



Published in final edited form as:

Radiother Oncol. 2017 February ; 122(2): 200–206. doi:10.1016/j.radonc.2016.11.005.

Internal and external generalizability of temporal dose-response relationships for xerostomia following IMRT for head and neck cancer

Maria Thor¹, Adepitan A Owosho², Haley D Clark^{3,4}, Jung Hun Oh¹, Nadeem Riaz⁵, Allan Hovan⁶, Jillian Tsai⁵, Steven D Thomas⁴, Sae Hee K Yom², Jonn S Wu⁶, Joseph M Huryn², Vitali Moiseenko⁷, Nancy Y Lee⁵, Cherry L Estilo², and Joseph O Deasy¹

¹Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, US

²Dept. of Surgery, Memorial Sloan Kettering Cancer Center, NY, US

³Dept. of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, CA

⁴Dept. of Medical Physics, British Columbia Cancer Agency-Vancouver Centre, Vancouver, British Columbia, CA

⁵Dept. of Radiation Oncology, Memorial Sloan Kettering Cancer Center, NY, US

⁶Radiation Oncology, British Columbia Cancer Agency-Vancouver Centre, Vancouver, British Columbia, CA

⁷Dept. of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, US

Abstract

Background and Purpose—To study internal and external generalizability of temporal dose-response relationships for xerostomia after intensity-modulated radiotherapy (IMRT) for head and neck cancer, and to investigate potential amendments of the QUANTEC guidelines.

Material and Methods—Objective xerostomia was assessed in 121 patients ($n_{Cohort1}=55$; $n_{Cohort2}=66$) treated to 70Gy@2Gy in 2006–2015. Univariate and multivariate analyses (UVA, MVA with 1000 bootstrap populations) were conducted in *Cohort1*, and generalizability of the best-performing MVA model was investigated in *Cohort2* (performance: AUC, p-values, and Hosmer-Lemeshow p-values (p_{HL})). Ultimately and for clinical guidance, minimum mean dose thresholds to the contralateral and the ipsilateral parotid glands ($D_{mean_{contra}}$, $D_{mean_{ipsi}}$) were estimated from the generated dose-response curves.

Corresponding author: Maria Thor, Ph. D., Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, New York, NY 10017, Tel: (+001) 646 888 8013, thorm@mskcc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest

The authors have no conflict of interest to declare.

Results—The observed xerostomia rate was 38%/47% (3 months) and 19%/23% (11–12 months) in *Cohort1/Cohort2*. Risk of xerostomia at 3 months increased for higher $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ (*Cohort1*: $0.17 \cdot D_{\text{mean}_{\text{contra}}} + 0.11 \cdot D_{\text{mean}_{\text{ipsi}}} - 8.13$; $AUC = 0.90 \pm 0.05$; $p = 0.0002 \pm 0.002$; $p_{\text{HL}} = 0.22 \pm 0.23$; *Cohort2*: $AUC = 0.81$; $p < 0.0001$; $p_{\text{HL}} = 0.27$). The identified minimum $D_{\text{mean}_{\text{contra}}}$ thresholds were lower than in the QUANTEC guidelines (*Cohort1/Cohort2*: $D_{\text{mean}_{\text{contra}}} = 12/19$ Gy; $D_{\text{mean}_{\text{contra}}}$, $D_{\text{mean}_{\text{ipsi}}} = 16, 25/20, 26$ Gy).

Conclusions—Increased $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ explain short-term xerostomia following IMRT. Our results also suggest decreasing $D_{\text{mean}_{\text{contra}}}$ to below 20 Gy, while keeping $D_{\text{mean}_{\text{ipsi}}}$ to around 25 Gy. Long-term xerostomia was less frequent, and no dose-response relationship was established for this follow-up time.

Keywords

Radiotherapy; Xerostomia; IMRT; Head and neck; dose response; QUANTEC

Introduction

Radiotherapy (RT) is the standard of care for the majority of individuals diagnosed with head and neck cancer (HNC) either within a primary setting with/without chemotherapy, or in an adjuvant setting following surgery [1]. Given an estimated five-year relative survival for localized HNC of around 80% [2], minimizing RT-induced oral complications is essential. Severe hyposalivation (xerostomia) results, in particular, from loss of stimulated saliva and is the most commonly reported RT-induced oral complication, and leads to dental caries, oral infections, pain, reduced mastication and swallowing ability, and speech difficulties [3]. Intensity-modulated RT (IMRT) as a primary treatment for HNC has proven superior over preceding two- or three-dimensional conformal RT techniques in terms of significantly reducing the number of patients suffering from moderate to severe xerostomia up to two years after completed treatment in two randomized controlled trials [4, 5]. Even after IMRT, however, patients may experience xerostomia to a degree that still compromises their quality of life [6].

Stimulated saliva is primarily derived from the parotid glands, and these are, thus, considered the key organs for salivary function [7]. In the salivary gland-specific Quantitative Evaluation of Normal Tissue Effects in the Clinic (QUANTEC) summary it was suggested that xerostomia would be reduced if either the mean dose to the contralateral parotid gland ($D_{\text{mean}_{\text{contra}}}$) would be kept below 20 Gy, or if neither $D_{\text{mean}_{\text{contra}}}$, nor $D_{\text{mean}_{\text{ipsi}}}$ would exceed 25 Gy [8]. The one-gland guideline has thereafter proven useful to prevent xerostomia after 3DCRT/IMRT [9, 10], and to some extent after IMRT [11]. Wider use of either guideline following IMRT remains unsettled. Furthermore, it has been suggested that the evolution of RT-induced xerostomia is described by distinct temporal phases with time-specific etiologies [12], and a recovery between around three to twelve months after completed RT has been observed [9, 10, 13, 14].

In this work, we hypothesized that the dose-response relationship for xerostomia depends on underlying temporal-specific patterns. Objectively measured xerostomia data, and dose

information were taken from two cohorts including patients treated with IMRT for HNC. Five potentially predisposing variables were addressed, and the study was furthermore performed in a training-test design to explore generalizability within and across cohorts. The ultimate goal was to investigate to what extent both QUANTEC guidelines apply to xerostomia, and to explore at what Dmean threshold(s) xerostomia starts to evolve.

Materials and Methods

Study design and participants

This retrospective Institutionally Review Board approved study included prospectively collected data for patients previously treated with primary IMRT for HNC to the pharynx and the neck at the Memorial Sloan Kettering Cancer Center (*MSKCC*), or at the British Columbia Cancer Agency-Vancouver Cancer Centre (*BCCA*) from March 2006 to May 2012 (*BCCA*) or to March 2015 (*MSKCC*) [9, 15]. Saliva collection is standard practice at *MSKCC*, whereas informed consent was received from all *BCCA* patients. Further inclusion criteria for the current study were: A minimum of three whole-mouth stimulated flow measurements (one WMSFM >1g/5mins assessed pre-IMRT to exclude potential predisposition of baseline xerostomia, and at least two WMSFM assessed within 24 months post-IMRT), and a reasonably high RT prescription dose (>50.4 Gy).

In total, 55 *MSKCC* patients and 66 *BCCA* patients fulfilled the inclusion criteria. These patients had been planned and treated based on Computed Tomography (CT) imaging, and the median prescribed dose to the primary tumor was 70.0 (range: 50.4–70.2) Gy delivered in 1.8 Gy or 2.0 Gy daily fractions (Eclipse, Varian, Palo Alto, CA, US). The average \pm standard deviation (SD) age at initiation of IMRT was 57 ± 10 years in the *MSKCC*, and 56 ± 13 years in the *BCCA* cohort. The predominant primary tumor site in the *MSKCC* cohort was tonsil (22%), followed by base of tongue (19%), nasopharynx (14%), and oral tongue (13%). The corresponding figure in the *BCCA* cohort was nasopharynx (22%), base of tongue (17%), tonsil (17%), and tumors of unknown primary (15%).

Stimulated whole mouth saliva flow measurements, and xerostomia definition

Patients refrained from consuming food, or drinking typically for at least one hour prior to WMSFM. The WMSFM were assessed over a five-minute period, and were triggered by administration of a citrate solution to both sides of the tongue every 30 seconds during a twominute period (*MSKCC*), or by chewing on a paraffin block (*BCCA*). Saliva was collected in a pre-weighed plastic cup, and patients were asked to spit into the cup after the triggering procedure. Xerostomia was defined as moderate to severe (Grade 4) according to the LENT SOMA tables [16] *i.e.* WMSFM <25% post-relative to pre-RT. The median time to WMSFM after completion of IMRT was 11 (range: 3–24) months in the *MSKCC* cohort. In the *BCCA* cohort WMSFM was conducted for all 66 patients at both 3 and 12 months after completion of IMRT.

Follow-up groups

Following previous findings on xerostomia fading in the range 3–12 months after RT [9, 10, 13, 14], we stratified patients in one short, and one long follow-up group. In the *BCCA*

cohort, a straightforward split was possible given complete WMSFM data at both 3 (short follow-up) and 12 (long follow-up) months with an observed xerostomia rate of 47% (n=31), and 23% (n=15), respectively. The corresponding split, with a reasonable number of patients in both follow-up groups for the *MSKCC* cohort was less than 6 months (short follow-up; n=55) and 6–24 months (long follow-up; n=53) after completion of IMRT, and the median follow-up time was 3 (range: 1–5) months, and 11 (range: 6–24) months with an observed xerostomia rate of 38% (n=15), and 19% (n=10), respectively. Within the short follow-up group, six patients had a follow-up time of five months, and within the long follow-up group only three patients had a follow-up time 18 months, and four a follow-up time between six and seven months.

Modeling approach

1. Candidate predictors (MSKCC cohort only)—Initially, candidate predictors in the *MSKCC* cohort were identified for each follow-up group by comparing the distribution of the available variables between patients with and without xerostomia using a Mann-Whitney U test with significance denoted at the two-sided 5% level. The contralateral and ipsilateral parotid glands were defined as the gland with the lower and the higher Dmean, respectively (population median of $Dmean_{ipsi} - Dmean_{contra}$: 7.1 Gy across both cohorts and follow-up groups).

2. Variable selection, internal generalizability, and model performance (MSKCC cohort only)—All analyses were performed separately for the short and the long follow-up group. Each candidate predictor identified in 1. was investigated using the following logistic regression-based function (F_{Logreg})

$$F_{Logreg}(Candidate_x) = \frac{\exp(\beta_{Candidate_x} \cdot Candidate_x + \beta_0)}{1 + \exp(\beta_{Candidate_x} \cdot Candidate_x + \beta_0)} \quad (\text{Eq. 1})$$

$Candidate_x$ denotes a candidate predictor x , $\beta_{Candidate_x}$ the related logistic regression coefficient, and β_0 the intercept. Univariate logistic regression analysis (UVA) with bootstrap resampling (1000 sample populations with replacement; each population having the same size as the original dataset) was applied to evaluate internal generalizability of model parameters [17]. Predictive ability on UVA was suggested by a two-sided p-value 0.05 (assessed as the average p-value over the 1000 bootstrap populations). If multiple variables were suggested, these were subject to multivariate logistic regression analysis (MVA). An analogous bootstrap resampling approach as used in the UVA was applied also in the MVA, and an MVA model was considered a candidate model if it was selected in 10% (in at least 100 of the 1000 possible models). A backward-forward stepwise selection was applied in the MVA with the objective of minimizing the Akaike Information Criterion [18]. The performance of all estimated UVA and MVA models was assessed by the area under the receiver-operating characteristics curve (AUC), Spearman's rank correlation coefficients (R_s), and p-values. Goodness-of-fit of the estimated relative to the observed rate of xerostomia was evaluated using the Hosmer-Lemeshow test (HL; 10 degrees of freedom)

[19], and good agreement was considered if p_{HL} was ≤ 0.05 . The AUC, R_s , p-values, and p_{HL} in the *MSKCC* cohort are all reported as the average \pm SD over the 1000 populations.

3. External generalizability (MSKCC models applied to BCCA cohort)—After the final model(s) with the highest AUC had been identified in the *MSKCC* cohort (*cf.* 2. above), this model was explored in the *BCCA* cohort *i.e.* the logistic regression coefficients (including the intercept) estimated for the *MSKCC* cohort were applied to the corresponding variables in the *BCCA* cohort. Also here, model performance was assessed using AUC, R_s and p-values, and goodness-of-fit of the estimated relative to the observed rate of xerostomia was evaluated by the HL test (good agreement: $p_{HL} \leq 0.05$). Since no re-fitting process was performed for the *BCCA* cohort, the AUC, R_s p-value, and p_{HL} are each represented by one value within this cohort. All analyses were conducted in MATLAB v.R2016a.

Estimation of minimum Dmean thresholds

In the dose-response curves, which were all based on the variable-specific regression coefficients from the *MSKCC* cohort, the minimum Dmean threshold at which xerostomia was observed to evolve was identified. At this threshold the Dmean value, the estimated rate, and the relative risk (RR) were assessed. The RR and 95% confidence intervals (95% CIs) were calculated according to the following expressions:

$$RR_{Threshold} = \frac{\left(\frac{N_{Xero, \geq Threshold}}{N_{Xero, \geq Threshold} + N_{Xero, < Threshold}} \right)}{\left(\frac{N_{Non-xero, \geq Threshold}}{N_{Non-xero, \geq Threshold} + N_{Non-xero, < Threshold}} \right)} \quad (Eq.2)$$

$$95\%CI_{RR_{Threshold}} = \exp(\ln(RR_{Threshold}) \pm 1.96 \cdot SE(\ln(RR_{Threshold}))) \quad (Eq.3)$$

N_{Xero} and $N_{Non-xero}$ are the number of patients experiencing, and not experiencing xerostomia, respectively; *Threshold*, and *<Threshold* refer to if the Dmean was equal/larger or lower than that of the Dmean threshold, respectively, and SE is the standard error. Similarly, the estimated rate, the RR and its 95%CI for the two QUANTEC guidelines were assessed and compared to those of the identified Dmean thresholds.

Results

Identified candidate predictors

Five candidate predictors were identified for short-term xerostomia: Dmean_{contra}, Dmean_{ipsi}, WMSFM pre-RT, and concurrent chemotherapy (all being higher/more frequent among patients with xerostomia; Table 1). Also, primary tumor site was a candidate predictor with tumors of unknown primary being present only among the xerostomia patients together with a slight increase in the number of patients treated for tonsil and nasopharyngeal cancer. For long-term xerostomia only Dmean_{contra}, and Dmean_{ipsi} were candidate predictors. Neither of these candidate predictors was highly correlated with one another (Pearson’s correlation

coefficient, $|P_r| < 0.68$ [20]). The median $|P_r|$ in the short follow-up group was 0.24 (range: 0.15–0.59) with the highest correlation observed between $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$, while $|P_r|$ was 0.53 between $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ in the long follow-up group. The distribution of WMSFM, $D_{\text{mean}_{\text{contra}}}$, and $D_{\text{mean}_{\text{ipsi}}}$ is demonstrated for each follow-up group and cohort in Figure S1.

The mean dose to both parotid glands predict short-term xerostomia

Within the internal generalizability, four of the five identified candidate predictors were found to explain short-term xerostomia in the *MSKCC* cohort: $D_{\text{mean}_{\text{contra}}}$ ($p=0.01\pm 0.07$) $D_{\text{mean}_{\text{ipsi}}}$ ($p=0.01\pm 0.04$), use of concurrent chemotherapy ($p=0.04\pm 0.05$), and WMSFM pre-RT ($p=0.02\pm 0.08$) with the two dose-related variables presenting with the overall highest discriminative ability ($D_{\text{mean}_{\text{contra}}}$: $\text{AUC}=0.78\pm 0.07$; $D_{\text{mean}_{\text{ipsi}}}$: $\text{AUC}=0.85\pm 0.06$; Table 2). The most frequently selected MVA model (Figure S2) with the highest AUC suggested that the combined contribution from $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ best explained the dose-response relationship for xerostomia ($\text{AUC}=0.90\pm 0.05$; $R_s=0.66\pm 0.12$; $p=0.0002\pm 0.002$), and $p_{\text{HL}} > 0.05$ ($p_{\text{HL}}=0.22\pm 0.23$) suggested good agreement between the estimated and the observed rate (Table 2). The difference between the most and the least risky quintile for the observed rate using this MVA model was 73%, while it was 55% for $D_{\text{mean}_{\text{contra}}}$ and 64% for $D_{\text{mean}_{\text{ipsi}}}$ (Figure 1). Neither of the two identified candidate predictors in the long follow-up group explained longterm xerostomia ($D_{\text{mean}_{\text{contra}}}$: $p=0.18\pm 0.21$; $D_{\text{mean}_{\text{ipsi}}}$: $p=0.25\pm 0.27$), and, consequently, no MVA models were generated for this follow-up group, nor was this follow-up group included for the analyses outlined below.

Dose-response relationships for short-term xerostomia are generalizable across cohorts

Only the *MSKCC* model parameters for $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ explained short-term xerostomia in the *BCCA* cohort ($D_{\text{mean}_{\text{contra}}}$: $p=0.001$; $D_{\text{mean}_{\text{ipsi}}}$: $p<0.0001$; Table 3) with discriminative abilities of a comparable magnitude as in the *MSKCC* cohort ($D_{\text{mean}_{\text{contra}}}$: $\text{AUC}=0.71$; $D_{\text{mean}_{\text{ipsi}}}$: $\text{AUC}=0.80$). Similarly, but less pronounced compared to the *MSKCC* cohort, the AUC increased for the MVA model that included both $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ ($\text{AUC}=0.81$) as opposed to the corresponding UVA models. In general, the AUC values were slightly lower than those of the *MSKCC* cohort. The difference between the most and the least risky quintile for the observed rate stratified according to this MVA model in the *BCCA* cohort was 86%, while the corresponding difference for the two UVA models was 70% (Figure 1).

Identified mean dose thresholds for short-term xerostomia are lower than in the QUANTEC guidelines

The identified minimum threshold for $D_{\text{mean}_{\text{contra}}}$ was 12.4 Gy/18.7 Gy in the *MSKCC*/*BCCA* cohort with related estimated xerostomia rates of 12%/25%, while the estimated rate of the one-gland QUANTEC guideline was 29% in both cohorts. In the *MSKCC* cohort, the RR was slightly lower for these thresholds compared to that of the QUANTEC guideline but the 95%CI was narrower for the former (1.44 (95%CI: 1.04–2.00) vs. 2.29 (95%CI: 1.40–3.74)). In the *BCCA* cohort, the 95%CI of the RR for the $D_{\text{mean}_{\text{contra}}}$ threshold was reasonable, whereas that of the QUANTEC guideline was not (1.47 (95%CI: 1.06–2.00) vs. 1.38 (95%CI: 0.93–2.04)). Studying both glands, the $D_{\text{mean}_{\text{contra}}}$ threshold was again

considerably lower than that of the two-gland QUANTEC guideline in both cohorts (*MSKCC/BCCA*: 15.5 Gy/19.7 Gy), while the $D_{\text{mean}_{\text{ipsi}}}$ threshold was in a similar range (*MSKCC/BCCA*: 25.0 Gy/25.9 Gy) with corresponding estimated xerostomia rates of 6%/16% in the *MSKCC/BCCA* cohort compared to 27% for the QUANTEC guideline in both cohorts. The RR for the identified thresholds of the two glands was slightly lower than that of the two-gland QUANTEC guideline but presented with narrower 95% CIs (*MSKCC*: 1.65 (95% CI: 1.28–2.14) vs. 3.62 (95% CI: 1.97–6.66); *BCCA*: 1.47 (95% CI: 1.06–2.03) vs. 1.81 (95% CI: 1.18–2.77)). Incorporating the identified minimum $D_{\text{mean}_{\text{contra}}}$, or $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ thresholds into (Eq.1), the corresponding expression for the F_{Logreg} function (*note: all regression coefficients come from the MSKCC models*) was:

$$MSKCC:F_{\text{Logreg}}(D_{\text{mean}_{\text{contra}=12.4\text{Gy}}}) = \frac{\exp(0.14 \cdot 12.4 - 3.7)}{1 + \exp(0.14 \cdot 12.4 - 3.7)}$$

$$BCCA:F_{\text{Logreg}}(D_{\text{mean}_{\text{contra}=18.7\text{Gy}}}) = \frac{\exp(0.14 \cdot 18.7 - 3.7)}{1 + \exp(0.14 \cdot 18.7 - 3.7)}$$

$$MSKCC:F_{\text{Logreg}}(D_{\text{mean}_{\text{contra}=15.5\text{Gy}}, D_{\text{mean}_{\text{ipsi}=25.0\text{Gy}}}) = \frac{\exp(0.17 \cdot 15.5 + 0.11 \cdot 25.0 - 8.1)}{1 + \exp(0.17 \cdot 15.5 + 0.11 \cdot 25.0 - 8.1)}$$

$$BCCA:F_{\text{Logreg}}(D_{\text{mean}_{\text{contra}=19.7\text{Gy}}, D_{\text{mean}_{\text{ipsi}=25.9\text{Gy}}}) = \frac{\exp(0.17 \cdot 19.7 + 0.11 \cdot 25.9 - 8.1)}{1 + \exp(0.17 \cdot 19.7 + 0.11 \cdot 25.9 - 8.1)}$$

Discussion

Addressing both internal and external generalizability, our results indicate that objectively measured severe hyposalivation (xerostomia) after primary IMRT for HNC is in particular observed at a median of three months after completed treatment, and that the mean dose to both the contralateral and the ipsilateral parotid gland ($D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$) increases the risk of developing xerostomia within this follow-up time. Furthermore, fewer patients experienced long-term compared to short-term xerostomia (*MSKCC*: n=10 vs. 15; *BCCA*: n=15 vs. 31), and this could explain why no dose-response relationship was established for long-term xerostomia. A similar pattern has previously been observed in several studies including various assessment methods and RT techniques [9, 10, 13, 14]. This indicates an overall recovery of xerostomia, but the exact temporal recovery between the shorter and the longer follow-up time investigated here remains unresolved.

The salivary gland-specific QUANTEC summary advocated that keeping $D_{\text{mean}_{\text{contra}}}$ below 20 Gy, or keeping both $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ below 25 Gy should be the objective in HNC RT in order to minimize RT-induced xerostomia [8]. In one of the few dose-response focused studies on xerostomia following IMRT published thereafter, Lee *et al* [12] found that higher $D_{\text{mean}_{\text{contra}}}$, $D_{\text{mean}_{\text{ipsi}}}$, and advanced age increased the risk of moderate to

severe patient-reported xerostomia three months after completed treatment with a similar AUC as observed in this study (AUC=0.86), and that xerostomia at one year was explained by increased $D_{\text{mean}_{\text{contra}}}$, $D_{\text{mean}_{\text{ipsi}}}$, low educational level, smoking, and decreased T-stage. Lee *et al*, however, neither addressed inter-variable correlation, nor conducted UVA. Also, educational level and T-stage were each represented by four presumably highly correlated sub variables, and the logistic regression coefficients for T-stage were negative. It is, thus, not surprising that their results, in particular for long-term xerostomia, differ from ours. In another study that to some extent addressed highly correlated variables, and conducted UVA, $D_{\text{mean}_{\text{contra}}}$ and baseline xerostomia were found to predict moderate to severe patient-reported xerostomia six months after IMRT [21] with an AUC of 0.68. The AUC of our UVA model for $D_{\text{mean}_{\text{contra}}}$ in the *MSKCC* cohort was considerably higher (AUC=0.78±0.07), while in a comparable range in the *BCCA* cohort (AUC=0.71). Baseline xerostomia rate was an individual predictor for short-term xerostomia in the *MSKCC* cohort but was not on MVA. In addition, our xerostomia definition was normalized to that of pre-IMRT and by such baseline xerostomia was directly incorporated in all analyses. In both of our cohorts, the AUC for the final MVA models including $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ was higher (*MSKCC*: AUC=0.90±0.05; *BCCA*: AUC=0.81) than that of the corresponding UVA models. Therefore, focusing on Dmean to both individual glands should be the goal in order to inhibit the development of xerostomia after IMRT. Although we recognize potential differences between our cohorts and that of Beetz *et al*, it is yet somewhat surprising to note that while the majority (86%) of their patients received bilateral neck irradiation [21], and since contralateral sparing techniques is assumed to result in higher $D_{\text{mean}_{\text{ipsi}}}$ and increased risk of xerostomia compared to ipsilateral sparing techniques [5, 14], $D_{\text{mean}_{\text{ipsi}}}$ was not present in their final MVA model.

Even though the administration of chemotherapy, gender, and WMSFM pre-RT were significantly different between the two included cohorts (Table S1), our final MVA model did not indicate that any of these variables predicted xerostomia. Furthermore, the rate of xerostomia was nine and four percentage points lower in the *MSKCC* cohort than in the *BCCA* cohort for short- and long-term xerostomia, respectively, and we do acknowledge that these differences could to some extent also be explained by the slightly different follow-up time splits across the two cohorts, and/or that gustatory stimulation was performed using a citrate solution in the *MSKCC* cohort and by chewing on a paraffin block in the *BCCA* cohort. All these aspects together with the *MSKCC* patients presenting with significantly lower $D_{\text{mean}_{\text{contra}}}$ could explain the observed differences in discriminative ability of the UVA models for $D_{\text{mean}_{\text{contra}}}$ across the two cohorts (*MSKCC/BCCA*: AUC=0.78/0.71), but the AUC for the *BCCA* cohort was yet within the standard deviation of the AUC for the *MSKCC* cohort (0.07). We acknowledge that all potentially predisposing variables could not be accounted for. Furthermore, intra-gland response variability was not addressed, and these aspects may be important forthcoming topics to increase understanding of the dose-response relationship for xerostomia following IMRT [6, 22–24]. This together with the discussion above may justify the overall slightly lower AUC values being observed for the *BCCA* cohort. In a subsequent effort we performed predictive modeling within the *BCCA* cohort based on the candidate predictors identified for this cohort (* for *BCCA* in Table 1). Interestingly, the generated models for short-term xerostomia presented with coefficients and

a discriminative ability of a similar magnitude as that of applying the *MSKCC* model parameters in the *BCCA* cohort, and the most frequently selected MVA model with the highest discriminative ability included also here $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ (Table S2). For long-term xerostomia, the only initial candidate predictor ($D_{\text{mean}_{\text{ipsi}}}$) did not describe the observed rate of xerostomia. This suggests that the models identified for short-term xerostomia in the *MSKCC* cohort are generalizable in a cohort of patients presenting with somewhat differently distributed patient characteristics but treated with a similar RT technique.

Beetz *et al* [22] previously found that the AUC of their 3DCRT-developed MVA model (age, baseline xerostomia, and the average of $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$; all with positive regression coefficients) to predict patient-reported xerostomia six months after treatment dropped from 0.82 to 0.66 when applied to an IMRT cohort of a similar size while an AUC of 0.77 was expected taking into account case-mix differences. The observed lack of generalizability across the two RT techniques could be due to both $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ being significantly lower in the IMRT arm than in the 3DCRT arm resulting in lower rates of xerostomia [5]. However, Beetz *et al* [21] did not observe a drop in xerostomia rates after IMRT compared to after 3DCRT (51% vs. 52%) potentially since a significantly higher fraction of the patients in the IMRT arm received bilateral neck irradiation, but they did find that both $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ were significantly different (higher/lower not indicated) across the two RT techniques, and were further less correlated in the IMRT- compared to in the 3DCRT cohort. It was, thus, concluded that $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ should be included as separate candidate predictors [22]. To this end, and considering that no dose-response focused study has included $D_{\text{mean}_{\text{contra}}}$ and $D_{\text{mean}_{\text{ipsi}}}$ as separate candidate predictors, the generalizability of our MVA model for xerostomia following other RT techniques than IMRT thus remains unclear.

For short-term xerostomia in our cohorts, both QUANTEC guidelines applied to the *MSKCC* cohort, whereas only the two-gland guideline applied to the *BCCA* cohort as judged by the fairly narrow 95% CIs of the RR. To the best of our knowledge, the applicability of the QUANTEC guidelines following IMRT have only been studied by Lee *et al* where the one-gland guideline was examined [13]. At their three months follow-up, the negative predicted value (NPV) *i.e.*, the avoidance rate of moderate to severe patient-reported xerostomia when the one-gland QUANTEC guideline was fulfilled was 83%. In our data, the NPV of the one-gland guideline was of a similar magnitude (*MSKCC/BCCA*: 81%/79%), while slightly higher for the two-gland guideline (*MSKCC/BCCA*: 86%/84%). Furthermore, the estimated rate at the one-gland guideline was 18% in [13], and thus considerably lower than that estimated in our study (one/two-gland: 29%/27%). More importantly, however, our results indicated that short-term xerostomia after IMRT starts evolving at much lower $D_{\text{mean}_{\text{contra}}}$ values than suggested by QUANTEC. Therefore, if focusing on either contralateral- or ipsilateral parotid gland sparing, we suggest decreasing $D_{\text{mean}_{\text{contra}}}$ to below 20 Gy, and keeping $D_{\text{mean}_{\text{ipsi}}}$ according to the recommended QUANTEC level of around 25 Gy. Reducing $D_{\text{mean}_{\text{contra}}}$ to as low as 12.4 Gy, as suggested by our lowest $D_{\text{mean}_{\text{contra}}}$ threshold in the *MSKCC* cohort, may become clinically impractical as any sparing procedure should foremost not deteriorate the objective of delivering the prescribed tumor dose.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded in part through the **NIH/NCI Cancer Center Support Grant P30 CA008748**.

References

1. Epstein JB, Thariat J, Bensadoun RJ, et al. Oral complications of cancer and therapy: from cancer treatment to survivorship. *CA Cancer J Clin*. 2012; 62:400–422. [PubMed: 22972543]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016; 66:7–30. [PubMed: 26742998]
3. Sciubba JJ, Goldenberg D. Oral complications of radiotherapy. *Lancet Oncol*. 2006; 7:175–183. [PubMed: 16455482]
4. Kam MK, Leung SF, Zee B, et al. Prospective randomized study of intensity-modulated radiotherapy on salivary gland function in early-stage nasopharyngeal carcinoma patients. *J Clin Oncol*. 2007; 25:4873–4879. [PubMed: 17971582]
5. Nutting CM, Morden KP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol*. 2011; 12:127–136. [PubMed: 21236730]
6. Vissink A, van Luijk P, Langdendijk JA, Coppes RP. Current ideas to reduce or salvage radiation damage to salivary glands. *Oral Dis*. 2015; 21:1–10. [PubMed: 25399654]
7. Dawes C, Wood CM. The contribution of oral minor mucous gland secretions to the volume of whole saliva in man. *Arch Oral Biol*. 1973; 18:337–342. [PubMed: 4515961]
8. Deasy JO, Moiseenko V, Marks L, Clifford Chao KS, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. *Int J Radiat Oncol Biol Phys*. 2010; 76(Suppl 3):58–63.
9. Moiseenko V, Wu J, Hovan A, Saleh Z, et al. Treatment planning constraints to avoid xerostomia in head-and-neck radiotherapy: an independent test of the QUANTEC criteria using a prospectively collected dataset. *Int J Radiat Oncol Biol Phys*. 2012; 82:1108–1114. [PubMed: 21640505]
10. Beetz I, Steenbakkers RJHM, Chouvalova O, et al. The QUANTEC criteria for parotid gland dose and their efficacy to prevent moderate to severe patient-rated xerostomia. *Acta Oncol*. 2014; 53:597–604. [PubMed: 23998646]
11. Coppes RP, Zeilstra LJ, Kampinga HH, Konings AW. Early to late sparing of radiation damage to the parotid gland by adrenergic and muscarinic receptor antagonists. *Br J Cancer*. 2001; 85:1055–1063. [PubMed: 11592779]
12. Lee TF, Chao PJ, Ting HM, et al. Using multivariate regression model with least absolute shrinkage and selection operator (LASSO) to predict the incidence of xerostomia after intensity-modulated radiotherapy for head and neck cancer. *PLOS ONE*. 2014; 9:1–11.
13. Lee TF, Fang FM. Quantitative analysis of normal tissue effects in the clinic (QUANTEC) guideline validation using quality of life questionnaire datasets for parotid gland constraints to avoid causing xerostomia during head-and-neck radiotherapy. *Radiother Oncol*. 2013; 106:352–358. [PubMed: 23333019]
14. Miah AB, Gulliford SL, Morden J, et al. Recovery of salivary function: Contralateral parotid-sparing intensity-modulated radiotherapy versus bilateral superficial lobe parotid-sparing intensity-modulated radiotherapy. *Clin Oncol*. 2016 Mar 17. [Epub ahead of print].
15. Clark H, Hovan A, Moiseenko V, Thomas S, Wu J, Reinsberg S. Regional radiation dose susceptibility within the parotid gland. *Med Phys*. 2015; 42:2064–2071. [PubMed: 25832096]
16. LENT SOMA tables. *Radiother Oncol*. 1995; 35:17–60. [PubMed: 7569012]
17. Deasy JO, Chao KS, Markman J. Uncertainties in model-based outcome predictions for treatment planning. *Int J Radiat Oncol Biol Phys*. 2001; 51:1389–1399. [PubMed: 11728700]
18. Collett, D. Modeling binary data. 2nd. NY: Chapman & Hall; 2002.

19. Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. *Commun in Stats.* 1980; 10:1043–1069.
20. Taylor R. Interpretation of the correlation coefficient. A Basic Review. *J Diag Med Sonog.* 1990; 1:35–39.
21. Beetz I, Schilstra C, van der Schaaf A, 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; 105:101–106. [PubMed: 22516776]
22. Beetz I, Schilstra C, van Luijk P, 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. [PubMed: 22169766]
23. van Luijk P, Pringle S, Deasy JO, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Cancer.* 2015; 7:1–8.
24. Buettner F, Miah AB, Gulliford SL, et al. Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial. *Radiother Oncol.* 2012; 103:82–87. [PubMed: 22444242]

Short-term xerostomia

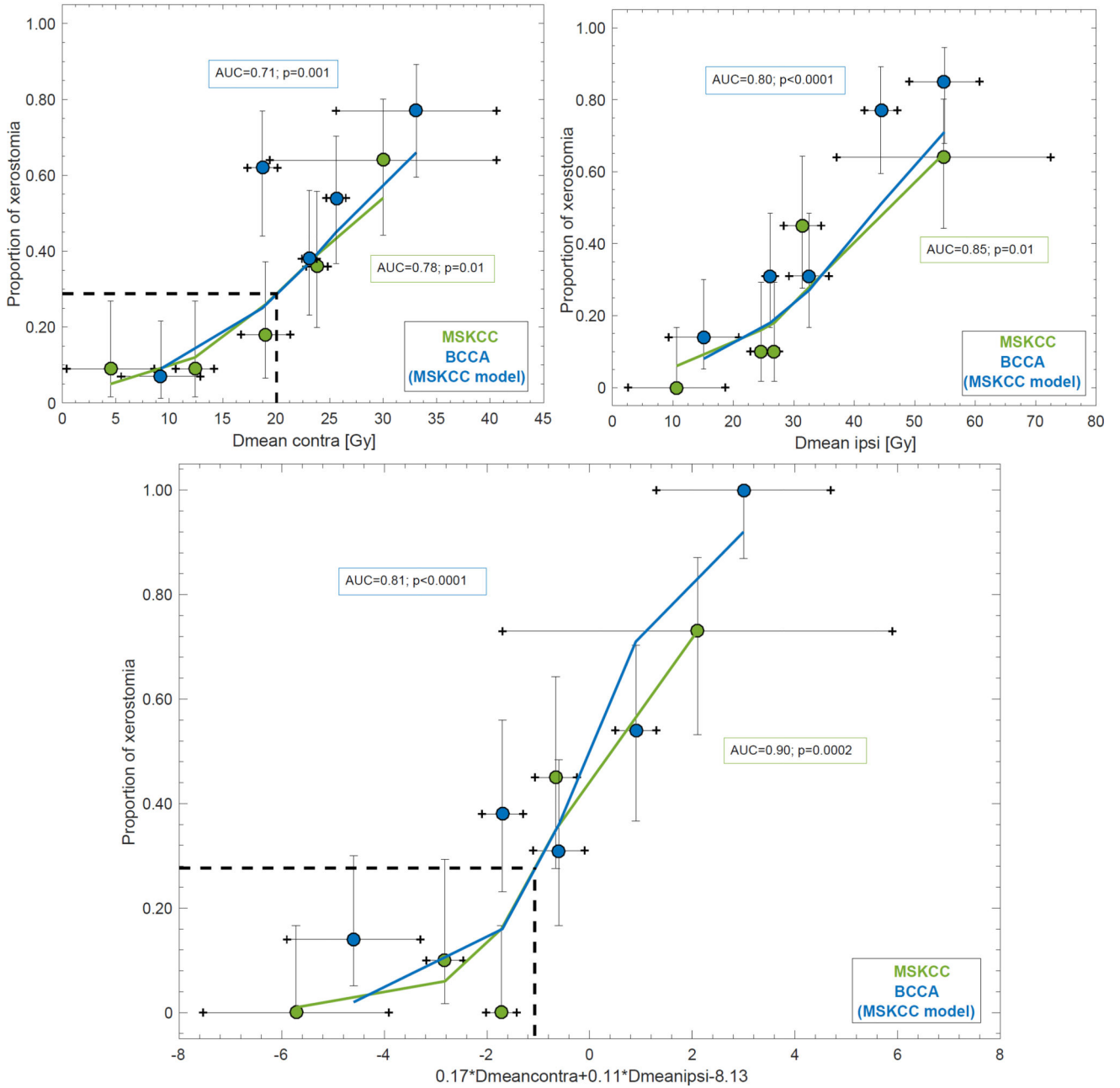


Figure 1. The UVA models for Dmean_{contra} (upper left), and Dmean_{ipsi} (upper right), and the MVA model for Dmean_{contra} and Dmean_{ipsi} (lower panel) for the MSKCC cohort (green) and as applied to the BCCA cohort (blue). Note: Solid lines denote the estimated xerostomia rate, the quintiles represent the observed xerostomia rate (x-axis: mean ± SD; y-axis: mean and 68% exact binomial confidence intervals), the dashed black lines are the estimated

xerostomia rate at the QUANTEC guidelines, and the AUC and p-values (MSKCC: average over the 1000 resamples) are inserted next to the corresponding curve.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Variable selection results and model performance for the two follow-up times in the *MSKCC* cohort based on the candidate predictors in this cohort (cf. Table 1).

Table 2

MSKCC		Follow-up: Long										
Follow-up: Short	Model	AUC	R _s	p	β	p _{HL}	Model	AUC	R _s	p	β	p _{HL}
	Dmean _{contra} [*]	0.78 (0.07)	0.46 (0.12)	0.01 (0.07)	0.14	0.51 (0.23)	Dmean _{contra}	0.80 (0.07)	0.28 (0.18)	0.18 (0.21)	0.07	0.39 (0.19)
	Dmean _{psi} [*]	0.85 (0.06)	0.50 (0.13)	0.01 (0.04)	0.09	0.24 (0.11)	Dmean _{psi}	0.73 (0.09)	0.24 (0.16)	0.25 (0.27)	0.03	0.43 (0.10)
	WMSFM Pre-RT [*]	0.71 (0.09)	0.45 (0.13)	0.02 (0.08)	0.84	0.51 (0.12)						
	Tumor site	0.69 (0.08)	0.28 (0.13)	0.13 (0.20)	0.14	0.46 (0.13)						
	Chemotherapy [*]	0.65 (0.04)	0.33 (0.07)	0.04 (0.05)	135	0.19 (0.07)						
	Dmean _{contra} +Dmean _{psi} ^{**}	0.90 (0.05)	0.66 (0.12)	0.0002 (0.002)	0.17/0.11	0.22 (0.23)						
	WMSFM Pre-RT+Dmean _{psi}	0.85 (0.06)	0.64 (0.12)	0.001 (0.01)	0.85/0.09	0.45 (0.24)						

Note: The AUC, R_s, p, and p_{HL} are given as the mean ± standard deviation (SD) over the 1000 bootstrap samples.

^{*} Qualified for MVA,

^{**} Most frequently selected model (38%).

External validation of the final models suggested for the *MSKCC* cohort as applied to the *BCCA* cohort.

Table 3

BCCA

Follow-up: Short

<i>Model</i>	<i>AUC</i>	<i>R_s</i>	<i>p</i>	β	<i>PHL</i>
Dmean _{contra}	0.71	0.40	0.001	0.14	0.48
Dmean _{ipsi}	0.80	0.53	<0.0001	0.09	0.67
WMSFM Pre-RT	0.50	0.01	0.96	-0.003	0.22
Chemotherapy	0.59	0.18	0.14	0.75	0.22
Dmean _{contra} +Dmean _{ipsi} *	0.81	0.53	<0.0001	0.17/0.11	0.27

*The MVA model including WMSFM Pre-RT+Dmean_{ipsi} was not included since WMSFM Pre-RT presented with p>0.05 on UVA.