Interobserver variation in parotid gland delineation: a study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature

S W LOO, FRCR, W M C MARTIN, FRCR, P SMITH, BSc, S CHERIAN, FRCR and T W ROQUES, FRCR

Objectives: This study evaluates the interobserver variation in parotid gland delineation and its impact on intensity-modulated radiotherapy (IMRT) solutions.

Methods: The CT volumetric data sets of 10 patients with oropharyngeal squamous cell carcinoma who had been treated with parotid-sparing IMRT were used. Four radiation oncologists and three radiologists delineated the parotid gland that had been spared using IMRT. The dose–volume histogram (DVH) for each study contour was calculated using the IMRT plan actually delivered for that patient. This was compared with the original DVH obtained when the plan was used clinically.

Results: 70 study contours were analysed. The mean parotid dose achieved during the actual treatment was within 10% of 24 Gy for all cases. Using the study contours, the mean parotid dose obtained was within 10% of 24 Gy for only 53% of volumes by radiation oncologists and 55% of volumes by radiologists. The parotid DVHs of 46% of the study contours were sufficiently different from those used clinically, such that a different IMRT plan would have been produced.

Conclusion: Interobserver variation in parotid gland delineation is significant. Further studies are required to determine ways of improving the interobserver consistency in parotid gland definition.

Permanent xerostomia is one of the most prevalent and debilitating long-term adverse effects of radiotherapy for head and neck squamous cell carcinoma (HNSCC) [1, 2]. It has a negative impact on patients' quality of life and oral health, and can lead to difficulties in chewing and swallowing [3–5]. It can also affect speech and taste, and predisposes these patients to dental caries, oral infections, mucosal ulcerations and osteoradionecrosis of the mandible [6]. The parotid glands are the largest of the salivary glands. In the stimulated state, they contribute more than two-thirds of the total salivary output. They are situated close to the Level II cervical lymph nodes, parapharyngeal space, tonsillar fossae and soft palate, and are likely to receive a significant dose when oropharyngeal cancers are treated with radiotherapy. Salivary flow from the parotid is affected by the radiation dose received and the volume of the gland irradiated. Several parameters of the parotid dose–volume–response relationship have been investigated. The one that seems to correlate best with long-term saliva production is the mean dose to the parotid [7–10]. An accepted target is to keep the mean dose below 24 Gy to preserve unstimulated salivary flow.

Intensity-modulated radiotherapy (IMRT) delivers highly conformal radiation to the planning target volumes (PTVs), while sparing adjacent uninvolved organs at risk (OARs) such as the parotid glands. Prospective randomised trials and non-randomised clinical studies have shown IMRT to be superior to conventional two-dimensional radiotherapy in the preservation of long-term parotid function [11, 12]. As a result, parotid-sparing IMRT has become the standard technique for delivering radiotherapy for oropharyngeal cancer.

Accurate delineation of target volumes and OARs is essential for the success of IMRT. Interobserver variation in gross tumour volume (GTV) definition has been shown to be large and clinically significant for many tumour types, including HNSCC [13–17]. There is variation between individuals and groups such as oncologists and radiologists. Variation in parotid gland delineation can potentially offset the benefits of parotid-sparing IMRT.

The objective of this study is to evaluate the interobserver variation in parotid gland delineation and to determine its impact on IMRT solutions.

Methods and materials

Patient selection

The CT volumetric data sets of 10 patients with Stage IV squamous cell carcinoma of the oropharynx were used in this study. All patients had N2a/b disease and were deemed at risk of harbouring microscopic disease in the contralateral neck. All had been treated with IMRT to a dose of 65 Gy in 30 fractions to the GTV and high-risk clinical target volume (CTV), and 54 Gy in 30 fractions to the lower-risk uninvolved nodal regions. The constraint for the parotid gland contralateral to the GTV was a mean dose of <24 Gy.
These data sets were selected from our database on the basis that the dose constraints to the PTVs and the spinal cord planning OAR volume (PRV) had been met, and the mean dose to the spared parotid gland was within 10% of the target of 24 Gy. In these cases, a small change in the parotid contour may potentially result in a significant difference to the plan accepted for treatment.

Imaging protocol

Planning images were obtained using a high-speed single-slice spiral CT scanner (GE Healthcare, Waukesha, WI). They were acquired with a 3-mm slice thickness reconstructed every 3 mm and subsequently transferred via DICOM (NEMA, Rosslyn, VA) to the Eclipse (Varian, Palo Alto, CA) radiotherapy treatment planning system. All data were made anonymous and stored under separate profiles.

Volume delineation

Four consultant radiation oncologists and three consultant radiologists with 2–15 years of experience in head and neck oncology participated in the study. A tutorial was provided for those participants who were unfamiliar with the treatment planning system. For each patient, participants were given relevant clinical information. All diagnostic radiological investigations, including contrast-enhanced MRI, were also made available electronically via PACS (Picture Archiving and Communications System). Participants were allowed access to an atlas of head and neck anatomy and imaging. They were asked to independently delineate the parotid gland that had been spared with IMRT on the axial CT images. This was done without reference to the PTV contours. The window levels for all CT images could be altered during volume delineation.

Data and statistical analysis

The dose distribution of the actual IMRT treatment plan was superimposed on each study contour to allow their DVHs to be derived. These were compared to the original DVH obtained when the plan was used clinically. The interobserver variation in the parotid contours was examined both quantitatively and qualitatively for each patient. The Jaccard coefficient is defined as the ratio of the volume of intersection or overlap to the encompassing volume for any two parotid contours (Figure 1). The conformity index was then calculated for each patient by averaging the Jaccard coefficient between all possible pairs of contours for that patient [18]. The conformity index could therefore range between zero (no overlap between contours) and one (identical contours), and was determined separately for radiation oncologists and radiologists.

Search strategy used for systematic review

A literature search was performed in PubMed using the keywords “parotid”, “radiotherapy”, and “variation” or “variability”. Articles published in English between 1990 and 2010 were reviewed.

Results

A total of 70 study contours were available for analysis. The mean conformity index was 0.66 (range 0.46–0.73) among radiologists and 0.52 (range 0.26–0.61) among oncologists. The mean parotid dose achieved during the actual treatment ranged from 22.1 to 25.2 Gy (median 23 Gy) and were all within 10% of 24 Gy. Using the study contours, the mean parotid dose obtained was within 10% of 24 Gy for only 53% of volumes by radiation oncologists and 55% of volumes by radiologists. Mean dose was within 20% of 24 Gy for 80 and 90% of the volumes, respectively. Overall, only 53% of all study contours achieved a mean parotid dose within 10% of 24 Gy, and 20% had mean doses >26.4 Gy (10% above 24 Gy). The mean doses obtained and the volumes of the “actual” treatment plans and all study parotid contours are shown in Tables 1 and 2, respectively. Figures 2 and 3 demonstrates the mean doses obtained by radiation oncologists and radiologists, respectively, for all 10 data sets.

When the study contours were analysed qualitatively, the greatest degree of variation consistently appeared in the medial edge of the deep lobe, the anterior border of the superficial lobe, and the superior and inferior margins of the parotid gland. Figure 4 is an axial CT image showing the study contours of all seven participants for one parotid gland.

Discussion and literature review

Radiotherapy is the main non-surgical treatment for patients with HNSCC. One of the commonest long-term adverse effects following such treatment is xerostomia. IMRT results in better dose conformation to the PTV and improved sparing of adjacent uninvolved normal tissues compared with conventional radiotherapy. Numerous studies showed a reduction in the incidence and severity of xerostomia following parotid-sparing IMRT [19–30], with post-treatment recovery of saliva production and resultant improvement in patient-reported quality of life [31]. Three randomised controlled trials comparing parotid-sparing IMRT and conventional radiotherapy were reported [12, 32, 33]. The PARSORT study [32] was a multicentre trial which recruited 94 patients with oropharyngeal or hypopharyngeal squamous cell carcinoma. Treatment consisted of either parotid-sparing
IMRT or CT-planned three-dimensional conformal radiotherapy using parallel-opposed lateral fields, with no concurrent chemotherapy. One of the planning constraints for parotid-sparing IMRT was a mean dose of \( <24 \) Gy to the entire contralateral parotid gland. In this study, IMRT resulted in a lower mean dose to the contralateral parotid, and this translated into a reduced incidence of Grade 2 or worse xerostomia at 12 and 24 months following treatment completion, assessed using the LENT SOMA (Late Effects of Normal Tissues Subjective–Objective Management Analytic) and the Radiation Therapy Oncology Group (RTOG) scoring systems. The IMRT arm also contained a higher proportion of patients with both stimulated and unstimulated saliva flow from the contralateral parotid post treatment. Patient-reported outcomes in favour of IMRT were observed. As a result, parotid-sparing IMRT has now become the current treatment of choice for patients with head and neck cancer who are likely to receive significant radiation doses to both parotid glands with conventional radiotherapy [34].

Accurate delineation of target volumes and OARs is critical to the success of IMRT. Its inverse planning process involves defining PTVs and OARs, stipulating constraints on the DVHs of each of these structures and optimising the treatment plan to meet these constraints with an iterative process. It creates high dose gradients between the PTVs and OARs. Interobserver variation in the delineation of target volumes and OARs may thus potentially result in underdosing of tumours, overdosing of OARs or both. For a relatively small structure such as the parotid gland with a steep dose gradient across, a slight change in its contour can lead to a significant alteration in its DVH and the ability of the treatment plan to meet the stipulated constraint.

This study demonstrates the presence of interobserver variation in parotid delineation and highlights its effect

<table>
<thead>
<tr>
<th>Table 1. Mean parotid doses for study contours and the doses in the actual treatment plans (Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Radiologist A</td>
</tr>
<tr>
<td>Radiologist B</td>
</tr>
<tr>
<td>Radiologist C</td>
</tr>
<tr>
<td>Oncologist A</td>
</tr>
<tr>
<td>Oncologist B</td>
</tr>
<tr>
<td>Oncologist C</td>
</tr>
<tr>
<td>Oncologist D</td>
</tr>
<tr>
<td>“Reference”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Volumes of study contours delineated by all seven observers (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Radiologist A</td>
</tr>
<tr>
<td>Radiologist B</td>
</tr>
<tr>
<td>Radiologist C</td>
</tr>
<tr>
<td>Oncologist A</td>
</tr>
<tr>
<td>Oncologist B</td>
</tr>
<tr>
<td>Oncologist C</td>
</tr>
<tr>
<td>Oncologist D</td>
</tr>
<tr>
<td>“Reference”</td>
</tr>
</tbody>
</table>

![Figure 2. Graph showing the mean parotid doses obtained by radiation oncologists for all ten data sets.](image-url)
on IMRT solutions. Only half of all study contours achieved a mean parotid dose within 10% of 24 Gy when they were evaluated using the actual IMRT plans delivered to the patients. The parotid DVHs of 46% of the study contours were sufficiently different from those used clinically, such that a different IMRT plan would have been produced. The mean doses for 18% of study contours were greater than 26.4 Gy.

Interobserver variation in GTV and CTV delineation in HNSCC has been extensively studied [13, 35–41]. By contrast, there have been few studies looking at interobserver variation in OAR definition in head and neck radiotherapy [42]. Geets et al [43] noted interobserver variation when parotid glands were delineated on CT and MRI. Even though MRI of the parotid gland provided excellent spatial resolution and soft-tissue contrast was not affected by dental artefact and allowed more accurate definition of the deep lobe, interobserver variation was found to be comparable between CT and MRI. Thus, coregistering the planning CT with MRI acquired in the treatment position may not offer any advantage, although the average parotid volumes were smaller when delineated with MRI than with CT [43]. These results are in contrast to those of another study which showed a reduction in interobserver variation with the addition of MRI, especially in patients with dental artefact on CT images [44]. Better reproducibility was noted in GTV delineation compared with that of the parotid [43], yet interobserver variation in GTV delineation has been much more extensively studied. This may be owing to clinicians’ perception that accurate definition of the GTV is more critical to the success of IMRT and that differences in parotid delineation may not be relevant. Ultimately, the significance of interobserver variation in parotid definition can best be demonstrated by its effect on the DVH when the treatment plan is applied. In this study, we established the importance of accurate delineation of the parotid gland given that differences in its contour may have impacted on IMRT dose solutions in almost half the cases.

Figure 3. Graph showing the mean parotid doses obtained by radiologists for all ten data sets.

Figure 4. Axial CT image showing the study contours of the seven observers for one parotid gland. (a) Radiologists; (b) radiation oncologists.
establish 24 Gy as the threshold mean parotid dose for the preservation of unstimulated saliva flow. In the pivotal study by Eisbruch et al [7] that defined the dose, volume and function relationships for the parotid, no details were provided on the methods used to outline the parotid glands. Uncertainties in parotid gland definition in that study may have led to an inaccurate correlation between the parotid DVH and xerostomia, and the threshold dose of 24 Gy thus obtained for unstimulated salivary flow may be misleading. A different dose threshold of 32 Gy was identified in a separate study [8]. Others failed to detect such a threshold, instead demonstrating a linear correlation between the mean parotid dose and post-radiotherapy salivary flow [49]. These discrepancies might have been caused by interstudy variation in parotid definition. In addition, variation in parotid delineation may also lead to a disparity between the true mean dose received in vivo and that calculated from the treatment plan, and therefore the actual sparing of the parotid glands may not be as effective as intended. Rather than assume that a mean dose of <24 Gy predicts parotid sparing, radiotherapy centres practising parotid-sparing IMRT should prospectively correlate parotid DVHs obtained from contours outlined at their centres with saliva measurements and evaluations using the RTOG/European Organisation for Research and Treatment of Cancer late radiation morbidity scoring for xerostomia [50].

Apart from interobserver variation in its delineation, other factors can also affect the radiation dose to the parotid gland. Volumetric and positional changes in patients' anatomy occur during head and neck radiotherapy as a result of weight loss, tissue oedema, and shrinkage of tumour and normal structures [51–54]. This, together with daily set-up variations, has an impact on the mean dose received by the parotid gland [55]. In a study by O'Daniel et al [56] using an integrated CT–linear accelerator and deformable image registration to compute cumulative dose distribution, the mean dose “delivered” to the contralateral parotid was found to be at least 5 Gy higher than the planned dose in 27% of patients, using alignment to mark on the immobilisation mask and weekly portal imaging. This increase in mean parotid dose was due to both uncertainties in patient set-up and radiation-induced reduction in tumour and parotid volumes, with a medial shift of the parotids' centre of volumes into the high-dose region [52, 56–58]. The effect of set-up uncertainties on the parotid dose distribution can be reduced via the use of daily image-guided radiotherapy (IGRT) or the addition of a PRV margin [59, 60]. Adaptive radiotherapy (ART) can also result in more effective sparing of the parotid gland by taking tumour and normal tissue changes into account, with subsequent modification of the original IMRT plans [61]. This technique is still under evaluation. Preliminary data from a prospective study conducted at MD Anderson Cancer Center suggest that ART is both feasible and reduces the mean dose to the contralateral parotid compared with IGRT alone [62].

Contrary to prevalent belief, the introduction and use of contouring guidelines did not consistently lead to a reduction in interobserver variation in target volume definition among radiation oncologists [63–65]. This may be secondary to interclinician differences in guideline and imaging interpretation. Although guidelines exist for delineation of the parotid gland, there is no published evidence to show that its use reduces interobserver variation [66]. Apart from contouring guidelines, other measures that may help improve interobserver consistency in OAR definition include introduction of protocol-based one-on-one training [67] and use of automatic atlas-based segmentation via deformable image registration [68–70]. In addition, evaluation and identification of features associated with low conformity between observers can be useful in educating clinicians [71] and improving existing delineation guidelines and protocols. Further studies are required to determine the effectiveness of these measures in reducing interobserver variation in parotid gland delineation.

This study is limited in using a relatively small number of patient data sets, with parotid delineations performed by clinicians from a single institution. While it would be interesting to replicate the study in other institutions, all the oncologists who participated in the current study were experienced in volume definition for conformal radiotherapy, and the radiologists were experienced in the interpretation of head and neck imaging. We believe the participants are likely to be representative of their peers.

Conclusions

This study demonstrates significant interobserver variation in parotid delineation for head and neck IMRT. Further studies are required to establish ways of reducing this variation, which will allow more accurate determination of the dose–volume–effect relationship for the parotid gland.

References

8. Chao KS, Deasy JO, Markman J, Haynie J, Perez CA, Purdy JA, et al. A prospective study of salivary function sparing in...


Dosimetric impact of interobserver variation in parotid gland delineation
