INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Normal Tissue Complication Probability Model for Acute Radiation Dermatitis in Patients With Head and Neck Cancer Treated With Carbon Ion Radiation Therapy

Yang Li, MD, PhD,[*,†] Makoto Sakai, PhD,[*] Anna Tsunoda,[‡] Nobuteru Kubo, MD, PhD,[*,‡] Yoko Kitada,[*] Yoshiki Kubota, PhD,[*] Akihiko Matsumura, PhD,[*] Yuan Zhou, MD,[‡] and Tatsuya Ohno, MD, PhD[*,‡]

[*]*Gunma University Heavy Ion Medical Center, Maebashi, Japan;* [†]*Department of Radiation Oncology, Harbin Medical University Cancer Hospital, Harbin, China; and* [‡]*Department of Radiation Oncology, Gunma University Graduate School of Medicine, Maebashi, Japan*

Received Sep 10, 2021; Accepted for publication Mar 2, 2022

Purpose: This study aimed to explore the prognostic factors associated with acute radiation dermatitis (ARD). A normal tissue complication probability (NTCP) model for ARD in patients with head and neck cancer treated with carbon ion radiation therapy was developed.

Materials and methods: A total of 187 patients were included in the analysis, and the endpoint was ≥grade 2 ARD. The biological and physical dose−surface parameters associated with ARD were used in the logistic regression model. The mean areas under the receiver operating characteristic curve in the internal cross-validation and Akaike's corrected Information Criterion were examined for model evaluation and selection. The multivariate logistic regression NTCP models were established based on factors with weak correlation.

Results: Tumor volume, planning target volume to the skin, radiation technique and all dose−surface parameters were significantly associated with ARD ($P < .05$). Models with high performance for grade 2 to 3 ARD were constructed. The most significant prognostic predictors were $S_{40\ Gy\ (relative\ biological\ effectiveness,\ RBE)}$ and $S_{20\ Gy}$ (absolute surface area receiving RBE-weighted dose of 40 Gy (RBE) or physical dose of 20 Gy). The internal cross-validation−based areas under the receiver operating characteristic curve for models with $S_{40\ Gy\ (RBE)}$ and $S_{20\ Gy}$ were 0.78 and 0.77, respectively. The biological and physical dose−surface parameters had similar performance at various dose levels. However, the performance of the multivariate NTCP models based on 2 factors was not better than that of the univariate models.

Conclusions: NTCP models for ARD may provide a basis for the development of individualized treatment strategies and reduce the incidence of severe ARD in patients with head and neck cancer receiving carbon ion radiation therapy. Furthermore, biological and physical dose−surface parameter-based models are comparable. However, further validation with more evaluation parameters is warranted.

Corresponding author: Makoto Sakai, PhD; E-mail: sakai-m@gunma-u.ac.jp

Disclosures: none.

All data are available for readers who need the data.

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijrobp.2022.03.002.

Acknowledgments—The authors thank the radiology technologists, medical doctors, and medical physicists at Gunma University Heavy Ion Medical Center for their valuable insights. We also thank Xiangdi Meng in the Department of Radiation Oncology at Weifang People's Hospital for his guidance in statistical analyses.

Int J Radiation Oncol Biol Phys, Vol. 113, No. 3, pp. 675−684, 2022
0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2022.03.002

Introduction

Acute radiation dermatitis (ARD) is a common side effect of radiation therapy (RT), particularly in head and neck cancer (HNC), because the tumor is closer to the skin than to other sites. Up to 95% of patients who receive photon RT may develop ARD, and severe ARD can cause evident discomfort, significantly affecting the quality of life and even the treatment.[1-3] Particle RT is correlated with a low incidence of radiation-induced late side effects because it can achieve high dose conformity while sparing a large number of healthy tissues surrounding the tumor.[4] Thus, compared with photon RT, particle RT is a good option.[5] However, carbon ion RT (CIRT) may also cause similar side effects, such as ARD. A phase 2 CIRT trial that assessed 236 patients with HNC reported that 93% of patients developed ARD.[6] Moreover, 44% of the patients presented with ≥grade 2 ARD according to the Radiation therapy and Oncology Group scoring system.[6] Based on the Radiation therapy and Oncology Group or the Common Terminology Criteria for Adverse Events scoring system, severe ARD (grade 3) was observed in 10%, 13%, and 15% of patients with major salivary gland carcinomas, sinonasal adenocarcinoma and locally advanced-stage squamous cell carcinoma of the ear, respectively.[7-9] The skin volume included in the target volume is a major indicator for determining the treatment protocol. A study reported that, to prevent skin or brain exposure to high-dose irradiation, the doses prescribed to the target area were compromised in 26 of 46 patients (57%) with parotid tumors receiving CIRT.[10] Although the available clinical data are still limited and the number of patients enrolled in these studies is small, particular attention must be paid to ARD in patients receiving CIRT.

Several risk factors are associated with ARD, including host factors such as obesity, age, and smoking, as well as treatment-related factors such as total skin dose, type of radiation, and chemotherapy.[11] Among these factors, the dose surface/volume parameter is the most important predictor of ARD, and several normal tissue complication probability (NTCP) models based on these predictors have been developed for various tumors.[12-15] NTCP models can be used to predict side effects based on specific risk factors, which may facilitate the development of personalized treatment strategies. Therefore, NTCP model-based treatment strategies should consider specific treatment modalities and clinical characteristics. However, NTCP models derived from photon or proton treatment may not be reliable for patients receiving CIRT, in which the relative biological effectiveness (RBE) should be considered in clinical dose calculations. Although the proton dose includes RBE, the CIRT beam usually has significantly higher RBE (approximately 3.0 vs 1.1). Thus, the biological effects significantly differ.[16,17] Moreover, several factors can affect RBE.[18,19] Considering early skin reaction as the endpoint, several studies showed that the RBE values for the skin ranged from 1.5 to 3.0, which are dependent on linear energy transfer and the number of fractions.[20-22] However, these data were obtained from small animal or in vitro studies, and there is no direct evidence of the clinical outcomes. In fact, RBE calculation in clinical settings may become more complicated as a result of the complex correlation between the tumor and skin in HNC (ie, the skin is near the irradiation field or even included in the target volume). Considering the limitations in RBE estimation, using biological dose-related parameters may introduce potential uncertainties in establishing an NTCP model. A comparison study considering both biological and physical doses must be performed. To the best of our knowledge, no study has used CIRT-based NTCP models for ARD in patients with HNC.

Therefore, this study analyzed the incidence of ARD in patients with HNC treated with CIRT and developed NTCP models for ARD based on multivariate logistic regression. ARD grades 2 and 3 (scored by Common Terminology Criteria for Adverse Events 4.0) were analyzed as the endpoint due to their significant effect on the quality of life. Moreover, the prognostic factors associated with the endpoint were evaluated, and these were used in the NTCP models to assess the individual risk or to guide clinical decision-making to minimize skin damage. Furthermore, the performances of the NTCP models based on biological and physical dose−surface parameters were compared.

Methods and Materials

Patients

We retrospectively evaluated 203 consecutive patients with pathologically confirmed HNC from June 2010 to August 2020. Recurrences were confirmed via tissue biopsy or clinically determined by the internal review board. At our facility, all patients are treated with CIRT using the passive irradiation method. Patients who received reirradiation to the same lesion and hypofractionated CIRT (<16 fractions), those with intracranial tumors, those aged <18 years, and those with interrupted treatment were excluded from the analysis. Finally, 187 patients were enrolled in this study. The tumor characteristics of the patients are shown in Table 1. This study was approved by the institutional review board of Gunma University (no.: HS2019-218).

Treatment planning and irradiation method

Treatment planning was established using the XiO-N system, in which a pencil beam algorithm is used (Elekta Sweden, Mitsubishi Electric, Japan).[23] We used Gy (RBE) as the unit of biological dose, which was calculated based on the physical dose (Gy) multiplied by RBE.[16] The patients were placed in the supine or prone position with customized cradles (MoldCare; ALCARE, Japan) and body shells (Shell Fitter; Sanyo Polymer Industrial, Japan). A mouthpiece was used in patients in whom the position of the lower jaw was required to be maintained. Two clinical target volumes

Table 1 Tumor characteristics

Tumor location	Tumor volume (mL) mean ± SD	Tumor to skin (mm) mean ± SD	Grade 0-1 no. (%)	Grade 2-3 no. (%)
Pharynx				
Nasopharynx	25.7 ± 36.9	44.9 ± 16.1	6 (3.2)	2 (1.1)
Oropharynx	40.7 ± 24.3	43.6 ± 13.5	6 (3.2)	0 (0.0)
Salivary gland				
Parotid gland	30.6 ± 20.9	7.5 ± 7.2	15 (8.0)	8 (4.3)
Submandibular gland	56.7 ± 39.4	21.1 ± 20.8	0 (0.0)	2 (1.1)
Sublingual gland	27.1 ± 16.9	17.0 ± 2.2	2 (1.1)	0 (0.0)
Nasal cavity	34.3 ± 27.7	17.2 ± 12.8	38 (20.3)	30 (16.0)
Paranasal sinus	68.7 ± 46.5	12.3 ± 9.1	12 (6.4)	18 (9.6)
Oral cavity	31.9 ± 32.0	23.2 ± 11.6	20 (10.7)	5 (2.7)
Neck	58.3 ± 110.7	6.7 ± 4.9	0 (0.0)	5 (2.7)
Others				
External auditory canal	39.4 ± 26.8	1.6 ± 3.2	3 (1.6)	1 (0.5)
Eye socket	23.5 ± 16.2	17.0 ± 12.9	2 (1.1)	2 (1.1)
Auricle	0.9	0.0	1 (0.5)	0 (0.0)
Bone	158.7 ± 121.4	5.8 ± 2.9	1 (0.5)	8 (4.3)

Abbreviation: SD = standard deviation.

(CTV1 and CTV2) were applied to achieve the desired dose distribution. CTV1 was used in the first 9 or 10 fractions (the latter was used in patients receiving a dose of 57.6 Gy [RBE]), which generally encompassed the entire anatomic site of the tumor origin (eg, nasal cavity and parotid grand). Meanwhile, CTV2 was used in all fractions with the total prescription dose. CTV2 was established by adding a 3-mm margin to the gross tumor volume or resection cavity, with consideration of contrast-enhanced magnetic resonance imaging results acquired before the irradiation. Two planning target volumes (PTV1 and PTV2) were generated accordingly by adding a 2-mm margin to CTV1 and CTV2, respectively. The prescribed doses of 57.6, 64, or 70.4 Gy (RBE) were delivered in 16 fractions according to the types of tumors and the specific correlation between the tumor and the surrounding organ at risk. Conventional irradiation and layer-stacking irradiation, which are passive irradiation methods, were applied (Fig. E1). Layer-stacking irradiation can provide a more conformal dose distribution with multiple thin-layer irradiation fields by stacking a finite number of mini spread-out Bragg peaks along the depth direction.[24] Our facility uses a fixed-beam port system with only the horizontal and vertical beam ports available. Thus, to achieve a satisfactory dose distribution, 2 to 4 beams at different directions were obtained by rotating the treatment couch within limited angles along its long axis. At least 95% of the prescribed dose should cover the PTV. To prevent late grade 3 skin toxicity, the surface area of the skin receiving 60 Gy (RBE) ($S_{60\ Gy\ (RBE)}$) should be <20 cm^2.[25]

Acute skin side effects and evaluation parameters

At our facility, skin and oral mucosal side effects are investigated before treatment, at every 4 fractions, at the end of CIRT and 3 months after CIRT. In this study, we focused on grades 0, 1, 2 and 3 ARD, which were prospectively scored by physicians or nurses using the CTCAE scoring system version 4.[26] However, none of the patients presented with grade 4 or 5 ARD. Before irradiation (baseline), the skin status of all patients was determined to be normal and was assigned with grade 0. The maximum severity scores of toxicities within 90 days of exposure to CIRT were analyzed. Most grade 2 or 3 ARD occurred at the end of treatment or during the follow-up period. We categorized the patients into two groups based on grade 0 to 1 and 2 to 3 ARD because the latter frequently results in significant discomfort and is correlated with late ARD.[25]

We investigated patient-specific and treatment-related factors that might be associated with acute skin conditions. These factors included age, sex, tumor volume, stage, surgery, chemotherapy, RT techniques, prescribed dose, and dose–surface histogram (DSH) parameters. Moreover, to evaluate the effect of tumor location on the endpoint, the distance from the most anterior side of PTV2 (on the center slice) to the skin surface along the irradiation path was assessed and defined as the PTV to skin. A negative value indicated that the anterior PTV margin was outside the body surface. A margin that extends outside the body is

usually used in cases in which the tumor is extremely close to or even invading the skin to compensate for any possible interfraction motion. MIM Maestro (version 7.1.3; MIM Software, OH) was used to calculate the skin volume, which was defined as the region within 0.2 cm under the skin surface.[12,27] The skin surface area (cm^2) was defined as the skin volume divided by 0.2.

To select an optimal model, both biological and physical surface doses were calculated. Biological doses ranging from 5 to 50 Gy (RBE) at increments of 5 Gy (RBE) and physical dose−surface parameters ranging from 2.5 to 25 Gy at increments of 2.5 Gy were used. $S_{X\ Gy/Gy\ (RBE)}$ was used as the evaluation metric to analyze the dose−skin surface area, which was defined as the skin surface area (cm^2) irradiated with at least X Gy/Gy (RBE).

Statistical analysis and construction of NTCP models

Multivariate logistic regression analysis was performed to evaluate the effect of both clinical factors and dose-surface parameters on acute skin side effects using the following formula:

$$NTCP = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots)}} \quad (1)$$

where $\beta_1, \beta_2 \ldots$ are the fit parameters of the corresponding variables (factors associated with ARD) of x_1, x_2. To compare the performance of the models with biological dose and physical dose−surface parameters, 3 groups were established based on the following doses: low dose (5-15 Gy (RBE)/2.5-7.5 Gy), middle dose (20−35 Gy (RBE)/10-17.5 Gy), and high dose (40-50 Gy (RBE)/20-25 Gy). A 10-fold internal cross-validation with 500 repetitions was performed. Subsequently, we calculated each model's mean area under the receiver operating characteristic curve (AUC). The AUC values were used to quantify classification performance and ranged between 0.5 and 1.0, with 0.5 denoting bad discrimination and 1.0 denoting excellent discrimination. A higher value indicates better model performance.[28] To obtain an accurate model for clinical use, the AUC should be >0.5. The prediction performance was assessed using Akaike's corrected Information Criterion (AICc), which considers the model deviance as well as the sample size and number of estimated parameters in the model.[29] The lower the value, the better the ability to forecast. Finally, the model with the lowest AICc and highest AUC (if possible) value was selected as the best NTCP model in each group. Accordingly, the most significant dose parameters associated with ARD were identified.

The independent clinical factors were determined using multivariate logistic regression analysis. Then, Spearman correlation analysis was performed to examine the correlations between the independent clinical factors and the dose−surface parameters associated with the endpoint. These analyses were based on bootstrapping with 1000 samples. In cases with a coefficient $|\rho| < 0.5$, multivariate logistic regression models were established.

The estimated fit parameters of β values with 95% confidence intervals were calculated using the profile likelihood method. The AUC and calibration plots were used to evaluate model performance. To visualize how well the predicted probabilities were calibrated, calibration curves were generated based on bootstraps with 1000 samples for each model. A perfectly accurate prediction model follows the 45° reference line. The goodness of fit was assessed using Nagelkerke's R^2, which was calculated as a pseudo measure of explained variance; it ranges in value from 0 to 1, with higher values indicating a better model fit.[30] The differences among dose−surface parameters were examined using the Mann−Whitney U test. Statistical Package for the Social Sciences software version 20 (IBM Corporation, Armonk, NY) and in-house written R software (version 4.1. 0) were used for statistical analyses. A P value of $< .05$ was considered statistically significant.

Results

Of the 187 patients, 107 experienced grade 0 (n = 9) and grade 1 (n = 98) ARD and 80 (42.8%) experienced grade 2 (n = 75) and grade 3 (n = 5) ARD during treatment. Approximately 60% of the patients presented with grade 2 to 3 ARD in the nasal cavity and paranasal sinus. However, in patients with tumors with a relatively deep location such as the oral cavity, the risk of grade 2 to 3 ARD was significantly lower than that of grade 0 to 1 ARD (2.7% vs. 10.7%) (Table 1). None of the patients developed >grade 3 ARD. Figure 1 shows the absolute biological and physical dose distributions for ARD in 4 tumor sites.

The average DSHs with standard deviations for all patients showed that the skin area exposed to irradiation in patients with grade 2 to 3 ARD was greater than that in those with grade 0 to 1 ARD ($P < .0001$) (Fig. 2). All dose−surface parameters proposed in this study, either biological or physical dose, had a significant effect on the endpoint ($P < .0001$) (Table 2). Three pairs of dose−surface parameters with the lowest AICc and largest AUC for each dose level were selected for comparison. The selected models' AUCs for biological and physical doses were 0.732 and 0.726 in the low-dose group, 0.779 and 0.776 in the middle-dose group, and 0.777 and 0.772 in the high-dose group, respectively. The performances of the NTCP models between the 2 types of doses were highly comparable (Fig. 3). $S_{40\ Gy\ (RBE)}$ and $S_{20\ Gy}$ derived from the high-dose group had the best model performance among all models. Although their AUC values were slightly lower than those of the middle group, the difference was negligible. In terms of the improved prediction ability (lower AICc), models based on $S_{40\ Gy\ (RBE)}$ and $S_{20\ Gy}$ were chosen as the final models. Calibration curves showed that both biological and physical dose-based models had similar predictive abilities. The middle- and high-dose models exhibited satisfactory

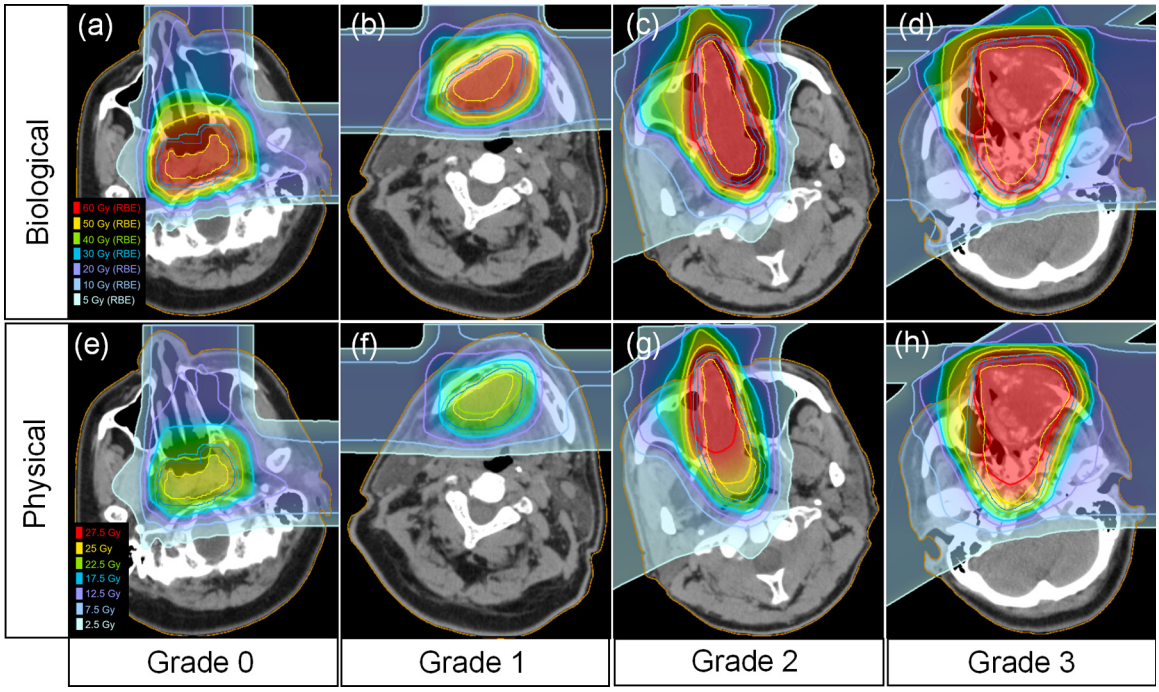

Fig. 1. Absolute dose distributions for acute radiation dermatitis in 4 tumor sites. (a), (b), (c), and (d) show the biological doses for grade 0, 1, 2, and 3 acute radiation dermatitis, respectively. (e), (f), (g), and (h) show the corresponding physical doses. Yellow = gross tumor volume; deep blue = clinical tumor volume 2; blue = planning target volume 2; brown = skin surface; filled colors = dose area.

accuracy of prediction, with a slight tendency to overpredict (<5% on average) when the predicted probability was >60%. Table 2 shows the best-fit parameters of β_0, β_1, and AICc for each model.

Based on multivariate logistic regression analysis, 3 independent prognostic clinical factors (tumor volume, PTV to skin, and radiation technique) were significantly associated with the endpoint ($P < .05$) (see Table E1). In contrast, there was no significant association between the endpoint and chemotherapy (odds ratio, 0.84; range, 0.39-1.82). Among the clinical factors, PTV to skin had the largest AUC (0.758). There was a strong correlation between most dose−surface parameters, except for those between low and high dose−surface parameters (Fig. E2). PTV to skin was significantly correlated with the middle and high dose−surface parameters. There was a weak correlation between tumor volume and high dose−surface parameters. Compared with the conventional broad beam method, the layer-stacking technique saved more low doses, but the middle- and high-dose area savings were comparable (Table E2). However, neither the multivariate logistic regression models' AUC in the internal cross-validation, AICc, nor Nagelkerke's R^2 improved, indicating that the model did not perform as well as the univariate models based on current data sets (Tables 3, E3, and E4).

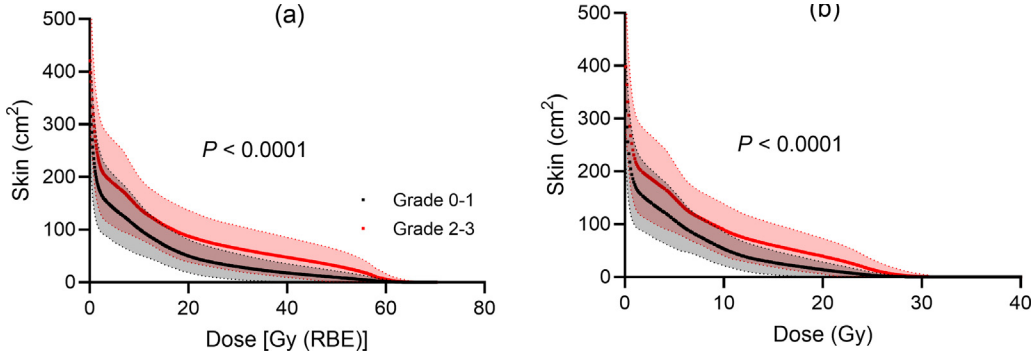

Fig. 2. Dose-surface histogram of all patients with biological doses (a) and physical doses (b). Solid red and black lines indicate the average value for grade 0 to 1 and 2 to 3 acute radiation dermatitis, respectively. The corresponding dotted line indicates the standard deviations.

Table 2 Performance of univariate logistic regression models with dose-surface parameters

Dose	Biological	Mean AUC (95% CI)	β_0 (95% CI)	β_1 (95% CI)	AICc	Physical	Mean AUC (95% CI)	β_0 (95% CI)	β_1 (95% CI)	AICc
Low	$S_{5\ Gy\ (RBE)}$	0.69 * (0.67-0.72)	−1.78 (−2.60 to −1.02)	0.01 (0.00-0.01)	240.6	$S_{2.5\ Gy}$	0.68 * (0.66-0.71)	−1.75 (−2.57 to −0.98)	0.01 (0.00-0.01)	241.8
	$S_{10\ Gy\ (RBE)}$	0.71 * (0.69-0.74)	−1.94 (−2.77 to −1.18)	0.01 (0.01-0.02)	235.8	$S_{5\ Gy}$	0.71 * (0.68-0.73)	-1.83 (−2.62 to −1.09)	0.01 (0.01-0.02)	237.4
	$S_{15\ Gy\ (RBE)}$	0.73 * (0.71-0.76)	−2.06 (−2.87 to −1.32)	0.02 (0.01-0.03)	228.0	**$S_{7.5\ Gy}$**	0.73 * (0.70-0.75)	−2.07 (−2.92 to −1.31)	0.02 (0.01-0.03)	230.8
Middle	$S_{20\ Gy\ (RBE)}$	0.75 * (0.73-0.78)	−1.98 (−2.74 to −1.31)	0.03 (0.02-0.04)	221.3	$S_{10\ Gy}$	0.74 * (0.72-0.77)	−1.93 (−2.67 to −1.25)	0.02 (0.02-0.03)	224.8
	$S_{25\ Gy\ (RBE)}$	0.77 * (0.75-0.79)	−1.96 (−2.692 to −1.31)	0.03 (0.02-0.05)	215.8	$S_{12.5\ Gy}$	0.77 * (0.75−0.80)	−1.95 (−2.68 to −1.30)	0.03 (0.02-0.05)	216.3
	$S_{30\ Gy\ (RBE)}$	0.78 * (0.75-0.80)	−1.87 (−2.56 to −1.26)	0.037 (0.02-0.05)	212.9	$S_{15\ Gy}$	0.78 * (0.75-0.80)	−1.85 (−2.53 to −1.24)	0.04 (0.03-0.06)	213.3
	$S_{35\ Gy\ (RBE)}$	0.78 * (0.76-0.80)	−1.81 (−2.46 to −1.22)	0.04 (0.03-0.06)	210.1	**$S_{17.5\ Gy}$**	0.78 * (0.75-0.80)	−1.70 (−2.32 to −1.14)	0.05 (0.03-0.06)	212.2
High	**$S_{40\ Gy\ (RBE)}$**	0.78 * (0.75-0.80)	−1.70 (−2.32 to −1.15)	0.05 (0.03-0.07)	208.8	**$S_{20\ Gy}$**	0.77 * (0.75-0.79)	−1.55 (−2.12 to −1.04)	0.06 (0.04-0.08)	210.4
	$S_{45\ Gy\ (RBE)}$	0.77 * (0.75 − 0.80)	−1.51 (−2.07 to −1.01)	0.05 (0.04-0.08)	210.7	$S_{22.5\ Gy}$	0.76 * (0.74-0.78)	−1.22 (−1.70 to −0.79)	0.06 (0.04-0.09)	216.6
	$S_{50\ Gy\ (RBE)}$	0.76 * (0.74-0.79)	−1.27 (−1.75 to −0.82)	0.06 (0.04-0.08)	216.1	$S_{25\ Gy}$	0.73 * (0.71-0.75)	−0.87 (−1.26 to −0.50)	0.09 (0.05-0.13)	227.3

Abbreviations: AUC = area under the receiver operating characteristic curve; CI = confidence interval; AICc = Akaike's corrected Information Criterion; RBE = relative biological effectiveness.

* $P < .0001$. The best model in each dose group is highlighted in bold.

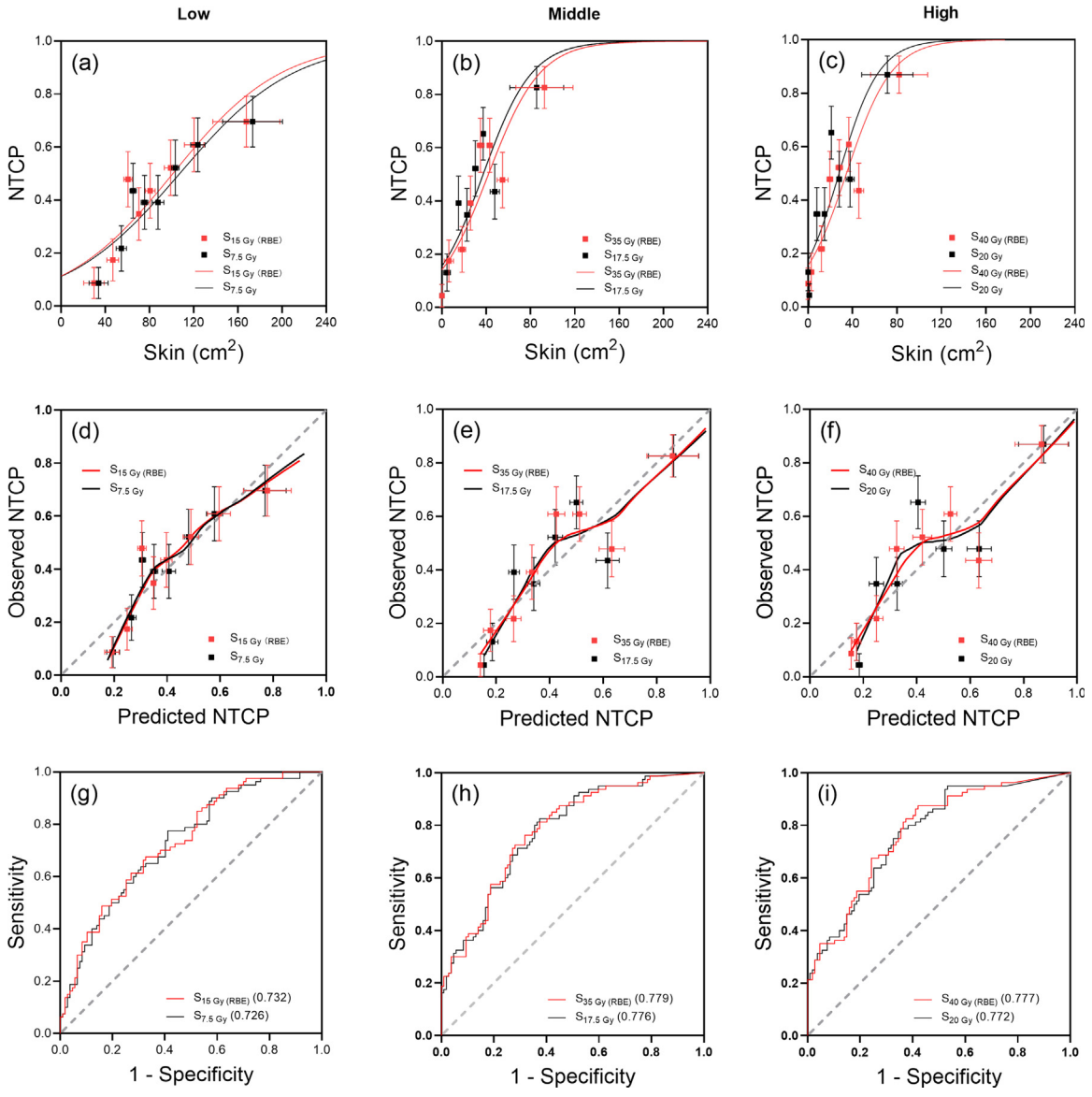

Fig. 3. Univariate logistic regression normal tissue complication probability model with low (a), middle (b) and high (c) dose−surface parameters. The points with error bars indicate observed normal tissue complication probability values with their standard deviation. (d), (e) and (f) show the corresponding calibration plots and curves. A perfectly calibrated model is represented by the gray dotted line. (g), (h), and (i) show the receiver operating characteristic curves. Areas under the curve, followed by the dose−surface parameters, are provided in parentheses.

Discussion

The exact radiobiological effects caused by carbon ion beams on healthy tissues remain unclear. Preclinical or clinical results are currently limited. Thus, the dose constraints for normal tissues have not been well established in CIRT.[31] To the best of our knowledge, this is the first study that constructed an NTCP model for ARD in patients with HNC treated with CIRT.

The incidence of ARD may differ between studies. In the present study, the incidence of grade 2 to 3 ARD was 42.8%. This result was comparable with that reported by Mizoe et al.[6] However, the finding was higher than that reported by Jensen et al, who enrolled 309 patients with HNC receiving intensity modulated RT (IMRT) and boost CIRT, and the incidence of grade 2 to 3 ARD was 24.3%.[32] The main reason may be the differences in cohort composition and radiation regimen. ARD commonly occurred in tumors with a larger volume and a smaller distance to the skin, which include those located at the paranasal sinus, neck and bone (see Table 1). Clinical factors such as tumor volume, distance of tumors to the skin and radiation technique were associated with ARD (see Table E1). Other studies reported similar results.[14,15,33,34] However, in contrast to other reports,[35,36] concurrent chemotherapy had limited effects on the development of ARD, and patient selection bias

Table 3 Performance of multivariate logistic regression models with both clinical factors and dose−surface parameters

Parameter	Mean AUC (95% CI)	β_0 (95% CI)	β_1 (95% CI)	β_2 (95% CI)	AICc
$S_{5\ Gy\ (RBE)}$	0.75 * (0.64-0.87)	−0.91 (−1.82 to −0.03)	0.01 (0.00-0.01)	−0.10 † (−0.15 to −0.06)	211.3
$S_{10\ Gy\ (RBE)}$	0.75 * (0.64-0.86)	−0.95 (−1.89 to −0.04)	0.01 (0.01-0.02)	−0.10 † (−0.15 to −0.06)	211.6
$S_{15\ Gy\ (RBE)}$	0.75 * (0.64-0.86)	−0.93 (−1.88 to −0.02)	0.01 (0.01-0.02)	−0.08 † (−0.13 to −0.04)	211.8
$S_{40\ Gy\ (RBE)}$	0.74 * (0.63-0.85)	−1.77 (−2.42 to −1.19)	0.05 (0.03-0.07)	0.00 ‡ (−0.01 to 0.01)	210.3
$S_{45\ Gy\ (RBE)}$	0.74 * (0.62-0.85)	−1.64 (−2.25 to −1.08)	0.05 (0.03-0.07)	0.01 ‡ (−0.00 to 0.02)	211.6
$S_{50\ Gy\ (RBE)}$	0.72 * (0.60-0.84)	−1.46 (−2.04 to −0.94)	0.05 (0.03-0.08)	0.01 ‡ (−0.00 to 0.02)	215.9
$S_{5\ Gy}$	0.75 * (0.64-0.87)	−0.92 (−1.82 to −0.06)	0.01 (0.01-0.02)	−0.10 † (−0.15 to −0.06)	210.2
$S_{7.5\ Gy}$	0.75 * (0.64-0.86)	−0.96 (−1.92 to −0.04)	0.01 (0.01-0.02)	−0.09 † (−0.14 to −0.05)	211.6

Abbreviations: AICc = Akaike's corrected Information Criterion; AUC = area under the receiver operating characteristic curve; CI = confidence interval; PTV = plannng target volume.
* $P < .0001$. The best model is highlighted in bold.
† PTV to skin.
‡ Tumor volume.

could have occurred due to the inclusion of a small number of patients (n = 33). The highest AUC was obtained from the curve using PTV to skin. This distance can be easily measured before treatment planning and dose distribution calculations. Therefore, it is possible to select an appropriate treatment method, and the patient can be informed about individual risks before treatment.

Moreover, radiation techniques may affect the incidence of ARD. In photon RT, IMRT can significantly reduce the dose to the skin compared with conventional RT in breast cancer.[37,38] Scanning irradiation of proton beams has a lower skin dose than passively scattered proton RT in prostate cancer.[39] Similarly, by providing a more conformable dose distribution, the layer-stacking method enables the saving of more skin doses during CIRT.[34] In our study, the layer-stacking irradiation method was associated with a lower incidence of grade 2 to 3 ARD than conventional CIRT (36.4% vs 51.3%). The risk of grade 2 to 3 ARD with the layer-stacking method was less than 50% of that with conventional CIRT. At our facility, the layer-stacking method is the standard treatment for patients with HNC because this method can spare the skin. However, whether treatment techniques can reduce ARD incidence in HNC remains controversial. Palma et al found no significant difference in the incidence of ARD in thoracic cancer treated using IMRT and passive scattered proton RT.[15] Dutz et al constructed NTCP-based modeling for ARD in brain tumors based on multiple-center cohorts and found significant differences in dose parameters only at the low-dose range between passively scattered and scanning proton RT.[14] This finding is consistent with our results (see Table E2). Although low doses contribute to ARD, middle and high doses clearly play a more significant role in skin toxicity. These results indicate that the use of layer-stacking or scanning techniques (the dose distributions of these 2 methods are similar) may not always reduce the incidence of ARD, especially when the tumor is very close to the skin. In such cases, radiation dose savings, especially in high-dose areas, may be comparable among irradiation methods. Therefore, the NTCP model is expected to generate similar results as the proposed model when performing the scanning technique. However, this hypothesis will need to be validated.

Considering the potential uncertainties in biological dose calculation in CIRT, the physical dose was calculated for analysis. $S_{40\ Gy\ (RBE)}$ and $S_{20\ Gy}$ were selected in the final NTCP model for ARD due to their high model prediction ability (see Table 2). Both NTCP models had comparable prediction and calibration performances in the low-, middle-, and high-dose groups (see Fig. 3). Thus, the RBE-weighted dose could be as reliable as the physical dose when used as an indicator for the NTCP model for skin reaction. However, the uncertainty of RBE in calculation may have been mitigated to some extent in the current study, which included multiple tumor sites. As a result, further research on estimating biological doses for skin tissue in cases of specific tumors is warranted. Although late skin reaction was not assessed in this study, the high prediction ability of the model for ARD can help physicians perform comprehensive risk assessment before treatment. This may provide guidance for planning optimization and toxicity management. Hence, the incidence of late skin toxicity can be reduced. For example, according to the model, if the $S_{40\ Gy\ (RBE)}$ in 1 treatment plan is 70 cm^2, the probability of developing

grade 2 to 3 ARD is approximately 85%. Thus, it is possible to optimize planning to reduce the $S_{40\ Gy\ (RBE)}$ or enhance skin care during treatment.

The average skin DSH had a parallel trend, which can explain the small variation in the performance of NTCP models between doses (see Fig. 2). The model performance presents a minute difference in doses ranging from $S_{25\ Gy\ (RBE)}/S_{12.5\ Gy}$ to $S_{45\ Gy\ (RBE)}/S_{22.5\ Gy}$ (see Table 2). However, the results did not clearly demonstrate the best indicator of NTCP for skin toxicity among the doses. In another study that included 22 patients with malignant bone and soft tissue tumors receiving CIRT, $S_{40\ Gy\ (RBE)}$ was the recommended predictor factor for grade 2 ARD.[33] $S_{20\ Gy}$ was the only factor included in the leave-one-out method-based NTCP model for grade 3 ARD in 166 patients with thoracic cancer treated with IMRT or proton RT.[15] $V_{35\ Gy}$ (skin volume receiving 35 Gy) was the optimum factor for the NTCP model for grade 1 to 2 erythema, which was internally validated via a 3-fold cross-validation in 113 patients with brain tumor treated with proton RT.[14] For tomotherapy, $V_{56\ Gy}$ was the most effective predictive parameter for the NTCP model without the validation approach for grade 2 to 3 ARD in 70 patients with HNC.[12]

The high predictive value in the aforementioned study may be attributed to the high incidence of severe ARD caused by the poor ability of tomotherapy in terms of skin sparing, which resulted in up to 19% of patients experiencing grade 3 ARD, compared with 2.7% in our study. Although a direct comparison of these studies may be difficult, the strong correlation of dose range with ARD in our study is similar to that reported in previous studies.

The DSH values for multiple doses may be attributed to prognosis. Therefore, we examined the NTCP models using multiple doses with or without clinical factors. Increasing the number of parameters in a limited number of cases may lead to over fitting. Thus, to prevent over fitting and improve the model generalizability, mean AUCs with 10-fold internal cross-validation were assessed. The multivariate model performances were not improved, particularly in those with 2 dose−surface parameters, indicating that over fitting may have been present.

Study limitations

This study had some limitations. First, the clinical outcomes of RBE models applied in different CIRT facilities may differ even if similar prescribed biological doses are administered. Therefore, applying the present model directly in other facilities, especially those in Europe, may be difficult. One approach to address this difficulty is to normalize the RBE-weight dose value among different facilities using conversion factors for biological dose correction between various RBE models.[40,41] Hence, our proposed model may be used as a reference after dose correction. However, the effectiveness of the model is not clear in cases involving a combination of RT procedures (eg, carbon ion RT combined with proton, photon, and electron). Further research is warranted to address this uncertainty. Second, DSH parameters could not reflect organ-specific spatial information. However, other studies with different definitions of skin surface or volume (mainly regarding skin thickness) can be easily compared using the estimation method of DSH. Moreover, the dose calculation uncertainties associated with sites located a few millimeters from the body surface should be considered and investigated further. However, these uncertainties seem to have negligible effect on logistic regression coefficients.[12] In a study based on passively scattered proton RT, the difference in skin dose between the average measured dose and average planning calculated dose was only 2%.[39] Additionally, potential electron fluence and secondary particles arising from the particle interactions in the snout and the scattering material in the path of the beam (eg, patient compensator) should be considered, especially in the passive RT technique.[42] These uncertainties could result in a 4% increase in the skin dose.[43] Adopting the Monte Carlo-based algorithm, which is generally considered an accurate tool for surface dose estimation,[44-46] or directly measuring the skin dose using ionization chambers may further improve the reliability of the results. Third, the dose changes caused by anatomic variations during RT were not considered. Fourth, the enrolment period was 10 years, which is relatively long. Hence, the treatment strategy (eg, beam arrangement and irradiation technique selection) and ARD management (eg, skin care) had changed, which might have affected the incidence of ARD. Further research with external data validation should be performed.

Conclusions

We successfully developed and internally validated NTCP models for grade 2 to 3 ARD in patients with HNC treated with CIRT. However, there were uncertainties in RBE calculations for skin tissues, which may affect toxicity assessment. Our results show that the NTCP model's prediction ability was highly comparable between biological and physical dose parameters. Although a validation study is warranted, our results may help improve planning optimization to provide individualized treatment and prevent severe skin toxicity in patients with HNC treated with CIRT.

References

1. Porock D. Factors influencing the severity of radiation skin and oral mucosal reactions: Development of a conceptual framework. *Eur J Cancer Care (Engl)* 2002;11:33–43.
2. Salvo N, Barnes E, van Draanen J, et al. Prophylaxis and management of acute radiation-induced skin reactions: A systematic review of the literature. *Curr Oncol* 2010;17:94–112.
3. Russi EG, Merlano MC, Numico G, et al. The effects on pain and activity of daily living caused by crusted exudation in patients with head and neck cancer treated with cetuximab and radiotherapy. *Support Care Cancer* 2012;20:2141–2147.

4. Tsujii H, Kamada T. A review of update clinical results of carbon ion radiotherapy. *Jpn J Clin Oncol* 2012;42:670–685.
5. Amirul Islam M, Yanagi T, Mizoe JE, Mizuno H, Tsujii H. Comparative study of dose distribution between carbon ion radiotherapy and photon radiotherapy for head and neck tumor. *Radiat Med* 2008;26:415–421.
6. Mizoe JE, Hasegawa A, Jingu K, et al. Results of carbon ion radiotherapy for head and neck cancer. *Radiother Oncol* 2012;103:32–37.
7. Hayashi K, Koto M, Demizu Y, et al. A retrospective multicenter study of carbon-ion radiotherapy for major salivary gland carcinomas: subanalysis of J-CROS 1402 HN. *Cancer Sci* 2018;109:1576–1582.
8. Koto M, Hasegawa A, Takagi R, et al. Feasibility of carbon ion radiotherapy for locally advanced sinonasal adenocarcinoma. *Radiother Oncol* 2014;113:60–65.
9. Koto M, Hasegawa A, Takagi R, et al. Carbon ion radiotherapy for locally advanced squamous cell carcinoma of the external auditory canal and middle ear. *Head Neck* 2016;38:512–516.
10. Koto M, Hasegawa A, Takagi R, et al. Definitive carbon-ion radiotherapy for locally advanced parotid gland carcinomas. *Head Neck* 2017;39:724–729.
11. Russi EG, Moretto F, Rampino M, et al. Acute skin toxicity management in head and neck cancer patients treated with radiotherapy and chemotherapy or EGFR inhibitors: Literature review and consensus. *Crit Rev Oncol Hematol* 2015;96:167–182.
12. Mori M, Cattaneo GM, Dell'Oca I, et al. Skin DVHs predict cutaneous toxicity in head and neck cancer patients treated with tomotherapy. *Phys Med* 2019;59:133–141.
13. Parekh A, Dholakia AD, Zabranksy DJ, et al. Predictors of radiation-induced acute skin toxicity in breast cancer at a single institution: Role of fractionation and treatment volume. *Adv Radiat Oncol* 2018;3:8–15.
14. Dutz A, Lühr A, Agolli L, et al. Development and validation of NTCP models for acute side effects resulting from proton beam therapy of brain tumors. *Radiother Oncol* 2019;130:164–171.
15. Palma G, Monti S, Conson M, et al. NTCP models for severe radiation induced dermatitis after IMRT or proton therapy for thoracic cancer patients. *Front Oncol* 2020;10:344.
16. Kanai T, Endo M, Minohara S, et al. Biophysical characteristics of HIMAC clinical irradiation system for heavy-ion radiation therapy. *Int J Radiat Oncol Biol Phys* 1999;44:201–210.
17. Paganetti H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Phys Med Biol* 2014;59:R419–R472.
18. Karger CP, Peschke P. RBE and related modeling in carbon-ion therapy. *Phys Med Biol* 2017;63:01TR02.
19. Suzuki M, Kase Y, Yamaguchi H, Kanai T, Ando K. Relative biological effectiveness for cell-killing effect on various human cell lines irradiated with heavy-ion medical accelerator in Chiba (HIMAC) carbon-ion beams. *Int J Radiat Oncol Biol Phys* 2000;48:241–250.
20. Leith JT, Powers-Risius P, Woodruff KH, McDonald M, Howard J. Response of the skin of hamsters to fractionated irradiation with x-rays or accelerated carbon ions. *Radiat Res* 1981;88:565–576.
21. Ando K, Koike S, Nojima K, et al. Mouse skin reactions following fractionated irradiation with carbon ions. *Int J Radiat Biol* 1998;74:129–138.
22. Sørensen BS, Horsman MR, Alsner J, et al. Relative biological effectiveness of carbon ions for tumor control, acute skin damage and late radiation-induced fibrosis in a mouse model. *Acta Oncol* 2015;54:1623–1630.
23. Kanematsu N. Dose calculation algorithm of fast fine-heterogeneity correction for heavy charged particle radiotherapy. *Phys Med* 2011;27:97–102.
24. Kanai T, Kanematsu N, Minohara S, et al. Commissioning of a conformal irradiation system for heavy-ion radiotherapy using a layer-stacking method. *Med Phys* 2006;33:2989–2997.
25. Yanagi T, Kamada T, Tsuji H, Imai R, Serizawa I, Tsujii H. Dose-volume histogram and dose-surface histogram analysis for skin reactions to carbon ion radiotherapy for bone and soft tissue sarcoma. *Radiother Oncol* 2010;95:60–65.
26. Chen AP, Setser A, Anadkat MJ, et al. Grading dermatologic adverse events of cancer treatments: The Common Terminology Criteria for Adverse Events Version 4.0. *J Am Acad Dermatol* 2012;67 10251-039.
27. Kim YS, Lee KW, Kim JS, et al. Regional thickness of facial skin and superficial fat: Application to the minimally invasive procedures. *Clin Anat* 2019;32:1008–1018.
28. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
29. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun Stat - Theory Methods* 1978;7:13–26.
30. Nagelkerke N. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–692.
31. Malouff TD, Mahajan A, Krishnan S, Beltran C, Seneviratne DS, Trifiletti DM. Carbon ion therapy: A modern review of an emerging technology. *Front Oncol* 2020;10:82.
32. Jensen AD, Poulakis M, Nikoghosyan AV, et al. High-LET radiotherapy for adenoid cystic carcinoma of the head and neck: 15 years' experience with raster-scanned carbon ion therapy. *Radiother Oncol* 2016;118:272–280.
33. Takakusagi Y, Saitoh JI, Kiyohara H, et al. Predictive factors of acute skin reactions to carbon ion radiotherapy for the treatment of malignant bone and soft tissue tumors. *Radiat Oncol* 2017;12:185.
34. Kubo N, Kubota Y, Oike T, et al. Skin dose reduction by layer-stacking irradiation in carbon ion radiotherapy for parotid tumors. *Front Oncol* 2020;10:1396.
35. Mei M, Chen YH, Meng T, Qu LH, Zhang ZY, Zhang X. Comparative efficacy and safety of radiotherapy/cetuximab versus radiotherapy/chemotherapy for locally advanced head and neck squamous cell carcinoma patients: A systematic review of published, primarily non-randomized, data. *Ther Adv Med Oncol* 2020;12 1758835920975355.
36. Bonomo P, Loi M, Desideri I, et al. Incidence of skin toxicity in squamous cell carcinoma of the head and neck treated with radiotherapy and cetuximab: A systematic review. *Crit Rev Oncol Hematol* 2017;120:98–110.
37. Freedman GM, Anderson PR, Li J, et al. Intensity modulated radiation therapy (IMRT) decreases acute skin toxicity for women receiving radiation for breast cancer. *Am J Clin Oncol* 2006;29:66–70.
38. Pignol JP, Olivotto I, Rakovitch E, et al. A multicenter randomized trial of breast intensity-modulated radiation therapy to reduce acute radiation dermatitis. *J Clin Oncol* 2008;26:2085–2092.
39. 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–N302.
40. Steinsträter O, Grün R, Scholz U, Friedrich T, Durante M, Scholz M. Mapping of RBE-weighted doses between HIMAC- and LEM-based treatment planning systems for carbon ion therapy. *Int J Radiat Oncol Biol Phys* 2012;84:854–860.
41. Molinelli S, Magro G, Mairani A, et al. Dose prescription in carbon ion radiotherapy: How to compare two different RBE-weighted dose calculation systems. *Radiother Oncol* 2016;120 307312.
42. Yonai S, Matsufuji N, Kanai T, et al. Measurement of neutron ambient dose equivalent in passive carbon-ion and proton radiotherapies. *Med Phys* 2008;35:4782–4792.
43. Kern A, Bäumer C, Kröninger K, Wulff J, Timmermann B. Impact of air gap, range shifter, and delivery technique on skin dose in proton therapy. *Med Phys* 2021;48:831–840.
44. Arbor N, Gasteuil J, Noblet C, Moreau M, Meyer P. A GATE/Geant4 Monte Carlo toolkit for surface dose calculation in VMAT breast cancer radiotherapy. *Phys Med* 2019;61:112–117.
45. Schreuder AN, Bridges DS, Rigsby L, et al. Validation of the RayStation Monte Carlo dose calculation algorithm using realistic animal tissue phantoms. *J Appl Clin Med Phys* 2019;20:160–171.
46. Devic S, Seuntjens J, Abdel-Rahman W, et al. Accurate skin dose measurements using radiochromic film in clinical applications. *Med Phys* 2006;33:1116–1124.