Geometric Image Biomarker Changes of the Parotid Gland Are Associated With Late Xerostomia

Lisanne V. van Dijk, MS,* Charlotte L. Brouwer, PhD,* Hans Paul van der Laan, PhD,* Johannes G.M. Burgerhof, PhD,† Johannes A. Langendijk, MD, PhD,* Roel J.H.M. Steenbakkers, MD, PhD,* and Nanna M. Sijtsema, PhD*

Departments of *Radiation Oncology, and †Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Received Dec 15, 2016, and in revised form Jul 17, 2017. Accepted for publication Aug 5, 2017.

Summary

The aim of this study was to identify a surrogate marker for late xerostomia 12 months after radiation therapy (Xer12m), according to information obtained shortly after treatment. Differences in parotid gland (PG) were quantified in image biomarkers (ΔIBMs) before and 6 weeks after radiation therapy in 107 patients. By performing stepwise forward selection, ΔIBMs that were associated with Xer12m were selected. Subsequently other variables, such as PG dose and acute xerostomia scores, were added to improve the prediction performance. All models were internally validated.

Purpose: To identify a surrogate marker for late xerostomia 12 months after radiation therapy (Xer12m), according to information obtained shortly after treatment.

Methods and Materials: Differences in parotid gland (PG) were quantified in image biomarkers (ΔIBMs) before and 6 weeks after radiation therapy in 107 patients. By performing stepwise forward selection, ΔIBMs that were associated with Xer12m were selected. Subsequently other variables, such as PG dose and acute xerostomia scores, were added to improve the prediction performance. All models were internally validated.

Results: Prediction of Xer12m based on PG surface reduction (ΔPG-surface) was good (area under the receiver operating characteristic curve, 0.82). Parotid gland dose was related to ΔPG-surface ($P<0.001$, $R^2 = 0.27$). The addition of acute xerostomia scores to the ΔPG-surface improved the prediction of Xer12m significantly, and vice versa. The final model including ΔPG-surface and acute xerostomia had outstanding performance in predicting Xer12m early after radiation therapy (area under the receiver operating characteristic curve, 0.90).

Conclusions: Parotid gland surface reduction was associated with late xerostomia. The early posttreatment model with ΔPG-surface and acute xerostomia scores can be considered as a surrogate marker for late xerostomia. © 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Introduction

Xerostomia is one of the most frequent side effects that affects many head and neck cancer (HNC) patients after radiation therapy and has a major impact on quality of life (1). Limiting the dose to the parotid glands (PGs) reduces the probability of developing xerostomia (2-4). Although multiple studies have investigated the relation between dose and the risk of xerostomia, substantial variability in this relationship remains unexplained (2, 3). A possible reason for this variation is that dosimetric parameters (and baseline xerostomia scores) are not the only explaining variables, but that patient-specific characteristics, such as intrinsic radiosensitivity, also affect the development of late xerostomia (5). Unexplained variability could, moreover, result from inconsistency in the assessment of xerostomia (ie, patient-rated xerostomia, because this is a subjective measure) (6). More specifically, the individual experience of a side effect with similar function loss varies widely among individual patients, depending on many aspects, such as interpretation of the questions and general quality of life (7). A more quantitative measure of late xerostomia may lead to improvement of prediction models by increasing the consistency of the endpoint.

A surrogate endpoint early after treatment to evaluate late xerostomia is not only interesting to understand the development of xerostomia better, but would also be desirable and beneficial to potentially improve the time and cost-effectiveness of future clinical studies in HNC patients. Additionally, this could also contribute to the physician—patient dialogue at the end of treatment to provide patients with a more reliable prognosis regarding the expected severity of xerostomia for the next few years. Moreover, it can support selection of patients who do not recover from acute xerostomia for potential future therapeutic strategies of xerostomia, such as adult stem cell—based therapy (8, 9).

Computed tomography (CT) image acquisition, which is routinely used for radiation therapy treatment planning and response assessment, would be an ideal modality to quantify changes of radiated tissues, because it is rapid, relatively cheap, and widely available.

To identify quantitative candidate surrogates for assessing xerostomia, PG characteristics were quantified by extracting image biomarkers (IBMs) of the PGs before and after radiation therapy and by calculating the differences (ΔIBMs). The main objective of this study was to test the hypothesis that ΔIBMs—either combined with other predictive factors or not—were associated with late xerostomia and to test whether an early posttreatment model based on these ΔIBMs could serve as a surrogate marker for late xerostomia.

Methods and Materials

Patients

The 107 HNC who were prospectively included in this study were treated with radiation therapy, either in combination with concurrent chemotherapy or cetuximab or not, between June 2008 and April 2012. Patients with salivary glands tumors, those who previously (or 1 year after) underwent surgery or radiation therapy in the head and neck area were excluded from this study. Moreover, patients without follow-up data 12 months after radiation therapy were excluded. For a detailed description of the radiation protocols we refer to the article of Christianen et al (10). Briefly, most patients were treated with intensity modulated radiation therapy that was optimized to spare the PGs without compromising the dose to the target volumes (11, 12), using a simultaneous integrated boost technique. Generally, 70 Gy (2 Gy per fraction) was administered to the primary tumor and pathologic lymph nodes over the course of 6 or 7 weeks (6 or 5 fractions per week, respectively). The majority of patients received elective radiation to the cervical lymph nodes of 54.25 Gy (1.55 Gy per fraction) (13). More patient characteristics are depicted in Table 1.

For all patients a standardized planning CT scan (Somatom Sensation Open, Siemens; voxel size, 0.94 × 0.94 × 2.0 mm³; 100-140 kV) was acquired 2 weeks before treatment. Six weeks after radiation therapy a second CT scan was acquired together with the last assessment of acute toxicity. Both scans were acquired with a thermoplastic mask in radiation therapy treatment positioning. This study was approved the medical ethics commission, and all participating patients gave informed consent.

Endpoints

Patient-rated xerostomia scores were evaluated prospectively on a routine basis, before radiation therapy, weekly during radiation therapy, and subsequently 6 weeks and 6 and 12 months after radiation therapy using the European Organization for Research and Treatment of Cancer QLQ-H&N35 questionnaire, as part of the standard follow-up program at the department of Radiation Oncology of the University Medical Center Groningen (2, 14). The primary endpoint of this study was moderate to severe patient-rated xerostomia at 12 months after radiation therapy (Xer12m). This corresponds to the 2 highest scores of the 4-point Likert scale (not, a bit, quite a bit, a lot).

Quantification of PG changes in ΔIBMs

The PGs were delineated on the planning CT scan according to the consensus guidelines of Brouwer et al (15). Using deformable image registration, delineations were warped to the repeat CT scan in Mirada RTx (Mirada Medical). The warped contours were manually corrected if necessary.

All image biomarkers were extracted from the planning and the repeat CT with in-house-developed software that was implemented in MATLAB (version R2014a; The MathWorks). Subtraction of the pre- from the posttreatment IBMs resulted in the ΔIBMs. Twenty geometric IBMs of
the PGs, such as volume and compactness, were extracted from the delineations directly. Additionally, 20 CT-intensity ΔIBMxs, were extracted from the CT data of the PGs. For a list of the ΔIBMxs and the additional 8 clinical variables, refer to Supplementary data 1 (available online at www.redjournal.org).

Reference model

A reference prediction model for late xerostomia based on the predictors found by Beetz et al (2) (mean dose to the contralateral PG and the baseline xerostomia) was fitted to the dataset (Fig. 1, “reference model”). The patient-reported xerostomia at start of radiation therapy (Xerbaseline) was dichotomized as none versus any. The PG that received the least amount of mean dose was considered contralateral.

ΔIBM selection for late xerostomia

To investigate the associations between the potential ΔIBMxs and Xer12m (① in Fig. 1), ΔIBMxs were considered as candidate variables in the variable selection process as described below. Subsequently, the modeling process was repeated by adding Xerbaseline first and subsequently the mean PG dose to the candidate variables. All individual patient variable values were normalized by subtracting each value by the sample mean and then dividing by the sample standard deviation of that IBM variable.

Introducing large numbers of highly correlated variables may have negative effects on variable selection, due to overfitting and multicollinearity (16, 17). Candidate ΔIBMxs were therefore preselected according to their (Pearson) correlation. If the correlation of 2 variables was larger than 0.80, only the ΔIBM with the highest association with Xer12m was selected for further analysis.

Stepwise forward selection, based on log-likelihood (18), was used to select the most important predictors (P<.01). The internal validity of the variable selection was estimated with a bootstrap procedure. The entire variable selection procedure (variable normalization, preselection, and forward selection) was repeated in 1000 bootstrapped samples (ie, with replacement). From the resulting models the most frequently selected variables were considered for the final models.

The selected model’s optimism was estimated by calculating the difference between the performance of the models in each bootstrap and in the original sample, as suggested by the TRIPOD statement (19).

The model’s performance was quantified in terms of discrimination with the area under the receiver operating characteristic curve (AUC), the Nagelkerke $R^2$, and the discrimination slope. Model calibration was tested with the Hosmer-Lemeshow test and by calculating the slope and intercept of a logistic regression model of the linear predictor derived from the predicted probability of moderate to severe late xerostomia (variable) against the actual xerostomia outcome (response). The coefficients were corrected for optimism accordingly. The R packages Regression Modeling Strategies (version 4.3-1) (20) were used for these purposes.

Relation PG dose and selected ΔIBMxs

Linear regression was performed to investigate the relation of mean PG dose to the selected ΔIBMxs (② in Fig. 1). Both PGs were considered separately in investigating this relation. Model performance was measured as the explained variance ($R^2$), and normality of the residuals of the regression models was checked.
**ΔIBMs and other predictive variables**

First, the addition of mean PG dose to the model with selected ΔIBMs and Xerbaseline was investigated.

Second, the relation between ΔIBMs and acute xerostomia 6 weeks after radiation therapy (Xer6w-post, moderate to severe) was investigated (③ in Fig. 1), to analyze whether the selected ΔIBMs were a direct substitute measure of acute xerostomia scores. If the assumption that acute and late xerostomia scores are related would be correct (④ in Fig. 1), then the selected ΔIBMs could actually be a measure of acute xerostomia rather than late xerostomia. Therefore, the presumed assumption was tested by investigating the logistic relation between Xer6w-post and Xer12m. Subsequently, a multivariable analysis and variable selection was performed to investigate whether the Xer6w-post contained additional information to the ΔIBMs to predict late xerostomia (⑤ in Fig. 1). Baseline xerostomia was also considered for these analyses.

Actual xerostomia incidences were depicted over time to illustrate the effects of the ΔIBMs and the final posttreatment model with the best prediction performance. Patients were classified according to their ΔIBM values (higher or lower than median) and their predicted risk calculated with the final prediction model (higher or lower than 50%).

**Results**

**Reference model**

Moderate to severe xerostomia 12 months after radiation therapy (Xer12m) was reported by 32 (30%) of the 107 patients. The reference model based on mean PG dose and Xerbaseline was fitted to the dataset. The model characteristics and the performance measures (AUC 0.76, $R^2 = 0.28$) are depicted in Table 2 (Reference model).

**ΔIBMs selection for late xerostomia**

The most frequently selected ΔIBM variable was ΔPG-surface (a visual representation is depicted in Fig. 2) of the contralateral PG, obtained with forward stepwise selection (468 times of 1000 bootstrapped samples; see Supplementary data 2 [available online at www.redjournal.org] for frequency plots). This variable was significantly associated to Xer12m (Table 2; $P<.001$; odds ratio [OR] 0.86, 95% confidence interval [CI] 0.79-0.92). The regression coefficient was negative, indicating that larger PG surface reduction or shrinkage relates to a higher risk of developing Xer12m. Xer baseline showed a significant predictive contribution in addition to ΔPG-surface (likelihood ratio test; $P<.001$). A model with ΔPG-surface and Xerbaseline was created with good performance (Table 2; AUC 0.82, 95% CI 0.72-0.91), also after internal validation, which includes the variable selection (Table 2; AUC 0.77). This model had better performance predicting Xer12m than the reference model (Table 2), which was based on mean contralateral PG dose and Xerbaseline. No variable selection was performed for the reference model, thus internal validation was calculated by refitting the variables in the bootstrapped samples.

It should be noted that the variable ΔPG-volume was highly correlated to ΔPG-surface ($ρ = 0.91$) and was therefore eliminated in the preselection. A model with ΔPG-volume and Xerbaseline had comparable, but slightly reduced, performance measures (AUC 0.80; $R^2 = 0.34$; see Supplementary material 3 [available online at www.redjournal.org] for model and performance characteristics).

**PG dose and surface change**

A significant linear relation was observed between contralateral ΔPG-surface and contralateral PG dose ($P<.001$),
but the explained variance was relatively low ($R^2 = 0.27$). On the basis of the scatterplot (Fig. 3A) a quadratic relation was fitted, and it proved significantly better than a linear fit ($R^2 = 0.34$, $P < 0.001$, analysis of variance $F$ test). This quadratic fit improved the fit even more for the ipsilateral PG (linear fit $R^2 = 0.19$, quadratic fit $R^2 = 0.38$, $P < 0.001$, analysis of variance $F$ test). Note that the mean dose levels received by the ipsilateral PG were larger (Fig. 3B). The residuals of the regression models were reasonably normally distributed.

**ΔIBMs and other predictive variables**

Initial (planning) PG mean dose did not significantly add to a model with ΔPG-surface in terms of predicting late Xer12m (likelihood ratio test; $P = 16$). Performance measures improved slightly, but no difference was seen after internal validation (Table 2).

A significant univariable logistic relation was found between ΔPG-surface and acute xerostomia scores at the same point in time, 6 weeks after radiation therapy (Xer6w-post) ($P = 0.17$; OR 0.93, 95% CI 0.87-0.99; 3 in Fig. 1). However, a stronger association between ΔPG-surface and Xer12m was observed ($P < 0.001$; OR 0.86, 95% CI 0.79-0.93; 3 in Fig. 1). Acute (Xer6w-post) and late xerostomia (Xer12m) were indeed related ($P < 0.001$; OR 14.29, 95% CI 5.20-39.27; 3 in Fig. 1). Moreover, acute xerostomia added significantly to ΔPG-surface in predicting Xer12m (likelihood ratio test; $P < 0.001$) and vice versa (likelihood ratio test; $P < 0.001$) (3 in Fig. 1). The performance measures of this model with ΔPG-surface and Xer6w-post further improved to an AUC of 0.90 (95% CI 0.84-0.96) and $R^2$ of 0.56 (Table 2). Again, mean PG dose could not improve the model (likelihood ratio test; $P = 0.27$). Calibration of all presented models was good (Table 2, Hosmer-Lemeshow test, calibration intercept and slope).

Depicting the actual moderate to severe xerostomia incidences, Figure 4 shows that ΔPG-surface was able to significantly differentiate between patients with high and low xerostomia incidence at 6 and 12 months.

Using the complete early posttreatment model (ΔPG-surface, Xerbaseline, and Xer6w-post) resulted in an even better distinction (Fig. 4B): the actual reported xerostomia differences of patients with high ($>50\%$) and low ($<50\%$) predicted risk were substantial.

Finally, using the same classification-based early posttreatment model for patients with moderate to severe xerostomia 6 weeks after radiation therapy (Xer6w-post = 1) showed that the predictions of the early posttreatment model could significantly differentiate between patients who recovered and those still suffering from xerostomia at 6 and 12 months (Fig. 4C). This suggests that ΔPG-surface contributes in differentiating between patients who have persistent xerostomia up to 12 months and those who recover. Two patients had no reported xerostomia scores at 6 months.

### Discussion

In this study a significant relationship was shown between the geometric ∆IBM (ΔPG-surface) and late xerostomia. ΔPG-surface added significantly and independently to acute toxicity scores in predicting late xerostomia. Moreover, the performance of the models based on ΔPG-surface (with or without acute toxicities) was better than the reference model based on PG dose. Those observations together suggest that ΔPG-surface contains additional information on patient-specific development of late xerostomia. Mean PG dose did not add significantly to any of the ΔPG-surface models in this cohort. A possible explanation could be that ΔPG-surface and Xer6w-post, which result from radiation dose, contain the same information as the PG mean dose; however, this should be confirmed in an external dataset.

High correlation between ΔPG-volume and ΔPG-surface was observed. Prediction of late xerostomia was good with both variables, but ΔPG-surface performed better than ΔPG-volume (Supplementary data 2; available online at www.redjournal.org). It can be hypothesized that surface change holds more information, because it also includes information on the shape of the PG. However, this observation may be limited to the current dataset; hence more research is necessary to investigate whether this can be confirmed in other datasets. Furthermore, in this study the absolute ΔPG-volume and ΔPG-surface were investigated; similar performance was achieved with proportional change.

A nonlinear (eg, quadratic) relation between mean PG dose and ΔPG-surface (or volume) was observed (ie, PG surface reduction increased with increasing the mean dose up to 30-40 Gy), but for PGs that received higher doses the PG surface reduction decreased again. This suggests that PGs react differently to higher doses, which might be due to direct necrosis of the PG cells inducing inflammatory swelling, instead of controlled apoptosis (21), which than compensates (partly) for the radiation-induced volume decrease.

Although there is no study to our knowledge that has investigated ΔPG-surface, many studies reported reductions of PG volumes after radiation therapy (22-27). In line with our results, the studies with adequate patient numbers observed a significant, but weak, relationship between mean PG dose and volume decrement ($r = 0.41$ [24], $r = 0.26$ [25]). This means that a large amount of unexplained variation remains. A possible explanation for these findings is variation in individual radiation sensitivity of PGs, or mean PG dose may not be the most optimal dosimetric parameter (9).

In the present study significant associations were found between: ΔPG-surface and late xerostomia (3 in Fig. 1), the PG dose and ΔPG-surface (3 in Fig. 1), and PG dose and late xerostomia (Fig. 1, “reference model”). This is the first study that verified all these relationships.
Parotid gland surface reduction was related to acute xerostomia scores. However, a stronger relation was observed between PG surface reduction and late xerostomia ($Xer_{12m}$). Moreover, not only did acute xerostomia add predictive information to $\Delta PG$-surface in predicting late xerostomia, also $\Delta PG$-surface added to acute xerostomia. These results suggest that $\Delta PG$-surface yields unique information about the patient-specific capability of the PG to recover from radiation damage (also see Fig. 4C) and is not only a quantitative substitute of $Xer_{6w-post}$.

The model with $\Delta PG$-surface, baseline, and acute xerostomia scores (6 weeks after treatment) predicts late xerostomia with an exceptionally good performance, reflected in the good discrimination measures ($AUC 0.90$, $95\% CI 0.84-0.96$; and $AUC_{bootstrapped} 0.86$). Early prediction of late xerostomia could improve effectiveness of future clinical studies, because the 1-year compliance is approximately 60\% (1-year overall survival of HNC approximately 70\% [29] together with other dropout factors). First, an early surrogate could increase the follow-up information and thereby the time and cost-effectiveness of clinical studies. Second, adequate

### Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta$</th>
<th>OR</th>
<th>$P$ (apparent)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$-2.613$</td>
<td>$-2.489^*$</td>
<td>0.76 (0.67-0.86)</td>
<td>0.75*</td>
</tr>
<tr>
<td>$Xer_{baseline}$</td>
<td>1.570</td>
<td>1.500*</td>
<td>4.81</td>
<td>4.48*</td>
</tr>
<tr>
<td>PG dose</td>
<td>0.049</td>
<td>0.047*</td>
<td>1.05</td>
<td>1.05*</td>
</tr>
<tr>
<td>$\Delta IBM$ models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$-2.180$</td>
<td>$-1.647$</td>
<td>0.76 (0.66-0.85)</td>
<td>0.72</td>
</tr>
<tr>
<td>$\Delta PG$-surface</td>
<td>$-0.155$</td>
<td>$-0.126$</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>Intercept</td>
<td>$-2.875$</td>
<td>$-2.112$</td>
<td>0.82 (0.72-0.91)</td>
<td>0.77</td>
</tr>
<tr>
<td>$Xer_{baseline}$</td>
<td>1.587</td>
<td>1.242</td>
<td>4.89</td>
<td>3.46</td>
</tr>
<tr>
<td>$\Delta PG$-surface</td>
<td>$-0.154$</td>
<td>$-0.121$</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>$\Delta IBM$ and dose model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$-3.295$</td>
<td>$-2.389$</td>
<td>0.82 (0.73-0.91)</td>
<td>0.77</td>
</tr>
<tr>
<td>$Xer_{baseline}$</td>
<td>1.533</td>
<td>1.175</td>
<td>4.63</td>
<td>3.24</td>
</tr>
<tr>
<td>$\Delta PG$-surface</td>
<td>$-0.130$</td>
<td>$-0.100$</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>PG dose</td>
<td>0.028</td>
<td>0.022</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>Acute xerostomia model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$-2.763$</td>
<td>$-1.982$</td>
<td>0.85 (0.77-0.93)</td>
<td>0.81</td>
</tr>
<tr>
<td>$Xer_{baseline}$</td>
<td>1.540</td>
<td>1.164</td>
<td>4.66</td>
<td>3.20</td>
</tr>
<tr>
<td>$Xer_{6w-post}$</td>
<td>2.581</td>
<td>1.950</td>
<td>13.21</td>
<td>7.03</td>
</tr>
<tr>
<td>$\Delta PG$-surface</td>
<td>$-0.169$</td>
<td>$-0.128$</td>
<td>0.84</td>
<td>0.88</td>
</tr>
</tbody>
</table>

**Abbreviations:** $\beta$ = regression coefficients; $\Delta PG$-surface = parotid gland surface difference after – before radiation therapy; AUC = area under the receiver operating characteristic curve; DS = discrimination slope; HL = Hosmer-Lemeshow test; OR = odds ratio; PG = parotid gland; $R^2$ = Nagelkerke $R^2$; $Xer_{baseline}$ and $Xer_{6w-post}$ = xerostomia score at baseline and 6 weeks after radiation therapy.

* No variable selection was performed for internal validation of the reference model.

simultaneously and developed a normal tissue complication probability model to predict late xerostomia with a quantitative measure from CT imaging. Belli et al (28) showed a relation between PG shrinkage and acute xerostomia scores. Another study (25) observed in a limited cohort ($n=20\sim24$, >12 months’ follow-up) that small mid-treatment PG volume loss was associated with a longer period of xerostomia recovery for patients receiving a relative high mean PG dose ($>35.7\text{ Gy}$, $n=11$). These counterintuitive findings might be explained by the nonlinear relation of the dose with $\Delta PG$-surface shown in the present study. This relation suggests that a high PG dose may result in small $\Delta PG$-surface (or volume), because this is potentially due to inflammatory PG swelling, which in turn might be related to the longer period of xerostomia recovery. Hence, in this specific group of patients, $\Delta PG$-surface alone might not optimally represent radiation therapy damage. In contrast, for patients with both a high PG dose and large PG change, $\Delta PG$-surface still indicated high risk to develop late xerostomia in the present study.
The relationship of ΔPG-surface with dose suggests that ΔPG-surface is a biomarker that measures physiologic response. However, a correlation like this does not necessarily make this biomarker a surrogate for a clinical marker (30). A candidate surrogate marker should also have a relationship with the clinical endpoint, which is patient-rated late xerostomia in this study. This study shows a significant association between ΔPG-surface and Xer12m. In addition, the model of ΔPG-surface together with acute xerostomia (Xer6w-post) also meets the criteria mentioned above. Therefore, this model can be considered as a candidate surrogate marker for late xerostomia. Subsequently, external validation or a clinical trial is needed to verify whether the model of ΔPG-surface together with Xer6w-post can be used as a validated surrogate marker (31, 30).

Unfortunately no contrast was used for the CT scan 6 weeks after treatment. Although this does not influence the geometric ΔIBMs, it could explain why no strong relation was observed between late xerostomia and CT intensity—based ΔIBMs, such as mean intensity/density change that has been reported in other studies (28). Univariable analysis, however, did show a significant relation between mean CT intensity and Xer12m (P = .019). Textural IBM changes (refer to Supplementary material 4 [available online at www.redjournal.org] for textural IBM details) were also tested in the present cohort, as described in a previous study (5). Univariable analysis showed that some textural IBMs were significantly associated with Xer12m; however, none gave a significant addition to ΔPG-surface. Textural IBM changes may yield similar information as ΔPG-surface or be biased owing to the presence of metal artefacts in some patients. These IBMs were not extensively discussed in this study, because they gave no conclusive rejectable results, owing to the above discussed limitations.

Furthermore, because the final models presented in our report may be susceptible to limitations of the chosen

<table>
<thead>
<tr>
<th>$R^2$</th>
<th>DS</th>
<th>$\chi^2 (P)$ (apparent)</th>
<th>Calibration intercept slope, corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apparent</td>
<td>Corrected</td>
<td>Apparent</td>
<td>Corrected</td>
</tr>
<tr>
<td>0.28</td>
<td>0.25*</td>
<td>0.21</td>
<td>0.20*</td>
</tr>
<tr>
<td>0.25</td>
<td>0.16</td>
<td>0.18</td>
<td>0.14</td>
</tr>
<tr>
<td>0.36</td>
<td>0.25</td>
<td>0.29</td>
<td>0.23</td>
</tr>
<tr>
<td>0.38</td>
<td>0.26</td>
<td>0.30</td>
<td>0.24</td>
</tr>
<tr>
<td>0.45</td>
<td>0.34</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>0.56</td>
<td>0.44</td>
<td>0.45</td>
<td>0.39</td>
</tr>
</tbody>
</table>
variable selection procedure, LASSO regularization, which is an alternative variable selection approach, was additionally performed and resulted in very comparable variable selection frequencies (ΔPG-surface was the most selected ΔIBM: 49% of the bootstrapped samples). This suggests a relatively large robustness of the associations found in this dataset, independent of the method of analysis. Additionally, modalities that provide functional information, such as positron emission tomography and magnetic resonance imaging, could contribute in determining functionality loss of the PG gland and could further improve quantifying and understanding the development of xerostomia.

**Conclusion**

Parotid gland surface reduction between start and 6 weeks after radiation therapy (ΔPG-surface) was significantly associated with the development of xerostomia 6 to 12 months after completing radiation therapy. Mean PG dose significantly correlated with ΔPG-surface and did not add information to the ΔPG-surface model in predicting late xerostomia in this cohort. The model with ΔPG-surface and acute xerostomia early after radiation therapy significantly improved model performance to predict late xerostomia (AUC 0.90, 95% CI 0.84-0.96; AUCbootstrapped 0.86) and can therefore be a good predictor of late xerostomia.

**Fig. 2.** Examples of parotid glands (PG) with a large negative ΔPG-surface (and ΔPG-volume) between the start of radiation therapy (green) and 6 weeks after radiation therapy (red).

**Fig. 3.** The relation dose and ΔPG-surface for both (A) contralateral and (B) ipsilateral parotid gland (PG). Linear (black line) and quadratic (pink curve) regression curves were plotted.
candidate surrogate marker for late xerostomia at subsequent time points.

References


