Radiotherapy and Oncology 199 (2024) 110462

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Incidence of alopecia in brain tumour patients treated with pencil scanning proton therapy and validation of existing NTCP models

Simona Gaito [a,b], Laura Cella [c,*], Anna France [a], Serena Monti [c], Gillian Whitfield [b,d], Peter Sitch [a,d], Neil Burnet [d], Ed Smith [a,b,d], Giuseppe Palma [e,1], Marianne Aznar [b,1]

[a] Proton Clinical Outcomes Unit, The Christie Proton Beam Therapy Center, Manchester, United Kingdom
[b] Division of Cancer Sciences, School of Medical Science, Faculty of Biology Medicine and Health, The University of Manchester, Manchester, United Kingdom
[c] Institute of Biostructures and Bioimaging, National Research Council, Naples, Italy
[d] Proton Beam Therapy Centre. The Christie NHS Foundation Trust, Manchester, United Kingdom
[e] Institute of Nanotechnology, National Research Council, Lecce, Italy

ARTICLE INFO

Keywords:
Brain tumor
Proton therapy: radiation-induced alopecia
NTCP

ABSTRACT

Background and purpose: Radiation-induced alopecia (RIA) is one of the most frequent and upsetting cosmetic side effects after radiotherapy (RT) for brain cancer. We report the incidence of RIA in a cohort of brain tumours patients treated with Proton Therapy (PT) and externally validate published NTCP models of grade 2 (G2) RIA for their implementation in clinical practice.

Methods: Data for patients treated for brain tumours with scanning beam PT between 2018 and 2022 were extracted. Acute, late and permanent RIA events were evaluated according to CTCAE 5.0. Lyman-Kutcher-Burman (LKB) and multivariable logistic regression (MLR) published models were computed from the relative dose-surface histogram of the scalp. External validity of models was assessed in terms of discrimination and calibration.

Results: In the 264 patients analysed, rates of any grade acute (≤90 days after PT completion), late (>90 days) and permanent RIA (persisting for> 12 months) were 61.8 %, 24.7 % and 14.4 %, respectively. In our independent cohort, LKB- and MLR-NTCP showed a good discrimination for G2 RIA (0.71≤ROC-AUC≤0.83) while model calibration was unsatisfactory possibly due to a different outcome evaluation between training and validation cohorts, as well as differences in clinical and treatment related variables between the two groups.

Conclusions: Despite the reasonable sensitivity and specificity of the NTCP models for RIA in the validation cohort, our study emphasizes the significance of differences between the cohorts utilized for model development and validation. Specifically, variations in the reporting of clinical outcomes inevitably jeopardize the validation of NTCP models. A standardize and objective RIA scoring system is essential.

Introduction

Cosmetic results from cancer therapies play a significant role in the psychological effects of the disease. Patients may endure societal stigmatisation, anxiety, and low self-esteem as a result of the cosmetic outcomes [1]. Nevertheless, these have traditionally been underestimated by health professionals as they may be considered aesthetic rather than "health-related" issues and therefore not clinical endpoints that should be prioritised.

Alopecia is reported to be the most common and distressing cosmetic outcome related to cancer treatments, being negatively associated with body-image and psychological well-being [2]. However, whilst chemotherapy-related alopecia is usually temporary, radiation-induced alopecia (RIA) can quite often be permanent [3]. The clinical presentation of RIA is usually that of a well-demarcated area of hair loss in correspondence of the treatment field. The severity and duration of RIA can be dose-dependent and higher radiation doses are more likely to cause more severe and longer-lasting hair loss. The acute manifestation of RIA typically occurs during the course of radiotherapy (RT). However, the duration of alopecia can vary among individuals. While for many

* Corresponding author at: Institute of Biostructures and Bioimaging, National Research Council, Napoli, Italy.
E-mail address: laura.cella@cnr.it (L. Cella).
[1] Joint last authors.

https://doi.org/10.1016/j.radonc.2024.110462
Received 5 April 2024; Received in revised form 19 July 2024; Accepted 22 July 2024
Available online 26 July 2024

patients, hair may begin to regrow within a few months after completing RT, in some cases, alopecia can persist beyond the 3-month mark. In rare instances, the hair loss may become permanent, lasting more than 12 months. The incidence of acute RIA is reported to be as high as 75–100 % in recipients of cranial RT, while up to 60 % of patients experience some degree of inadequate hair regrowth six months after radiation therapy is finished [3].

An Australian survey showed that there are wide differences among consultants when it comes to estimating the chances of normal tissue complications from RT, including hair loss [4]. The physicians' experience may contribute to some of the heterogeneity, which is also probably caused by the lack of robust data to use as a basis for estimates. Because of this, even after carefully examining the plan's metrics, clinicians are often unable to accurately advise patients about the likelihood of a particular side effect.

A substantive body of research has examined the dose–response relationship for the development of temporary or permanent hair loss following cranial radiation therapy, but the different techniques, patients' demographics and endpoint definition included in the different studies make risk estimates difficult.

Following cranial irradiation using conventional photon RT, Lawenda et al. [5] reported the first dose–response relationship indicating the effect of the follicular dose on the future development of permanent RIA. After analysing other possible confounding factors that might be involved in the RIA, the authors concluded that the only factor that was connected to the probability of permanent RIA was the dose to the scalp follicles, with a threshold dose of 43 Gy (which gives 50 % risk of permanent RIA). There was no discernible correlation between age, gender, family history, beam energy, chemotherapy treatment, or personal history of alopecia. On a cohort of 101 patients treated with VMAT from brain tumours, Scoccianti et al. [6] found that the volumes receiving radiation doses of at least 20 Gy or 40 Gy, or V20Gy and V40Gy, respectively, were determined to be the best indicators of acute and chronic (i.e. persisting after 9 months) grade 2 hair loss. Large regions of acute but temporary alopecia were correlated with the low-dose bath that is typical of VMAT, whilst doses exceeding 40 Gy seem correlated with chronic G2 RIA. Dutz et al. [7] identified D2% to the scalp to be the best predictor of G$\geq$1 RIA at 12 months after cranial PT, reporting an incidence of G $\geq$1 RIA of 34 % and 22 % in the training and validation cohorts, respectively. The same dosimetric variable was also reported by Palma et al. [8] as the best predictor of permanent RIA from multivariable model selection.

Due to differences in the physical properties between XRT and Proton Therapy (PT), models of toxicity prediction developed on XRT-treated cohorts might not be applicable in the PT setting [9]. XRT can benefit of a skin-sparing effect thanks to the so-called "build-up" [8]. This skin-sparing effect is much less pronounced with PT, and therefore a possible increase in the incidence and severity of skin toxicities (including alopecia) has been postulated. However, comparative study published by our group has shown no meaningful clinical differences in acute and late skin toxicities between 2 cohorts of children treated with PT and XRT for paediatric sarcomas [10].

These factors have led research groups from different institutions to develop PT-specific Normal Complication Probability Models (NTCP) in order to estimate the risk of this side effect [7,8]. In fact, a reliable PT-specific NTCP model for alopecia, especially permanent alopecia, would enhance the patient-centred care of pencil beam scanning (PBS) −PT. As such, Palma et al. [8], developed NTCP models of acute, late and permanent grade 2 RIA in a cohort of brain cancer patients treated with PBS PT.

In this work, we aim to evaluate the incidence of RIA in a cohort of patients treated with PBS-PT at the Christie Proton Beam Therapy Centre (Manchester, UK) and to independently validate the NTCP models previously published by Palma et al.[8]. The availability of accurate NTCP models for predicting alopecia could improve plan optimization and could be easily exploited for scalp-sparing RT in brain tumor patients, consequently improving patients' quality of life.

As a general rule, NTCP model validation is essential for their implementation in clinical practice for the following reasons: *i*) to quantify the risk of RIA for a given patient for plan optimisation purposes (i.e. beam arrangement); *ii*) to provide evidence supporting the inclusion of scalp as an OAR in plan optimisation; *iii*) to accurately estimate the risk of RIA for a given patient for consent purposes.

Materials and methods

Patient data/ cohort

Data for patients of all age groups treated for brain tumours at The Christie Proton Beam Therapy centre between December 2018 and April 2022 were extracted from the Hospital electronic patient record system (Clinical Web Portal −CWP). These patients had satisfied the UK NHS commissioning criteria to receive PT. In the UK, the primary justification for using PT in the central nervous system (CNS) for patients younger than 25 is the reduction of late side effects. For this reason, PT is routinely used in this age group to treat good-prognosis CNS malignancies (estimated life expectancy >5 years, non-metastatic disease, no comorbidities likely to reduce the life expectancy, adequate performance status to travel to receive PT). The UK's PT indications for adults over 25 are restricted to extremely rare cases that require dose-escalation close to dose-limiting normal tissues.

Patients were treated with PBS PT technology. As part of the planning preparation, all patients underwent a Computed Tomography (CT) planning scan of 1–2 mm thickness, covering the whole head at least to the C1/C2 junction. For delineation purposes, all patients underwent Magnetic Resonance (MR) imaging with contrast. Immobilisation in the treatment position (using a 3-points thermoplastic mask) always preceded planning imaging. After the acquisition, all planning images were uploaded to the proton treatment planning system (TPS) Eclipse v16.1 (Varian Palo Alto, Ca) to check treatment planning suitability. PT plan generation adhered to Institutional Proton treatment planning protocols, with either single-field or multi-field optimisation, and all patients were subject to patient-specific quality assurance. The dose calculation algorithm was Proton Convolution Superposition v16.1 (PCS), and the dose isotropic grid resolution was 2.5 mm. The PT plans were optimised assuming a constant relative biological effectiveness (RBE) of 1.1. Throughout treatment, patients received daily imaging for setup verification, usually with Cone Beam CT scan for the first 3 fractions and once a week for the following weeks, and orthogonal KV imaging on the remaining fractions.

Patients were followed up according to the national PT follow-up schedule, being reviewed at the Proton Centre where feasible or via video consultations at 6 weeks, 6 months and 12 months post-treatment, and annually thereafter.

Clinically related data along with dosimetric data that were relevant for the model validation were extracted from the TPS. Throughout the process, data handling in an anonymised format as per NHS policies was ensured [11]. The project received Institutional board approval by the Proton Research Committee (Ref. 2021-025).

Alopecia was defined as per CTCAE 5.0 grading system as:

- ***Grade 1 (G1):*** Hair loss of <50 % of the normal scalp for that individual that is not obvious from a distance but only on close inspection; a different hairstyle may be required to cover the hair loss but it does not require a wig or hair piece to camouflage.
- ***Grade 2 (G2):*** Hair loss of ≥50 % of the normal scalp for that individual that is readily apparent to others; a wig or hairpiece is necessary if the patient desires to completely camouflage the hair loss; associated with psychosocial impact.

Acute RIA was defined as hair loss observed ≤90 days since PBT completion, late RIA as that occurring >90 days and permanent RIA as

hair loss persisting >12 months.

Alopecia that was present at baseline (most likely caused by chemotherapy) was not considered RIA if it disappeared in the year following the completion of PT.

NTCP models

In the present study, the NTCP models described by Palma et al. [8] were applied to predict acute, late, and permanent G2 RIA within the cohort of patient treated at the Christie Proton Beam Therapy Centre (hereinafter validation cohort). Those models were developed based on data from patients undergoing PBS-PT for brain tumours at the PT facility in Trento, Italy (hereinafter training cohort). In the above study, the decision on RIA scoring was achieved consensually by two radiation oncologists through the analysis of pictures acquired during the follow-up, aiming to limit intra- and inter-observer variability.

The original research [8] employed two distinct modelling strategies, the Lyman-Kutcher-Burman (LKB) analysis, based on pure dosimetric parameters, and the multivariable logistic regression (MLR) methodology, based on both clinical and dosimetric parameters. Both LKB and MLR models were based on the extraction of the relative scalp dose-surface histograms (DSHs). As reported in [8], in the MLR model, the most predictive dosimetric factors included the relative scalp surface receiving 21 Gy (S_{21Gy}) for acute RIA and S_{25Gy}, for late G2-RIA. Younger age was selected as a risk factor for acute G2-RIA while surgery was selected as a risk factor for late G2-RIA. The near maximum dose $D_{2\%}$ was the only variable selected for permanent RIA. Both LKB and MLR models exhibited high predictive performances in the training cohort (cross-validated ROC-AUCs range 0.86–0.90).

To maintain consistency, in the validation cohort, the same algorithm for the extraction of the scalp region of interest as in the original work [8] was adopted. This method standardizes, through a fully automated implementation, the extraction of the scalp structure thus easing the process of unbiased validation on external cohorts of patients [12,13].

Statistical analysis and validation

In order to externally validate the previously published RIA NTCP models, for each patient in the validation cohort, individual DICOM RT plans were imported, and the DSHs of the scalp were computed using a recently developed tool for Matlab [13,14]. Predicted RIA probabilities were then estimated using the LKB and MLR models described in the previous paragraph.

For the purpose of model validation, late, acute and permanent G2-RIA endpoints were considered. Descriptive statistics were used to report information about the proportions of acute, late and permanent rates of RIA in this cohort. Model performance in the validation cohort was assessed by the calculated area under the receiver operating characteristic (ROC) curve (AUC). The discrimination value on the ROC curve, i.e. the cut-off point optimally classifying patients in a binary prediction problem, was determined by Youden's J statistic. Calibration plots comparing predicted to observed probabilities were also generated. Statistical analysis was performed using R. 4.3.1 and in house developed tool in Matlab [15].

Results

A total of 264 patients started treatment for brain tumours at the Christie Proton Beam Therapy centre with PBS-PT. Clinical and treatment-related characteristics, as well as incidence of RIA, are reported in Table 1. For comparison purposes, RIA incidence for the training cohort is also reported in Table 1.

Of note, 14 patients had reported alopecia at baseline, likely resulting from induction chemotherapy. Of these, 13/14 gradually subsided and resolved in the 12 months following PBT and therefore were not considered in the analysis, but one who persisted, after initial improvement, at 12 months was included in the group of patients with RIA. After accounting for baseline alopecia, rates of any grade acute, late and permanent RIA in our cohort were 61.8 %, 24.7 % and 14.4 %, respectively. Most of the permanent RIA (35/38; 92 %) were preceded by acute or late RIA, or both (Fig. 1). Moreover, the rates of acute, late

Table 1
Patient and treatment characteristics of the validation cohort. Incidence of Radiation Induced Alopecia (RIA) as per CTCAE 5.0 grading system is also reported.

Clinical and treatment characteristics		
	Median	**Range**
Age (years)	12	1–74
Radiotherapy dose (GyRBE)	54	5.4–73.8
	Count	**Percentage**
Gender		
Male	156	59,1
Female	108	40,9
Indication Category		
Ependymoma	54	20,5
Craniopharyngioma	51	19,3
Low Grade Glioma	46	17,4
Germ Cell Tumour (CNS)	36	13,6
Medulloblastoma	31	11,7
Pituitary Adenoma	11	4,2
ATRT	6	2,3
NGGCTs	6	2,3
Chordoma	5	1,9
Others	18	6,8
Disease site		
Supratentorial	189	71,6
Infratentorial	75	28,4
Chemotherapy		
Yes	72	27,3
No	192	72,7
Surgery		
Yes	222	84,1
No	42	15,9
Treatment site		
Focal	194	73,5
Craniospinal	52	19,7
Whole Ventricular	18	6,8
Radiation Induced Alopecia (validation cohort)		
Acute		
Grade 0	101	38,3
Grade 1	109	41,3
Grade 2	54	20,5
Late		
Grade 0	199	75,4
Grade 1	39	14,8
Grade 2	26	9,9
Permanent		
Grade 0	226	85,6
Grade 1	29	11,0
Grade 2	9	3,4
Radiation Induced Alopecia (training cohort)		
Acute		
Grade 0	40	34,5
Grade 1	16	13,8
Grade 2	60	51,7
Late		
Grade 0	41	35,3
Grade 1	21	18,1
Grade 2	41	35,3
NA	13	11,2
Permanent		
Grade 0	80	69,0
Grade 1	3	2,6
Grade 2	20	17,2
NA	13	11,2

Abbreviations: Gy: Gray; RBE: relative biological effectiveness; CNS: central nervous system; ATRT: atypical teratoid rhabdoid tumor; NGGCTs: Non germinomatous germ cell tumours.

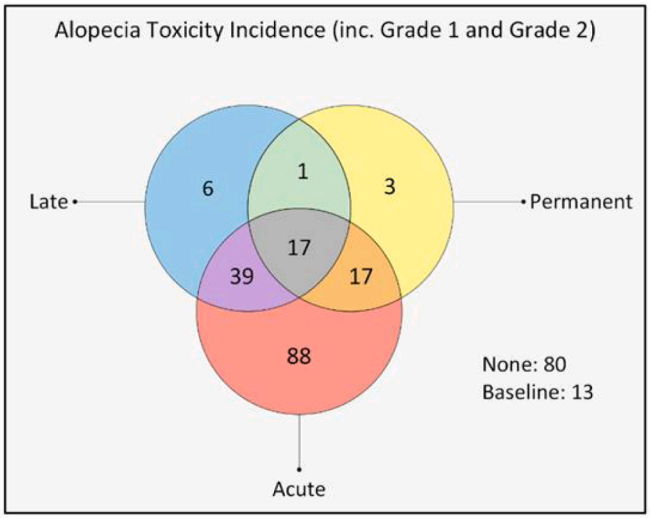

Fig. 1. Venn-diagram reporting the incidence of Grade 1 and Grade 2 alopecia, which includes the patients with baseline alopecia.

and permanent G2-RIA in our cohort were 20.5 %, 9.9 % and 3.4 % respectively.

The validation ROC curves, AUC values and optimal cut-off points of the original MLR- and LKB-NTCP models for acute, late, and permanent RIA are displayed in Fig. 2 and reported in Table 2, respectively.

Despite exhibiting a reasonable discriminative ability [16], all models consistently overestimated the risk estimates of acute, late and permanent RIA. This is evidenced by the slope and intercept values of the calibration curves presented in Fig. 3.

Lastly, we performed a dose–response analysis for the effect of $D_{2\%}$ on the risk of permanent RIA in our patient population, defining the risk categories for the given endpoint as follows:

0–5 %: very low risk.
5–15 %: low risk.
15–25 %: intermediate risk.
25–50 %: high risk.
>50 %: very high risk.

Dose thresholds for each risk category are presented in Fig. 4, illustrating the predicted NTCP of permanent RIA as determined by the LKB model compared to the observed G1 and G2 RIA in our patient population, for given values of $D_{2\%}$ to the scalp surface. It's noteworthy that as the value of $D_{2\%}$ increases, the risk of permanent RIA also increases. This risk is higher for the G1 compared to G2 and it tends to be closer to the predicted NTCP value.

Table 2
Area under the receiver operating characteristic (ROC) curve (AUC) and ROC optimal cut-off points of the Multivariable Logistic Regression (MLR) model and of the Lyman-Kutcher-Burman (LKB) model for acute, late, and permanent Grade 2 Radiation-Induced Alopecia (RIA).

Model	RIA	AUC (95 %CI)	Cut-off points
MLR	Acute	0.81 (0.74–0.88)	0.997
	Late	0.83 (0.74–0.81)	0.635
	Permanent	0.71 (0.54–0.88)	0.087
LKB	Acute	0.79 (0.73–0.85)	0.232
	Late	0.78 (0.71–0.85)	0.348
	Permanent	0.71 (0.53–0.89)	0.068

Discussion

In the Christie-treated validation cohort, over 60 % of patients had some degree of RIA in the first few months after starting treatment. However, in a significant portion of these patients, the hair starts growing back, and the incidence of RIA gradually decreases at the late and permanent time points with the combined incidence of G1 and G2 permanent RIA dropping to 14.4 %. Of all the endpoints examined, permanent RIA is indeed the most significant, as it impacts the patients' quality of life, particularly in terms of their psychological and social well-being and is the one that clinicians should make an effort to reduce through planning optimisation. In our patient population, a value of $D_{2\%}$ above 53 Gy confers a risk >50 % of G2 permanent alopecia, with the risk of the same endpoint falling below 5 % for a D2 % below 26 Gy.

Whilst the permanent RIA is clinically the most relevant and the most dreaded from a patient perspective, the Venn diagram shows how acute RIA plays a role in the development of late and permanent RIA, indicating that the prevention of acute RIA (by using the appropriate NTCP models) could be a helpful strategy to reduce adverse effects down the road.

We first applied the LKB approach, which takes advantage of the average generalised equivalent uniform dose (gEUD) across different dose levels.

In our validation cohort, the three LKB models for the acute, late and permanent RIA showed accurate discrimination between patients with and without RIA.

Palma et al.[8] also looked into the possible influence of clinical parameters on the G2 RIA using a MLR technique. Aside from DSH measures, younger age at RT was identified as a risk factor for acute G2 RIA, whereas surgery was linked to a higher risk of late RIA.

The MLR-NTCP models performed well based on the AUC values for clinician-rated G2 alopecia in our independent cohort, despite the differences in the median age between the training and validation cohorts

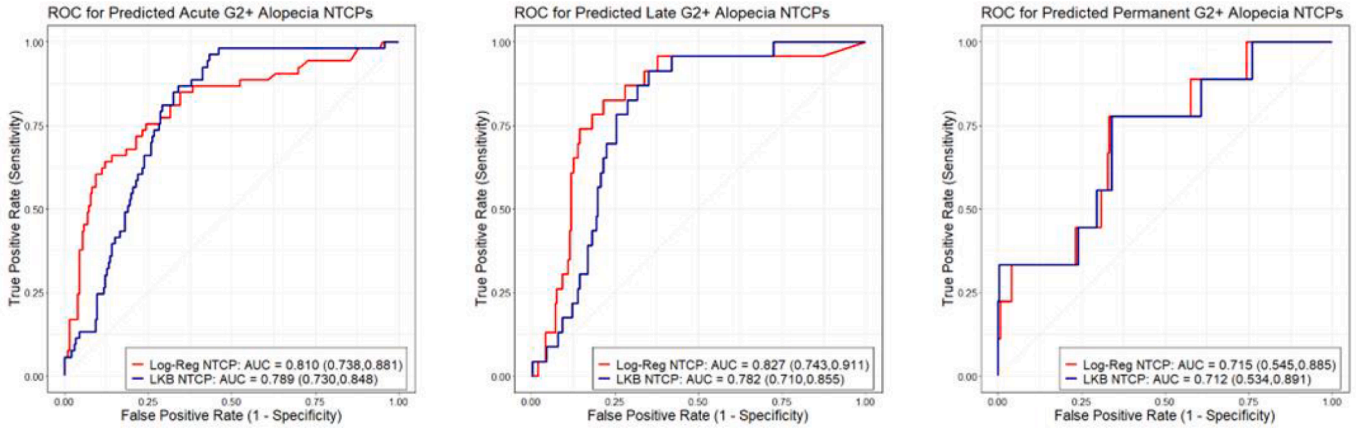

Fig. 2. Acute, Late and Permanent Grade 2 Radiation Induced Alopecia (RIA) modelling. Receiving Operator Characteristic (ROC) curves of Lyman-Kutcher-Burman (LKB) and Multivariable Logistic Regression (MLR) models.

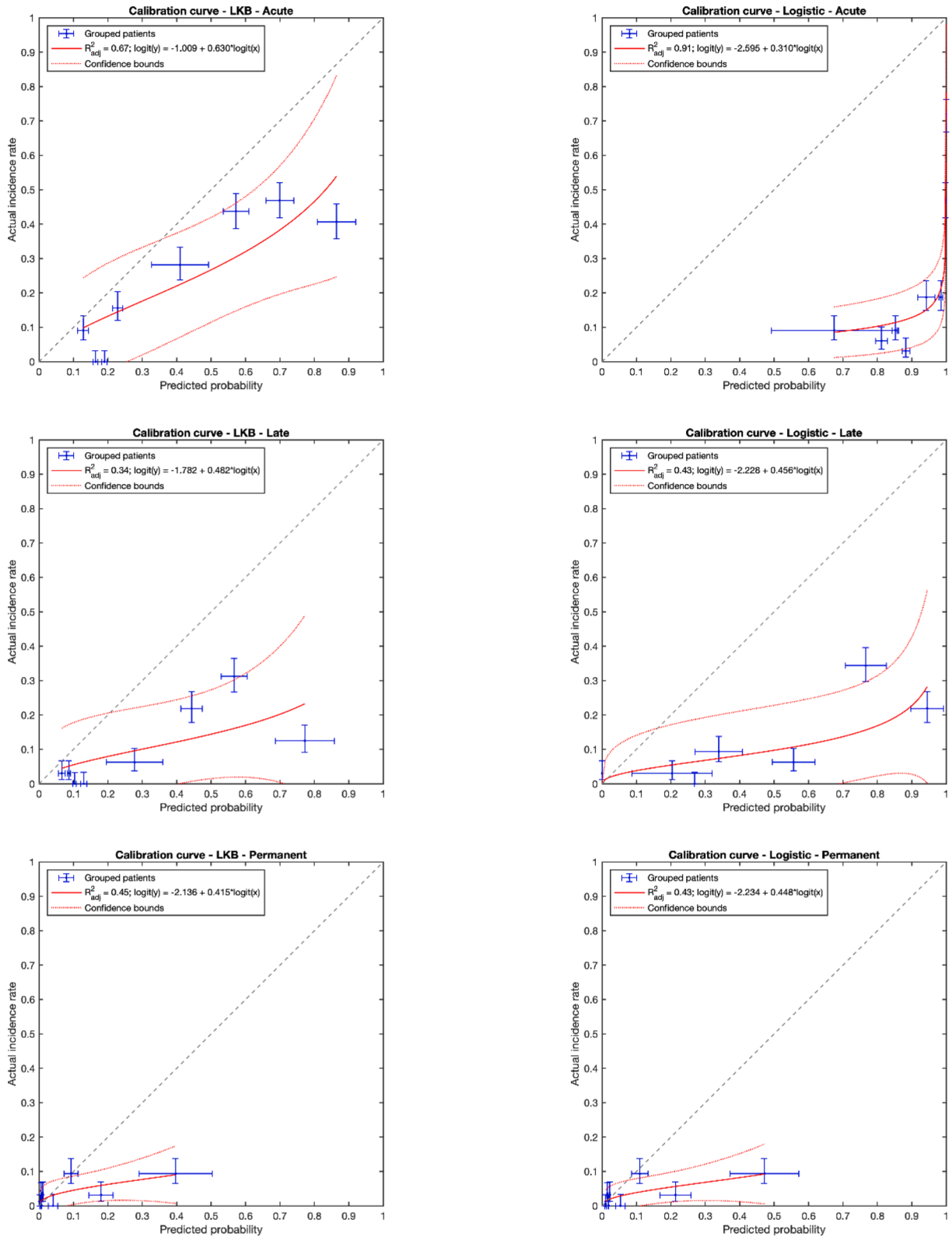

Fig. 3. Calibration curves of Lyman-Kutcher-Burman (LKB) and Multivariable Logistic Regression (MLR) models for acute, Late and Permanent Grade 2 Radiation Induced Alopecia (RIA) modelling.

(56 vs. 17 years, respectively). However, the issue identified in the external validation of RIA models pertain to weak calibration, resulting in a systematic overestimation of G2 RIA predictions within our patient population.

Several factors can introduce distortions in the calibration of risk predictions unrelated to the development of the algorithm itself.

Of note, in the training cohort reported by Palma et al. the combined incidence rates of G1 and G2 acute, late, and permanent RIA were 65.5 %, 53.4 % and 19.8 % whilst the corresponding rates observed in our validation cohort were 61.8 %, 24.7 % and 14.4 % respectively. This

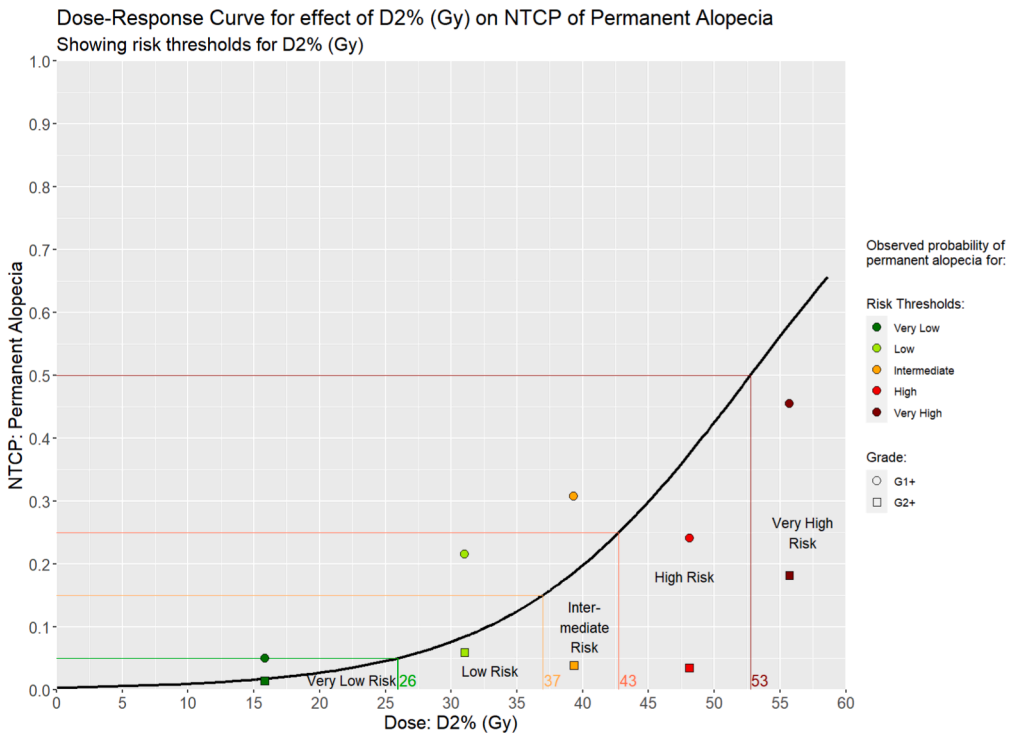

Fig. 4. Dose-Response Curve correlating the effect of D2% (Gy) to the scalp structure with the risk of Grade 1 and Grade 2 permanent radiation-induced alopecia (RIA). Dose thresholds are given for the risk categories listed in the text: very low risk, low risk, intermediate risk, high risk, very high risk.

difference is more evident when the rates of G2 acute, late, and permanent RIA are considered separately, being 51.7 %, 35.3 % and 17.2 % in the training cohort and 20.5 %, 9.9 % and 3.4 % in the validation cohort, respectively.

This could be a reflection of different clinical and treatment-related characteristics between the two cohorts. In fact, whilst the median RT prescription dose is the same between the training and validation cohorts (54 GyRBE), some clinical and treatment characteristics were significantly different between the two. These differences likely originate from the different commissioning policies for PT in the two countries. Firstly, the median age was 56 years in the training and 17 years in the validation cohort. In the training cohort, the prevalent tumour histologies were Gliomas (49 %), followed by Meningiomas (29 %), base of skull tumours (13 %) and other rare tumours (9 %). Conversely, the validation cohort is dominated by tumour histologies that are prevalent in the paediatric/ young adults population, the 5 most common being Ependymomas (20.45 %), Craniopharyngiomas (19.32 %), Low Grade Gliomas (17.42 %), Germ Cell tumours (13.64 %) and Medulloblastomas (11.74 %). In addition, whilst the proportion of patients receiving surgery is slightly higher in the validation compared to the training cohort (84.09 % vs 79.3 %), the proportion of patients receiving chemotherapy is significantly higher in the training compared to the validation cohort (44.8 % vs 27.27 %), as a reflection of the different prevalent histologies (higher proportion of adult high-grade gliomas in the former compared to the latter). Other two noticeable differences are that 36.2 % of patients in the training cohort have received re-irradiation for recurrent disease (subgroup not represented in the validation cohort), whereas in the validation cohort 26.5 % of patients have received non-focal radiotherapy (craniospinal or whole ventricular) for specific diagnoses.

Nevertheless, the significant impact of varying interpretations and grading of RIA among operators from different institutions cannot be disregarded. Specifically, a distinct methodology for patient follow-up is apparent between the two Institutions, particularly regarding remote consultations utilizing video follow-ups through a dedicated platform in the external cohort.

Model generalizability has been a long-debated topic, and it is likely related to different factors.

Establishing reliable risk estimates in clinical outcomes data collection requires comprehensive national databases and standardized scoring systems. The UK's national registry for proton outcomes data addresses this need, yet systemic variations in data recording and reporting may hinder model generalizability. Patient population differences between nations affect proton therapy commissioning guidelines and patient selection, leading to demographic and disease-related variations between training and validation cohorts [17,18]. In addition, uncertainty in TPS dose computations, particularly in the build-up region, poses a challenge to model robustness. Recent research by Wang et al. [19] demonstrated the limitations of the Eclipse TPS in terms of estimating patient skin dose, which was determined as the mean dose to a contoured structure of 0.5 cm thickness from the surface. When the patient's external body contour begins at their skin, there has been a reported underestimation of skin doses by as much as 14 % of the prescribed dose. Even though in this model's definition the scalp region does not reach the body contour, but remains 2 mm below this, this uncertainty around dose computations is only partially mitigated. NTCP models, based on dose-volume histograms (DVHs) or DSHs, have intrinsic limitations due to their reliance on artificial models and the exclusion of organ-specific spatial information, impacting their accuracy in representing dose distributions within organs [20,21].

The results of our LKB and MLR NTCP model validation for RIA offer a clear picture of the limited transferability of NTCP models that rely on clinical parameters across patient populations with different characteristics, less so for NTCP models that rely on dosimetric parameters only. Moreover, discrepancies in the grading of toxicities across Institutions might also hinder model generalizability, since granular and consistent outcome data collection is paramount to build robust NTCP models of toxicities. Therefore, it is essential to ensure that outcomes data are consistently captured and reported using standardized methodologies based on photographs/ video. Implementing protocols for collecting RIA data should involve utilizing more accurate grading scale, such as the Sinclair scale. This scale better describes the magnitude and severity of the RIA considering factors like hair density and thickness, leading to

improved reproducibility and objectivity in the evaluation [22].

In cases where we identify poorly calibrated predictions during validation, one strategy could contemplate algorithm updating. This process aims to enhance the accuracy of predictions for new patients within the validation setting. Updating strategies are gaining in popularity. An example is provided by the closed testing procedure (CTP) described by Vergouwe et al. [23] and implemented by a Dutch group for the validation of an NTCP dysphagia model [24].

Nowadays, PBS PT is the most widely used PT technique, with improved modulation capabilities and flexibility in treatment planning due to increased opportunities to tune the energy and spot size, and the opportunity to overcome the limitations of the older passive scattering PBT techniques. In this setting, the value of NTCP models for selected toxicity endpoints cannot be understated.

Conclusions

NTCP models represent a key strategy to maximise the benefit of technological advances in radiotherapy, such as PBS-PT [25]. The implementation of NTCP models for different toxicity endpoints can allow better prediction and prevention of toxicities, for consent purposes and, whenever possible, planning optimisation [26]. A crucial step in integrating prediction models into clinical practice is model validation. However, differences between the cohorts used for model creation and validation may affect their generalizability. It is crucial to assess the differences between different Institutional practices in toxicity grading, as differences in the way clinical outcomes are reported inevitably jeopardize the validation of RIA NTCP models and therefore their implementation in clinical practice. Implementing a standardized and objective RIA scoring system is essential.

CRediT authorship contribution statement

Simona Gaito: Writing – original draft, Data curation, Conceptualization. **Laura Cella:** Writing – original draft, Supervision, Conceptualization. **Anna France:** Formal analysis. **Serena Monti:** Writing – review & editing, Software. **Gillian Whitfield:** Writing – review & editing. **Peter Sitch:** Data curation. **Neil Burnet:** Writing – review & editing. **Ed Smith:** Data curation. **Giuseppe Palma:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Conceptualization. **Marianne Aznar:** Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgements

Dr Marianne Aznar acknowledges the support of the Engineering and Physical Research Council (Grant number EP/T028017/1), Cancer Research UK RadNet Manchester (C1994/A28701) and the NIHR Manchester Biomedical Research Centre (NIHR203308).

References

[1] Yeshua-Katz D, Shvarts S, Segal-Engelchin D. Hierarchy of hair loss stigma: Media portrayals of cancer, alopecia areata, and cancer in Israeli newspapers. Isr J Health Policy Res 2019;8. https://doi.org/10.1186/s13584-019-0338-0.
[2] Quesada S, Guichard A, Fiteni F. Cancer-related alopecia: From etiologies to global management. Cancers (Basel) 2021;13. https://doi.org/10.3390/cancers13215556.
[3] Phillips GS, Freret ME, Friedman DN, Trelles S, Kukoyi O, Freites-Martinez A, et al. Assessment and treatment outcomes of persistent radiation-induced alopecia in patients with cancer. JAMA Dermatol 2020;156. https://doi.org/10.1001/jamadermatol.2020.2127.
[4] Shakespeare TP, Dwyer M, Mukherjee R, Yeghiaian-Alvandi R, Gebski V. Estimating risks of radiotherapy complications as part of informed consent: The high degree of variability between radiation oncologists may be related to experience. Int J Radiat Oncol Biol Phys 2002;54. https://doi.org/10.1016/S0360-3016(02)02996-6.
[5] Lawenda BD, Gagne HM, Gierga DP, Niemierko A, Wong WM, Tarbell NJ, et al. Permanent alopecia after cranial irradiation: Dose-response relationship. Int J Radiat Oncol Biol Phys 2004;60. https://doi.org/10.1016/j.ijrobp.2004.04.031.
[6] Scoccianti S, Simontacchi G, Greto D, Perna M, Terziani F, Talamonti C, et al. Dosimetric predictors of acute and chronic alopecia in primary brain cancer patients treated with volumetric modulated arc therapy. Front Oncol 2020;10. https://doi.org/10.3389/fonc.2020.00467.
[7] Dutz A, Lühr A, Agolli L, Bütof R, Valentini C, Troost EGC, et al. Modelling of late side-effects following cranial proton beam therapy. Radiother Oncol 2021;157. https://doi.org/10.1016/j.radonc.2021.01.004.
[8] Palma G, Taffelli A, Fellin F, D'Avino V, Scartoni D, Tommasino F, et al. Modelling the risk of radiation induced alopecia in brain tumor patients treated with scanned proton beams. Radiother Oncol 2020;144. https://doi.org/10.1016/j.radonc.2019.11.013.
[9] Pedersen J, Flampouri S, Bryant C, Liang X, Mendenhall N, Li Z, et al. Cross-modality applicability of rectal normal tissue complication probability models from photon- to proton-based radiotherapy. Radiotherapy Oncol 2020;142. https://doi.org/10.1016/j.radonc.2019.09.017.
[10] Gaito S, Abravan A, Richardson J, Lowe M, Indelicato DJ, Burnet N, et al. Skin toxicity profile of photon radiotherapy versus proton beam therapy in paediatric and young adult patients with sarcomas. Clin Oncol 2021;33. https://doi.org/10.1016/j.clon.2021.03.009.
[11] CRUK. https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines n.d.
[12] Martel JL, Miao JH, Badri T. Anatomy, Hair Follicle. 2022.
[13] Palma G, Cella L. A new formalism of Dose Surface Histograms for robust modeling of skin toxicity in radiation therapy. Physica Medica 2019;59. https://doi.org/10.1016/j.ejmp.2019.02.005.
[14] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. Med Phys 2003;30. https://doi.org/10.1118/1.1568978.
[15] Monti S, Xu T, Mohan R, Liao Z, Palma G, Cella L. Radiation-Induced esophagitis in non-small-cell lung cancer patients: voxel-based analysis and NTCP modeling. Cancers (Basel) 2022;14. https://doi.org/10.3390/cancers14071833.
[16] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5. https://doi.org/10.1097/JTO.0b013e3181ec173d.
[17] Hwang E, Gaito S, France A, Crellin AM, Thwaites DI, Ahern V, et al. Outcomes of patients treated in the UK proton overseas programme: non-central nervous system group. Clin Oncol 2023. https://doi.org/10.1016/j.clon.2023.02.009.
[18] Gaito S, Hwang EJ, France A, Aznar MC, Burnet N, Crellin A, et al. Outcomes of patients treated in the UK proton overseas programme: central nervous system group. Clin Oncol 2023. https://doi.org/10.1016/j.clon.2023.01.024.
[19] Wang L, Cmelak AJ, Ding GX. A simple technique to improve calculated skin dose accuracy in a commercial treatment planning system. J Appl Clin Med Phys 2018;19. https://doi.org/10.1002/acm2.12275.
[20] Amols HI, Zaider M, Hayes MK, Schiff PB. Physician/patient-driven risk assignment in radiation oncology: reality or fancy? Int J Radiat Oncol Biol Phys 1997;38. https://doi.org/10.1016/S0360-3016(97)00093-X.
[21] Gaito S, Burnet N, Aznar M, Crellin A, Indelicato DJ, Ingram S, et al. Normal tissue complication probability modelling for toxicity prediction and patient selection in proton beam therapy to the central nervous system: a literature review. Clin Oncol 2022;34. https://doi.org/10.1016/j.clon.2021.12.015.
[22] Kasprzak M, Sicińska J, Sinclair R. The trichoscopy derived sinclair scale: enhancing visual assessment through quantitative trichoscopy. Australas J Dermatol 2019;60. https://doi.org/10.1111/ajd.12964.
[23] Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 2017;36. https://doi.org/10.1002/sim.7179.
[24] Kalendralis P, Sloep M, Moni George N, Snel J, Veugen J, Hoebers F, et al. Independent validation of a dysphagia dose response model for the selection of head and neck cancer patients to proton therapy. Phys Imaging. Radiat Oncol 2022;24. https://doi.org/10.1016/j.phro.2022.09.005.
[25] Pacelli R, Caroprese M, Palma G, Oliviero C, Clemente S, Cella L, et al. Technological evolution of radiation treatment: Implications for clinical applications. Semin Oncol 2019;46. https://doi.org/10.1053/j.seminoncol.2019.07.004.
[26] Palma G, Monti S, Conson M, Pacelli R, Cella L. Normal tissue complication probability (NTCP) models for modern radiation therapy. Semin Oncol 2019;46. https://doi.org/10.1053/j.seminoncol.2019.07.006.