



NTCP Modeling of Late Effects for Head and Neck Cancer: A Systematic Review

Sonja Stieb, MD^{1,2}; Anna Lee, MD, MPH¹; Lisanne V. van Dijk, PhD^{1,3}; Steven Frank, MD¹; Clifton David Fuller, MD, PhD¹; Pierre Blanchard, MD, PhD^{1,4}

¹Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

²Center for Radiation Oncology KSA-KSB, Kantonsspital Aarau, Aarau, Switzerland

³Department of Radiation Oncology, University Medical Center–Groningen, Groningen, the Netherlands

⁴Department of Radiotherapy, Gustave Roussy Cancer Campus, Université Paris-Saclay, Villejuif, France

Introduction

Head and neck cancer survival rates are increasing as a result of more advanced treatment regimens [1–7]. The increasing number of head and neck cancer survivors stresses the importance of preventing late radiation-induced toxicities, which may persist or occur years after treatment, gravely impacting the quality of life [8]. Owing to the intrinsic properties of particles, proton therapy has the ability to deliver dose more conformal to the tumor, consequently sparing more normal tissue surrounding it [9–11]. While proton therapy arguably can be beneficial for a large proportion of head and neck cancer patients, it is currently still limited in availability and there is a benefit gradient with regard to the estimated toxicities [12, 13]. Introduced by Langendijk et al [14], the model-based approach is a systematic way to identify patients that may benefit most from proton therapy based on the predicted toxicity risk, that is, normal tissue complication probability (NTCP). In other words, NTCP can be considered the individual percentile risk for a patient to develop a certain radiation-associated toxicity (eg, risk of 40% to develop feeding tube dependence).

NTCP-guided treatment decision support requires reliable NTCP models, which generally embody the toxicity's association with organ-at-risk (OAR)-specific dose-volume histogram parameters. The development of (semi)auto-contouring of head and neck OARs has facilitated more time-efficient and robust extraction of dose-volume histogram information, particularly from OARs that are not contoured for clinical planning purposes [15–18]. In addition, clinical factors (eg, chemotherapy, smoking status, age, xerostomia before radiotherapy) can contribute to or interact with the effect of radiation dose in developing radiation-induced toxicities [15–17, 19–21].

The aims of this systematic review were (1) to describe how NTCP models are developed and validated, (2) to perform and present a systematic review of existing NTCP models for photon or proton therapy for head and neck cancers, and (3) to explore and propose future directions, concerning novel NTCP model types, as well as methodologic development.

Normal Tissue Complication Probability Model Development and Validation

NTCP models are classifiers, meaning they aim to stratify between patients at high risk and low risk for developing a toxicity. In contrast to the historically used Lyman–Kutcher–Burman NTCP models [18, 22, 23], which are only based on a single-dose variable, multivariable logistic regression represents the current modeling preference as shown in the following equation:

$$\text{NTCP} = \frac{1}{1 - e^{-s}},$$

where $s = \beta_0 + \beta_1 \cdot \text{variable}_1 + \beta_2 \cdot \text{variable}_2 \dots + \beta_n \cdot \text{variable}_n$.

Submitted 30 Nov 2020

Accepted 08 Feb 2021

Published 25 June 2021

Corresponding Author:

Pierre Blanchard, MD, PhD

Department of Radiotherapy

Gustave Roussy Cancer

Campus

Université Paris-Saclay

Villejuif, France

Phone: +41767567799

pierre.blanchard@

gustaveroussy.fr

Review Article

DOI

10.14338/20-00092

© Copyright

2021 The Author(s)

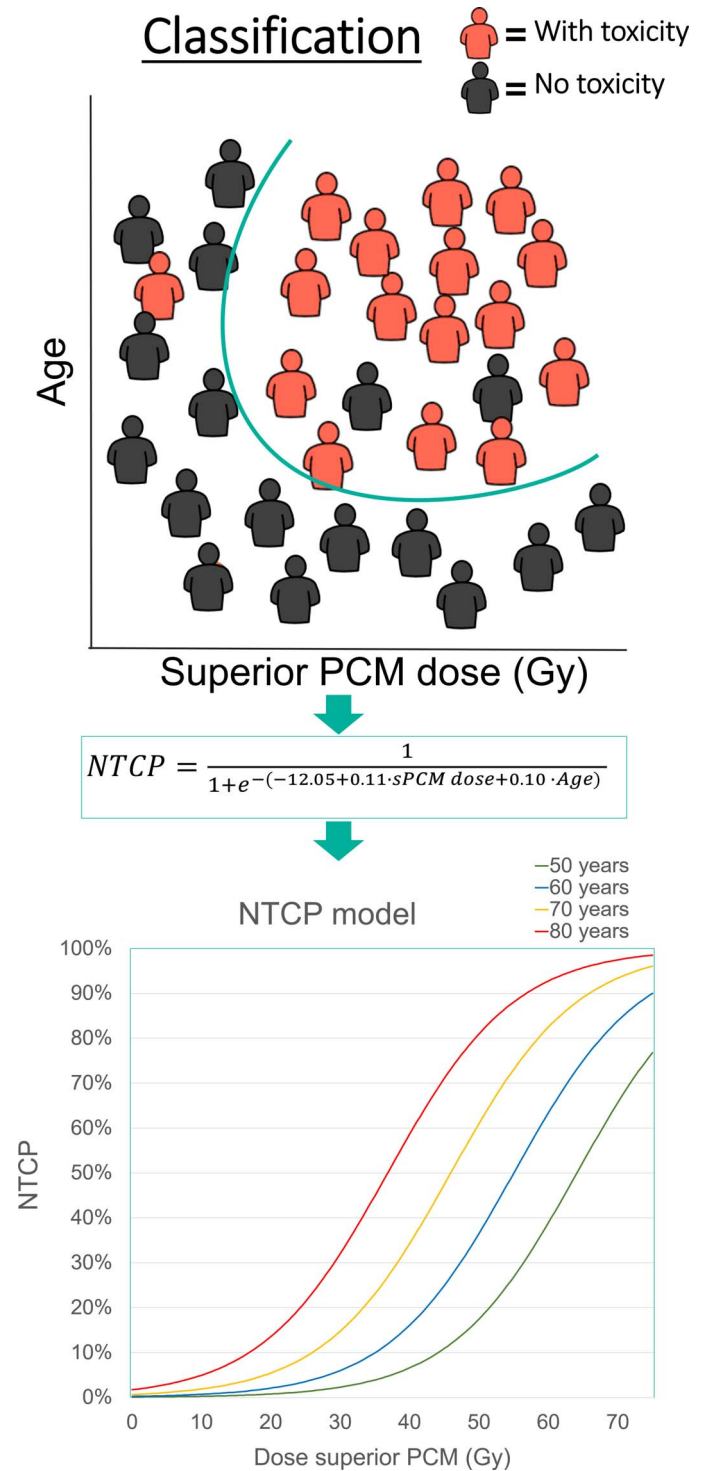
Distributed under

Creative Commons CC-BY

OPEN ACCESS

<http://theijpt.org>

Figure 1. Normal tissue complication probability model optimization.



The β_0 is the intercept coefficient, which is a constant; the other β are the coefficients multiplied by, thus linked to, a specific variable, which could be for example the mean contralateral parotid gland dose or baseline xerostomia complaints.

Model development is performed by “training” on training data to (1) select an optimal set of variables that are significantly associated with the predicted toxicity, and (2) estimate the model coefficients (β) to fit the model to the data. **Figure 1** illustrates

Table 1. Level of evidence for NTCP-models, based on TRIPOD [28] and NVRO [29].

Level	Description
1a	NTCP model that is externally validated on an independent multi-institute dataset with different treatment modality (proton therapy)
1b	NTCP model that is externally validated on independent data from another institute
2a	NTCP model trained and externally validated on nonrandomly split of single-center data
2b	NTCP model trained and externally validated on randomly split of single-center data
3	NTCP model developed with internal validation
4a	Multivariable NTCP model without internal/external validation
4b	Univariable NTCP model without internal/external validation

Abbreviations: NTCP, normal tissue complication probability; TRIPOD, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; NVRO, Nederlandse Vereniging voor Radiotherapie en Oncologie.

a hypothetical example of an NTCP classification (ie, the equation); with a so-called loss function, the most optimal distinction is sought to separate patients with and without a toxicity. The linear prediction (s in the equation) that is obtained from this optimization process is the basis of the resulting logistic regression NTCP model (visual representation in **Figure 1**).

One of the commonly used variable selection approaches is the forward (or Step-Up) Stepwise selection, which adds variables to the model, while testing the significance of model improvement. Alternatively, backward Stepwise selection operates in a reverse manner, by entering the full pool of variables to the model and subsequently removing variables. This significant improvement is often tested with a likelihood-ratio test, or alternatively with Akaike Information Criterion. Another popular variable selection method is the L1, or Lasso, regularization where, like backward selection, all variables are presented to the model and by increasing the penalization term λ , the regularization shrinks the coefficients of the variables. Consequently, the effect size is reduced for the variables that do not contribute to the model, and thereby procedurally excluding variables when their coefficient approaches zero [22, 24].

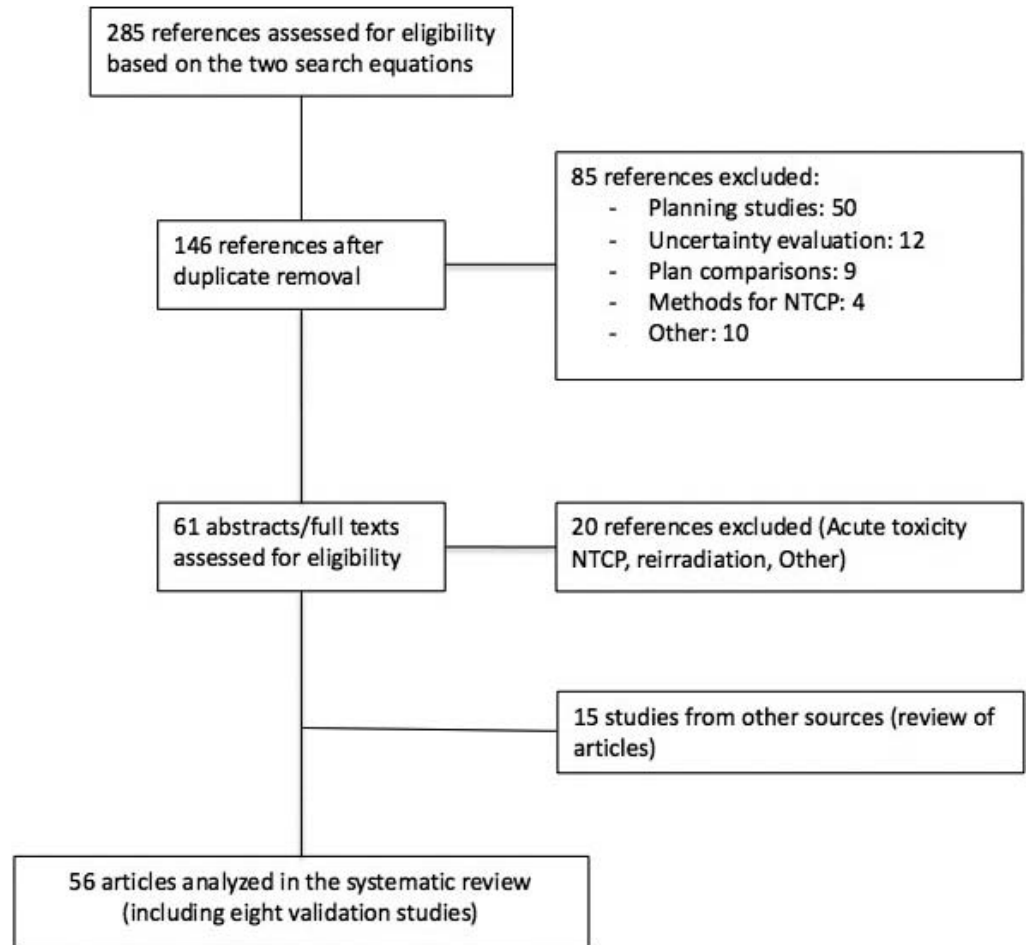
The inherent aim of developing an NTCP model is that the model is “generalizable” to new data. Internal validation can improve the robustness of the NTCP model by iteratively subsampling the training data and repeating the entire training process, by which robust variable selection and coefficient can be obtained. The most common variants are cross-validation and bootstrapping [25]. The Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement group has excellent modeling guidelines, and is a supply source to consult for consistent train and test modeling [26, 27].

Ultimately, the highest form of validation is obtained by performing external validation on data from multiple centers. Following the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement classification [28] and the Dutch proton indication protocol [29], **Table 1** details the different levels of evidence for NTCP models.

Materials and Methods

For the systematic review of existing NTCP models, the inclusion criteria were NTCP models that predict late toxicity after radiotherapy in head and neck cancer patients. The article search was performed on PubMed, by 2 board-certified radiation oncologists (PB and SS), using the following 2 search equations: “(‘head neck neoplasms’[MeSH Terms]) AND [(NTCP)[MeSH Terms] or ‘normal tissue complication’]” and “(‘head neck neoplasms’[MeSH Terms]) AND ‘normal tissue complication probability’.” After removal of duplicates the references were reviewed by the 2 investigators (PB and SS) and discrepancies were discussed with a third radiation oncologist (AL). The flow chart is presented in **Figure 2** [30]. Most articles excluded were using NTCP models to quantify the clinical meaning of differences according to variation in treatment planning systems. Sixty-one articles were retrieved that presented NTCP modeling. Twenty were finally excluded because of focus on acute toxicity, reirradiation, or because of focus on planning. The overall quality of the NTCP models was ranked as poor, fair, good, or excellent according to criteria pertaining to endpoint definition and collection, sample size, and data analysis/validation.

Figure 2. PRISMA flow chart.



Review of Current Normal Tissue Complication Probability Models

In total, 48 studies and 8 validation studies were found with NTCP modeling to predict late radiation-associated toxicity after irradiation of tumors in the head and neck area (**Figure 2, Supplemental Table**). The majority of these studies developed models for the prediction of late xerostomia ($n = 25$), followed by dysphagia/feeding tube dependency/esophageal stricture ($n = 7$), and brain necrosis/nerve palsy ($n = 6$) fewer studies investigated prediction of toxicities, such as hypothyroidism ($n = 3$), hearing loss/tinnitus ($n = 3$), trismus ($n = 2$), taste impairment ($n = 1$), dry eye ($n = 1$), laryngeal edema ($n = 1$), and hypopituitarism ($n = 1$) (2 studies developed models for 2 different toxicities, respectively). No study was found to report on models predicting late fibrosis/skin changes, osteoradionecrosis, hoarseness, aspiration/choking, weight loss, carotid artery calcification, fatigue, or second malignancies. The mean dose parameters that were generally selected for late toxicities were related to glandular structures (eg, parotid, submandibular, or thyroid gland) and muscular structures (eg, superior pharyngeal constrictor muscles or masseter). The maximum dose, or robust representation (eg, D_{100}) were selected in the final models of severe neurologic injury (eg, brain necrosis or nerve palsy), while this was not the case for hearing loss. Some of the final NTCP models included clinical variables, that is, for xerostomia, the baseline scores for dry mouth showed to be of importance [12, 20], while for tube feeding weight loss and chemotherapy were included in the final model [31]. Nevertheless, most studies did not select clinical variables in the NTCP model, which may be owing to most being Lyman–Kutcher–Burman models, the relatively small number of patients, or that potential variables were not included in the variable selection pool (eg, many studies did not include baseline symptom scores).

While many of the included studies were based on low patient numbers, 28 studies (29 models) (58%) had cohorts of over 100 patients. In addition, most studies lacked internal and/or external validation as follows: only 9 studies performed internal, 8 external validation, and 5 studies both, internal and external validation. Ten studies included patients treated with

nonadvanced treatment techniques only, and in 10 other studies at least a certain proportion of patients were included with 3-dimensional conformal radiation therapy. Only 4 studies (5 models) included patients with proton therapy. With these considerations, an overview of the most complete studies is depicted per symptom category in **Table 2**. Most of the models were rated as good or fair; none of the models was rated as excellent due to the reasons described above.

Normal Tissue Complication Probability Models for the Selection of Proton Therapy

The patients that benefit the most from proton therapy would ideally be identified with NTCP models that are trained on larger intensity modulated radiation therapy/volumetric modulated arc therapy-treated patient cohorts and subsequently validated in proton therapy patients. Unfortunately, these models do not currently exist (**Supplemental Table**). Two studies have developed NTCP models predicting brain necrosis in a proton therapy cohort only [46, 47]. Nevertheless, even with current challenges of determining the generalizability of NTCP models for both treatment modalities, NTCP differences (Δ NTCP) are still anticipated to be a better representation of the clinical importance (ie, expected clinical symptom burden reduction) than direct dose-volume histogram parameter differences. Consequently, Δ NTCP are considered a better marker to identify patients that will benefit most from proton compared with photon therapy. This is the foundation of the model-based approach [14], where validated models can serve as a tool to assess the late toxicity Δ NTCP between a photon and proton therapy plan. If the NTCP or Δ NTCP of a specific or multiple OARs exceed a clinically relevant toxicity-specific threshold, proton therapy might be the preferred treatment modality for that patient. Herein lies a weighting factor of the severity of toxicities, and some toxicities may be negligible of influence, such as hypothyroidism, which is treatable with medical substitution.

A recently published paper by Tambas et al [48] describes a practical institutional workflow of proton-photon plan comparison decision-making in a patient cohort from the University Medical Center Utrecht. Patients who have a $\geq 10\%$ higher risk of developing grade 2+ xerostomia, grade 2+ dysphagia, $\geq 5\%$ risk for feeding tube dependency, or a $\geq 15\%$ higher risk in the combination of those toxicities with photon compared with proton radiotherapy will be considered for proton therapy. The models used for this estimate [19, 20, 31] are predicting the toxicities at 6 months postradiotherapy and are externally validated. There is still a need to develop models for later time points, as Janssens et al [49], for example, showed a decrease in incidence of xerostomia from over 40% at 6 months postradiotherapy to approximately 25% 24 months postradiotherapy.

Despite the usefulness of NTCP models in guidance of patient selection for proton therapy, the final decision regarding the treatment modality may still be influenced by individual factors of the patient (ie, cost, travel, scheduling) and proton therapy capacity; thus, clinical prioritization and insurance issues should ultimately be left to the discretion of the patient and treating physician, taking into account the patient's goals and limitations of the NTCP models.

Normal Tissue Complication Probability Model Considerations

While dose-volume histogram parameters extraction comes with its challenges (eg, OAR definitions [50, 51], corrupt dose information, treatment plan system differences), the limiting factor in NTCP model data accumulation is the availability of qualitative toxicity data. While standardized follow-up programs are slowly being introduced [52, 53], many radiotherapy centers do not systematically collect toxicity information, leading to sparsely available and retrospectively collected data. Introduction of proton therapy has provoked improved toxicity collection, owing to the need to show the beneficial OAR sparing with proton therapy compared with photon therapy. Therefore, NTCP model development based on big data cohort is expected in years to follow.

The inclusion of a wide spectrum of different head and neck tumor subsites and stages into the modeling process can improve the model performance, as it provides a greater variability in dose administered to specific OARs. The downside of wide-inclusion criteria is that therapy patients receive may differ, such as receipt of induction and/or concurrent chemotherapy. While chemotherapy as a variable is often not selected in the final model, it may in fact affect the toxicity development, but may be too rudimentarily examined (**Supplemental Table**; eg, chemotherapy yes/no, no number of cycles), not adequately represented (eg, adaptation of agent dosing), or without sufficient patient numbers to measure the effect. Furthermore, the biological dose of OARs might be different according to the corresponding alpha/beta ratio, which might translate to a different toxicity development; for example, in the case of hypofractionated/accelerated radiotherapy treatment schedules. Current NTCP models are not designed to include altered fractionation, nevertheless some studies investigated the fractionation effect [31, 39, 54] (**Supplemental Table**), whereas others did not [38]. Other studies showed the potential of improving NTCP models with the addition of baseline imaging biomarkers [33, 55, 56], which represent patient-specific tissue characteristics. These image biomarkers in NTCP models are still novel and are hampered by the availability of magnetic resonance imaging or positron emission tomography-computed tomography for treatment planning in some institutions. None of the studies,

Table 2. Selection of the best available NTCP models per symptom category in head and neck cancer patients.

Author	Advanced RT			Model	Variables included in final model	Prospective toxicity collection	Prediction probability	Validation	Others	Overall quality
	Patients, n, (OARs)/events	techniques only	Endpoint ≥ 12 m post-RT							
Xerostomia										
Beetz 2012 [20]	17/83	Yes (IMRT)	No (6 m)	Logistic regression with bootstrapping	Xerostomia: Dmean PG contra, BL xerostomia Sticky saliva: Dmean SMG contra, Dmean SLG both, Dmean soft palate	Yes	AUC 0.68 (xerostomia), 0.70 (sticky saliva)	Yes (internal)		Good
Lee 2014 [32]	158/52	Yes (IMRT)	Yes (3 m and 12 m)	LASSO logistic regression with bootstrapping	12 m: Dmean PG ipsi/contra	Yes	AUC 12 m: 0.98 (HNSCC), 0.96 (NPC)	Yes (internal)		Good
van Dijk 2017 [33]	249/63 (sticky saliva)/100 (xerostomia)	No (IMRT, 3DCRT, VMAT)	Yes (1 y)	LASSO multivariate logistic regression	Xerostomia: Dmean PG contra, BL xerostomia, SRE GLRLM PG contra; Sticky saliva: Dmean SMG (both), BL sticky saliva, max HU both SMG	Yes	AUC xerostomia \pm imaging biomarker 0.77 and 0.75, sticky saliva \pm imaging biomarker 0.77 and 0.74	Yes (internal)	Limited added value of imaging biomarkers	Good
Dysphagia										
Christianen 2012 [19]	354/NA	No (3DRT, IMRT)	No (6 m)	Logistic regression	Dmean SPC, Dmean supraglottic larynx	Yes	AUC 0.80	Yes (external)		Good
Feeding tube dependency										
Wopken 2014 [31]	355/38	Yes (IMRT)	No (6 m), but clinically relevant	Logistic regression	T stage, weight loss, accelerated RT, chemo, Cetuximab, Dmean inf. PCM, Dmean PG contra, Dmean cricopharyngeal muscle	Yes	AUC 0.88	Yes (internal)		Good
Dysgeusia										
Sapir 2016 [34]	73/26	Yes (IMRT)	No (3 m)	LKB	Dmean oral cavity	Yes	NA	No	Oral cavity as OAR	Fair
Esophageal stricture										
Mavroidis 2003 [35]	82/26	No (3DCRT)	No 1–40 m (median 7 m)	LKB	Dmean esophagus	No	ROC = 0.84, X ² test = 0.95	No		Poor

Table 2. Continued.

Author	Patients, n, (OARs)/ events	Advanced RT techniques only	Endpoint ≥ 12 m post-RT	Model	Variables included in final model	Prospective toxicity collection	Prediction probability	Validation	Others	Overall quality
Brain necrosis										
Wang 2019 [36]	749/38	Yes (IMRT)	No (3.5–75 m, median 49 m)	Lasso binary regression	D0.5 cc and D10 selected for final model	No	AUC 0.68 (testing set)	Yes (internal)		Good
Zeng 2015 [37]	351/29	Yes (IMRT)	No (6–100 m, median 76 m)	Logistic regression	D _{1cc}	No	NA	No		Fair
Nerve palsy										
Chow 2019 [38]	330 nerves/46	Yes (IMRT)	No (min. FU 6 m, median 8.1 y)	Logistic regression	D _{1cc}	No	AUC 0.83	No	False high rate of palsies: min. FU all patients 6 m/ healthy control 8 y	Fair
Trismus										
Morimoto 2019 [39]	132/30	No (3DCRT, IMRT (percentage unclear)	No (6 m)	Logistic regression	Dmean TMJ contra, max. intercisial opening at BL	Yes	$P = .182$ (Hosmer and Lemeshow test)	No	Collinearity check; acc. fx in 95 patients	Good
Lindblom 2014 [40]	121/50	No (3DCRT, IMRT)	Yes for MID (21–127 m, median 66 m), unclear for QoL scores	Logistic regression	Different models with different variables studies; best fit for ipsi masseter for both endpoints	No (MID), yes (QoL)	0.77 and 0.73 for model with endpoint MID and QoL, respectively and ipsi masseter as variable	Yes (internal)	70 patients with acc. fx	Good
Hypothyroidism										
Ronjom 2013 [41]	203/35	Yes (IMRT)	Yes (1 y, 2 y)	Logistic regression	Dmean thyroid gland, thyroid gland volume	No (but objective criterion)	NA	No		Good
Hearing loss										
Marzi 2015 [42]	280 ears/73	Yes (PT)	Yes (median FU 26 m)	LKB	Dmean inner ear	Yes	AUC 0.86	No		Fair
Tinnitus										
Lee 2015 [43]	422 ears/49	Yes (IMRT)	Yes (51 m, range 36–77 m)	LKB Logistic regression	Dmean cochlea ipsi	No	LKB: 0.76, Logistic: 0.76	No	Influence of chemotherapy not accounted for	Fair
Dry eye										
Bhandare 2012 [44]	78/40	No (EBRT 1996–2000)	No (mean 0.9 y)	Logistic regression with bootstrapping	Dmax lacrimal gland	No	NA	No		Fair
Laryngeal edema										
Rancati 2009 [45]	48/25	Yes (IMRT)	Yes (15 m)	Lyman Logit	Dmean larynx	No (but objective criterion)	NA	No		Fair

Table 2. Continued.

Author	Patients, n, (OARs)/ events	Advanced RT techniques only	Endpoint ≥ 12 m post-RT	Model	Variables included in final model	Prospective toxicity collection	Prediction probability	Validation	Others	Overall quality
Hypopituitarism										
Marzi 2015 [42]	103/45	Yes (PT)	Yes (median 26 m)	LKB	Dmean pituitary gland	Yes	AUC 0.86	No		Fair

Abbreviations: NTCP, normal tissue complication probability; OAR, organ at risk; RT, radiotherapy; prosp., prospective; IMRT, intensity modulated radiation therapy; Dmean, mean dose; PG, parotid gland; Contra, contralateral; BL, baseline; SMG, submandibular gland; SLG, sublingual gland; AUC, area under curve; ipsi, ipsilateral; HNSCC, head neck squamous cell cancer; NPC, nasopharyngeal cancer; 3DCRT, 3-dimensional conformal radiotherapy; VMAT, volumetric modulated arc therapy; SRE GLRLM, short run emphasis gray level co-occurrence; HU, Hounsfield unit; NA, not assessed; PCM, Pharyngeal Constrictor Muscle; LKB, Lyman–Kutcher–Burman; ROC, receiver operator curve; min, minimum; FU, follow-up; TMJ, temporomandibular joint; acc. fx, accelerated fractionation; MID, maximal interincisal distance; QoL, quality of life; PT, proton therapy; EBRT, external beam radiotherapy.

however, has taken into account if the patient had major surgery before radiotherapy. For example, if a salivary gland had to be removed because of tumor infiltration, this adds a significant risk to the development of xerostomia.

Future Directions

Radiation-induced toxicities are often difficult to predict before treatment, as they are patient-specific, complex, interdependent, and nondeterministic. Currently, the vast majority of NTCP models rely on minimal input variables, making the estimated risk of radiation-induced side effects relatively simplistic. New modeling approaches, such as machine and deep learning, have the capacity to handle high-dimensional data, such as imaging, spatial localization of radiation dose, and potentially alterations in intertoxicity interactions over time. Nevertheless, there will always remain patient-specific factors, that are difficult to account for in modeling, like genetic susceptibility to radiotherapy toxicity, patient adherence with medication and abstaining from smoking/alcohol, tolerability of concurrent chemotherapy or targeted therapy, unexpected treatment breaks, and so on.

With modern photon techniques, like intensity modulated radiation therapy and volumetric modulated arc therapy, there is increased heterogeneity of the dose within OARs [57], which can improve the NTCP, albeit with higher integral dose [58]. In other words, the need to include the information from the entire dose-volume histogram curve or maybe even the full 3-dimensional spatial distribution of dose may be needed to improve the performance of NTCP models in the current and future era. Moreover, the applicability of intensity modulated radiation therapy studies to patients receiving proton therapy is limited as the relative biological effectiveness function needs to be considered. Indeed, it has been demonstrated that relative biological effectiveness varies along the beam path and depending on where the Bragg peaks are distributed, and this could lead to increased toxicity. In a series of 34 children treated using proton therapy, Peeler and colleagues [59] showed that showed T2-FLAIR hyperintensity was dependent on linear energy transfer dose, which indicated a possible variable biological dose effectiveness with clinical implications. For photon models to be incorporated to proton therapy, relative biological effectiveness–based weighting would ideally be needed although the validation of such models will require a lot of prospectively followed patients. Furthermore, range uncertainty in proton therapy needs to be considered and anatomic changes, such as paranasal sinus filling, swelling of irradiated tissue, or weight loss under therapy can lead to a lack of robustness in proton plans. This effect is minimized in photon therapy but can become quite significant in proton therapy, especially if OARs are lying just behind the target/Bragg peak, and can vary from institution to institution and with differing forms of proton delivery technique (passively beam scattered protons versus intensity modulated proton therapy). It has been previously shown that NTCP models developed for photon therapy did retain discriminatory properties for proton therapy and could hence be used to select patients [60]. Furthermore, the conformality afforded by proton therapy decreases concerns such that an NTCP model can reliably score a radiation plan for both photons and protons [61, 62].

As dedicated NTCP modeling for particle therapy is being developed, the issue of modeling for reirradiation becomes particularly salient as many centers are using proton therapy in this setting. The specific challenges in these cases include incorporation of different fractionation schemes, accounting for normal tissue recovery and dose-volume overlap [32].

The various challenges with developing adequate models are numerous but one important step with the right study design is through a multicenter approach. Van den Bosch and colleagues [58] from the Netherlands have discussed that consistent definitions of predictor and outcomes variables as well as use of standardized scoring systems are needed to reduce

heterogeneity of data. Uniform delineation guidelines should also be used to improve consistency of OARs. To achieve this aim, a multicenter prospective trial is needed so that the final models can be applicable in a variety of settings. Currently, there are 2 ongoing head and neck multicenter, randomized trials (MD Anderson IMPT vs IMRT Trial, NCT01893307 [63] and UK TORPEDO Trial, ISRCTN16424014 [64]) that have the potential to validate and expand these NTCP models to achieve a more robust prediction of clinical benefit for appropriate patient selection and access to proton therapy. Novel NTCP modeling and validation has significant health policy considerations for future head and neck cancer patients, and buy-in from the scientific community is needed with government level funding to achieve this aim.

However, NTCP modeling will always be an estimation of risk, rather than a real prediction of toxicity. With all the techniques described above (machine/deep learning, multicenter validation, etc.) models will significantly improve in the next decades and can help as a tool to select patients most likely to benefit from proton therapy or to define organs at risk, which should be spared with higher priority. However, these models will never incorporate all factors, which influence the outcome of every single patient and should therefore never replace the physician's clinical judgement.

In conclusion, we presented a systematic review of NTCP modeling studies performed for head and neck cancers. These studies have improved in quality over time and will help define how to select a treatment modality for a given patient. Future investigations with large number of cases are needed to apply NTCP models in practice and should also include reduction in acute toxicity (ie, mucositis) to aid with improved selection of patients.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflicts of Interest: Sonja Stieb, MD received funding from the Swiss Cancer League (BIL KLS-4300-08-2017). Anna Lee, MD, MPH received funding and salary support unrelated to this project from the National Institutes of Health (NIH) National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Education Programs for Residents and Clinical Fellows Grant (R25EB025787-01). Steven J. Frank, MD is an Associate Editor of the *International Journal of Particle Therapy*. Dr. Frank is a paid consultant/advisory board member for Varian and advisory board member for Hitachi and Breakthrough Chronic Care. He is founder and director of and has received grants and personal fees from C4 Imaging. He reports grants from Eli Lilly, Elekta, grants and honoraria from Hitachi and personal fees from Boston Scientific and the National Comprehensive Cancer Center (NCCN). Clifton David Fuller, MD, PhD has received direct industry grant support, speaking honoraria and travel funding from Elekta AB. Dr. Fuller received funding from the National Institute for Dental and Craniofacial Research Award (1R01DE025248-01/R56DE025248) and Academic-Industrial Partnership Award (R01DE028290), the NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148) outside the scope of this project. The authors have no relevant conflicts of interest to disclose.

Funding: Clifton David Fuller, MD, PhD received funding in support of this project from the Division of Mathematical Sciences, Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant (NSF1557679), the National Institutes of Health (NIH) Big Data to Knowledge (BD2K) Program of the National Cancer Institute (NCI) Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825), the NIH/NCI Cancer Center Support Grant (CCSG) Pilot Research Program Award from the UT MD Anderson CCSG Radiation Oncology and Cancer Imaging Program (P30CA016672), the NIH/NCI Head and Neck Specialized Programs of Research Excellence (SPORE) Developmental Research Program Award (P50CA097007), and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Education Program (R25EB025787).

Ethical approval: This review did not involve human subjects/participants and was exempt from institutional review board approval.

References

1. Beadle BM, Liao KP, Elting LS, Buchholz TA, Ang KK, Garden AS, Guadagnolo BA. Improved survival using intensity-modulated radiation therapy in head and neck cancers: a SEER-Medicare analysis. *Cancer*. 2014;120:702–10.
2. Pignon JP, le Maitre A, Maillard E, Bourhis J, Group M-NC. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17, 346 patients. *Radiother Oncol*. 2009;92:4–14.
3. Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. *Oncologist*. 2010;15:994–1001.

4. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW. Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer*. 2008;44:1345–89.
5. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol*. 2010;11:781–9.
6. Braakhuis BJ, Visser O, Leemans CR. Oral and oropharyngeal cancer in the Netherlands between 1989 and 2006: increasing incidence, but not in young adults. *Oral Oncol*. 2009;45:e85–9.
7. Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol*. 2014;50:565–74.
8. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol*. 2008;26:3770–6.
9. Leeman JE, Romesser PB, Zhou Y, McBride S, Riaz N, Sherman E, Cohen MA, Cahlon O, Lee N. Proton therapy for head and neck cancer: expanding the therapeutic window. *Lancet Oncol*. 2017;18:e254–e65.
10. Steneker M, Lomax A, Schneider U. Intensity modulated photon and proton therapy for the treatment of head and neck tumors. *Radiother Oncol*. 2006;80:263–7.
11. van de Water TA, Bijl HP, Schilstra C, Pijls-Johannesma M, Langendijk JA. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: a systematic review of literature. *Oncologist*. 2011;16:366–77.
12. van Dijk LV, Steenbakkers RJ, ten Haken B, van der Laan HP, van 't Veld AA, Langendijk JA, Korevaar EW. Robust intensity modulated proton therapy (IMPT) increases estimated clinical benefit in head and neck cancer patients. *PLoS One*. 2016;11:e0152477.
13. Jakobi A, Bandurska-Luque A, Stutzer K, Haase R, Lock S, Wack LJ, Monnich D, Thorwarth D, Perez D, Luhr A, Zips D, Krause M, Baumann M, Perrin R, Richter C. Identification of patient benefit from proton therapy for advanced head and neck cancer patients based on individual and subgroup normal tissue complication probability analysis. *Int J Radiat Oncol Biol Phys*. 2015;92:1165–74.
14. 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;107:267–73.
15. Teguh DN, Levendag PC, Voet PW, Al-Mamgani A, Han X, Wolf TK, Hibbard LS, Nowak P, Akhlat H, Dirks ML, Heijmen BJ, Hoogeman MS. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys*. 2011;81:950–7.
16. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017;44:547–57.
17. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, R JHMS, Langendijk JA, Gooding MJ, Brouwer CL. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiother Oncol*. 2020;142:115–23.
18. Houweling AC, Philippens ME, Dijkema T, Roesink JM, Terhaard CH, Schilstra C, Ten Haken RK, Eisbruch A, Raaijmakers CP. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys*. 2010;76:1259–65.
19. Christianen ME, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, Doornaert P, Koken PW, Leemans CR, Rinkel RN, de Bruijn MJ, de Bock GH, Roodenburg JL, van der Laan BF, Slotman BJ, Verdonck-de Leeuw IM, Bijl HP, Langendijk JA. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. *Radiother Oncol*. 2012;105:107–14.
20. Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, Vissink A, van der Laan BF, Leemans CR, Bijl HP, Christianen ME, Steenbakkers RJ, Langendijk JA. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. *Radiother Oncol*. 2012;105:101–6.
21. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, Yang J. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41:050902.

22. van der Schaaf A, Xu CJ, van Luijk P, Van't Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size. *Radiother Oncol.* 2012;105:115–21.
23. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl.* 1985;8:S13–9.
24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33:1–22.
25. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001;54:774–81.
26. Collins GS, Reitsma JB, Altman DG, Moons KGM, members of the Tg. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Eur Urol.* 2015;67:1142–51.
27. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162:W1–73.
28. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg.* 2015;102:148–58.
29. NVRO (Nederlandse Vereniging voor Radiotherapie en Oncologie), Langendijk JA. Langendijk indicatieprotocol protontherapie (versie 2.2). Bijlage 5: Level of evidence voor NTCP-modellen. https://nvro.nl/images/documenten/rapporten/2019-08-15__Landelijk_Indicatieprotocol_Protontherapie_Hoofdhals_v2.2.pdf. Accessed March 25, 2021.
30. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
31. Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkens RJ, Doornaert P, Slotman BJ, Oosting SF, Christianen ME, van der Laan BF, Roodenburg JL, Leemans CR, Verdonck-de Leeuw IM, Langendijk JA. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiother Oncol.* 2014;113:95–101.
32. Lee TF, Liou MH, Huang YJ, Chao PJ, Ting HM, Lee HY, Fang FM. LASSO NTCP predictors for the incidence of xerostomia in patients with head and neck squamous cell carcinoma and nasopharyngeal carcinoma. *Sci Rep.* 2014 Aug 28;4:6217. doi: 10.1038/srep06217. PMID: 25163814; PMCID: PMC5385804.
33. van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, Sijtsma NM, Steenbakkens R. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother Oncol.* 2017;122:185–91.
34. Sapir E, Tao Y, Feng F, Samuels S, El Naqa I, Murdoch-Kinch CA, Feng M, Schipper M, Eisbruch A. Predictors of dysgeusia in patients with oropharyngeal cancer treated with chemotherapy and intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys.* 2016;96:354–61.
35. Mavroidis P, Laurell G, Kraepelien T, Fernberg JO, Lind BK, Brahme A. Determination and clinical verification of dose-response parameters for esophageal stricture from head and neck radiotherapy. *Acta Oncol.* 2003;42(8):865–81. doi: 10.1080/02841860310012833. PMID: 14968948.
36. Wang J, Miao Y, Ou X, Wang X, He X, Shen C, Ying H, Hu W, Hu C. Development and validation of a model for temporal lobe necrosis for nasopharyngeal carcinoma patients with intensity modulated radiation therapy. *Radiat Oncol.* 2019;14:42. doi:10.1186/s13014-019-1250-z
37. Zeng L, Huang SM, Tian YM, Sun XM, Han F, Lu TX, Deng XW. Normal tissue complication probability model for radiation-induced temporal lobe injury after intensity-modulated radiation therapy for nasopharyngeal carcinoma. *Radiology.* 2015;276:243–9.
38. Chow JCH, Cheung KM, Au KH, Zee BCY, Lee J, Ngan RKC, Lee AWM, Yiu HHY, Li KWS, Leung AKC, Chan JCH, Lee FKH, Wong KH. Radiation-induced hypoglossal nerve palsy after definitive radiotherapy for nasopharyngeal carcinoma: clinical predictors and dose-toxicity relationship. *Radiother Oncol.* 2019;138:93–8.
39. Morimoto M, Bijl HP, A VDS, Xu CJ, Steenbakkens R, Chouvalova O, Yoshioka Y, Teshima T, Langendijk JA. Development of normal tissue complication probability model for trismus in head and neck cancer patients treated with radiotherapy: the role of dosimetric and clinical factors. *Anticancer Res.* 2019;39:6787–98.

40. Lindblom U, Gärskog O, Kjellén E, Laurell G, Levring Jäghagen E, Wahlberg P, Zackrisson B, Nilsson P. Radiation-induced trismus in the ARTSCAN head and neck trial. *Acta Oncol.* 2014;53:620–7.
41. Rønjom MF, Brink C, Bentzen SM, Hegedüs L, Overgaard J, Johansen J. Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: normal tissue complication probability modeling with latent time correction. *Radiother Oncol.* 2013;109:317–22.
42. De Marzi L, Feuvret L, Boulé T, Habrand JL, Martin F, Calugaru V, Fournier-Bidoz N, Ferrand R, Mazal A. Use of gEUD for predicting ear and pituitary gland damage following proton and photon radiation therapy. *Br J Radiol.* 2015Apr;88(1048):20140413. doi: 10.1259/bjr.20140413. Epub 2015 Feb 11. PMID: 25671247; PMCID: PMC4651261.
43. Lee TF, Yeh SA, Chao PJ, Chang L, Chiu CL, Ting HM, Wang HY, Huang YJ. Normal tissue complication probability modeling for cochlea constraints to avoid causing tinnitus after head-and-neck intensity-modulated radiation therapy. *Radiat Oncol.* 2015;10:194. doi: 10.1186/s13014-015-0501-x
44. Bhandare N, Moiseenko V, Song WY, Morris CG, Bhatti MT, Mendenhall WM. Severe dry eye syndrome after radiotherapy for head-and-neck tumors. *Int J Radiat Oncol Biol Phys.* 2012;82:1501–8.
45. Rancati T, Fiorino C, Sanguineti G. NTCP modeling of subacute/late laryngeal edema scored by fiberoptic examination. *Int J Radiat Oncol Biol Phys.* 2009;75:915–23.
46. Niyazi M, Niemierko A, Paganetti H, Sohn M, Schapira E, Goldberg S, Adams J, Kim V, Oh KS, Hwang WL, Lu HM, Belka C, Busse PM, Loeffler JS, Shih HA. Volumetric and actuarial analysis of brain necrosis in proton therapy using a novel mixture cure model. *Radiother Oncol.* 2020;142:154–61.
47. Engeseth GM, Stieb S, Mohamed ASR, He R, Stokkevåg CH, Brydoy M, Fuller CD, Garden AS, Rosenthal DI, Phan J, Morrison WH, Reddy JP, Wu R, Zhang X, Frank SJ, Brandon Gunn G. Outcomes and patterns of radiation associated brain image changes after proton therapy for head and neck skull base cancers. *Radiother Oncol.* 2020;151:119–25.
48. Tambas M, Steenbakkens RJHM, van der Laan HP, Wolters AM, Kierkels RGJ, Scandurra D, Korevaar EW, Oldehinkel E, van Zon-Meijer TWH, Both S, van den Hoek JGM, Langendijk JA. First experience with model-based selection of head and neck cancer patients for proton therapy. *Radiother Oncol.* 2020 Oct;151:206–13.
49. Janssens GO, Langendijk JA, Terhaard CH, Doornaert PA, van den Ende P, de Jong MA, Takes RP, Span PN, Kaanders JH. Quality-of-life after radiotherapy for advanced laryngeal cancer: results of a phase III trial of the Dutch Head and Neck Society. *Radiother Oncol.* 2016;119:213–20.
50. Brouwer CL, Steenbakkens RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, Chouvalova O, Burlage FR, Meertens H, Langendijk JA, van 't Veld AA. 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol.* 2012;7:32.
51. Brouwer CL, Steenbakkens RJ, Bourhis J, Budach W, Grau C, Gregoire V, van Herk M, Lee A, Maingon P, Nutting C, O'Sullivan B, Porceddu SV, Rosenthal DI, Sijtsema NM, Langendijk JA. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol.* 2015;117:83–90.
52. Matuszak MM, Fuller CD, Yock TI, Hess CB, McNutt T, Jolly S, Gabriel P, Mayo CS, Thor M, Caissie A, Rao A, Owen D, Smith W, Palta J, Kapoor R, Hayman J, Waddle M, Rosenstein B, Miller R, Choi S, Moreno A, Herman J, Feng M. Performance/outcomes data and physician process challenges for practical big data efforts in radiation oncology. *Med Phys.* 2018;45:e811–19.
53. Vergeer MR, Doornaert PA, Rietveld DH, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int J Radiat Oncol Biol Phys.* 2009;74:1–8.
54. Boomsma MJ, Bijl HP, Christianen ME, Beetz I, Chouvalova O, Steenbakkens RJ, van der Laan BF, Wolffenbuttel BH, Oosting SF, Schilstra C, Langendijk JA. A prospective cohort study on radiation-induced hypothyroidism: development of an NTCP model. *Int J Radiat Oncol Biol Phys.* 2012;84:e351–6.
55. van Dijk LV, Thor M, Steenbakkens R, Apte A, Zhai TT, Borra R, Noordzij W, Estilo C, Lee N, Langendijk JA, Deasy JO, Sijtsema NM. Parotid gland fat related magnetic resonance image biomarkers improve prediction of late radiation-induced xerostomia. *Radiother Oncol.* 2018;128:459–66.
56. van Dijk LV, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, Sijtsema NM, Steenbakkens R. (18)F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. *Radiother Oncol.* 2018;126:89–95.

57. Robinson M, Sabbagh A, Muirhead R, Durrant L, Van den Heuvel F, Hawkins M. Modeling early haematologic adverse events in conformal and intensity-modulated pelvic radiotherapy in anal cancer. *Radiother Oncol.* 2015;117:246–51.
58. Van den Bosch L, Schuit E, van der Laan HP, Reitsma JB, Moons KGM, Steenbakkers R, Hoebbers FJP, Langendijk JA, van der Schaaf A. Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy. *Radiother Oncol.* 2020;148:151–6.
59. Peeler CR, Mirkovic D, Titt U, Blanchard P, Gunther JR, Mahajan A, Mohan R, Grosshans DR. Clinical evidence of variable proton biological effectiveness in pediatric patients treated for ependymoma. *Radiother Oncol.* 2016;121:395–401.
60. Blanchard P, Wong AJ, Gunn GB, Garden AS, Mohamed ASR, Rosenthal DI, Crutison J, Wu R, Zhang X, Zhu XR, Mohan R, Amin MV, Fuller CD, Frank SJ. Toward a model-based patient selection strategy for proton therapy: external validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort. *Radiother Oncol.* 2016;121:381–6.
61. Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van der Schaaf A, Schuit E. Clinical trial strategies to compare protons with photons. *Semin Radiat Oncol.* 2018;28:79–87.
62. Prayongrat A, Umegaki K, van der Schaaf A, Koong AC, Lin SH, Whitaker T, McNutt T, Matsufuji N, Graves E, Mizuta M, Ogawa K, Date H, Moriwaki K, Ito YM, Kobashi K, Dekura Y, Shimizu S, Shirato H. Present developments in reaching an international consensus for a model-based approach to particle beam therapy. *J Radiat Res.* 2018;59:i72–i6.
63. Randomized trial of intensity-modulated proton Beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for the treatment of oropharyngeal cancer of the head and neck. ClinicalTrials.gov.identifier: NCT01893307. <https://clinicaltrials.gov/ct2/show/NCT01893307>. Updated October 1, 2020. Accessed March 25, 2021.
64. Price J, Hall E, West C, Thomson D. TORPEdO – a phase III trial of intensity-modulated proton beam therapy versus intensity-modulated radiotherapy for multi-toxicity reduction in oropharyngeal cancer. *Clin Oncol (R Coll Radiol).* 2020;32:80–4.