CLINICAL INVESTIGATION

GRADING XEROSTOMIA BY PHYSICIANS OR BY PATIENTS AFTER INTENSITY-MODULATED RADIOTHERAPY OF HEAD-AND-NECK CANCER

AMICHAY MEIROVITZ, M.D.,* CAROL ANNE MURDOCH-KINCH, D.D.S., Ph.D.,‡ MATHEW SCHIPPER, Ph.D.,*† CHARLIE PAN, M.D.,* AND AVRAHAM EISBRUCH, M.D.*

Departments of *Radiation Oncology, and of †Biostatistics, and ‡School of Dentistry, University of Michigan, Ann Arbor, MI

Purpose: To assess observer-based vs. patient self-reported scoring of xerostomia after intensity-modulated radiotherapy (IMRT) of head-and-neck (HN) cancer.

Methods: A total of 38 patients who had received IMRT for HN cancer underwent xerostomia evaluations 6 to 24 months after completion of therapy using three methods each time: (1) Grading by 3 observers according to the Radiotherapy Oncology Group/European Organization for Research and Therapy of Cancer (RTOG/EORTC) system; (2) patient self-reported validated xerostomia questionnaire (XQ); and (3) major salivary gland flow measurements.

Results: The interobserver agreement regarding the RTOG/EORTC grades was moderate: /H9260-coefficient 0.54 (95% CI = 0.31–0.76). The correlations between the average RTOG/EORTC grades and the salivary flow rates were not statistically significant. A trend for significant correlation was observed between these grades and the percent (relative to the pretherapy) nonstimulated salivary flow rates (p = 0.07), but not with the percent stimulated flow rates. Better correlations were found between grading made more than the median time (15 min) after the last liquid sipping and the nonstimulated (but not the stimulated) flows compared with grading made shortly after sipping. In contrast, significant correlations were found between the XQ scores and the nonstimulated (p < 0.005) and the stimulated (p < 0.005) salivary flow rates, as well as with the percentages of the corresponding pretherapy values (p = 0.002 and 0.038, respectively). No significant correlation was found between the RTOG/EORTC grades and the XQ scores. The observer-based grades underestimated the severity of xerostomia compared with the patient self-reported scores.

Conclusions: Patient self-reported, rather than physician-assessed scores, should be the main end points in evaluating xerostomia. © 2006 Elsevier Inc.

INTRODUCTION

Various strategies have been used in recent years to reduce xerostomia after irradiation of head-and-neck cancer, including radiation protectors, salivary stimulants, surgical transfer of a submandibular salivary gland away from the radiation fields, and intensity-modulated radiotherapy (IMRT) aimed at the sparing of the parotid salivary glands (1). To evaluate and compare the merits of each of these strategies, it is necessary to score reliably and reproducibly the degree of the resulting post-therapy xerostomia. To this end, several methods have been used: observer-based grading, patient self-reported scoring, and objective measurements of the production of saliva such as salivary flow rate measurements or salivary gland scintigraphy. It is not yet clear which method reflects the “true” state of postradiation xerostomia, and few studies examined the correlations among these methods, yielding conflicting results (2).

Our PubMed search of articles with key words included “xerostomia” and “radiotherapy” published between 2000 and 2005, in which xerostomia evaluation was a major end point and which contained an observer-based or patient-reported evaluation tool (or both), yielded a total of 51 studies. Of these studies, 24 relied on an observer-rated evaluation (3–26), 21 relied on patient self-reported scores (27–47), and 6 included both observer-rated and patient self-reported evaluations (48–53). Of the 51 studies, 18 included an objective measure of the salivary output in addition to the patient self-reported or observer-based scores (21, 24, 29–34, 37–38, 41–42, 44, 47–48, 50–52).
Almost all of the studies that relied on an observer-based xerostomia evaluation used the Radiotherapy Oncology Group/European Organization for Research and Therapy of Cancer (RTOG/EORTC) grading scale. This scale ranks chronic xerostomia according to the degree of mouth dryness and the response to stimulus. Investigators using the RTOG/EORTC xerostomia scoring system have rarely detailed how scoring has been done, nor are details available in the original RTOG/EORTC grading publication (54). Specifically, the “response to stimulus” has not been defined, and no formal validation of this grading system has been reported.

As xerostomia is primarily a symptom, patient self-reporting may be meaningful in assessing its severity. To facilitate patient self-reporting, a xerostomia questionnaire (XQ) has been developed at the University of Michigan (32). It consists of eight questions, divided into four questions related to dryness while eating/talking and four related to dryness at rest (Table 1). The XQ was found to be reliable, valid, and reproducible in measuring patient-reported xerostomia (32, 34, 37). It has been independently validated by investigators at the University of Florida, who found that it distinguished accurately between patient groups according to their parotid gland doses (40).

To evaluate the adequacy of the observer-rated RTOG/EORTC late xerostomia scoring, we assessed its reproducibility among several observers and its correlation with the patient self-reported XQ and with the major salivary glands flow rates.

### METHODS AND MATERIALS

This was a prospective cross-sectional study approved by the Institutional Review Board of the University of Michigan. The study population consisted of patients with head-and-neck cancer who had been treated with IMRT at the University of Michigan and had participated in a separate longitudinal study of quality of life. Treatment included forward or inverse-planned IMRT aimed at the sparing of the parotid salivary glands while adequately treating the targets with homogeneous dose distributions, according to previously published methods (55, 56).

At follow-up periods ranging from 6 to 24 months after the completion of therapy, patients were assessed using the observer-rated RTOG/EORTC scoring system for late xerostomia, the patient-reported XQ instrument, and the salivary flow rates from the major salivary glands.

The RTOG/EORTC late xerostomia grade was determined during follow-up visits at the clinic by 3 physicians at each visit. Each of the observers assigned the score independently and was blinded to the scores given by the others. The scores ranged from 0 to 3 (0 = none; 1 = slight dryness of the mouth, good response on stimulation; 2 = moderate dryness, poor response on stimulation; and 3 = complete dryness, no response on stimulation). The observers were instructed to determine the scores based on an assessment of the patients’ oral mucosa moistness and based on asking patients how dry they were, how often they had to sip liquids, whether they needed salivary stimulation by any means (such as chewing gum), and whether stimulation helped. A mean RTOG/EORTC score was calculated from the scores provided by the 3 observers and was used for correlations with the patient-reported scores and with the salivary flow rates. At each observer-based evaluation, patients were also asked how long ago they had last sipped liquids.

The patient self-reported XQ instrument is detailed in Table 1. The questions are equally divided into four items asking about dryness while eating or chewing, and four items about dryness while not eating. Subjects rated each symptom on an 11-point ordinal Likert scale from 0 to 10, with higher scores indicating greater dryness or discomfort because of dryness. Each item score was added, and the sum was transformed linearly to produce the final summary score ranging between 0 and 100, with higher scores representing greater levels of xerostomia as detailed elsewhere (32).

Measurements of the flow rates from the major salivary glands were made according to previously reported methods (57). In brief, parotid saliva, unstimulated and stimulated, was collected separately from each parotid gland duct orifice with a collecting cup attached to the mucosa surrounding the orifice. Patients did not eat or drink for 2 h before collection. The unstimulated saliva was collected for 2 min, after which the salivary flow was stimulated by swabbing 2% citric acid on the dorsolateral surfaces of the tongue at 30-s intervals for 2 min. This was followed by a 2-min collection period during which gustatory stimulation was maintained. After collection, the volume of saliva was gravimetrically determined, assuming a specific gravity of 1.0, and the flow rate (ml/min) was recorded. The submandibular and sublingual saliva was collected from Wharton’s duct orifices with a micropipet attached to gentle suction. Unstimulated and stimulated saliva volumes were determined by methods identical to those of the parotid gland. The total unstimulated and stimulated salivary flow rates were calculated by adding the respective contributions from all of the glands. The percentage salivary flow rates were calculated relative to the pretherapy baseline flow rates, which were available from a longitudinal quality of life study in which these patients had participated.

We previously reported that both the salivary flow rates and the XQ scores improved continuously throughout the first 2 years after IMRT (32, 34, 37). Therefore the correlations were tested only among the observer-rated grades, the XQ scores, and the saliva output measurements (or their percentage relative to baseline), which were made at about the same time point (i.e., within 1 month).

### Statistical analysis

The agreement between the RTOG grades among the 3 observers at each observation time point was assessed using the weighted κ-coefficient. Spearman rank-based correlations were computed.

---

**Table 1. The xerostomia questionnaire (XQ)**

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Rate your difficulty in talking due to dryness</td>
<td>0 to 10</td>
</tr>
<tr>
<td>2. Rate your difficulty in chewing due to dryness</td>
<td>0 to 10</td>
</tr>
<tr>
<td>3. Rate your difficulty in swallowing solid food due to dryness</td>
<td>0 to 10</td>
</tr>
<tr>
<td>4. Rate the frequency of your sleeping problems due to dryness</td>
<td>0 to 10</td>
</tr>
<tr>
<td>5. Rate your mouth or throat dryness when eating food</td>
<td>0 to 10</td>
</tr>
<tr>
<td>6. Rate your mouth or throat dryness while not eating</td>
<td>0 to 10</td>
</tr>
<tr>
<td>7. Rate the frequency of sipping liquids to aid swallowing food</td>
<td>0 to 10</td>
</tr>
<tr>
<td>8. Rate the frequency of sipping liquids for oral comfort when not eating</td>
<td>0 to 10</td>
</tr>
</tbody>
</table>

Patients rate each item on a scale from 0 to 10. Higher score denotes worse xerostomia.
to measure correlation between the different grading systems and between each grading system and the salivary flow rates at each time point and their percentages of the pretherapy baseline flow rates. Basing the correlations on ranks allowed the correlation of variables that are measured on different scales. The usual methods for testing whether the correlation is significantly different from zero are not valid, as they assume that all pairs of observations are independent, whereas in our data there were several pairs of observations from each subject. We therefore used a bootstrap method to assess the significance of the correlations. The bootstrap method provides confidence intervals that do not require independence of the observations. It consists of re-sampling sets of data multiple times such that the distribution of a random sample can guide to the distribution in the parent population (58). Statistical significance was determined using two-tailed tests at the level of \( p < 0.05 \).

**RESULTS**

A total of 38 patients participated in this study from November 2001 through October 2003. They included 30 men and 8 women with primary tumor sites in the oropharynx (26 patients), oral cavity (9 patients), and 1 patient each with larynx, hypopharynx, and unknown primary cancer. The American Joint Committee on Cancer tumor stages were stage I in 2 patients, II in 4 patients, III in 6 patients, IVA in 24 patients, and IVB in 2 patients. A total of 20 patients were treated with definitive radiotherapy (1 patient had concurrent chemotherapy) and 18 received postoperative radiotherapy (1 patient had concurrent chemotherapy). In all patients the targets involved the bilateral neck. The mean doses to the contralateral and ipsilateral parotid glands were on average 22 Gy (SD 5 Gy) and 53 Gy (SD 7 Gy), respectively, and the mean doses to the contralateral and ipsilateral submandibular glands were on average 57 Gy (SD 8 Gy) and 65 Gy (SD 7 Gy), respectively.

Xerostomia evaluations consisted of observer-rated grades, patient self-reported XQ scores, and salivary flow rate measurement, performed within 1 month. A total of 64 evaluations were made (range, 1 to 3 per patient; median, 2), at 6 to 24 months (median, 12 months) after the completion of therapy. The results of the evaluations are detailed in Table 2.

Assessment of the agreement among the 3 observer-rated scores taken at each clinic visit showed that the \( \kappa \)-coefficient, measuring inter-rated agreement, was 0.54 (95% CI = 0.31–0.76), where 1 indicates perfect agreement and 0 indicates only chance agreement. The correlations between the XQ scores, the observer-rated scores, and the salivary flow rates and their percentage relative to pretherapy are detailed in Table 3. Statistically significant correlations were found between the XQ scores and the nonstimulated and the stimulated salivary flow rates, with Spearman correlation coefficients of \(-0.59\) and \(-0.38\), respectively. Statistically significant correlations, with somewhat lower correlation coefficients, were also observed between the XQ scores and the percent (of the corresponding pretherapy baseline) stimulated and nonstimulated salivary flow rates. In contrast, the correlations between the RTOG/EORTC scores and the stimulated or the unstimulated salivary flow rates were much lower and were not statistically significant. A trend was noted for a correlation between the RTOG/EORTC grades and the percent nonstimulated saliva (\( p = 0.07 \)) but not with the percent stimulated saliva.

The median time interval between the observer-based RTOG/EORTC grading and the most recent liquid sipping, reported by the patients at the time of grading, was 15 min (range, 0–360 min). There was no statistically significant difference in the observer-rated grades made \( \leq 15 \) min or \( >15 \) min after sipping liquids. However, different patterns were observed in the correlations between the grades and the salivary flow rates in the two subgroups (Table 3). The correlations between the grades and the nonstimulated saliva flow rates, and their percentages relative to baseline, were found to be statistically significant when patients sipped liquids \( >15 \) min before grading (\( p = 0.025 \) and 0.004, respectively) but not when patients sipped \( \leq 15 \) min. The correlations between the observer-rated grades and the stimulated saliva, or its percentage of baseline, remained nonsignificant in either subgroup.

No significant correlations were observed between the XQ scores and the RTOG/EORTC grades in all patients (Table 3), nor between the XQ scores and the observer-rated grades in each of the two subgroups separated according to the time since the last liquid sipping. Examination of the relationship between the XQ scores and the RTOG/EORTC grades at each time point (Fig. 1)

---

**Table 2. Summary of the RTOG/EORTC grades, xerostomia questionnaire (XQ) scores, major salivary gland (parotid and submandibular) flow rates, and their percentage of the pretherapy flow rates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTOG/EORTC score (average of 3 observers)</td>
<td>0.34</td>
<td>0.48</td>
<td>0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>XQ score</td>
<td>37.3</td>
<td>24.4</td>
<td>35</td>
<td>0</td>
<td>86</td>
</tr>
<tr>
<td>Stimulated saliva flow rate (ml/min)</td>
<td>0.55</td>
<td>0.27</td>
<td>0.57</td>
<td>0.01</td>
<td>2.42</td>
</tr>
<tr>
<td>Unstimulated saliva flow rate (ml/min)</td>
<td>0.10</td>
<td>0.16</td>
<td>0.13</td>
<td>0</td>
<td>0.96</td>
</tr>
<tr>
<td>% Stimulated</td>
<td>40</td>
<td>32</td>
<td>30</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>% Unstimulated</td>
<td>32</td>
<td>27</td>
<td>18</td>
<td>0</td>
<td>243</td>
</tr>
</tbody>
</table>

*Abbreviations: RTOG/EORTC = Radiation Therapy Oncology Group/European Organization for Research and Treatment of Cancer; SD = standard deviation.*
revealed that the observers tended to underestimate the severity of xerostomia compared with the patient-reported scores. The physicians rarely used the worst grades of the scale: the RTOG/EORTC grade was >1 only at five time points, and the grade distribution was skewed toward the lowest grades (median, Grade 0). In comparison, the distribution of the XQ scores was relatively even around a median of 35. Of the four evaluations in which patients reported the worst xerostomia (XQ score ≥80, where the worst possible score is 100), only one was assigned by the observers as an RTOG/EORTC xerostomia grade >1. Choosing as cut-off points the grades that represent the middle of the scales (scales of 0–3 for the RTOG/EORTC and 0–100 for the XQ), only three (5%) of the average RTOG/EORTC grades were worse (higher) than 1.5. In comparison, 22 (34%) of the XQ scores were worse (higher) than 50 (p = 0.0001). The underestimation of the severity of xerostomia by the observers vs. that reported by the patients was similar in the two observer-based grading subgroups (separated according to the time since the last liquid sipping).

### DISCUSSION

The main findings of this study are moderate agreement in the RTOG/EORTC scores among various observers, low correlation with the salivary output, and an underestimation by the observers of the severity of xerostomia compared with that reported by the patients. The lack of robust agreement is likely caused by the inherent uncertainties faced by the observers when they are required to assess the severity of someone else’s symptoms.

How exactly should the observer-based grades be determined? Despite an extensive use of the RTOG/EORTC late xerostomia scale, only a few papers provided details about the exact way that the observer-based evaluations were made and how grades were assigned. These papers reported a wide range of methods. Lauve et al. stated that “the treating physician assessed the degree of xerostomia by recording an oral moisture score . . . the oral moisture score was based on visible oral moisture and scored according to the objective measures of the RTOG late effects on normal tissue scales” (50). In contrast, Jen et al. stated that “the grading was assessed by subjective complaints by the patients during interview” (16). Sultanem et al. combined both approaches: their assignment of RTOG xerostomia grades was “based on subjective evaluation of patient symptoms and clinical examination” (25). This process was similar to that used in the current study, in which the observers took into account both patients’ mucosal moisture and their symptoms. As demonstrated in our study these evaluations are limited by poor reproducibility, likely caused by different thresholds set by different observers in defining “mild” or “moderate” dryness, or in assessing whether “response to stimulus” did or did not exist.

The finding that the observer-based grades in patients who sipped liquids >15 min before scoring correlated better with the nonstimulated salivary output than the grades assigned when patients sipped liquids closer to the time of assessment was unexpected. It is possible that sipping liquids shortly before assessment increased oral mucosal moisture and “improved” observers’ assessment of the severity of xerostomia at the moment, reducing the correlations between the xerostomia grades and the salivary output. However no significant differences were detected in the observer-rated xerostomia scores between the two subsets.
of assessments or in their lack of correlation with the XQ scores, possibly because of relatively small patient numbers in the subgroups. These findings require independent validation. They suggest that observer-rated grading of xerostomia is inadequate when patients sip liquids shortly before assessment, and motivate standardization of the observer-based evaluations to address this issue. Furthermore when observer-based evaluations were made long after sipping liquids, high and statistically significant correlations were found between the scores and the nonstimulated, but not the stimulated, salivary output. The lack of significant correlation with the percent stimulated output cannot be explained just by small patient numbers, as the correlation coefficient \( R \) was positive (negative \( R \) was expected as higher RTOG/EORTC scores denoted worse xerostomia) (Table 3). The contrast with the XQ scores in the correlations with the stimulated saliva may be explained by the inclusion in the XQ of questions about dryness while eating, which are relevant to the stimulated salivary output.

We have found only two other studies that examined the correlations between observer-based xerostomia grades and the salivary output or patient-reported xerostomia. Brizel et al. reported that the correlation between RTOG/EORTC grade \( \geq 2 \) xerostomia and the unstimulated whole-mouth salivary production was statistically significant; however the correlation coefficient was relatively low (\( r = 0.31 \)) (48). A similar correlation was found in our study only for patients who sipped liquids \( > 15 \) min before evaluation. Haddad et al. used an observer-based “objective” grading of xerostomia which graded the degree of mucosal moisture (15). They found a statistically significant correlation between the scores according to this scale and the scores of a patient-reported questionnaire. Similar correlation was not found in our study, even though the evaluation of the degree of mucosal moisture was included as an important part of our observer-based grading.

Compared with the RTOG/EORTC system, a step forward in forming an observer-based evaluation of xerostomia has been made in the LENT-SOMA (Late Effects on Normal Tissue–Subjective, Objective, Management, Analytic) scoring system (59). In this system the observer scores separately patients’ reported degree of dryness, oral moisture, and the frequency and efficacy of water and saliva substitutes or sialogogues. In addition, the salivary flow rates have been assigned as objective measure of xerostomia, with arbitrary values assigned for each grade. This evaluation process is more detailed and specific than the RTOG/EORTC system, potentially improving its reproducibility and validity. One study reported that its grades correlated well with the patients’ self-reported xerostomia (51). However, it remains a predominantly an observer-based system of patient symptom severity, prone to similar difficulties inherent in the RTOG/EORTC system. The LENT-SOMA system has thus far been infrequently used: only two studies (13, 51) among the 51 studies of xerostomia and radiotherapy published between 2000 and 2005 and cited in the Introduction used it.

Most recently, the observer-rated Common Toxicity Criteria (CTC), version 3.0, has been developed by the National Cancer Institute (60). It represents a comprehensive, multimodality grading system for reporting the adverse effects of cancer therapy. Notably, it does not distinguish between acute and late side effects. The dry mouth/salivary gland (xerostomia) adverse event in CTC 3.0 includes three grades according to symptoms and salivary flow rates. The symptoms are as follows: Grade 1 = symptomatic (dry or thick saliva) without significant dietary alterations; Grade 2 = symptomatic and significant oral intake alteration (e.g., copious water, other lubricants, diet limited to purees and/or soft, moist foods); Grade 3 = symptoms leading to inability to achieve adequate alimentation by the oral route. The CTC version 3.0 is intended to replace the RTOG grading system, and its use to evaluate xerostomia has not yet been reported. Earlier versions of the CTC have been used by few investigators of xerostomia (13–14, 26). Had we used the CTC 3.0 criteria retrospectively for our patients, they would have been assigned the status of either no xerostomia or Grade 1 xerostomia (none of our patients reported significant oral intake alterations caused by xerostomia). This would have narrowed the range of the grades and would have likely resulted in reduced sensitivity to modest variations in the severity of xerostomia, leaving potentially important variations in its severity unrecorded. We therefore recommend that the CTC, if used, be supplemented by a patient self-reported xerostomia instrument.

Significant correlations were observed in this study between the XQ scores and both the stimulated and the nonstimulated salivary flow rates and the percentages of their respective baselines. Although statistically significant, the correlation coefficients were modest, meaning that a substantial variability in the scores could not be explained by the salivary flow rates alone. This was also the case in other studies that reported statistically significant correlations between patient-reported xerostomia scores and the salivary output (28, 31, 41, 42, 48). On the other hand, some studies found no similar significant correlations (61, 62). In addition several studies of interventions aiming at improving xerostomia reported dichotomous results regarding patient-reported symptoms vs. salivary output: in each of two randomized studies of pilocarpine after RT, patient-reported xerostomia improved significantly in the pilocarpine arm but there was little and inconsistent improvement in the salivary output compared with placebo (63, 64). Conversely, in a randomized study of pilocarpine concurrent with RT, significant improvement of the salivary flow but no improvement in patient-reported symptoms were found in the treatment vs. the placebo groups (33). Similarly, a randomized study of amifostine concurrent with radiotherapy (RT) showed significant and substantial salivary flow preservation in the treatment group; however the improvement in patient-reported symptoms, although statistically significant, was clinically trivial (−0.6 on a scale of 0–10) (48). A recently presented randomized study of IMRT vs. conventional RT for nasopharyngeal cancer re-
ported significantly higher salivary flows in the IMRT group, but there was no difference between the groups in patient-reported symptoms (65).

The discrepancies in these results may be explained by several factors. One is the importance of the minor salivary glands, the output of which contains large part of the salivary mucins, in the perception of xerostomia by the patient (2). Because of their low volume (only 10% of the total saliva), their true contribution is unlikely to be accounted for in whole-mouth salivary volume measurements, used by most studies, and has not been taken into account in studies such as the current one that have measured only the major salivary gland output. Direct measurements of the minor salivary gland output, either their volume or mucin content, are expected to improve the correlations between the salivary output and patient-reported xerostomia. In a previous study we used the mean dose to the oral cavity as a surrogate to the loss of function of the minor salivary glands, because a direct measure of their output was not available (32). In that study, the predictors of xerostomia after IMRT in 132 patients were assessed and statistically significant correlations were found between the XQ scores and the salivary flow rates from the major salivary glands (correlation coefficients ranging from −0.27 to −0.41 at various time points). However, after controlling for the radiation doses to the major salivary glands, the salivary flow rates were not independently predictive of the XQ scores. Rather, the doses to the major salivary glands, oral cavity mean dose, the time since therapy, and the baseline pretherapy scores, were found to be independent predictors (32). The current study contained much smaller numbers of patients and evaluations per patient, and a detailed analysis of predictors was outside its scope. An additional issue is the importance of the submandibular major salivary glands, whose saliva is more mucinous than the parotid gland secretions. In IMRT studies the submandibular glands cannot usually be spared substantially because of their close proximity to level II neck nodes, which are typically included in the targets when both sides of the neck are at risk. Thus the parotid glands alone are spared, contributing to almost all the output measured by either selective salivary gland or whole mouth saliva measurements. Sparing just the parotid glands by IMRT results in improvements mainly in the production of serous saliva, the effect of which on the patient symptoms may not be proportional to its measured amount.

Another confounding issue is the adequacy of the patient-reported xerostomia instrument. Various xerostomia questionnaires and other instruments have been used by different investigators. They include Likert-type questions similar to the one that we have used, and visual analog scales (VAS) in which the patient draws a line the length of which represents the severity of the symptom. Some of these instruments have been tested for reproducibility and accuracy in the same way QOL instruments are typically validated, while the validity of other instruments, including individual questions about xerostomia selected out from larger validated QOL instruments, have not been tested. We do not know which instrument is most adequate. Fox et al. reported no correlation between patient-reported xerostomia scores and saliva, however the instrument they used contained only yes/no possible answers, which may have reduced its sensitivity (61). Wasserman et al. detailed the correlations between whole-mouth unstimulated saliva and patient-benefit questionnaire in a randomized study of amifostine during radiotherapy (53). Although the mean overall scores were not significantly different between the treatment or control groups at several time points, patients in the amifostine group reported significant improvement in few questions (“dry mouth,” “use of oral comfort aids”) but not in others (impaired speaking, taste, swallowing). It is possible that only some of the questions in that instrument adequately reflected patient-perceived xerostomia. Several studies that reported correlations between patient self-reported questionnaire scores and the salivary output or the salivary gland doses consisted of two or more distinct patient populations that differed widely in the salivary gland radiation doses and in the salivary output. The studies from Washington University and the University of Florida consisted of patients who had received either IMRT or non-parotid-sparing conformal RT (31, 40), and our previous study consisted of patients who had received either bilateral neck IMRT or unilateral neck RT (32). The patient-reported instruments tested in these studies likely discriminated well among patient populations that had received very different parotid gland doses and widely divergent salivary flow rates. In comparison, the current study consisted of patients who all received bilateral neck, parotid-sparing IMRT, and in whom the ranges of the parotid gland doses and salivary output were relatively narrow. The significant correlations of the XQ scores with the salivary flow rates in these patients attest to the sensitivity of this instrument.

Another aspect of patient-reported xerostomia is our previously reported findings that the XQ scores are significantly correlated with general head-and-neck cancer–related QOL after IMRT (37), and that improvements in XQ scores over time parallel improvements in QOL after IMRT, although neither improve over time after conventional radiotherapy (34). These results suggest that the potential benefit of IMRT are broader than just reduction of xerostomia. This issue requires further study, notwithstanding the results of a recent randomized study of early nasopharyngeal cancer treated with IMRT or conventional radiotherapy, which did not find a significant difference in QOL between the arms despite higher saliva output in the IMRT arm (65).

**CONCLUSION**

In conclusion, our findings that the observers underestimated the severity of xerostomia compared with the patient self-reported scores, and that observer-based xerostomia grading is only moderately reproducible and lacks a high correlation with the salivary output (especially if patients sipped liquids shortly before evaluation), highlight the need for cau-
tion in interpreting the results of studies in which observer-based grading was used as the main evaluator of xerostomia. In this respect, xerostomia is similar to other domains of QOL in which patient self-reported scores seem to be more reliable in predicting the outcome of head-and-neck cancer therapy compared with physician-based assessments (66). Patient self-reported measures should therefore be used whenever the assessment of xerostomia is a major study goal.

REFERENCES


