HHS Public Access

Author manuscript

Radiother Oncol. Author manuscript; available in PMC 2017 June 18.

Published in final edited form as:

Radiother Oncol. 2016 December ; 121(3): 381–386. doi:10.1016/j.radonc.2016.08.022.

Towards 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

P Blanchard, MD PhD[1,3], **AJ Wong, BS**[1], **GB Gunn, MD**[1], **AS Garden, MD**[1], **ASR Mohamed, MD PhD**[1], **DI Rosenthal, MD**[1], **J Crutison**[1], **R Wu, MS**[2], **X Zhang, PhD**[2], **XR Zhu, PhD**[2], **R Mohan, PhD**[2], **MV Amin, CMD**[2], **CD Fuller, MD PhD**[1], and **SJ Frank, MD**[1]

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

[2]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[3]Department of Radiation Oncology, Gustave Roussy Cancer Campus, Villejuif, France

Abstract

Objective—To externally validate head and neck cancer (HNC) photon-derived normal tissue complication probability (NTCP) models in patients treated with proton beam therapy (PBT).

Methods—This prospective cohort consisted of HNC patients treated with PBT at a single institution. NTCP models were selected based on the availability of data for validation and evaluated using the leave-one-out cross-validated area under the curve (AUC) for the receiver operating characteristics curve.

Results—192 patients were included. The most prevalent tumor site was oropharynx (n=86, 45%), followed by sinonasal (n=28), nasopharyngeal (n=27) or parotid (n=27) tumors. Apart from the prediction of acute mucositis (reduction of AUC of 0.17), the models overall performed well. The validation (PBT) AUC and the published AUC were respectively 0.90 versus 0.88 for feeding tube 6 months post-PBT; 0.70 versus 0.80 for physician rated dysphagia 6 months post-PBT; 0.70 versus 0.80 for dry mouth 6 months post-PBT; and 0.73 versus 0.85 for hypothyroidism 12 months post-PBT.

Conclusion—While the drop in NTCP model performance was expected in PBT patients, the models showed robustness and remained valid. Further work is warranted, but these results support the validity of the model-based approach for treatment selection for HNC patients.

Corresponding author: Steven J. Frank, MD, Associate Professor, Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 853, Houston, TX 77030, Phone: 713/563-2300; Fax: 713/563-1521; sjfrank@mdanderson.org.

Conflict of Interest Statement: Dr. Frank receives honoraria from Hitachi, Varian and IBA and research grant from Elekta. The other authors declare to have no conflict of interest regarding the topic under consideration.

Previous publication: This study has never been presented and has not been submitted for publication elsewhere. It has however been submitted for next ASTRO 2016 meeting and will be presented as an oral poster discussion.

Keywords

Intensity-modulated proton therapy; head and neck cancer; radiation therapy; complication; prediction; late toxicity; model

Introduction

A recognized advantage of proton therapy over photon therapy is achieving higher dose conformity and thereby reducing dose to normal tissue outside target volumes [1]. However, while randomized trials comparing protons and photons have been and are being performed for a few selected clinical indications, the demonstrable clinical benefit derived from proton therapy is mostly derived from single center series. There remains an ongoing debate as to whether randomized controlled trials should be the preferred evidence-based investigation regarding proton beam therapy (PBT) [2].

Traditionally, evidence-based medicine would call for the randomization of the study population into photon and proton treatment arms. However, this approach only answers the question of whether proton or photon therapy is superior – an “all-or-none” question rather than evaluating the applicable benefit to an individual patient, and is also considered inefficient in a setting where patients might refuse randomization, as in the case of strong technological appeal [2]. Moreover, there may be some patients for whom there is certainty that they will not benefit from protons and some others there is certainty that they will. The alternate approach, model-based, would estimate the potential clinical benefit for protons over photons in terms of reduction in normal tissue complication probability (NTCP) for each individual patient and assign the patient to PBT only if the reduction in toxicity is above a specified threshold [3]. This is a desirable approach in the context of precision medicine in clinical oncology. It would also allow for a more informed selection of candidates to receive PBT, which is more costly and not as widely accessible as photon therapy [4]. A model-based approach is thus potentially cost-effective through the estimation of total costs in the light of quality-adjusted life years after each treatment approach [5]

The availability of well-fitted models for both proton and photon treated patients is a prerequisite of the model based approach for treatment selection [6]. However, NTCP models have mostly been developed in single center cohorts of patients that were all treated with photon therapy. There is a paucity of literature regarding NTCP models for patients receiving PBT, although dose-volume-toxicity relationships might differ between photons and protons given their different physical and biological properties. Therefore, there is a pressing need for valid NTCP models in a PBT treated population. To this end, we aimed to externally validate established photon-derived NTCP models in head and neck cancer patients who were treated with PBT.

Materials and Methods

Literature search and model selection

A literature search for NTCP models for head and neck cancer patients was performed on September 29, 2015 in PubMed using the keywords “head and neck cancer”, “normal tissue

complication probability", "radiotherapy" and variations on these keywords, as well as hand searching articles [7] and reviews for additional references. Overall, 133 references were retrieved and analyzed. The twelve identified models considered for external validation along with specific information on model development and validation are described in supplementary table 1. Models were examined for feasibility of analysis based upon the availability of the endpoint in our prospectively collected database. Five models were eventually selected (Supplementary table 1) [8–12].

Patient Selection

All adult patients (n = 359) treated with proton therapy for head and neck cancer between 2011 and 2015 at the Proton Therapy Center at The University of Texas MD Anderson Cancer Center were considered for inclusion. Patients were all included in one of two prospective cohorts of patients treated with proton therapy, the first one being open to all tumor sites (PCR05-0207, normal tissues, NCT00991094) while the second was specifically designed for head and neck malignancies (PA11-0803, NCT01627093). Patients had provided informed consent for data recording and analysis. Head and neck tumors involving mucosa (nasopharynx, oropharynx, oral cavity, hypopharynx, larynx), sinus and nose, or parotid and with at least a neck irradiation were considered. Patients treated for a periorbital (n=25), base of skull (n=13), skin (n=6) or "other" (n=61) tumor were excluded, as were patients that did not complete treatment (n=27), withdrew consent (n=2), or lost to follow-up (n=55). Patients previously irradiated in the head and neck region (n=29) were excluded from all models except the one related to acute mucositis [8].

Treatment planning and delivery

The general management of head and neck cancer patients undergoing proton therapy has been previously described in detail [1,13]. All but 7 patients were treated using intensity modulated proton therapy (IMPT). Briefly, all patients underwent non-contrast simulation computed tomography (CT) in immobilized supine position using head, neck, and shoulder thermoplastic mask, bite block with or without an oral stent and a posterior customized head, neck and shoulder mold. Target volumes and organs at risk were manually delineated and peer-reviewed for quality assurance purposes [14]. Doses were prescribed using a relative biological effectiveness (RBE) value of 1.1 and ranged from 60 Gy(RBE) for adjuvant treatment to 66–70 Gy(RBE) for definitive, while the elective regions received 54 to 63 Gy(RBE). Patients with well-lateralized cancers underwent ipsilateral neck irradiation [15]. Treatment planning was performed with Eclipse proton treatment planning system (version 8.9 and 11, Varian Medical Systems, Palo Alto, California), typically using 3 beams for whole-field bilateral neck IMPT plans: a left and right anterior oblique and a single posterior beam. The robustness of each treatment plan was considered in order to evaluate the sensitivity to uncertainties in patient setup and proton beam range [16,17]. Plan-specific quality assurance measurements were made prior to treatment delivery [18]. Daily kilovoltage orthogonal image guidance was used. Verification CT simulation were performed at weeks 1 and 4 of therapy and adaptive re-planning was performed if deemed necessary.

Clinical and dosimetric data collection

Clinical information and the primary endpoints of the selected models had been prospectively recorded. Additional clinical information was retrospectively retrieved as required. For each NTCP model, we used the same endpoint definition and inclusion criteria as described in published models. The five NTCP models considered for validation were: 1) persistence of feeding tube 6 months after treatment [9]; 2) physician-rated grade 2+ dysphagia 6 months following treatment [10]; 3) patient-rated dry mouth 6 months after treatment [12]; 4) hypothyroidism 12 months after treatment [11]; and 5) physician-rated grade 3+ acute mucositis during treatment [8].

Patient population for each analysis

Patients were selected according to each NTCP model inclusion criteria and the availability of the clinical endpoint. For acute mucositis, patients had to have a tumor arising from the oropharynx, nasopharynx or hypopharynx, but had no follow-up limit after treatment and could have been previously irradiated in the head and neck region. For swallowing endpoints at 6 months after treatment (feeding tube, grade 2+ dysphagia), patients had to be followed for at least 6 months, received neck radiotherapy, to not have been previously irradiated and to have had neither feeding tube nor severe weight loss prior to radiotherapy. For dry mouth at 6 months after treatment, patients had to have been followed for at least 6 months, received neck radiotherapy, and to not have been previously irradiated. For hypothyroidism 12 months post treatment, patients had to have been followed for at least 12 months, received low neck radiotherapy (unilateral radiotherapy allowed), to not have been previously irradiated, to not have had pre-treatment hypothyroidism (defined as elevated TSH prior to RT or at first follow-up visit), to not have had pre-treatment thyroid medication or thyroidectomy, and to have had available TSH and T4 measurements during follow-up.

Extraction of Dosimetric Parameters

For each patient who met inclusion criteria for an endpoint, the planning CT and dose grid were retrieved, then pertinent organs-at-risk (OAR) were auto-segmented according to published contouring guidelines [19] using a previously validated atlas-based auto-segmentation software ADMIRE® 1.12.0.0 (Elekta AB, Stockholm, Sweden) [20]. OARs consisted of pharyngeal constrictor muscles, contralateral parotid gland, cricopharyngeal muscle, supraglottic larynx, thyroid gland and buccal mucosa. The contours were individually checked by two experienced head and neck radiation oncologists (PB, ASRM) and appropriate modifications were made only if these were deemed substantial. Relevant doses to each pertinent OAR, usually mean dose, were obtained through the dose-volume histogram.

Statistical Methods

The models' discriminating ability was evaluated using the leave-one-out cross-validated area under the curve (AUC) of the receiver operating characteristics (ROC) curve. Values below 0.7 were considered as indicators of poor model performance. Model accuracy were compared to random prediction using the Wald test. To look for differences between the regression coefficients in the photon and proton cohorts, the regression models were also

refitted with the PBT data using the same covariates as described in the literature, but without specifying any initial value. Goodness of fit was measured using the same metric as in the development articles, that is the Hosmer-Lemeshow test or the R^2 statistics. Statistical analyses were conducted using SAS (version 9.3, SAS Institute Inc., Cary, NC).

Results

One hundred and ninety-two patients were selected. Patient, tumor and treatment characteristics are described in table 1. Median age was 60 years. Most patients were male (n=134, 70%), and most tumors were oropharyngeal cancers (n=86, 45%). The majority received full neck irradiation (n=145, 75%), median dose to the CTV1 was 66 Gy(RBE) (range: 60–70), delivered in 2 Gy(RBE) per fraction. Prior radiotherapy to the head and neck had been delivered to 35 patients (18%). Two-thirds of the patients received concomitant systemic therapy and 24% received induction chemotherapy (n=47). Surgery was performed prior to proton therapy in 40% of the patients, in general for tumors of the sinonasal region or parotid, or tumors arising in previously irradiated areas. In the dysphagia analysis cohort, only 3 patients had mucosal surgery prior to proton therapy, the others classified as surgery had either diagnostic tonsillectomy or neck sampling/dissection.

The number of patients and events included in each analysis is presented, along with cross validated AUC values, in table 2. The number of patients available for each analysis ranged from 58 (hypothyroidism) to 113 (acute mucositis). The number of events ranged from four (feeding tube six months after treatment) to 40 (acute mucositis or hypothyroidism). When looking at the cross validated AUC values, they were similar between the validation cohort and the published photon cohort for feeding tube 6 months post treatment (0.90 vs 0.88). For the other four endpoints there was an average 0.1 overall decrease in AUC value, the largest decrease being for acute mucositis (0.17). All these models remained significantly capable of predicting the outcomes, as shown by p-values compared to random prediction < 0.001 for each. Goodness of fit was correct for feeding tube and dysphagia, but was significantly lower in proton patients than in the development models for hypothyroidism, xerostomia and acute mucositis (table 2).

The regression coefficients were estimated for the covariates of three models for which discrimination capability on proton patients remained good and with a sufficient number of events, that is excluding the models for prediction of acute mucositis (decrease of AUC by 0.17) and for prediction of feeding tube dependence at 6 months after treatment (only four events and eight covariates in the published model). Results are presented in table 3 and show the overall qualitative and quantitative coherence between the published and the estimated coefficients, as well as a significant increase in goodness of fit estimated by the Hosmer-Lemeshow test.

Discussion

The external validity of photon-derived NTCP models for head and neck cancer patients in a population treated with PBT showed, as expected, a drop in model performance, with around a 10% decrease in AUC value. However, most of the models had an AUC ≥ 0.7, and

therefore showed robustness and remained valid for proton therapy patients. These results serve as a necessary first step toward supporting the validity of the model-based approach for allocation of proton- or photon-based treatment for head and neck cancer patients, provided continuous refinement of the predictive models is undertaken.

There are plausible explanations why the external validity was the lowest for the models predicting acute mucositis and thyroid dysfunction. Indeed grade 3 or higher acute mucositis is a subjective endpoint, and while its occurrence is very high in patients receiving definitive doses to a mucosal surface, it would be more relevant to know the extent of this mucositis in terms of percentage of mucosa involved. Besides this model was developed based on data for a relatively small number of patients. Regarding thyroid dysfunction, thyroid dose constraints were not routinely applied at our institution and the frequent treatment of prelaryngeal and paratracheal nodal stations along with bilateral levels three and four would likely preclude significant thyroid sparing. This resulted in a bimodal dose distribution for thyroid, with the majority of patients receiving 50 Gy(RBE) or higher doses, while a minority received very low mean doses, usually due to unilateral neck treatment, that might explain the lower external validity for this model.

This study was the first to evaluate NTCP models in PBT patients, and has several other strengths. First, the use of prospectively collected data from patients included in clinical registries ensures the quality of data, especially for baseline characteristics and toxicity endpoints. Second, the use of cross-validation for the computation of ROC curves and AUC values has produced estimations and confidence intervals that are less overestimated and more easily generalizable. The fact that model discrimination remains acceptable between different institutions using different protocols and ionizing radiations, offers further support for the model-based approach. As an example in the initial publication that developed the NTCP model for feeding tube dependency [9], all patients in the development cohort had prophylactic tube placement while none in our validation cohort had one, and post-operative patients were in theory excluded (although 19 had oral cavity cancer, which is usually surgically excised). The external validity of this NTCP model was however good, suggesting that it is not dependent on the use of a prophylactic or reactive feeding tube placement strategy.

The limitations of this work are on one hand related to our patient cohort and on the other hand related to the current NTCP models. Although our initial literature review has identified twelve NTCP models, only five could be evaluated in our patients, because the clinical endpoints of the other seven were not prospectively collected or because patient follow-up was not yet sufficient. However we believe that the models we have tested were largely developed using clinically relevant and easily assessable endpoints, using a rigorous methodology by an experienced team, and have been used to develop a model-based approach for allocation of PBT. These models were therefore the most important to externally validate. Besides the lower event rates in our cohort compared to the development cohorts might invalidate the statistical analysis, especially in the feeding tube model where only 4 events occurred. The second set of limitations are inherent to models themselves, either due to patient case mix or statistical methods employed. For example a recent publication showed that late dysphagia was related to a different set of muscle dose

constraints than in the models used here [9,10], which could be related to the inclusion of only oropharyngeal cancer and the use of a recursive partitioning approach in this latter publication [21]. Indeed the majority of NTCP models investigated in the present study are based on logistic regression, which makes assumptions with regards to dose-toxicity relationship that are valid only for a subset of toxicities, e.g. dry mouth or hypothyroidism, but not for many others, e.g. osteoradionecrosis, dysphagia or trismus. The use of non-parametric or machine learning approaches in the future generation of NTCP models would certainly help develop models that rely less on statistical assumptions [22]. Last, these models are all based on the dose planned but none have been adapted to the dose actually delivered and received by organs at risk. Uncertainties in target contouring, treatment planning and delivery of proton therapy and uncertainties in RBE are not taken into account in these models, other than in the confidence intervals around the estimated values.

Allocating scarce or costly resources to individuals who will likely benefit from them is of paramount importance, both to tailor treatments to each patient's needs, to rationalize the access to innovative treatments and to limit the rise of healthcare expenditure [5]. To this end one needs first to be able to predict outcomes reliably, and then individualize treatment allocation based on each patient's probability of developing the outcome. The outcomes considered can be efficacy, toxicity, or ideally a combination of these two. The benefits of proton therapy for oropharyngeal cancers are being investigated in a classical prospective randomized trial, (NCT01893307), but the question of patient selection remains of crucial importance. Proton therapy, provided it is used in adequately selected patients, could be cost-effective compared to standard treatment, the increase in costs related to PBT delivery being mitigated by a reduction in costs related to the management of side effects. This was suggested in a case report of head and neck cancers [23] as well as in a cost-effectiveness analysis based on NTCP models and planning studies [5]. With this in mind, the present study represents a necessary first step towards the selection of patients for either proton or photon therapy.

Several improvements to this work can be envisioned. The drop in discrimination capabilities and goodness of fit in PBT patients is similar as when models developed in patients treated with 3D conformal radiotherapy were applied to IMRT patients [24]. When trying to allocate treatment based on predicted toxicity, one does not need a model with a perfect fit, but rather a model that detects patients with large and therefore clinically relevant differences in toxicity between two treatment strategies. However, NTCP models should be refined further, preferably using non-parametric rapid learning modeling and multicenter databases [25], to iteratively improve model fit and prediction accuracy. The toxicity cut-off used for treatment allocation could vary from country to country due to the variety of healthcare systems, but the models would be similar. Given the uncertainties in the models and the possibility of erroneous predictions, a generous leeway should be observed in making decisions. These models should also include patient reported outcomes and focus on patient priorities [26], to improve the value of treatments [27]. Last, this approach should be extended for other tumor types where proton therapy is promising, such as lung, liver or esophageal cancers.

In conclusion, our study demonstrates the external validity of a set of NTCP models in head and neck cancer patients treated with proton therapy. Although improvement in model fit is warranted, our study provides a proof of concept for the use of the model-based approach as a strategy to select patients for proton therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant or financial support: Supported in part by the National Institutes of Health (NIH)/National Cancer Institute (NCI) Cancer Center Support (Core) Grant CA016672 and a U19 CA021239 to The University of Texas MD Anderson Cancer Center. Dr Blanchard received funding from The Foundation Nuovo Soldati for Medical Research, the Philippe Foundation and the FRM grant SPE20150331822. Dr. Fuller received/receives grant and/or salary support from the National Science Foundation, Division of Mathematical Sciences Quantitative Approaches to Biomedical Big Data (QuBBD)/Big Data to Knowledge (BD2K) Program (Paul Calabresi Clinical Oncology Award Program (K12 CA088084-06) and Clinician Scientist Loan Repayment Program (L30 CA136381-02); a General Electric Healthcare/MD Anderson Center for Advanced Biomedical Imaging In-Kind Award; an Elekta AB/MD Anderson Department of Radiation Oncology Seed Grant; the Center for Radiation Oncology Research at MD Anderson Cancer Center Seed Grant; and the MD Anderson Institutional Research Grant Program.

References

1. Frank SJ, Cox JD, Gillin M, Mohan R, Garden AS, Rosenthal DI, et al. Multifield optimization intensity modulated proton therapy for head and neck tumors: A translation to practice. Int J Radiat Oncol Biol Phys. 2014; 89:846–53. DOI: 10.1016/j.ijrobp.2014.04.019 [PubMed: 24867532]
2. Cox JD. Impediments to Comparative Clinical Trials With Proton Therapy. Int J Radiat Oncol. 2016; 95:4–8. DOI: 10.1016/j.ijrobp.2015.09.036
3. Langendijk JA, Lambin P, De Ruysscher 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. DOI: 10.1016/j.radonc.2013.05.007 [PubMed: 23759662]
4. Lievens Y, Pijls-Johannesma M. Health economic controversy and cost-effectiveness of proton therapy. Semin Radiat Oncol. 2013; 23:134–41. DOI: 10.1016/j.semradonc.2012.11.005 [PubMed: 23473691]
5. Ramaekers BLT, Grutters JPC, Pijls-Johannesma M, Lambin P, Joore MA, Langendijk JA. Protons in Head-and-Neck Cancer: Bridging the Gap of Evidence. Int J Radiat Oncol. 2013; 85:1282–8. DOI: 10.1016/j.ijrobp.2012.11.006
6. Widder J, van der Schaaf A, Lambin P, Marijnen CAM, Pignol J-P, Rasch CR, et al. The Quest for Evidence for Proton Therapy: Model-Based Approach and Precision Medicine. Int J Radiat Oncol. 2016; 95:30–6. DOI: 10.1016/j.ijrobp.2015.10.004
7. Jakobi A, Bandurska-Luque A, Stützer K, Haase R, Löck S, Wack L-J, et al. 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. DOI: 10.1016/j.ijrobp.2015.04.031 [PubMed: 26194685]
8. Bhide SA, Gulliford S, Schick U, Miah A, Zaidi S, Newbold K, et al. Dose-response analysis of acute oral mucositis and pharyngeal dysphagia in patients receiving induction chemotherapy followed by concomitant chemo-IMRT for head and neck cancer. Radiother Oncol. 2012; 103:88–91. DOI: 10.1016/j.radonc.2011.12.027 [PubMed: 22280809]
9. Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJHM, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. Radiother Oncol. 2014; 113:95–101. DOI: 10.1016/j.radonc.2014.09.013 [PubMed: 25443500]
10. Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective

observational study. Radiother Oncol. 2012; 105:107–14. DOI: 10.1016/j.radonc.2011.08.009 [PubMed: 21907437]

11. Boomsma MJ, Bijl HP, Christianen MEMC, Beetz I, Chouvalova O, Steenbakkers RJHM, et al. A prospective cohort study on radiation-induced hypothyroidism: development of an NTCP model. Int J Radiat Oncol Biol Phys. 2012; 84:e351–356. DOI: 10.1016/j.ijrobp.2012.05.020 [PubMed: 22717243]
12. Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol. 2012; 105:101–6. DOI: 10.1016/j.radonc.2012.03.004 [PubMed: 22516776]
13. Gunn GB, Blanchard P, Garden AS, Zhu XR, Fuller CD, Mohamed AS, et al. Clinical Outcomes and Patterns of Disease Recurrence After Intensity Modulated Proton Therapy for Oropharyngeal Squamous Carcinoma. Int J Radiat Oncol Biol Phys. 2016; 95:360–7. DOI: 10.1016/j.ijrobp.2016.02.021 [PubMed: 27084653]
14. Rosenthal DI, Asper JA, Barker JL, Garden AS, Chao KSC, Morrison WH, et al. Importance of patient examination to clinical quality assurance in head and neck radiation oncology. Head Neck. 2006; 28:967–73. DOI: 10.1002/hed.20446 [PubMed: 16823872]
15. Chronowski GM, Garden AS, Morrison WH, Frank SJ, Schwartz DL, Shah SJ, et al. Unilateral radiotherapy for the treatment of tonsil cancer. Int J Radiat Oncol Biol Phys. 2012; 83:204–9. DOI: 10.1016/j.ijrobp.2011.06.1975 [PubMed: 22019242]
16. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: The potential effects of calculational uncertainties. Phys Med Biol. 2008; 53:1027–42. DOI: 10.1088/0031-9155/53/4/014 [PubMed: 18263956]
17. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions. Phys Med Biol. 2008; 53:1043–56. DOI: 10.1088/0031-9155/53/4/015 [PubMed: 18263957]
18. Zhu XR, Li Y, Mackin D, Li H, Poenisch F, Lee AK, et al. Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. Cancers. 2015; 7:631–47. DOI: 10.3390/cancers7020631 [PubMed: 25867000]
19. Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. 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; doi: 10.1016/j.radonc.2015.07.041
20. Mohamed ASR, Ruangskul M-N, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, 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; 274:752–63. DOI: 10.1148/radiol.14132871 [PubMed: 25380454]
21. Dale T, Hutcheson K, Mohamed ASR, Lewin JS, Gunn GB, Rao AUK, et al. Beyond mean pharyngeal constrictor dose for beam path toxicity in non-target swallowing muscles: Dose–volume correlates of chronic radiation-associated dysphagia (RAD) after oropharyngeal intensity modulated radiotherapy. Radiother Oncol. 2016; 118:304–14. DOI: 10.1016/j.radonc.2016.01.019 [PubMed: 26897515]
22. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. Int J Radiat Oncol Biol Phys. 2015; 93:1127–35. DOI: 10.1016/j.ijrobp.2015.07.2286 [PubMed: 26581149]
23. Thaker NG, Frank SJ, Feeley TW. Comparative costs of advanced proton and photon radiation therapies: lessons from time-driven activity-based costing in head and neck cancer. J Comp Eff Res. 2015; 4:297–301. DOI: 10.2217/cer.15.32 [PubMed: 26274791]
24. Beetz I, Schilstra C, van Luijk P, Christianen MEMC, Doornaert P, Bijl HP, et al. External validation of three dimensional conformal radiotherapy based NTCP models for patient-rated xerostomia and sticky saliva among patients treated with intensity modulated radiotherapy. Radiother Oncol. 2012; 105:94–100. DOI: 10.1016/j.radonc.2011.11.006 [PubMed: 22169766]

25. Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. Adv Drug Deliv Rev. 2016; doi: 10.1016/j.addr.2016.01.006

26. List MA, Stracks J, Colangelo L, Butler P, Ganzenko N, Lundy D, et al. How Do head and neck cancer patients prioritize treatment outcomes before initiating treatment? J Clin Oncol. 2000; 18:877–84. [PubMed: 10673531]

27. Porter ME. What is value in health care? N Engl J Med. 2010; 363:2477–81. DOI: 10.1056/NEJMp1011024 [PubMed: 21142528]

28. 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. DOI: 10.1016/j.ijrobp.2009.04.087 [PubMed: 19747783]

29. Eisbruch A, Kim HM, Feng FY, Lyden TH, Haxer MJ, Feng M, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. Int J Radiat Oncol Biol Phys. 2011; 81:e93–99. DOI: 10.1016/j.ijrobp.2010.12.067 [PubMed: 21592678]

30. Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. 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. DOI: 10.1016/j.ijrobp.2009.07.1685 [PubMed: 20005639]

31. Lindblom U, Garskog O, Kjellen E, Laurell G, Levring Jaghagen E, Wahlberg P, et al. Radiation-induced trismus in the ARTSCAN head and neck trial. Acta Oncol. 2014; 53:620–7. DOI: 10.3109/0284186X.2014.892209 [PubMed: 24669774]

32. Lee T-F, Yeh S-A, Chao P-J, Chang L, Chiu C-L, Ting H-M, et al. Normal tissue complication probability modeling for cochlea constraints to avoid causing tinnitus after head-and-neck intensity-modulated radiation therapy. Radiat Oncol. 2015; 10:194.doi: 10.1186/s13014-015-0501-x [PubMed: 26377924]

Table 1

Patient and Treatment Characteristics

Characteristics		N (%)
Age	Median (range)	60 (19–92)
Gender	Female	58 (30%)
	Male	134 (70%)
Tumor site	Oropharynx	86 (45%)
	Nasopharynx	27 (14%)
	Sinonasal	28 (15%)
	Parotid	27 (14%)
	Neck only	14 (7%)
	Hypopharynx/Larynx	5 (2.5%)
	Oral cavity	5 (2.5%)
T-stage	T0–T2	102 (53%)
	T3–T4	50 (26%)
	Recurrent or no staging appropriate	40 (21%)
N-stage	N0–N1	72 (37.5%)
	N2–N3	80 (41.5%)
	Recurrent or no staging appropriate	40 (21%)
Prior radiotherapy	No	157 (82%)
	Yes	35 (18%)
Surgery	No	116 (60%)
	Yes	76 (40%)
Neck radiotherapy	Full neck	145 (75%)
	Upper neck	11 (6%)
	Not performed	36 (19%)
RT dose (Gy CGE)	Median (range)	66 (60–70)
Fractions (n)	Median (range)	33 (28–35)
Proton delivery	Intensity modulated proton therapy	185 (96%)
	Passive scattered proton therapy	7 (4%)
Concurrent CT	Platinum based	108 (56%)
	Cetuximab	19 (10%)
	Other	2 (1%)
	No	63 (33%)
Induction CT	Yes	47 (24%)
	No	145 (76%)

Table 2

Discrimination properties of photon-derived NTCP models in proton treated patients

Model		Development set	IMPT validation	IMPT cross-validation
Persistence of feeding tube 6 months post treatment [9]	N patients	355	89	89
	N events	38	4	4
	AUC [95% CI]	0.88	0.947 [0.85–1.00]	0.90 [0.75–1.00]
	Model fit (HL)	0.70	0.43	
Physician rated grade 2+ dysphagia 6 months post treatment [10]	N patients	354	89	89
	N events	NA	27	27
	AUC [95% CI]	0.8	0.708 [0.59–0.82]	0.697 [0.58–0.80]
	Model fit (HL)	NR	0.23	
Patient rated dry mouth 6 months post treatment [12]	N patients	161	94	94
	N events	83	36	36
	AUC [95% CI]	0.8	0.735 [0.63–0.83]	0.704 [0.59–0.81]
	Model fit (HL)	0.84	0.05	
Hypothyroidism 12 months post treatment [11]	N patients	105	58	58
	N events	35	40	40
	AUC [95% CI]	0.85	0.743 [0.57–0.91]	0.728 [0.55–0.90]
	Model fit (HL)	NR (“good”)	0.01	
Acute mucositis [8]	N patients	148	113	113
	N events	NP	40	40
	AUC [95% CI]	0.85	0.700 [0.60–0.80]	0.68 [0.58–.78]
	Model fit (R^2)	0.8	0.22	

Abbreviations: AUC, area under the ROC curve; HL, Hosmer-Lemeshow test; NR, no value reported; ROC, receiver operator characteristics The Hosmer-Lemeshow test evaluates the model fit, a p-value of 0.05 or lower indicates a poor fit to the data. The R^2 is another measure of goodness of fit, and values close to one are considered an indicator of a good fit.

Table 3

Comparison of published coefficients for NTCP models derived on photon patients and coefficients estimated using the same covariates on proton patients.

Endpoint	Coefficients	Published for photon cohort	Estimated for the proton cohort
Physician rated grade 2+ dysphagia 6 months post treatment [10]	Intercept	−6.09	−3.85 [SE: 1.19]
	Dmean SPC	0.057	0.04 [SE: 0.02]
	Dmean Supraglottic Larynx	0.037	0.018 [SE: 0.017]
	Model fit (Hosmer-Lemeshow test)	NR	p=0.66
Patient rated dry mouth 6 months post treatment [12] *	Intercept	−1.443	−1.79 [SE: 0.45]
	Dmean CL Parotid	0.047	0.075 [SE: 0.02]
	Model fit (Hosmer-Lemeshow test)	p=0.84	p=0.08
Hypothyroidism 12 months post treatment [11]	Intercept	0.11	1.93 [SE: 1.33]
	Dmean thyroid	0.062	0.025 [SE: 0.024]
	Thyroid Volume	−0.19	−0.17 [SE: 0.057]
	Model fit (Hosmer-Lemeshow test)	NR	p=0.53

Abbreviations: CL, contralateral; Dmean, mean dose; NR, not reported; SE, standard error; SPC, superior pharyngeal constrictor

* baseline dry mouth not included in the model as no patient had dry mouth at baseline

The Hosmer-Lemeshow test evaluates the model fit, a p-value of 0.05 or lower indicates a poor fit to the data