

University of Groningen

Normal Tissue Complication Probability (NTCP) Prediction Model for Osteoradionecrosis of the Mandible in Patients With Head and Neck Cancer After Radiation Therapy

MD Anderson Symptom Working Grp; van Dijk, Lisanne; Abusaif, Abdelrahman A.; Rigert, Jillian; Naser, Mohamed A.; Hutcheson, Katherine A.; Lai, Stephen Y.; Fuller, Clifton D.; Mohamed, Abdallah S. R.

Published in:

International Journal of Radiation Oncology Biology Physics

DOI:

[10.1016/j.ijrobp.2021.04.042](https://doi.org/10.1016/j.ijrobp.2021.04.042)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

MD Anderson Symptom Working Grp, van Dijk, L., Abusaif, A. A., Rigert, J., Naser, M. A., Hutcheson, K. A., Lai, S. Y., Fuller, C. D., & Mohamed, A. S. R. (2021). Normal Tissue Complication Probability (NTCP) Prediction Model for Osteoradionecrosis of the Mandible in Patients With Head and Neck Cancer After Radiation Therapy: Large-Scale Observational Cohort. *International Journal of Radiation Oncology Biology Physics*, 111(2), 549-558. <https://doi.org/10.1016/j.ijrobp.2021.04.042>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Physics Contribution

Normal Tissue Complication Probability (NTCP) Prediction Model for Osteoradionecrosis of the Mandible in Patients With Head and Neck Cancer After Radiation Therapy: Large-Scale Observational Cohort

Lisanne V. van Dijk, PhD,^{*,†} Abdelrahman A. Abusaif, MD,^{*}
Jillian Rigert, DMD, MD,^{*} Mohamed A. Naser, PhD,^{*}
Katherine A. Hutcheson, MD, PhD,^{*,‡} Stephen Y. Lai, MD, PhD,^{*,‡}
Clifton D. Fuller, MD, PhD,^{*} and Abdallah S.R. Mohamed, MD^{*}
on behalf on the MD Anderson Symptom Working Group

^{*}Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; [†]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, NL; and [‡]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas

Received Feb 8, 2021; Accepted for publication Apr 26, 2021

Purpose: Osteoradionecrosis (ORN) of the mandible represents a severe, debilitating complication of radiation therapy (RT) for head and neck cancer (HNC). At present, no normal tissue complication probability (NTCP) models for risk of ORN exist. The aim of this study was to develop a multivariable clinical/dose-based NTCP model for the prediction of ORN any grade (ORN_{I-IV}) and grade IV (ORN_{IV}) after RT (\pm chemotherapy) in patients with HNC.

Corresponding author: Lisanne V. van Dijk, PhD;;
E-mail: lvvan@mdanderson.org

Sources of support: L.vD. received/receives funding and salary support from the Dutch organization NWO ZonMw during the period of study execution via the Rubicon Individual career development grant. A.S.R.M. and A.A.A. were funded for this work by The University of Texas MD Anderson Cancer Center-Oropharynx Cancer Program generously supported by Mr. and Mrs. Charles W. Stiefel. K.A.H., A.S.R.M., S.Y.L., and C.D.F. received/receives funding and salary support related to this project during the period of study execution from the NIH National Cancer Institute (NCI) Early Phase Clinical Trials in Imaging and Image Guided Interventions Program (R01CA218148). S.Y.L., A.S.R.M., and C.D.F. received/receives funding and salary support related to this project during the period of study execution from the National Institutes of Health (NIH) National Institute for Dental and Craniofacial Research (NIDCR) Establishing

Outcome Measures Award (R01DE025248/R56DE025248). C.D.F. received funding unrelated to this project during the period of study execution from NIH/NCI Cancer Center (P30CA016672, P50 CA097007, and R01CA2148250), from NIH/NIBIB (R25EB025787-01), from NIH/NSF (NSF1557679), NSF-CMMI grant (NSF1933369), and the Sabin Family Foundation.

Disclosures: none.

Data sharing statement: An anonymized version of the data set, including DVH and clinical variables with ORN grades is available at doi:10.6084/m9.figshare.13568207.

Supplementary material for this article can be found at 10.1016/j.ijrobp.2021.04.042.

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijrobp.2021.04.042.

Methods and Materials: Included patients with HNC were treated with (chemo-)RT between 2005 and 2015. Mandible bone radiation dose-volume parameters and clinical variables (ie, age, sex, tumor site, pre-RT dental extractions, chemotherapy history, postoperative RT, and smoking status) were considered as potential predictors. The patient cohort was randomly divided into a training (70%) and independent test (30%) cohort. Bootstrapped forward variable selection was performed in the training cohort to select the predictors for the NTCP models. Final NTCP model(s) were validated on the holdback test subset.

Results: Of 1259 included patients with HNC, 13.7% ($n = 173$ patients) developed any grade ORN (ORN_{I-IV} primary endpoint) and 5% ($n = 65$) ORN_{IV} (secondary endpoint). All dose and volume parameters of the mandible bone were significantly associated with the development of ORN in univariable models. Multivariable analyses identified D_{30%} and pre-RT dental extraction as independent predictors for both ORN_{I-IV} and ORN_{IV} best-performing NTCP models with an area under the curve (AUC) of 0.78 (AUC_{validation} = 0.75 [0.69-0.82]) and 0.81 (AUC_{validation} = 0.82 [0.74-0.89]), respectively.

Conclusions: This study presented NTCP models based on mandible bone D_{30%} and pre-RT dental extraction that predict ORN_{I-IV} and ORN_{IV} (ie, needing invasive surgical intervention) after HNC RT. Our results suggest that less than 30% of the mandible should receive a dose of 35 Gy or more for an ORN_{I-IV} risk lower than 5%. These NTCP models can improve ORN prevention and management by identifying patients at risk of ORN. © 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

Osteoradionecrosis (ORN) of the mandible is a severe, late toxicity after chemo-radiation for head and neck cancer (HNC) with a reported incidence between 1% to 16%.¹⁻⁴ Although ORN is less prevalent relative to other radiation-attributable HNC toxicities, ORN is often extremely debilitating, requires intense resource requirements for management, and contributes to a substantial negative effect on the quality of life.⁵ With the rising incidence of human papillomavirus (HPV) associated subtypes of HNC,⁶ survival rates have improved, as HPV-associated tumors are more sensitive to radiation therapy (RT) than HPV-negative tumors and exhibit improved tumor control.⁷⁻⁹ Moreover, because patients with HPV-positive tumors are typically younger and healthier,¹⁰ the longer life-years expectancy postradiation and expected chronic compromise to bone healing for these patients result in a higher cumulative lifetime risk for ORN development,¹¹ highlighting the importance of dedicated strategies aimed to prevent ORN in modern practice.

ORN is characterized by nonhealing bone and mucosal insult after radiation treatment, and the condition may present with variable severity.^{2,12} Some cases of ORN may clinically heal spontaneously over time (grade I), while other presentations of ORN may require minor debridement of the injured tissue (grade II), hyperbaric therapy (grade III), or major invasive mandible surgery (grade IV).¹³ Due to the characteristic presence of devitalized bone and reduced blood supply, successful treatment for ORN may be challenging and unpredictable, thus the optimal management for the condition is prevention. Normal tissue complication probability (NTCP) prediction of ORN based on dose-volume parameters can guide RT mandibular dose constraints in an attempt to prevent the development of ORN in patients with HNC.^{14,15} NTCP may also be used to guide alternative selections for treatment modalities with less distal beam-path toxicity, such as proton therapy.¹⁶ Furthermore, NTCP prediction may be used to identify patients at medium-high risk of ORN to prescribe dedicated follow-up

imaging for early detection of ORN and intervention before advanced stages.¹⁷

At present, no NTCP model has been developed for ORN, yet previous case-control studies have identified a significant relationship between mandibular dose and the development of ORN.^{1,12,13,18,19} Many identified the mandible bone volume receiving 50 Gy (V_{50Gy}) as the most important volume parameter (V_{xGy}); the related dose parameters (D_x) were not investigated in these studies.^{1,12,13,18,19} Moreover, pre-RT dental extractions have also been identified as a risk factor for ORN development.^{13,20} Some studies observed a significant association between smoking status and ORN^{12,13,20}; however, others did not observe this correlation.^{1,3} Consequently, in the absence of a formal NTCP model with clinical variables, monolithic nonpatient-specific dose constraints are used in general practice. Without a usable NTCP model, the confounding effect of clinical variables on dose-toxicity may be obscured.

To this end, the aim of this study was to develop a multivariable NTCP model for the prediction of development of any grade of ORN after RT in patients with HNC. The model building considers both dose-volume parameters and clinical risk factors to provide an optimized pretreatment ORN risk assessment. Secondary study analysis aimed at development of an NTCP model for the prediction of advanced (grade IV) ORN.

Methods and Materials

Patients

Subsequent to institutional review board approval (RCR03-0800), retrospective data of patient information for cases with proven squamous cell carcinoma HNC were included if patients received RT alone, in combination with surgery, or with chemotherapy with curative intent between 2005 and 2015 at a single institution, MD Anderson Cancer

Center. These patients were part of a larger “big data RT HNC” collection effort that is currently being constructed. Patients with previously documented head and neck irradiation, history of salivary gland cancer, and patients with a survival or follow-up time of less than a year were excluded from the study. Generally, the prescribed dose to primary tumor range was 68 to 72 Gy for definitive treatment (typically, 2.12 Gy in 33 fractions 5 times per week), 60 to 66 Gy for postoperative indications (typically, 2 Gy in 30-33 fractions 5 times per week), and 57 Gy to the elective lymph node levels (1.72 Gy in 33 fractions). Generally, a radiation source of 6 MV, a traditional beam, and a nominal dose rate of 600 Monitor-Units/min were used. In the study period, for primary tumor and upper neck nodal disease, the vast majority received a split-field technique matching a lower anterior neck field and larynx midline block. Alternatively, “whole-field” intensity modulated radiation therapy (IMRT) was deployed when tumors were located more inferiorly to avoid underdosing.

Data extraction and processing

Planned dose distribution and corresponding planning computed tomography (CT) were extracted from various planning systems (Pinnacle, Philips Radiation Oncology Systems; Eclipse, Varian Medical Systems; Raystation, RaySearch Laboratories) to standardized DICOM-RT format. The mandibular bone was subsequently auto-segmented with a previously validated multiatlas-based auto-segmentation using commercial software ADMIRE (research version 1.1; Elekta AB, Stockholm, Sweden).²¹ Dose-volume histogram (DVH) parameters were extracted with bulk extraction using an in-house developed software script in MATLAB (version R2014a).

NTCP endpoints

The primary NTCP endpoint of this study was binary ORN (ORN_{I-IV}) development any time point after treatment in patients with a minimum of 12 months of post-RT follow-up. The secondary NTCP endpoint was the development of ORN_{IV} (ORN_{IV}) at any time point after treatment. The ORN grades are defined as follows¹³: grade I, minimal bone exposure requiring conservative management; grade II, bone exposure requiring and receiving minor debridement; grade III, hyperbaric oxygen needed; grade IV, major invasive surgery required. ORN cases and grades were identified through querying radiology HNC RT CT scan reports from the radiology information systems, together with a thorough manual inspection of the electronic health record for ORN diagnosis.

Candidate predictors

Candidate DVH parameters of the mandibular bone were mean, minimum, and maximum dose; $D_{2\%}$; from $D_{5\%}$ – in

increments of 5% – to $D_{95\%}$; $D_{97\%}$; $D_{98\%}$; $D_{99\%}$; from V_{5Gy} – in increments of 5 Gy – to V_{70Gy} . The following clinical variables were considered: age; sex (female vs male); tumor subsite (oral cavity vs oropharynx vs hypopharynx/unknown-primary/larynx/nasopharynx: as discrete ordinal 1, 2, 3); smoking status (current vs former/never); smoking pack-years (continuous); postoperative RT (PORT) (definitive vs PORT); dental extraction (no/edentulous vs dental extractions); and chemotherapy (no vs chemotherapy). Only dental extractions within 6 weeks before treatment were considered; preradiation dental extractions are typically performed 4 to 6 weeks before RT at our institution.

Statistical modeling

The complete retrospective collected data were randomly divided into a training set and an independent test set with a 70:30 ratio. Univariable logistic regression analysis was performed on the training set to investigate statistically significant DVH and clinical variables ($P < .05$). Multivariable NTCP model development was performed with all candidate variables with step-wise forward selection with ranking based on Akaike information criterion (AIC) score while testing per variable selection “step” for significance of $P < .01$ with likelihood-ratio test for nested model comparison. The internal validity of the variable selection was estimated by repeating the variable selection 5000 times with a bootstrap procedure (ie, with replacement), as suggested by the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.²² Internal model robustness of variable selection was confirmed if variables were serially selected in the bootstrapped samples. These analyses were performed for the primary (ORN_{I-IV}) and secondary endpoint (ORN_{IV}) separately on the training cohort. Final models were independently validated (ie, not changing variables and coefficients) using the embargoed test subset. Model performance used area under the receiver operating characteristic curve (area under the curve [AUC]), Nagelkerke’s R^2 , and the discrimination slope as evaluative criteria. In addition, nested model improvement was determined with AIC score difference (Δ), which was considered “significant” when $\Delta AIC > 2$, and “strong” discriminatory/informativeness assertion can be made when $\Delta AIC > 5$. The R-packages Regression Modeling Strategies (version 4.3-1)²³ were implemented for these purposes.

Results

Patients

Of the total 1789 patients with HNC, 1259 patients were included in this study after screening (inclusion diagram in [Supplementary Materials](#)). They were randomly split in a training set of 882 patients (70%) and a validation set of

Table 1 Patient demographics for the full cohort and for training and valuation set separate

n (%)	Total 1259 (100)		Training set 882 (70)		Validation set 377 (30)		P value
ORN grade (%)							.681
Any	173	(14)	124	(14)	49	(13)	
G1	22	(2)	14	(2)	8	(2)	.44
G2	36	(3)	30	(3)	6	(2)	
G3	50	(4)	36	(4)	14	(4)	
G4	65	(5)	44	(5)	21	(6)	
Mean mandible dose (SD)	37.74	(12.51)	37.45	(12.86)	38.41	(11.67)	.217
Sex (%)							.021
Female	215	(17)	136	(15)	79	(21)	
Male	1044	(83)	746	(85)	298	(79)	
Age (SD)	60.72	(10.07)	60.82	(9.89)	60.68	(10.15)	.824
Tumor site (%)							.375
Oral cavity	190	(15)	54	(6)	136	(36)	
Oropharynx	826	(66)	249	(28)	577	(153)	
Larynx	159	(13)	42	(5)	117	(31)	
Hypopharynx	24	(2)	7	(1)	17	(5)	
Nasopharynx	22	(2)	10	(1)	12	(3)	
Unknown primary	38	(3)	15	(2)	23	(6)	
T stage (%)							.748
Tx	10	(1)	3	(0)	7	(2)	
T0	38	(3)	15	(2)	23	(6)	
T1	268	(21)	76	(9)	192	(51)	
T2	416	(33)	121	(14)	295	(78)	
T3	272	(22)	87	(10)	185	(49)	
T4	255	(20)	75	(9)	180	(48)	
N stage (%)							.496
N0	248	(20)	69	(8)	179	(47)	
N1	146	(12)	51	(6)	95	(25)	
N2	834	(66)	247	(28)	587	(156)	
N3	31	(2)	10	(1)	21	(6)	
p16 HPV positive (%)							.921
Positive	397	(32)	281	(32)	116	(31)	
Negative	71	(6)	50	(6)	21	(6)	
Unknown	791	(63)	551	(62)	240	(64)	
Technique (%)							.053
3D-CRT	123	(10)	88	(10)	35	(9)	
IMRT	891	(71)	608	(69)	283	(75)	
VMAT	224	(18)	173	(20)	51	(14)	
IMPT	21	(2)	13	(1)	8	(2)	
Chemotherapy (%)							.965
No chemotherapy	233	(19)	164	(19)	69	(18)	
Concurrent	624	(50)	436	(49)	188	(50)	
Induction + concurrent	285	(23)	197	(22)	88	(23)	
Induction	97	(8)	70	(8)	27	(7)	
missing	20	(2)	15	(2)	5	(1)	
Surgery (%)							1.000
Definitive	1043	(83)	731	(83)	312	(83)	
Postoperative	216	(17)	151	(17)	65	(17)	
Dental status (%)							.212
No extraction	707	(56)	506	(57)	201	(53)	
Edentulous	210	(17)	137	(16)	73	(19)	
Dental extraction	342	(27)	239	(27)	103	(27)	
Smoking status (%)							.49
Current	180	(14)	121	(14)	59	(16)	
Former	607	(48)	434	(49)	173	(46)	
Never	472	(37)	327	(37)	145	(38)	
Pack years (SD)	20.39	(28.62)	21.84	(28.89)	19.77	(28.50)	.238

Abbreviations: 3D-CRT = three-dimensional conformal radiotherapy; HPV = human papilloma virus; IMPT = intensity modulated proton therapy; IMRT = intensity modulated radiation therapy; ORN = osteoradionecrosis; SD = standard deviation; VMAT = volumetric-modulated arc therapy.

Significant difference levels comparing train and test set are given. Chi-squared test was used for categorical variables; *t* test for continuous variables.

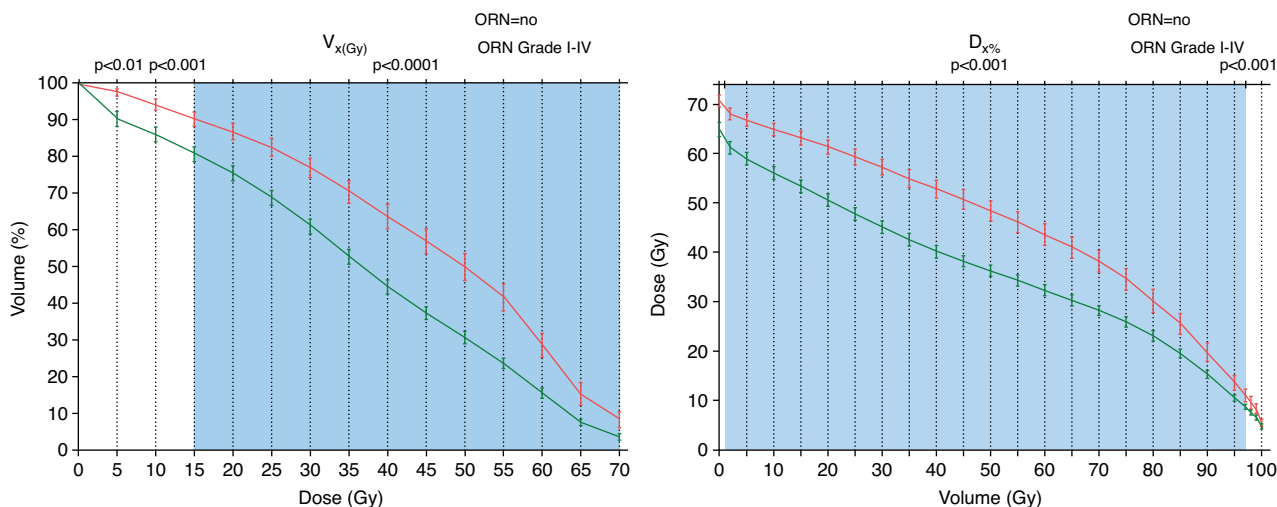


Fig. 1. Average dose-volume histogram (DVH) for patients who develop osteoradionecrosis (ORN) (red) versus those who do not (green) for volume (V_{xGy}) (left) and dose ($D_{x\%}$) parameters (right). Error bars represent 95% confidence interval. Color shading indicates the univariable significance of parameters, indicating that $D_{2\%}$ to $D_{98\%}$ and V_{15Gy} to V_{70Gy} were significant with a $P < .0001$. Notably, the curves are inversions of each other, but show the univariable separation ability between the groups and significance per V_{xGy} and $D_{x\%}$ variable.

377 patients (30%). Median follow-up time for all patients was 57 months (range, 12-174). Patient characteristics and demographics are detailed in Table 1. Briefly, the vast majority were patients with oropharyngeal cancer (OPC) (66%), followed by patients with oral cavity cancer (15%) and patients with laryngeal cancer (13%). The majority were male (83%) treated with IMRT (71%). Demographics were not significantly different between the training and validation set with the exception of sex ($P = .02$). From the total cohort, 13.7% ($n = 173$ patients) developed any grade ORN (primary endpoint) and 5% ORN_{IV} (secondary endpoint). Median time to development of ORN was 17 months (range, 2-142) post-RT. The distribution of ORN grades was as follows: grade I (12.7%), grade II (20.8%), grade III (28.9%), and grade IV ($n = 37.6\%$).

Univariable analyses

All DVH parameters were significantly associated with the development of ORN (any grade) in univariable analyses. The parameters ranging between $D_{2\%}$ and $D_{98\%}$ and V_{15Gy} and V_{70Gy} were highly significant ($P < .0001$) (Supplementary Materials). The $D_{30\%}$ and V_{50Gy} showed the best classification performance with AUCs of 0.76 (95% confidence interval, 0.72-0.80). Notably, DVH parameters ranging from $D_{15\%}$ to $D_{55\%}$, V_{40Gy} to V_{60Gy} , and mean mandible dose performed similarly (AUC, 0.74-0.76). Figure 1 depicts the dose ($D_{x\%}$) and volume (V_{xGy}) distinction of patients who do and do not develop ORN and the DVH parameter significance level. For example, Figure 1 shows that patients who did not develop ORN

Table 2 Univariable results of best performing dose and volume parameter and all clinical variables

Variable	Any grade ORN					Grade IV ORN				
	β	OR (95% CI)	AIC	AUC	P value	β	OR (95% CI)	AIC	AUC	P value
D30%*	0.09	1.10 (1.07-1.12)	627	0.76	<.0001	0.11	1.12 (1.08-1.16)	306	0.80	<.0001
V50Gy*	0.04	1.04 (1.03-1.05)	643	0.76	<.0001	0.05	1.05 (1.03-1.06)	307	0.82	<.0001
Tumor site	-0.92	0.40 (0.28-0.56)	691	0.63	<.0001	-1.34	0.26 (0.15-0.45)	329	0.68	<.0001
Dental extraction	0.51	1.67 (1.35-2.06)	698	0.62	<.0001	0.51	1.67 (1.19-2.33)	345	0.62	0.003
PORT	0.52	1.68 (1.07-2.65)	716	0.54	.025	1.30	3.67 (1.96-6.88)	339	0.63	<.0001
Chemotherapy	0.61	1.85 (1.06-3.21)	715	0.54	.029	0.14	1.15 (0.53-2.53)	353	0.51	.721
Smoking status	0.08	1.09 (0.62-1.91)	720	0.50	.776	1.25	3.48 (0.83-14.55)	349	0.55	.088
Gender	0.60	1.83 (0.98-3.41)	716	0.53	.059	-0.04	0.96 (0.42-2.2)	354	0.50	.926
Age	-0.01	0.99 (0.97-1.01)	720	0.52	.449	-0.01	0.99 (0.96-1.02)	353	0.53	.369
Pack years	0.00	1.00 (0.99-1.01)	720	0.50	.699	-0.01	0.99 (0.98-1.01)	352	0.52	.251

Abbreviations: AIC = Aikake Information Criterion; AUC = area under the receiver operator characteristic curve; CI = confidence interval; OR = odds ratio; ORN = osteoradionecrosis; PORT = postoperative radiation therapy.

* Best performing mandible bone dose volume histogram (DVH) variables are shown; note all candidate DVH parameters were significant (refer to Supplementary Materials). β : model coefficient.

Table 3 Model parameters for any grade (I-IV) and grade IV ORN NTCP models

Variables	Grade I-IV ORN			Grade IV ORN					
	β	OR	<i>P</i> value	β	OR	<i>P</i> value	β	OR	<i>P</i> value
Intercept	-6.85			-9.16			-12.27		
D30	0.09	1.1 (1.07-1.12)	<.0001	0.11	1.12 (1.07-1.16)	<.0001	0.12	1.13 (1.08-1.17)	<.0001
Dental extractions	0.66	1.93 (1.28-2.92)	.002	0.62	1.85 (0.98-3.49)	.057			
Smoking status							1.51	4.54 (1.05-19.68)	.043

Abbreviations: NTCP = normal tissue complication probability; OR = odds ratio; ORN = osteoradionecrosis.
 β : model coefficient.

received an average $D_{30\%}$ of 46 ± 16 Gy, whereas this was 57 ± 9 Gy for those who did develop ORN. Additionally, dental extraction (odds ratio [OR], 1.67 [1.35-2.06]; $P < .0001$), PORT (OR, 1.68 [1.07-2.65]; $P = .0253$), and chemotherapy (OR, 1.85 [1.06-3.21]; $P = .0293$) were significantly associated with the development of ORN (Table 2). Specifically for tumor site, compared with oral cavity (ie, as reference), ORs were 0.58 (0.37-0.92) ($P = .021$) for OPC and 0.08 (0.03-0.23) ($P < .0001$) for others. Both smoking status and pack-years did not show a significant relationship with ORN development.

Multivariable NTCP model development and validation

AIC-ranked forward selection in the training set step-wise identified $D_{30\%}$ first ($P < .0001$) followed by pre-RT dental extraction (likelihood-ratio test; $P = .005$) with a “significant” Δ AIC of 5.96. Bootstrapped forward variable selection in the training cohort also showed that $D_{30\%}$ was the most frequently selected first variable (50% of the bootstrapped samples; note, $D_{25\%}$ in 23%), and the clinical variable dental extraction was the second variable (47%; Supplementary Materials). The positive regression

coefficients reveal that higher $D_{30\%}$ (OR, 1.10 [1.07-1.12]) and dental extraction (OR, 1.67 [1.35-2.06]) are associated with higher risk of developing ORN (Table 3). The model performance was good with an AUC of 0.78 (0.74-0.82) and R^2 of 0.20 (Table 4). Validation of the performance of the NTCP model with D_{30} and dental extraction tested on the independent test set ($n = 377$) was also good (AUC_{validation} = 0.75 [0.69-0.82]; $R^2 = 0.17$). The calibration plot (Supplementary Materials) showed that the predicted NTCP values were an underestimation compared with the actual observed ORN_{I-IV} rate in the validation cohort.

For the secondary NTCP endpoint ORN_{IV} (ie, needing major surgical intervention), forward selection selected the dose variable $D_{30\%}$. Disregarding dose variables ($V_{70/65Gy}$) that flipped to negative coefficient in multivariable analyses (ie, suggesting over/incorrect-fitting), smoking status was the next most-associated variable but did not meet our pre-specified significance level (likelihood-ratio test; $P = .013$), nor did dental extraction ($P = .06$). Bootstrapped variable selection selected V_{55Gy} (32%) over $D_{25\%}$ (20%), $D_{30\%}$ (18%), and $D_{40\%}$ (14%) (Supplementary Materials), together with the clinical variables smoking status (29%) and/or dental extraction (21%) in multiple “runs.” In training, ORN_{IV} model performance was nearly identical for NTCP models with $D_{30\%}$ or V_{55Gy} alone or combined with

Table 4 Performance of the NTCP models

	Any grade ORN		Grade IV ORN	
	D30 Dental extractions		D30 Dental extractions	D30 Smoking status
Training (n = 882)				
AIC	619.2		304.9	302.3
AUC _{training}	0.78 (0.74-0.82)		0.81 (0.76-0.86)	0.81 (0.76-0.86)
Nagelkerke R^2 _{training}	0.20		0.17	0.18
Discrimination slope	0.12		0.06	0.06
HL test X^2 (<i>P</i> value)	8.44 (.39)		10.82 (.21)	11.7 (.17)
Validation (n = 377)				
AUC _{validation}	0.75 (0.69-0.82)		0.82 (0.74-0.89)	0.75 (0.64-0.86)
Nagelkerke R^2 _{validation}	0.17		0.20	0.14

Abbreviations: AIC = Aikake Information Criterion; AUC = area under the receiver operator characteristic curve; HL = Hosmer-Lemeshow; NTCP = normal tissue complication probability; ORN = osteoradionecrosis.
 n: number of patients.

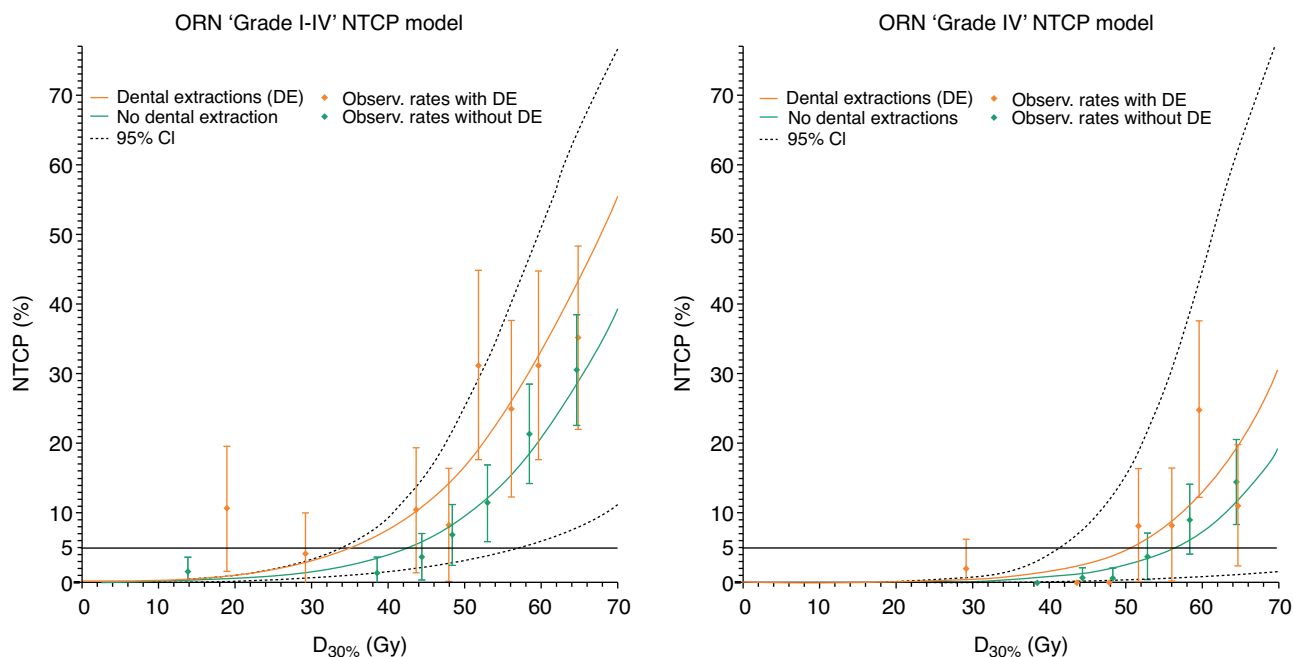


Fig. 2. Final osteoradionecrosis (ORN) normal tissue complication probability (NTCP) models. NTCP curves are plotted against $D_{30\%}$ split into patients with pretreatment dental extraction (orange lines) and those without (green lines) for NTCP models for ORN any grade (left plot) and grade IV (right). The dotted line indicates the most outer 95% confidence interval (CI) limits of the NTCP curves. Point with 95% CI error bars of the actual observed ORN rates by binning all patients sorted by dose in X percentiles; points are positioned at the average dose per bin.

smoking status or dental extraction (AUC range, 0.80-0.82; R^2 range, 0.16-0.18). However, external validation showed that the model with $D_{30\%}$ and dental extraction (AUC_{validation} = 0.82 [0.74-0.89]; $R^2_{validation}$ = 0.21) performed significantly better than the model with V_{55Gy} and smoking status (AUC_{validation} = 0.71 [0.60-0.83]; Z-test P = .02) (refer to [Supplementary Materials](#) for alternative models). Performance improved with a model $D_{30\%}$ and dental extraction compared with the model with $D_{30\%}$ alone; even though this improvement is limited, for consistency with ORN_{I-IV}, we selected the same 2 variables in the final ORN_{IV} NTCP model (note: coefficients deviate). Moreover, the calibration cohort ([Supplementary Materials](#)) of the model also showed an underestimation of the NTCP values compared with the ORN_{IV} observed rates, yet this was less pronounced as for ORN_{I-IV}.

Final NTCP models that were developed in the training cohort (model coefficients in [Table 3](#)) and validated in the unseen/embargoed test cohort are plotted in [Figure 2](#). Binned actual observed ORN proportions, represented by points with error bars, correspond with the NTCP models. The horizontal gray lines in [Figure 2](#) indicate the 5% ORN threshold risks.

Subcohort analyses

The final NTCP model ([Table 3](#)) performed similarly for patients with OPC only (n = 826; AUC_{OPC-cohort} = 0.76 [0.71-0.80]), for larynx/hypo/nasopharynx/unknown-

primary patients with cancer (ie, others; n = 243; AUC_{Other-cohort} = 0.76 [0.53-0.98]), and combined cohorts (ie, OPC + other patients; n = 1069; AUC_{OPC + other-cohort} = 0.79 [0.75-0.83]). In contrast, performance in the patients with oral cavity cancer was poor (n = 190; AUC_{Oral cavity} = 0.59 [0.50-0.68]). A similar trend was seen for ORN_{IV} (AUC_{OPC-cohort} = 0.80 [0.74-0.86], AUC_{OPC + other-cohort} = 0.84 [0.79-0.89], AUC_{Oral cavity} = 0.57 [0.46-0.68]), except that in the “others cohort” no ORN_{IV} was present. Refer to [Supplementary Materials](#) for subanalyses test results per tumor site and for definitive and PORT patients.

Discussion

Although ORN rates are relatively low (~5%-15%), the consequences for patients experiencing ORN are highly disabling with a substantial effect on the health care utilization and quality of life.⁵ Once ORN develops, treatment is complicated by the lack of regenerative bone and tissue cells needed for healing and repair. Advanced stage ORN requires extensive surgery associated with significant perioperative morbidity.²⁴ Given the potential severity of ORN and limitations in treatment once ORN has developed, improved pretreatment risk assessment tools aimed at identifying high-risk patients and guiding strategies for prevention and early intervention of ORN represent an important unmet need.

Due to the low relative prevalence of ORN among HNC survivors, a large data set is needed to design a robust NTCP model, which is particularly challenging in this case as large-scale radiation dose plans with matching late toxicity scores for each patient are rarely readily available. Previous studies assessing radiation dose to the mandible and development of ORN have, at best, 200 to 600 patients.^{1,12,13,18,19} In response to the unmet need for prediction models for the development of ORN and ORN severity validated across data from a sufficient patient cohort, this study developed NTCP models for the prediction of ORN of any grade and grade IV in a large cohort of 1259 patients with HNC treated with definitive or postoperative (chemo-) RT.

The association between ORN development and mandible radiation dose was clearly observed with the univariable significance of all DVH parameters (Fig. 1). The final NTCP models were based on $D_{30\%}$ and pretreatment dental extraction. This NTCP model had good performance in both the training and validation cohort for ORN_{I-IV} ($AUC_{train/validation} = 0.78/0.75$) and ORN_{IV} ($AUC_{train/validation} = 0.81/0.82$). These models are clinically useful tools to determine appropriate dose constraints for the mandibular bone when feasible (ie, when tumor coverage is not compromised).¹⁶ Additionally, the models identify patients at high risk for ORN development who may benefit from more intensive clinical surveillance programs with dedicated imaging follow-up²⁵ and/or earlier intervention, whether conservative or surgical, to prevent ORN progression.

Our results demonstrate that mandible dose constraints can be distilled from these NTCP models to optimize patients' IMRT plans. For example, our models suggest that mandibular $D_{30\%}$ of patients without pretreatment dental extraction should be kept below 42 Gy to achieve <5% risk of ORN development, while a $D_{30\%} < 35$ Gy is required for patients with dental extractions to achieve the same level of risk (Fig. 2). Alternatively, for a more conservative risk threshold of 1%, $D_{30\%}$ should be <25 Gy (without dental extractions) and <17 Gy (with dental extractions). With respect to ORN_{IV} only, maintaining $D_{30\%} < 56$ Gy without pre-RT dental extractions or $D_{30\%} < 50$ Gy with pre-RT dental extractions may be sufficient to achieve <5% risk of ORN_{IV} development.

Our findings of significant association between ORN and several DVH parameters as well as with pre-dental status match the results of several recent publications.^{1,12,13,18,19,26} For instance, a recent publication from a Danish group showed that several DVH parameters in the intermediate- and high-dose range including D_{mean} were associated with ORN in a cohort of patients with HNC with 56 ORN cases and 112 controls.²⁶ Another study from the Princess Margaret Cancer Centre reported that V_{50} and V_{60} were significantly higher in 71 patients with ORN compared with 142 patients with no ORN.¹² In addition, another group previously reported that maximum radiation dose to the mandible as a single dose constraint was a poor

correlate of ORN in patients with OPC, and mandibular volumes receiving 44 Gy (V_{44Gy}) and 58 Gy (V_{58Gy}) were comparatively more discriminatory of patients with ORN versus non-ORN patients.¹ However, these studies were case-control studies, based on a limited number of patients and did not design a multivariable NTCP model.

Our multivariable NTCP models showed that a combination of mandibular dosimetric parameters ($D_{30\%}$) with the pre-RT dental extraction status achieved the best performing model for ORN risk prediction. A study by the Memorial Sloan Kettering Cancer Center group showed that, in addition to mandibular radiation doses, the presence of mild-severe periodontal bone loss was associated with increased ORN risk. However, in this study, pre-RT radiographs were only available for 18 patients with ORN who were matched with 36 controls.³ In concurrence with our study results, several recent studies have demonstrated that pre-RT dental extractions are a significant risk factor for ORN development.^{12,20,27-29}

Whether pre-RT dental extraction is a direct incipient insult preceding ORN development or merely a surrogate for poor dentition remains unclear. Although all patients receive pretherapy dental oncology assessment, we do not routinely deploy asymptomatic dental surveillance posttherapy, referring these cases to their community dentists. Consequently, our data set lacked significant prospective postradiation dental assessment variables, surveillance of radiation caries, and posttherapy dental extractions that may have been completed outside our facility. Consequently, there remains a significant need to undertake prospective assessment of orodental health with developed instruments (eg, formal sialometry, radiation caries monitoring with DMFS160 (grading system for post-radiation caries),³⁰ and patient-reported outcomes) to determine whether the observed association of ORN with pretherapy dental extractions can be related to 1 or more mechanisms. In particular, we plan to expand the current research to investigate the relationship between the location of the pretherapy dental extraction and posttherapy ORN with dental reports and pre-/posttherapy CT and magnetic resonance images.

In contrast to previous studies,^{12,13,20} smoking status was not found to be significantly associated with all grades of ORN_{I-IV} in the current study, but smoking status was frequently identified on variable selection with higher ORN grade (ie, grade IV). Notably, our validation showed reduced performance of models with smoking status included compared with that in the training cohort (in contrast to the model with dental extraction). Other groups have shown similar ambiguity as to the role of tobacco in development of ORN, with other publications also showing no association between smoking status and ORN.^{1,3} These contradictory findings may be due to intercohort variables inherent in different studies' populations. A second possible explanation is that smoking continuation during and after treatment may be of more influence for the development of ORN compared with patients who elect to stop smoking

before treatment as in our data set, which had limited active smokers. More research is needed to investigate the discordance between our findings and other group reports.^{12,13,20}

Subanalyses showed that the NTCP model performance was poor when tested in the patients with oral cavity cancer only ($AUC_{\text{Oral-cavity}} = 0.59$), especially relative to the performance in the patients with nonoral cavity cancer ($AUC_{\text{OPC} + \text{Other-cohort}} = 0.79$). For the patients with oral cavity cancer, both the mandible dose ($D_{\text{mean}} = 46.5 \pm 6.5$ Gy) and ORN prevalence (23%) were higher compared with the rest of the cohort (36.2 ± 12.7 Gy; prevalence, 12%). Although a relatively small sample size of patients with oral cavity cancer ($n = 190$) could explain the limited significance of the dose variables (ie, only $D_{30\%}$, $D_{35\%}$, and $D_{40\%}$), the poor performance of the NTCP models suggests that there is an effect in these patients not captured in the present data set. One consideration is that patients with oral cavity cancer typically received PORT (89%), whereas patients with other tumor sites were generally treated with primary RT (94%). Across tumor locations, NTCP model performance was better in patients treated with definitive RT ($AUC = 0.78$) than those in the PORT group ($AUC = 0.65$) (Supplementary Materials). Although PORT was significant in univariable analysis, it did not perform well in the multivariable analyses. Further research with specific focus on the role of pre-RT surgical intervention and/or other oral cavity-specific factors is needed to better explain ORN development in patients with oral cavity cancer. Additionally, although patients with oral cavity cancer generally receive radiation to greater volumes of the mandible, the gradients across the mandible were more homogenous. The current NTCP approach treats DVH dose-volume “bins” as discrete independent constructs, which may obfuscate discriminatory signal in organs with more homogenous cohort dose distributions and suggests further investigation with alternative normal tissue injury approaches are warranted. Previous studies have proposed and approach to investigate a spatial dose-toxicity association by warping the dose distribution with deformable registration techniques of the patients to a reference CT scan.^{31,32} This may allow for voxel-based identification of ORN significantly associated mandible areas, which are projected on the reference patient.

Although the validation NTCP model performance measures were good, the calibration plots in the validation cohort suggested that the predicted NTCP values were underestimated, that is, the model coefficients should have been larger. Additional external validation is needed to improve the estimation of the model coefficient according to the closed-testing procedure.³³

Though this study is based on an extensive retrospective cohort of patients with HNC treated between 2005 and 2015 at MD Anderson Cancer Center, limitations include that this sample represents a fraction of all patients in the study time frame, an estimated 25%. Nevertheless, we are convinced that the included patient cohort is likely a fair representation of our institutional HNC population. In

addition, other variables that are not included in this study may be related to ORN development and potentially improve the NTCP models. For instance, posttreatment alcohol use was associated with development of ORN in the study by Owosho et al³; nevertheless, alcohol use history had no relation with ORN in other studies.^{1,20} Alcohol use may also act as a surrogate variable for general oral health, and the same can be reasoned for social-economic status, insurance status, ethnicity, and smoking status. More extensive research is needed to identify the role of general oral health in the development of ORN. Moreover, we considered ORN development as a binary variable, leading (as in most NTCP studies) to potential limitations with regard to right-censored event prediction. For simplicity, we used conventional NTCP model approaches, but efforts for dynamic time-incorporating risk models (eg, partially observed Markov decision processes) are ongoing.

Despite these limitations, to our knowledge, this study represents the largest extant ORN survey of dose-response data and the first published ORN NTCP model. To ensure findable, accessible, interoperable and reusable (FAIR) data³⁴ and allow external validation, an anonymized version of the data set, including DVH and clinical variables with ORN grades, has been deposited at doi:<https://doi.org/10.6084/m9.figshare.13568207>. Our hope is that this can afford others the opportunity to validate our approach, generate institutional-specific models, and engender further cross-platform research for ORN toxicity modeling and multi-institutional dose constraints.

Conclusions

The developed NTCP models performed well in predicting ORN_{I-IV} (primary NTCP endpoint) and ORN_{IV} (secondary NTCP endpoint) in both the patient with HNC training and independent test cohorts. NTCP models were based on mandible bone $D_{30\%}$ and pretreatment dental extraction. Our results show a distinct association between planned mandible bone radiation dose and ORN development and suggest that less than 30% of the mandible should receive a dose of 35 Gy or more for an ORN_{I-IV} risk lower than 5%. These NTCP models may be used to improve prevention of ORN as well as guide ORN surveillance/management strategies by identifying and stratifying patients at risk of ORN.

References

1. Mohamed ASR, Hobbs BP, Hutcheson KA, et al. Dose-volume correlates of mandibular osteoradionecrosis in oropharynx cancer patients receiving intensity-modulated radiotherapy: Results from a case-matched comparison. *Radiother Oncol* 2017;124:232–239.
2. Chronopoulos A, Zarra T, Ehrenfeld M, Otto S. Osteoradionecrosis of the jaws: Definition, epidemiology, staging and clinical and radiological findings. A concise review. *Int Dent J* 2018;68:22–30.

3. Owosho AA, Tsai CJ, Lee RS, et al. The prevalence and risk factors associated with osteoradionecrosis of the jaw in oral and oropharyngeal cancer patients treated with intensity-modulated radiation therapy (IMRT): The Memorial Sloan Kettering Cancer Center experience. *Oral Oncol* 2017;64:44–51.
4. Beadle BM, Liao K-P, Chambers MS, et al. Evaluating the impact of patient, tumor, and treatment characteristics on the development of jaw complications in patients treated for oral cancers: A SEER-Medicare analysis. *Head Neck* 2013;35:1599–1605.
5. Wong ATT, Lai SY, Gunn GB, et al. Symptom burden and dysphagia associated with osteoradionecrosis in long-term oropharynx cancer survivors: A cohort analysis. *Oral Oncol* 2017;66:75–80.
6. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70:7–30.
7. Licitra L, Perrone F, Bossi P, et al. High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. *J Clin Oncol* 2006;24:5630–5636.
8. Sedaghat AR, Zhang Z, Begum S, et al. Prognostic significance of human papillomavirus in oropharyngeal squamous cell carcinomas. *Laryngoscope* 2009;119:1542–1549.
9. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010;363:24–35.
10. Young D, Xiao CC, Murphy B, Moore M, Fakhry C, Day TA. Increase in head and neck cancer in younger patients due to human papillomavirus (HPV). *Oral Oncol* 2015;51:727–730.
11. Patel D, Haria S, Patel V. Oropharyngeal cancer and osteoradionecrosis in a novel radiation era: A single institution analysis. *Oral Surg* 2020;14:113–121.
12. Caparrotti F, Huang SH, Lu L, et al. Osteoradionecrosis of the mandible in patients with oropharyngeal carcinoma treated with intensity-modulated radiotherapy. *Cancer* 2017;123:3691–3700.
13. Tsai CJ, Hofstede TM, Sturgis EM, et al. Osteoradionecrosis and radiation dose to the mandible in patients with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2013;85:415–420.
14. Kierkels RGJ, Korevaar EW, Steenbakkers RJHM, et al. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. *Radiother Oncol* 2014;112:430–436.
15. Witte MG, Van Der Geer J, Schneider C, Lebesque JV, Alber M, Van Herk M. IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP. *Med Phys* 2007;34:3544–3555.
16. Langendijk JA, Lambin P, De Ruyscher D, 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;107:267–273.
17. Sandulache VC, Hobbs BP, Mohamed ASR, et al. Dynamic contrast-enhanced MRI detects acute radiotherapy-induced alterations in mandibular microvasculature: Prospective assessment of imaging biomarkers of normal tissue injury. *Sci Rep* 2016;6:29864.
18. Zhang W, Zhang X, Yang P, et al. Intensity-modulated proton therapy and osteoradionecrosis in oropharyngeal cancer. *Radiother Oncol* 2017;123:401–405.
19. Brodin NP, Tomé WA. Revisiting the dose constraints for head and neck OARs in the current era of IMRT. *Oral Oncol* 2018;86:8–18.
20. Moon DH, Moon SH, Wang K, et al. Incidence of, and risk factors for, mandibular osteoradionecrosis in patients with oral cavity and oropharynx cancers. *Oral Oncol* 2017;72:98–103.
21. Mohamed ASR, Ruangskul M-N, Awan MJ, et al. Quality assurance assessment of diagnostic and radiation therapy—simulation CT image registration for head and neck radiation therapy: Anatomic region of interest—based comparison of rigid and deformable algorithms. *Radiology* 2015;274:752–763.
22. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 2015;162:W1.
23. R Development Core Team. *R: A language and environment for statistical computing*; Vienna, Austria: The R Foundation for Statistical Computing; 2011. Available at: <http://www.R-project.org/>. Accessed 28 May 2021.
24. Lambade PN, Lambade D, Goel M. Osteoradionecrosis of the mandible: A review. *Oral Maxillofac Surg* 2013;17:243–249.
25. Joint Head and Neck Radiotherapy-MRI Development Cooperative. Dynamic contrast-enhanced MRI detects acute radiotherapy-induced alterations in mandibular microvasculature: Prospective assessment of imaging biomarkers of normal tissue injury. *Sci Rep* 2016;6:29864.
26. Aarup-Kristensen S, Hansen CR, Forner L, Brink C, Eriksen JG, Johansen J. Osteoradionecrosis of the mandible after radiotherapy for head and neck cancer: Risk factors and dose-volume correlations. *Acta Oncol (Madr)* 2019;58:1373–1377.
27. Beech NM, Porceddu S, Batstone MD. Radiotherapy-associated dental extractions and osteoradionecrosis. *Head Neck* 2017;39:128–132.
28. Sathasivam HP, Davies GR, Boyd NM. Predictive factors for osteoradionecrosis of the jaws: A retrospective study. *Head Neck* 2018;40:46–54.
29. Kojima Y, Yanamoto S, Umeda M, et al. Relationship between dental status and development of osteoradionecrosis of the jaw: A multicenter retrospective study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2017;124:139–145.
30. Watson E, Eason B, Kreher M, Glogauer M. The DMFS160: A new index for measuring post-radiation caries. *Oral Oncol* 2020;108:104823.
31. Monti S, Palma G, D’Avino V, et al. Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients. *Sci Rep* 2017;7:7220.
32. Beasley W, Thor M, McWilliam A, et al. Image-based data mining to probe dosimetric correlates of radiation-induced trismus. *Int J Radiat Oncol* 2018;102:1330–1338.
33. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med* 2017;36:4529–4539.
34. Goble C, Cohen-Boulakia S, Soiland-Reyes S, Garijo D, Gil Y, Crusoe MR. FAIR computational workflows. *Data Intell* 2019;23:2.