37] Williamson D, Gonzalez M, Finlay A. The effect of hair loss on quality of life. J Eur Acad Dermatol Venereol 2001;15:137–9.
[38] Erol O, Can G, Aydner A. Effects of alopecia on body image and quality of life of Turkish cancer women with or without headscarf. Support Care Cancer 2012;20:2349–56.
[39] Carelle N, Piotto E, Bellanger A, Germanaud J, Thuillier A, Khayat D. Changing patient perceptions of the side effects of cancer chemotherapy. Cancer 2002;95:155–63.
[40] Fisher J, Scott C, Stevens R, et al. Randomized phase III study comparing best supportive care to biafine as a prophylactic agent for radiation-induced skin toxicity for women undergoing breast irradiation: Radiation therapy oncology group (RTOG) 97-13. Int J Radiat Oncol Biol Phys 2000;48:1307–10.
[41] Horiot JC, Aapro M. Treatment implications for radiation-induced nausea and vomiting in specific patient groups. Eur J Cancer 2004;40:979–87.
[42] Brown AP, Barney CL, Grosshans DR, et al. Proton beam craniospinal irradiation reduces acute toxicity for adults with medulloblastoma. Int J Radiat Oncol Biol Phys 2013;86:277–84.
[43] Lawenda BD, Gagne HM, Gierga DP, et al. Permanent alopecia after cranial irradiation: dose-response relationship. Int J Radiat Oncol Biol Phys 2004;60:879–87.
[44] Mendelsohn FA, Divino CM, Reis ED, Kerstein MD. Wound care after radiation therapy. Adv Skin Wound Care 2002;15:216–24.
[45] Ginot A, Doyen J, Hannoun-Lévi JM, Courdi A. Dose de tolérance des tissus sains: la peau et les phanères. Cancer/Radiothérapie 2010;14:379–85.
[46] Arjomandy B, Sahoo N, Cox J, Lee A, Gillin M. Comparison of surface doses from spot scanning and passively scattered proton therapy beams. Phys Med Biol 2009;54:N295.
[47] Di Maio M, Gallo C, Leighl NB, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. JCO 2015;33:910–5.
[48] Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Natl Cancer Inst 2014;106. dju244–dju244.
[49] Pater J, Zee B, Palmer M, Johnston D, Osoba D. Fatigue in patients with cancer: results with National Cancer Institute of Canada Clinical Trials Group studies employing the EORTC QLQ-C30. Support Care Cancer 1997;5:410–3.
[50] Miaskowski C. Gender differences in pain, fatigue, and depression in patients with cancer. JNCI Monographs 2004;2004:139–43.