# A Model-Based Approach to Predict Short-Term Toxicity Benefits With Proton Therapy for Oropharyngeal Cancer

Rwigema, Jean-Claude M.; Langendijk, Johannes A.; van der Laan, Hans Paul; Lukens, John N.; Swisher-McClure, Samuel D.; Lin, Alexander

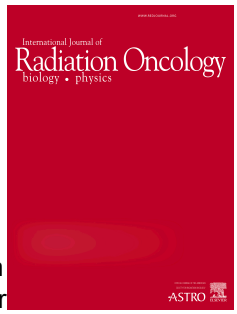
*Published in:*
International Journal of Radiation Oncology, Biology, Physics

*DOI:*
10.1016/j.ijrobp.2018.12.055



*Publication date:*
2019

# Accepted Manuscript




A model-based approach to predict short-term toxicity benefits with proton therapy for oropharyngeal cancer

Jean-Claude M. Rwigema, M.D., Johannes A. Langendijk, M.D. Ph.D., Hans Paul van der Laan, Ph.D., John N. Lukens, M.D., Samuel D. Swisher-McClure, M.D., Alexander Lin, M.D.

PII: S0360-3016(19)30002-1

DOI: https://doi.org/10.1016/j.ijrobp.2018.12.055

Reference: ROB 25487

To appear in: *International Journal of Radiation Oncology • Biology • Physics*

Received Date: 29 May 2018

Revised Date: 11 December 2018

Accepted Date: 20 December 2018

Please cite this article as: Rwigema J-CM, Langendijk JA, Paul van der Laan H, Lukens JN, Swisher-McClure SD, Lin A, A model-based approach to predict short-term toxicity benefits with proton therapy for oropharyngeal cancer, *International Journal of Radiation Oncology • Biology • Physics* (2019), doi: https://doi.org/10.1016/j.ijrobp.2018.12.055.

# A model-based approach to predict short-term toxicity benefits with proton therapy for oropharyngeal cancer

Jean-Claude M. Rwigema M.D.[1,2], Johannes A. Langendijk M.D. Ph.D.[3], Hans Paul van der Laan Ph.D.[3], John N. Lukens M.D.[1], Samuel D. Swisher-McClure M.D.[1], Alexander Lin M.D.[1]

*[1]Perelman School of Medicine, University of Pennsylvania, Department of Radiation Oncology, Philadelphia, PA; [2]Mayo Clinic, Department of Radiation Oncology, Phoenix, AZ; [3] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands*

**Short title:** Predictive outcomes model for proton therapy in oropharyngeal cancer.

**Corresponding Author:**

Alexander Lin M.D.

Department of Radiation Oncology

Perelman School of Medicine, University of Pennsylvania

3400 Civic Center Blvd

Philadelphia, PA 19104

Email: alexander.lin@uphs.upenn.edu.

**Conflicts of interest:** None

**Abstract**

**Purpose**: The aim of this study was to generate normal tissue complication probability (NTCP) models in patients treated with either proton beam therapy (PBT) or intensity-modulated radiotherapy (IMRT) for oropharynx cancer, and to use a model-based approach to investigate the added value of PBT in preventing treatment complications.

**Methods**: For patients with advanced-stage oropharynx cancer, treated with curative intent (PBT, n=30; IMRT, n=175), NTCP models were developed using multivariable logistic regression analysis with backward selection. For PBT-treated patients, an equivalent IMRT plan was generated, to serve as a reference to determine the benefit of PBT in terms of NTCP. The models were then applied to the PBT treated patients to compare predicted and observed clinical outcomes (calibration-in-the large). Five binary endpoints were analyzed at 6-months post-treatment: dysphagia ≥ grade 2, dysphagia ≥ grade 3, xerostomia ≥ grade 2, salivary duct inflammation ≥ grade 2, and feeding tube dependence. Corresponding toxicity grading was based on CTCAEv4. Paired t-tests and Wilcoxon rank tests were used to compare mean NTCP results for endpoints between PBT and IMRT.

**Results:** NTCP models developed based on outcomes from all patients were applied to those receiving PBT. NTCP-values were calculated for the equivalent IMRT plans for all PBT treated patients, revealing significantly higher NTCP-values with IMRT. PBT was associated with statistically significant reductions in the mean NTCP values for each endpoint at 6-months post treatment, with the largest absolute differences in rates of ≥ grade 2 dysphagia and ≥ grade 2 xerostomia.

**Conclusion**: NTCP models predict significant improvements in the probability of short-term, treatment-related toxicity with PBT compared to IMRT for oropharyngeal cancer. This study demonstrated an NTCP model-based approach to compare predicted patient outcomes when randomized data are not available.

**Key words:** Oropharyngeal cancer, NTCP, toxicity, IMRT, proton therapy, head and neck cancer

## INTRODUCTION

Currently most patients diagnosed with oropharyngeal head and neck carcinoma are cured after undergoing definitive multimodality therapy [1, 2]. Despite technological advances in head and neck radiotherapy, many patients experience long-term severe toxicities that negatively impact quality of life [3-7].

Data from single institution series have demonstrated advantages of proton beam therapy (PBT) over intensity-modulated radiotherapy (IMRT), due to PBT's favorable dose deposition beam profile that improves sparing of organs at risk and reduces integral dose to the patient [8-12]. As a result, randomized trials are ongoing to provide level I evidence regarding the clinical benefit of PBT [13]. Completing comparative randomized trials for new treatment technology remains challenging due to pre-existing patient preferences for selected treatments, high costs of conducting research, and potential ethical considerations related to clinician equipoise [14]. Moreover, in an era of personalized medicine with ever-increasing patient and tumor data heterogeneity, traditional level I evidence may not always adequately support individualized clinical decision making [15]. Data derived from statistical modeling of clinical outcomes for individual patients, can provide complementary data regarding the comparative effectiveness of treatment approaches in question. A model-based approach may be a cost-effective strategy to quantify clinical gains with PBT via estimation of potential reduction in normal tissue complication probability (NTCP) [16]. Such an approach may be optimal in informing patient eligibility for a chosen therapy to enhance clinical outcomes and cost efficiency [17].

To date, only one study evaluated NTCP models for PBT in a heterogeneous group of head and neck patients [18]. The aim of this study was to generate multivariable normal tissue complication probability (NTCP) models in patients treated with either PBT or IMRT for oropharynx cancer. We

hypothesize that improvements in dosimetric normal tissue sparing with PBT will translate to lower toxicity compared to treatment with IMRT.

## METHODS AND MATERIALS

### *Study population*

This (institution review board approved) study included patients with locally-advanced oropharyngeal carcinoma treated with curative intent multi-modality therapy from two institutions who had at least one year of follow-up. The cohort from the XXXXX consisted of 30 patients with oropharyngeal carcinoma treated with surgery followed by adjuvant proton radiotherapy, with or without chemotherapy (decision to offer chemotherapy was consistent with standard of care, such as the presence of positive margin and/or extranodal extension [19]) between 2013 and 2016. The cohort from the XXXXXX consisted of 175 patients mainly with locally-advanced oropharyngeal carcinoma treated with definitive photon radiotherapy with or without chemotherapy..

### *Treatment*

Patients in the postoperative cohort from XXXXX underwent radiotherapy planning at approximately 3-4 weeks after surgery. The process of computed tomography (CT) simulation acquisition, target delineation and treatment planning has been previously described [9]. For these patients, PBT plans (which were the ones clinically delivered to the patients) were generated for treatment delivery using pencil-beam scanning (PBS) via single field uniform dosing, plus an accompanying IMRT (VMAT) plan which was clinically reviewed and deemed acceptable for treatment (but not delivered, as they were reserved as a contingency plan only in case of unexpected proton beam unavailability) [9]. These accompanying IMRT plans needed to meet all of the coverage goals and organ sparing constraints similar to patients who receive the entirety of their radiotherapy via IMRT. For patients receiving organ-preservation RT at XXXXX, photon plans were generated and delivered using IMRT as previously described [20].

***Dosimetric data collection and extraction***

For each patient, relevant organs-at-risk (OARs) were contoured as previously described [21, 22]. Target delineation was consistent with standard of care in both postoperative and definitive RT patients, and no patients were enrolled or treated on protocols involving omission or reduction of standardly defined clinical targets. OARs were bilateral parotid glands, inferior, middle and superior pharyngeal constrictor muscles (PCM), supraglottic larynx, and oral cavity. All plans and structures were centrally reviewed and modified as needed to reflect uniformity and consistency across both institutions. The following dose volume histogram (DVH) parameters were collected for OARs: minimum dose, maximum dose, mean dose, V5Gy, V10Gy, V20Gy, V30Gy, V40Gy, V50Gy, V60Gy and V70Gy (percent volume receiving 5Gy, 10 Gy, 20 Gy, 30 Gy, 40 Gy, 50 Gy, 60 Gy and 70 Gy). DVH parameters were extracted by MIRADA-software (Oxford Centre for Innovation UK) from both the XXXX PBT and IMRT plans and then combined with XXXX IMRT plans for analysis.

***Follow-up***

After completion of therapy, patients were followed with clinical examinations and head and neck imaging, initially with a 3-month post-treatment PET-CT, then PET-CT or CT every 3-6 months for the first 2 years, and then every 6-12 months thereafter. Toxicity data was collected before the start of radiotherapy and at every follow-up visit and graded using CTCAE version 4.0.

***Endpoints***

The following toxicity endpoints were defined at 6 months from treatment completion: (1) Dysphagia ≥ grade 2; (2) dysphagia ≥ grade 3; (3) xerostomia ≥ grade 2; (4) salivary duct inflammation ≥ grade 2; and (5) tube feeding dependence. Salivary duct inflammation toxicity was graded as such: Grade

1 = slightly thickened saliva; slightly altered taste; Grade 2 = thick, ropy, sticky saliva; markedly altered taste; alteration in diet indicated; secretion-induced symptoms; limiting instrumental activities of daily living (ADL); Grade 3 = acute salivary gland necrosis; sever secretion-induced symptoms; tube feeding indicated; limiting self-care ADL; disabling; Grade 4 = life-threatening consequences; urgent intervention indicated; Grade 5 = death.

The 6-month endpoint was chosen, given that this was the time point for which the largest amount of toxicity data existed for all patients. All toxicity endpoints were collected and documented prospectively.

***Statistics***

Patients who had one of the endpoints already at baseline, were excluded from the analyses regarding that particular endpoint. For each endpoint, multivariable NTCP models were created. Univariable logistic regression analyses and correlation statistics were performed to select candidate predictors for each endpoint that were significantly associated with the endpoints in univariable analysis ($p < 0.05$), but not mutually correlated ($r < 0.80$). Then, a stepwise backward multivariable logistic regression procedure was used to exclude the variables with $p > 0.157$ from the model. The resulting model was then manually explored further in two ways: 1) by testing whether the models would significantly deteriorate when one or more variables would be removed; 2) by exchanging the selected dose volume variables by other potentially relevant dose variables that were highly correlated to the selected dose variable and therefore discarded in an earlier stage. The final best model was chosen primarily by the applying the likelihood-ratio test, but also by evaluating the general model performance measures, i.e., ROC-area under the curve, discrimination slope, explained variance and calibration. For each endpoint, the final model was subjected to internal validation with a bootstrapping procedure to

correct (shrink) the models (slope and intercept) for optimism. This was done to obtain realistic regression coefficients for the model variables that are representative for populations like the development sample. A figure summarizing the steps in model generation is shown in **Figure 1**.

Candidate variables that were initially entered in the model were: gender (male versus female), age (as continuous variable), concomitant chemotherapy (no vs. yes), weight loss at baseline (0-10 vs. >10%), accelerated radiotherapy (no vs. yes), T-stage (stage 1-2 vs. 3-4), N-stage (negative vs. positive), target volume (local/unilateral vs. bilateral neck irradiation), surgery (no vs. yes), and baseline toxicities (grade 0 vs. grade 1). Paired t-tests and Wilcoxon rank tests were used to compare mean NTCP results for endpoints between PBT and IMRT. Data were analyzed using SPSS Statistics for Windows, version 23.

***NTCP calculation***

NTCP values were determined for each patient at all endpoints in both PBT and photon plans using the NTCP formulae [23]: $\mathrm{NTCP} = \frac{1}{(1+e^{-S})}$ with the linear predictor (S) for complications defined as: $S = \beta_0 + \sum_i \beta_i \cdot x_i$

where $\beta_0$ (intercept) and $\beta_i$ (variable coefficients)were the model parameters and $x_i$ the predictor variables.

**RESULTS**

Patient and treatment characteristics, observed rates of toxicities at 6-months with corresponding OAR mean doses for relevant endpoints of the 2 patient cohorts are shown in **Table 1**. The difference in increased sparing of an organ at risk, such as the oral cavity, is shown in **Figure 2**. A summary of all model performance results for all endpoints is shown in **Table 2,** which presents the uncorrected (apparent) modeling results, with uncorrected regression coefficients.

The final NTCP models for the endpoints below were developed based on outcomes of each endpoint from all patients, and include the corrected coefficients (after internal validation).

(1) **Dysphagia ≥ grade 2 at 6 months:** S= -4.3477 + (0.0345 * contralateral parotid mean dose (Gy)) + (0.0524 * oral cavity mean dose (Gy)).

(2) **Dysphagia ≥ grade 3 at 6 months:** S= -4.3188 + (1.3744 * T3 or 4) + (1.0222 * baseline weight loss >10%) + (0.0385 * oral cavity mean dose (Gy)).

(3) **Xerostomia ≥ grade 2 at 6 months:** S= -3.6891 + (0.8639 * baseline xerostomia grade 1) + (0.6423* concomitant chemotherapy) + (0.0520 * oral cavity mean dose (Gy)).

(4) **Salivary Duct Inflammation ≥ grade 2 at 6 months:** S= -6.3436 + (0.0389 * Age (years)) + (1.0231* accelerated radiotherapy) + (0.0367 * oral cavity mean dose (Gy)).

(5) **Tube feeding dependence at 6 months:** S= -10.3690 + (1.3848 * T3 or 4) + (1.3805* baseline weight loss >10%) + (0.0364 * PCM inferior mean dose (Gy)) + (0.0939* PCM superior mean dose (Gy)).

The NTCP-values were calculated for the equivalent IMRT plans for all PBT treated patients, revealing significantly higher NTCP-values for the IMRT plans for all endpoints **(Table 3).** PBT was associated with statistically significant reductions in the paired mean NTCP values for each endpoint at 6 months

post treatment, with the largest absolute differences in rates of ≥ grade 2 dysphagia and xerostomia (**Table 3)**. The absolute reductions in individual patient NTCP by PBT as compared to IMRT ranged from 2 to 14% for grade 2 dysphagia, 1 to 8% for grade 3 dysphagia, 2 to 17% for grade 2 xerostomia, 1 to 8% for salivary duct inflammation and 1 to 7% for tube dependence (**Figure 2A-E**).

## DISCUSSION

Although IMRT has led to reduction of radiation induced side effects with improved global quality of life from 3D-conformal techniques, efforts are still needed to further enhance the therapeutic ratio in oropharyngeal carcinoma after multimodality curative therapy [24-27]. It is for this reason that proton therapy, with its ability to improve normal tissue sparing when compared to IMRT, may help to improve patient toxicity outcomes and long-term quality of life. However, precise estimates of the clinical impact of PBT are lacking with the current absence of randomized data. The present study evaluates toxicity outcomes between IMRT to PBT using normal tissue complication probability models in order to estimate potential clinical benefits of PBT using a large cohort of patients receiving radiotherapy for oropharyngeal carcinoma.

Our study extends the existing literature regarding the comparative effectiveness of PBT for head and neck radiotherapy and is the first report of such a comparative analysis limited to patients with oropharynx cancer, in whom high rates of long-term survival emphasize a focus on toxicity mitigation to preserve quality of life [1, 28]. Treatment-related late complications that commonly affect quality of life in these patients are mainly dysphagia, gastrostomy-tube dependence, and xerostomia [29, 30]. In this study, NTCP models were developed using patient cohorts from 2 institutions treated with IMRT and PBT, respectively. The NTCP models were then applied to all patients receiving PBT, for whom each had a treatment-approved 'backup' IMRT plan. Thus, each PBT patient served as an internal control when comparing estimated toxicity from PBT vs IMRT, which we believe to be a unique strength of the study. Four toxicity domains (ie. dysphagia, xerostomia, salivary duct inflammation, and G-tube dependence) were evaluated and modeled at 6 months from completion of PBT. Results herein demonstrated significant reduction of predicted complications in all evaluated head and neck treatment-related toxicities, with the greatest differences observed favoring PBT for grade $\geq$ 2 dysphagia and xerostomia.

With the introduction of new technologies in radiation delivery, coupled with its potential significant costs, it is important to assess and confirm that new technologies for radiation delivery will lead to meaningful gains for patients. The gold standard for such an effort remains a prospective, randomized trial; however, barriers to successful implementation of such a trial exist, and will likely remain for current and future efforts. Our study represents a novel approach that can be used currently to assess potential benefits while we await the results of prospective trials.

This study has some limitations to warrant mention. First, even though our data suggests that proton therapy may be a method by which treatment-related toxicity can be improved, it does not specifically address the issue of cost effectiveness. The issue of cost effectiveness and justification of new technologies is a much more complex issue, which is outside the limits of this study, and will have to be addressed by future, collective efforts. Second, the PBT cohort was limited to only 30 patients, and the cohorts from each institution were dissimilar in that one institution largely treated patients with an initial surgical approach followed by adjuvant radiotherapy (+/- chemotherapy), while the other institution largely treated patients with multimodality organ preservation. We acknowledge that the different approaches may itself affect patient toxicity outcomes. However, our model was generated using data and outcomes from the entire cohort from both institutions, incorporating patients receiving a range of accepted treatment approaches, which may allow for this model to be generalizable for allowable treatment approaches. Finally, our model overpredicted the rate of xerostomia compared to observed prevalence for patients receiving proton therapy. While we would prefer, given a choice, that such models overpredict rather than underpredict toxicity for proton therapy, it is clear that clinical validation of this model in a larger group of patients, receiving a range of accepted treatment approaches, is needed. This is already in progress, as patients at one of the participating institutions in this study is selecting and treating patients with proton therapy for oropharynx cancer based on these models. The results and clinical validation from current patient treatments will be a natural follow-up to this initial effort, and will be reported in the near future.

In summary, this study demonstrates the potential value of NTCP model based approaches in comparing predicted patient outcomes. Such a tool may be highly useful when randomized data are not available, or when deciding on which patients may be most likely to benefit from the use of a limited resource. Results of the current study may serve as a guide to patient selection, and provide complementary data regarding estimated clinical effectiveness of PTV when results from properly conducted phase III trials are not available [13, 14]. A model-based approach can also be incorporated into the context of a prospective, randomized trial, either as a potential outcome biomarker, or as criteria to best select patients for trial enrollment. In the end, we as radiation oncologists believe that our mission is to improve the lives of our patients, and to apply advances in our field in a feasible and judicious manner. Our manuscript is a reflection of that belief.

## REFERENCES


1. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, Westra WH, Chung CH, Jordan RC, Lu C, Kim H, Axelrod R, Silverman CC, Redmond KP, Gillison ML. Human papillomavirus and survival of patients with oropharyngeal cancer. *The New England journal of medicine.* 2010;363:24-35.
2. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruse S, Anderson WF, Rosenberg PS, Gillison ML. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2011;29:4294-301.
3. Ling DC, Kabolizadeh P, Heron DE, Ohr JP, Wang H, Johnson J, Kubicek GJ. Incidence of hospitalization in patients with head and neck cancer treated with intensity-modulated radiation therapy. *Head & neck.* 2015;37:1750-5.
4. 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2008;26:3770-6.
5. Hunter KU, Lee OE, Lyden TH, Haxer MJ, Feng FY, Schipper M, Worden F, Prince ME, McLean SA, Wolf GT, Bradford CR, Chepeha DB, Eisbruch A. Aspiration pneumonia after chemo-intensity-modulated radiation therapy of oropharyngeal carcinoma and its clinical and dysphagia-related predictors. *Head & neck.* 2014;36:120-5.
6. Tsai CJ, Hofstede TM, Sturgis EM, Garden AS, Lindberg ME, Wei Q, Tucker SL, Dong L. Osteoradionecrosis and radiation dose to the mandible in patients with oropharyngeal cancer. *International journal of radiation oncology, biology, physics.* 2013;85:415-20.
7. Ringash J. Survivorship and Quality of Life in Head and Neck Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015;33:3322-7.
8. van der Laan HP, van de Water TA, van Herpt HE, Christianen ME, Bijl HP, Korevaar EW, Rasch CR, van 't Veld AA, van der Schaaf A, Schilstra C, Langendijk JA. The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. *Acta oncologica (Stockholm, Sweden).* 2013;52:561-9.
9. Apinorasethkul O, Kirk M, Teo K, Swisher-McClure S, Lukens JN, Lin A. Pencil beam scanning proton therapy vs rotational arc radiation therapy: A treatment planning comparison for postoperative oropharyngeal cancer. *Medical dosimetry : official journal of the American Association of Medical Dosimetrists.* 2017;42:7-11.
10. Gunn GB, Frank SJ. Advances in radiation oncology for the management of oropharyngeal tumors. *Otolaryngologic clinics of North America.* 2013;46:629-43.
11. Gregoire V, Langendijk JA, Nuyts S. Advances in Radiotherapy for Head and Neck Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015;33:3277-84.
12. Sio TT, Lin HK, Shi Q, Gunn GB, Cleeland CS, Lee JJ, Hernandez M, Blanchard P, Thaker NG, Phan J, Rosenthal DI, Garden AS, Morrison WH, Fuller CD, Mendoza TR, Mohan R, Wang XS, Frank SJ. Intensity Modulated Proton Therapy Versus Intensity Modulated Photon Radiation Therapy for Oropharyngeal Cancer: First Comparative Results of Patient-Reported Outcomes. *International journal of radiation oncology, biology, physics.* 2016;95:1107-14.

13. M.D. Anderson Cancer Center. 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. Retrieved from https://clinicaltrials.gov/ct2/show/NCT01893307 (ClinicalTrials.gov Identifier: NCT01893307).
14. Cox JD. Impediments to Comparative Clinical Trials With Proton Therapy. *International journal of radiation oncology, biology, physics.* 2016;95:4-8.
15. Lambin P, Zindler J, Vanneste BG, De Voorde LV, Eekers D, Compter I, Panth KM, Peerlings J, Larue RT, Deist TM, Jochems A, Lustberg T, van Soest J, de Jong EE, Even AJ, Reymen B, Rekers N, van Gisbergen M, Roelofs E, Carvalho S, Leijenaar RT, Zegers CM, Jacobs M, van Timmeren J, Brouwers P, Lal JA, Dubois L, Yaromina A, Van Limbergen EJ, Berbee M, van Elmpt W, Oberije C, Ramaekers B, Dekker A, Boersma LJ, Hoebers F, Smits KM, Berlanga AJ, Walsh S. Decision support systems for personalized and participative radiation oncology. *Advanced drug delivery reviews.* 2017;109:131-53.
16. 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. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2013;107:267-73.
17. Lievens Y, Pijls-Johannesma M. Health economic controversy and cost-effectiveness of proton therapy. *Seminars in radiation oncology.* 2013;23:134-41.
18. 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. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2016;121:381-6.
19. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, Ozsahin EM, Jacobs JR, Jassem J, Ang KK, Lefebvre JL. Defining risk levels in locally advanced head and neck cancers: a comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (# 9501). *Head & neck.* 2005;27:843-50.
20. van der Laan HP, Christianen ME, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2012;103:76-81.
21. Christianen ME, Langendijk JA, Westerlaan HE, van de Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2011;101:394-402.
22. van de Water TA, Bijl HP, Westerlaan HE, Langendijk JA. Delineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2009;93:545-52.
23. Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, Chouvalova O, Leemans CR, de Bock GH, Christianen ME, van der Laan BF, Vissink A, Steenbakkers RJ, Langendijk JA. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2012;105:86-93.
24. Eisbruch A, Schwartz M, Rasch C, Vineberg K, Damen E, Van As CJ, Marsh R, Pameijer FA, Balm AJ. Dysphagia and aspiration after chemoradiotherapy for head-and-neck cancer: which

anatomic structures are affected and can they be spared by IMRT? *International journal of radiation oncology, biology, physics.* 2004;60:1425-39.

25. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *International journal of radiation oncology, biology, physics.* 2001;50:695-704.
26. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, Miles EA, Miah AB, Newbold K, Tanay M, Adab F, Jefferies SJ, Scrase C, Yap BK, A'Hern RP, Sydenham MA, Emson M, Hall E. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *The Lancet. Oncology.* 2011;12:127-36.
27. Feng FY, Kim HM, Lyden TH, Haxer MJ, Worden FP, Feng M, Moyer JS, Prince ME, Carey TE, Wolf GT, Bradford CR, Chepeha DB, Eisbruch A. Intensity-modulated chemoradiotherapy aiming to reduce dysphagia in patients with oropharyngeal cancer: clinical and functional results. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2010;28:2732-8.
28. Hoxbroe Michaelsen S, Gronhoj C, Hoxbroe Michaelsen J, Friborg J, von Buchwald C. Quality of life in survivors of oropharyngeal cancer: A systematic review and meta-analysis of 1366 patients. *European journal of cancer (Oxford, England : 1990).* 2017;78:91-102.
29. Hunter KU, Schipper M, Feng FY, Lyden T, Haxer M, Murdoch-Kinch CA, Cornwall B, Lee CS, Chepeha DB, Eisbruch A. Toxicities affecting quality of life after chemo-IMRT of oropharyngeal cancer: prospective study of patient-reported, observer-rated, and objective outcomes. *International journal of radiation oncology, biology, physics.* 2013;85:935-40.
30. Terrell JE, Ronis DL, Fowler KE, Bradford CR, Chepeha DB, Prince ME, Teknos TN, Wolf GT, Duffy SA. Clinical predictors of quality of life in patients with head and neck cancer. *Archives of otolaryngology--head & neck surgery.* 2004;130:401-8.

**Figure legends**

Figure 1: Variable selection and logistic regression modeling for each endpoint

Figure 2: IMRT vs. PBT Comparison: axial (left) and sagittal (right) slices of representative radiation plans for adjuvant radiation therapy in a patient with T1N2aM0 stage IVA ($7^{th}$ edition) base of tongue carcinoma, showing IMRT and PBT radiation plans (60Gy in 30 fractions) for the same patient. The PBT plan demonstrates lower dose to oral cavity structures compared to IMRT.

Figure 3. Waterfall plots showing illustrating individual reduction in NTCP for (A) Dysphagia grade ≥2, (B) Dysphagia grade ≥3, (C) Xerostomia ≥grade 2, (D) Salivary duct inflammation grade ≥2, and (E) Tube dependence grade ≥2.

**Table 1**

| | Proton cohort (n=30) | | Photon cohort (n=175) | | | |
|---|---|---|---|---|---|---|
| | **Number** | **%** | **Number** | **%** | ***SMD** | **p value** |
| Neck RT Bilateral | 29 | 96.70% | 155 | 88.60% | -0.31 | 0.05 |
| Male | 26 | 86.70% | 112 | 64.00% | -0.55 | 0.003 |
| Robotic Surgery Primary Site | 29 | 96.70% | 0 | 0.00% | -7.66 | < 0.001 |
| Extensive Surgery Primary Site* | 1 | 3.30% | 1 | 0.60% | -0.20 | 0.42 |
| Surgery neck* | 30 | 100.00% | 3 | 1.70% | -10.75 | < 0.001 |
| Concomitant chemotherapy | 7 | 23.30% | 101 | 57.70% | 0.75 | < 0.001 |
| T3 or T4 | 5 | 16.70% | 90 | 51.40% | 0.79 | < 0.001 |
| Node positive | 29 | 96.70% | 135 | 77.60% | -0.60 | < 0.001 |
| Pre-Treatment Weight loss >10% | 2 | 6.70% | 14 | 8.40% | 0.06 | 0.75 |
| Accelerated (6 fraction per week) RT | 3 | 10.00% | 54 | 30.90% | 0.54 | 0.002 |
| Dysphagia CTCAEv4 ≥G2 at Baseline | 0 | 0.00% | 46 | 26.30% | 0.84 | < 0.001 |
| Dysphagia CTCAEv4 ≥G2 at 6 Months | 2 | 6.70% | 84 | 48.00% | 1.04 | < 0.001 |
| Dysphagia CTCAEv4 ≥G3 at 6 Months | 1 | 3.30% | 47 | 26.90% | 0.70 | < 0.001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Xerostomia CTCAEv4 ≥G1 at Baseline | 4 | 13.30% | 24 | 13.70% | 0.01 | 0.96 |
| Xerostomia CTCAEv4 ≥G2 at 6 Months | 0 | 0.00% | 80 | 46.20% | 1.31 | < 0.001 |
| Salivary Duct Inflammation CTCAEv4 ≥G1 at Baseline | 1 | 3.30% | 23 | 13.10% | 0.36 | 0.02 |
| Salivary Duct Inflammation CTCAEv4 ≥G2 at 6 Months | 1 | 3.30% | 31 | 17.90% | 0.49 | 0.001 |
| Tube feeding dependence at 6 months | 0 | 0.00% | 36 | 20.60% | 0.72 | < 0.001 |
| | | | | | | |
| **AVERAGE VALUES** | | | | | | |
| Age (years) | 58.2 ± 11.8 | | 60.1 ± 8.7 | | 0.18 | 0.40 |
| High Risk PTV Prescribed dose (Gy) | 62.2 ± 2.7 | | 69.8 ± 2.2 | | 3.09 | < 0.001 |
| Parotid ipsilateral mean dose (Gy) | 32.4 ± 7.0 | | 41.2 ± 11.9 | | 0.90 | < 0.001 |
| Parotid contralateral mean dose (Gy) | 13.6 ± 5.7 | | 29.1 ± 10.6 | | 1.82 | < 0.001 |
| PCM inferior mean dose (Gy) | 29.1 ± 6.6 | | 47.0 ± 13.0 | | 1.74 | < 0.001 |
| PCM superior mean dose (Gy) | 43.0 ± 6.8 | | 62.7 ± 7.1 | | 2.83 | < 0.001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Oral cavity mean dose (Gy) | 22.3 ± 9.5 | | 56.3 ± 7.6 | | 3.95 | < 0.001 |

RT=radiotherapy; G2/3=grade 2 or 3 common toxicity Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0; PTV=planning tumor volume; PCM=pharyngeal constrictor muscles; SMD = standardized mean difference

* In the photon cohort, 1 patient received open (non-robotic ) surgery to the primary site (without neck surgery), 3 patients received neck surgery without surgery to the primary tumor.

**TABLE 2. Model Summary Results**

| **Measures (at 6 months)** | | **≥G2 Dysphagia** | **≥G3 Dysphagia** | **≥G2 Xerostomia** | **≥G2 Salivary Duct Inflammation** | **Tube Dependence** |
|---|---|---|---|---|---|---|
| Events | Controls | 107 | 147 | 118 | 163 | 162 |
| | Events* | 52 (33%) | 37 (20%) | 76 (39%) | 31 (16%) | 34 (17%) |
| | | | | | | |
| Overall | Nagelkerke adjusted $R^2$ | 0.206 | 0.248 | 0.204 | 0.140 | 0.403 |
| | | | | | | |
| Discrimination | ROC-curve AUC (95% CI) | 0.750 (0.669-0.830) | 0.783 (0.701-0.864) | 0.713 (0.642-0.785) | 0.710 (0.606-0.808) | 0.864 (0.800-0.928) |
| | Discrimination slope | 0.152 | 0.179 | 0.142 | 0.089 | 0.298 |
| | | | | | | |
| Calibration | Hosmer-Lemeshow test | $X^2$ = 4.499 (p = 0.810) | $X^2$ = 10.769 (p = 0.216) | $X^2$ = 6.718 (p = 0.577) | $X^2$ = 4.483 (p = 0.810) | $X^2$ = 5.845 (p = 0.701) |
| | | | | | | |
| Validation / bootstrap | Model slope correction | 0.966 | 0.912 | 0.937 | 0.914 | 0.905 |
| | | | | | | |
| Model coefficients (uncorrected) | Intercept | -4.482 | -4.595 | -3.935 | -6.813 | -11.478 |
| | Oral cavity MD (Gy) | 0.054 | 0.042 | 0.056 | 0.040 | X |
| | Parotid cont. MD (Gy) | 0.036 | x | x | x | X |
| | PCM sup MD (Gy) | x | x | x | x | 0.105 |
| | PCM inf MD (Gy) | x | x | x | x | 0.041 |
| | T3 or T4 | x | 1.492 | x | x | 1.549 |
| | Weight loss BL >10% | x | 1.110 | x | x | 1.544 |
| | Dry mouth BL Gr 1 | x | x | 0.927 | x | x |
| | Concomitant chemo | x | x | 0.689 | x | x |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Age (years) | x | x | x | 0.043 | x |
| | Accelerated RT | x | x | x | 1.122 | x |

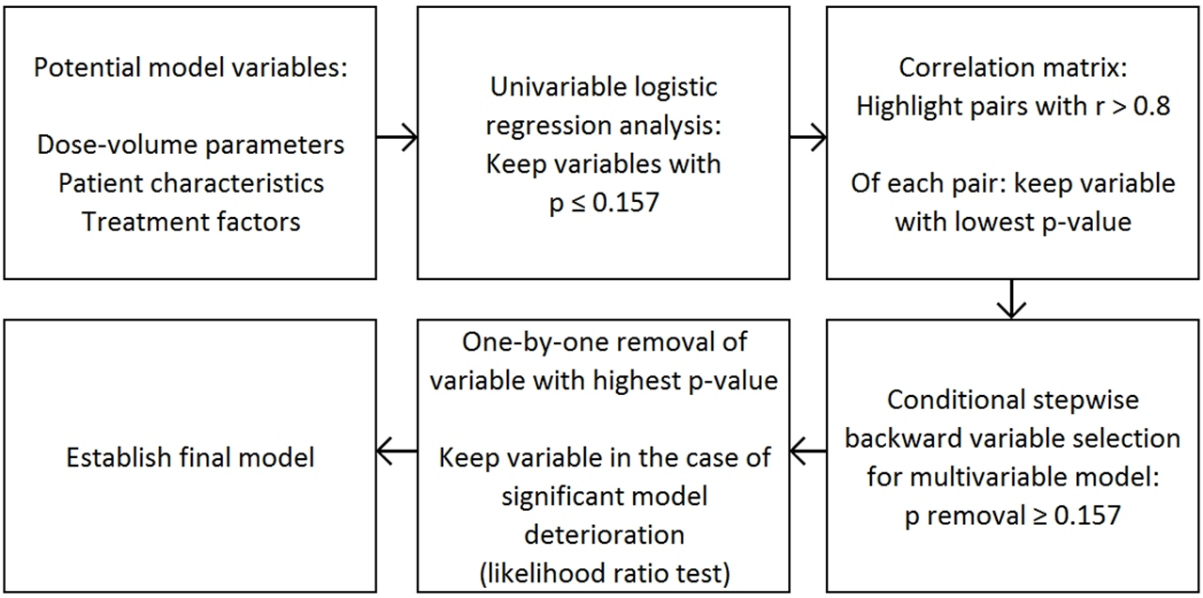
ROC=receiver operating curve; AUC=are under curve; $R^2$=linear regression coefficient squared. MD=mean dose; PCM=pharyngeal constrictor muscles; sup=superior; inf=inferior; T=T stage; BL=at baseline; RT=radiotherapy; CI=confidence interval.

*Events are the patients in the whole dataset that had the endpoint at 6 months (i.e., the prevalence in the whole group). Both controls and those with events were part of the NTCP modeling, not just those with events.

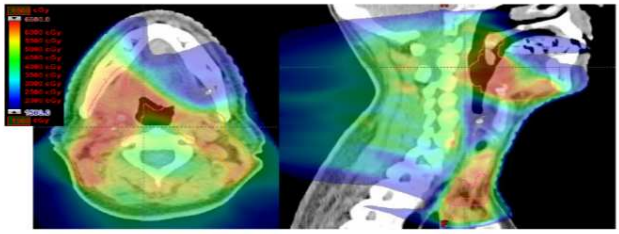
**TABLE 3**

| Endpoint | Observed Prevalence (%) | NTCP (Protons) | | NTCP (IMRT) | | Difference in mean/ gain (%) | 95% Confidence Interval of the Difference in mean (gain) (%) | | t-test | Wilcoxon |
|---|---|---|---|---|---|---|---|---|---|---|
| | | (mean % ± SD) | (median %, range) | (mean % ± SD) | (median %, range) | | | | Sig. (2-tailed) | Sig. |
| Dysphagia grade ≥2 | 6.7 | 6.7 ± 3.6 | 5.6 (2.2 - 17.7) | 14.9 ± 5.8 | 14.2 (5.2 - 31.2) | -8.318 | -9.431 | -7.205 | < 0.001 | < 0.001 |
| Dysphagia grade ≥3 | 3.3 | 4.9 ± 4.4 | 3 (1.5 - 16.3) | 7.6 ± 5.7 | 5.3 (3.4 - 22) | -2.694 | -3.250 | -2.137 | < 0.001 | < 0.001 |
| Xerostomia grade ≥2 | 0 | 10.3 ± 7.1 | 8.4 (3 - 39.7) | 18.6 ± 9.1 | 17.3 (8.4 - 50.4) | -8.315 | -9.613 | -7.016 | < 0.001 | < 0.001 |
| Salivary duct inflammation grade ≥2 | 3.3 | 4.7 ± 3.3 | 3.8 (1.3 - 15.4) | 7.6 ± 4.7 | 6.1 (2.6 - 23.2) | -2.857 | -3.534 | -2.180 | < 0.001 | < 0.001 |
| Tube dependence | 0 | 1.3 ± 1.7 | 0.6 (0.1 - 6.2) | 1.7 ± 2.5 | 0.7 (0.1 - 11.1) | -0.419 | -0.890 | 0.052 | 0.079 | 0.005 |

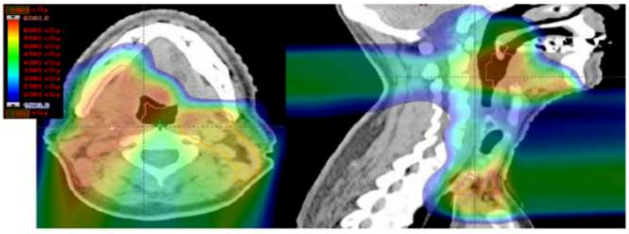
NTCP= reduction normal tissue complication probability; SD= standard deviation;

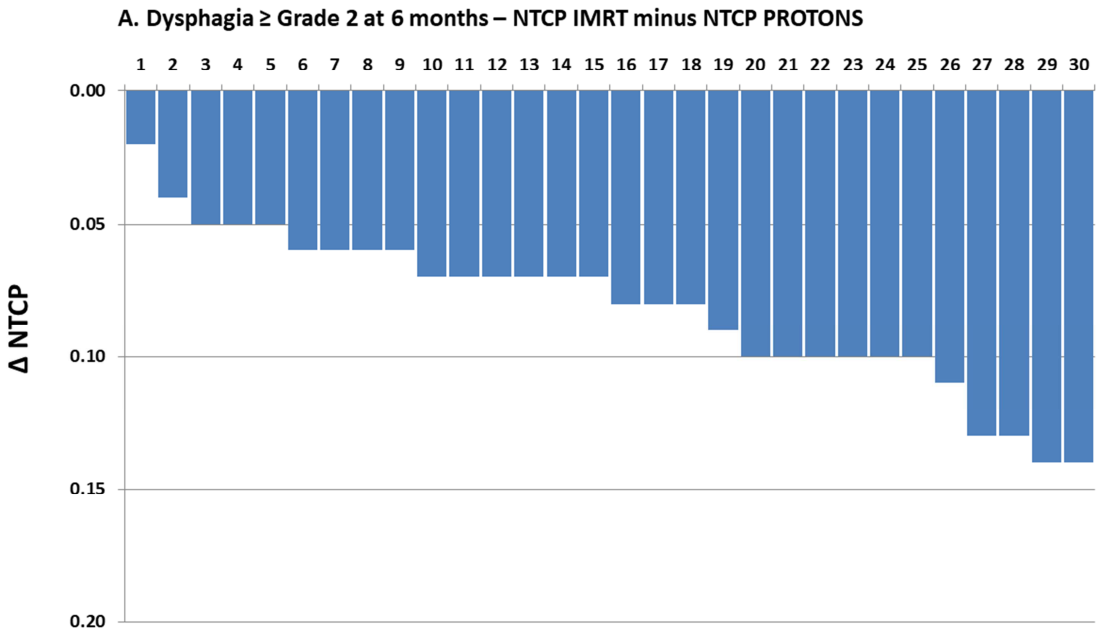


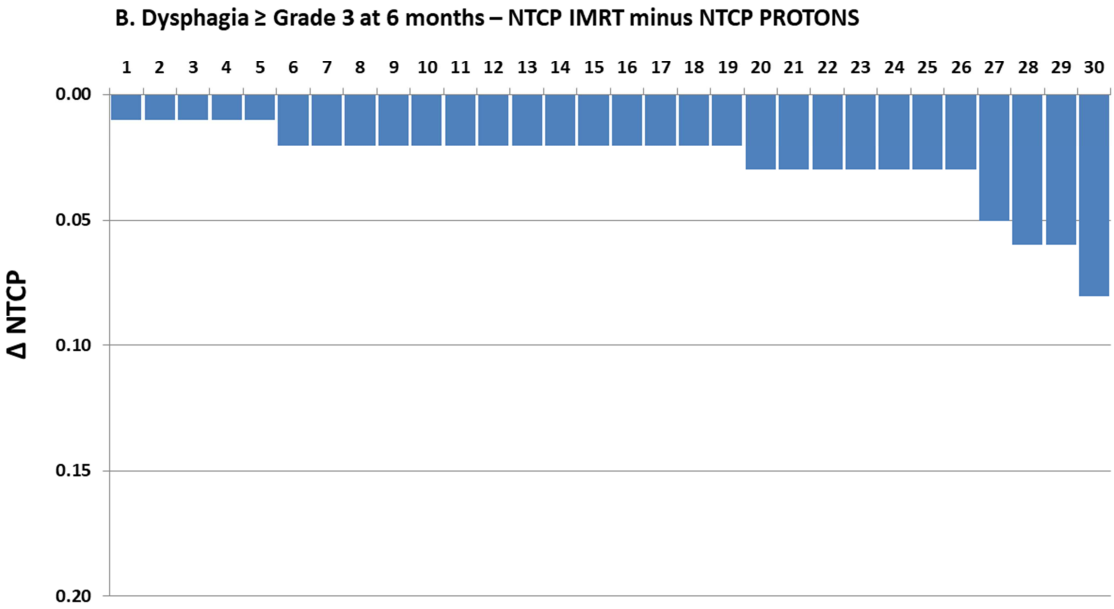
IMRT

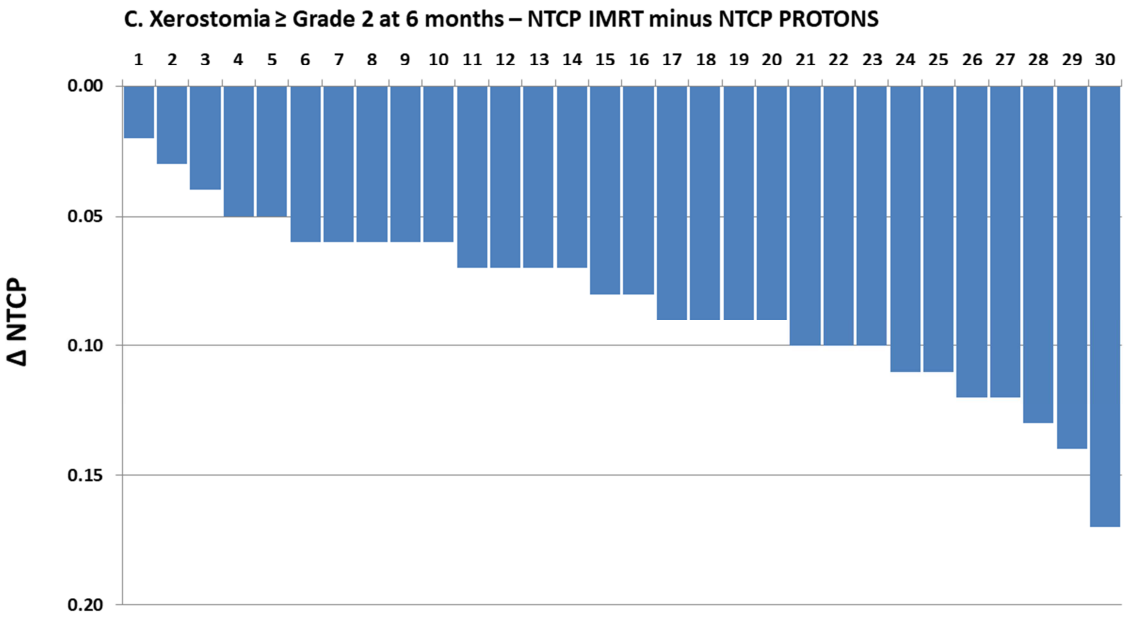


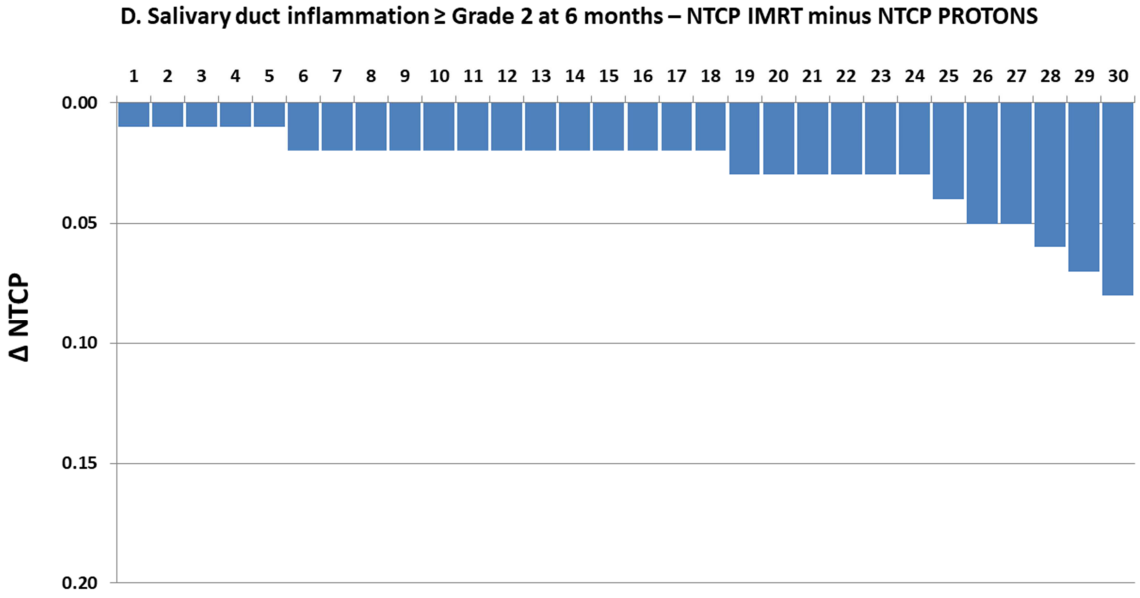
PBT

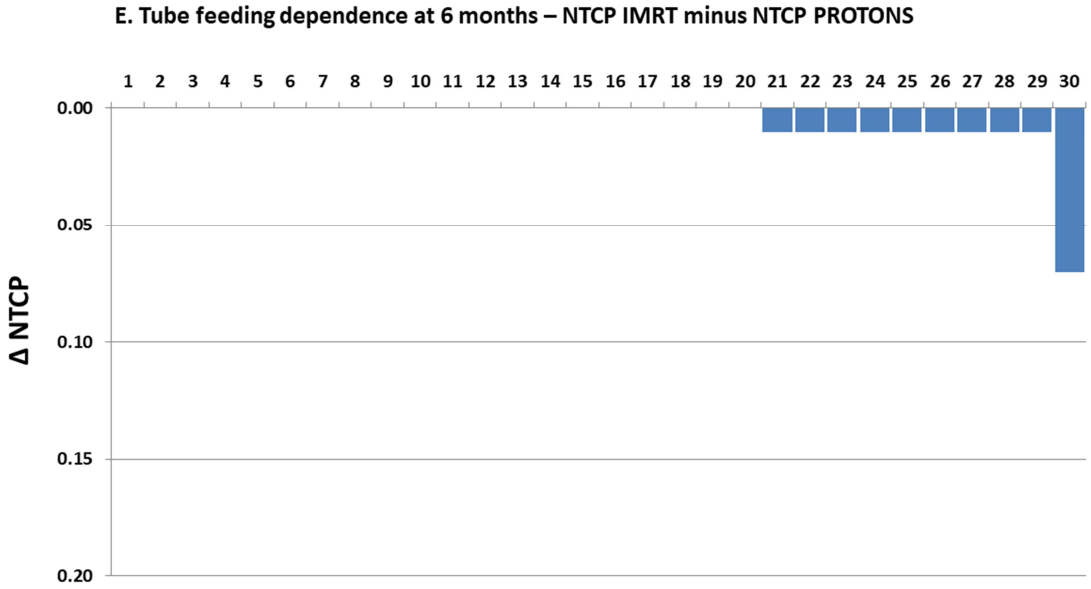












Approaches to predict upfront the potential clinical gains of a new technology or approach in radiation delivery are needed in a rapidly advancing field. This study reports on an outcomes-based predictive model of anticipated gains (xerostomia and dysphagia) for proton therapy in the treatment of oropharynx cancer. These results and this approach can be used to complement prospective trials, or to rationalize novel treatment approaches when randomized data are not yet available.