

Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?

¹M MUKESH, MB BS, FRCR, ¹R BENSON, MB BS, FRCR, ¹R JENA, MD, FRCR, ²A HOOLE, PhD, MIPEM, ³T ROQUES, BMBCh, FRCR, ⁴C SCRASE, MB BS, FRCR, ³C MARTIN, PhD, FRCR, ¹G A WHITFIELD, PhD, FRCR, ¹J GEMMILL, MSc and ¹S JEFFERIES, PhD, FRCR

¹Department of Oncology, Addenbrooke's Hospital, Cambridge, UK, ²Department of Medical Physics, Addenbrooke's Hospital, Cambridge, UK, ³Department of Oncology, Norfolk and Norwich University Hospital, Norwich, UK, and ⁴Department of Oncology, Ipswich Hospital, Ipswich, UK

Objective: A study of interobserver variation in the segmentation of the post-operative clinical target volume (CTV) and organs at risk (OARs) for parotid tumours was undertaken. The segmentation exercise was performed as a baseline, and repeated after 3 months using a segmentation protocol to assess whether CTV conformity improved.

Methods: Four head and neck oncologists independently segmented CTVs and OARs (contralateral parotid, spinal cord and brain stem) on CT data sets of five patients post parotidectomy. For each CTV or OAR delineation, total volume was calculated. The conformity level (CL) between different clinicians' outlines was measured using a validated outline analysis tool. The data for CTVs were reanalysed after using the cochlear sparing therapy and conventional radiation segmentation protocol.

Results: Significant differences in CTV morphology were observed at baseline, yielding a mean CL of 30% (range 25–39%). The CL improved after using the segmentation protocol with a mean CL of 54% (range 50–65%). For OARs, the mean CL was 60% (range 53–68%) for the contralateral parotid gland, 23% (range 13–27%) for the brain stem and 25% (range 22–31%) for the spinal cord.

Conclusions: There was low conformity for CTVs and OARs between different clinicians. The CL for CTVs improved with use of a segmentation protocol, but the CLs remained lower than expected. This study supports the need for clear guidelines for segmentation of target and OARs to compare and interpret the results of head and neck cancer radiation studies.

Received 29 March 2011
Revised 22 June 2011
Accepted 5 July 2011

DOI: 10.1259/bjr/66693547

© 2012 The British Institute of
Radiology

Carcinoma of the salivary glands constitutes 1–3% of head and neck cancer (HNC), with most tumours arising in the parotid glands. The mainstay of treatment is surgery, with post-operative radiotherapy in selected cases. Post-operative radiotherapy reduces local recurrence rates from 30% to 10% but has not been shown to impact on overall survival [1]. The aim of radiation therapy is to deliver homogeneous high-dose radiation to the target area with relative sparing of the healthy surrounding tissues. The physical delivery of radiotherapy has improved with the introduction of intensity-modulated radiotherapy (IMRT), which allows greater sparing of the surrounding healthy tissue and reduces the risk of long-term treatment morbidity [2], and image-guided radiotherapy (IGRT), which helps to improve geometric localisation of the high-dose volume to the target.

Sources of geometric uncertainty in dose delivery have been well characterised in modern IMRT workflows [3].

Despite this, geometric uncertainty due to target volume and normal tissue structure segmentation has escaped close scrutiny. It has been suggested in a number of tumour sites that the largest systematic error in radiotherapy planning today is related to the radiation oncologist's ability to localise and delineate the target volume [4–15]. When using IMRT and inverse planning, accurate and consistent segmentation of target and normal tissues impacts not only on the spatial localisation of the high-dose volume, but also on accuracy of dose calculation to organs at risk (OARs). Inaccuracies in segmentation can compromise treatment, either by unnecessarily treating the healthy surrounding tissue (which may increase treatment morbidity) or by inadequate target segmentation risking marginal tumour recurrence.

Previously published studies have assessed techniques to reduce interobserver and intraobserver variation in segmentation of gross tumour volume (GTV) and clinical target volume (CTV). These include the use of multimodality imaging such as MRI and positron emission tomography (PET)–CT [10, 16–20], multidisciplinary team contouring [5, 21], use of segmentation guidelines [22] and using auto-delineation software [23,

Address correspondence to: Dr Mukesh Mukesh, Oncology Centre, Box 193, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, UK. E-mail: drmukesh12@doctors.net.uk

24]. Studies have looked at interobserver and intraobserver variation in GTV (primary tumour) segmentation for HNC [10, 13, 25, 26]. Changes in the patient anatomy post surgery introduce difficulties for post-operative CTV segmentation. There is limited data on interobserver variation in CTV segmentation in HNC. Rasch et al [9] found significant interobserver variation in CTV segmentation in patients with paranasal sinus tumour, especially in the anterior border of the nasal cavity, possibly due to lack of anatomical landmarks to define the edge of the volume.

We studied interobserver variation in the segmentation of CTVs and OARs for patients receiving post-operative radiotherapy after parotidectomy and the impact of introducing segmentation guidelines for the CTV. We tried to replicate the clinical setting by asking radiation oncologists to outline both the primary tumour bed CTV and high-risk nodal groups (if appropriate). Currently, there are well-established guidelines for segmentation of CTV nodal volumes in HNC [27], but there are no standard segmentation guidelines for patients receiving post-operative radiotherapy to the parotid bed. Cochlear sparing therapy and conventional radiation (COSTAR) is a UK National Cancer Research Institute-supported head and neck trial, designed to demonstrate a difference in sensori-neural hearing loss with cochlear sparing IMRT compared with conventional radiotherapy [28]. The COSTAR trial management group created a segmentation protocol for the parotid bed CTV, which includes landmarks for defining extent of microscopic disease and recommends guidelines for elective and post-operative nodal irradiation. We obtained written permission to use the COSTAR segmentation guidelines for this study.

Methods and materials

Imaging data set

Planning CT data sets of five patients were randomly selected from the Addenbrooke's radiotherapy database. All five patients had been diagnosed with malignant parotid tumours (two acinic cell carcinoma, one pleomorphic carcinoma, one adenocarcinoma and one neuroendocrine tumour) and had completed post-operative radiotherapy. Each patient had been immobilised in a perspex shell, and non-contrast CT images were acquired on a CT scanner (GE Healthcare, Waukesha, WI) at 120 kV energy with a 50-cm field of view, 3-mm slice thickness and pixel size of 0.97 mm. Each data set was anonymised and distributed directly to the participating oncologists in the three regional cancer centres (Cambridge, Norwich and Ipswich). This study was registered as a service evaluation in the Anglia Cancer Network.

Observers

Four consultant radiation oncologists with a special interest in HNC were invited to participate. Each oncologist sees nearly 100 new patients with HNC per annum and has more than 5 years of experience in the use of IMRT for HNC. Radiation oncology trainees were not included in the study to minimise variability due to

lack of experience. Each observer was identified by a letter (A–D) in order to preserve anonymity.

Contouring exercise

The four observers were asked to segment the CTVs and OARs on axial slices of the planning CT data sets. The baseline segmentation exercise was performed according to the oncologist's routine practice. The CT window levels were adjustable to suit individual preferences for segmentation. Information on clinical history, surgical findings, staging and histology report (Table 1) were provided. All four observers were later provided with the COSTAR segmentation protocol and asked to resegment the CTV on the same CT data sets with a minimum interval of 3 months to avoid recall bias. The COSTAR protocol defines the CTV of the parotid bed and the indication for ipsilateral neck irradiation. The CTV1 defines the high-risk volume, which includes the parotid bed, adjacent tissues at risk of microscopic spread and ipsilateral lymph node levels Ib, IIa and IIb. CTV2 defines the low-risk volume, which includes ipsilateral node levels III, IV and V. CTV1 and CTV2 were fused together as CTV for the purpose of the analysis. The OARs were not resegmented by the observers as the COSTAR guidelines have no segmentation protocol for OARs.

These measurements were grouped together and compared to assess:

- interobserver variability in segmentation of CTVs
- interobserver variability in segmentation of OARs
- effect of contouring protocol on interobserver variability in segmentation of CTVs.

Method for analysis

We calculated the total volume outlined by each of the four clinicians for the CTV and OARs. The degree of agreement between observers' delineations was quantified using the conformity level (CL) proposed by Kouwenhoven et al [29]. This CL is a generalisation of the Jaccard coefficient for more than two delineated volumes (the Jaccard coefficient for two delineated volumes A and B is equal to the ratio of the common and encompassing volumes, *i.e.* $|A \cap B| / |A \cup B|$). It has the desirable properties that it is unbiased with respect to the number of delineations and equals the Jaccard coefficient in the case of two delineations. The CL ranges between 0% and 100%, where 100% is full concordance and 0% is no concordance. Hence, the CL can express conformity of a group of contours without requiring the identification of a gold standard contour and is therefore particularly appropriate for assessing conformity of a group of outlines.

Some variation in the CL for spinal cord and brain stem OARs was due simply to variation in the superior and inferior extent of the OARs outlined, because of variation in the superior–inferior extent of the CTV. Such variation would have minimal clinical impact as the radiotherapy plans are optimised based on maximum

Table 1. Type of surgery, histological report and TNM staging

Case	Surgery	Histology	TNM staging
1	Right parotidectomy and level 1–4 neck dissection	28-mm large-cell neuro-endocrine tumour with 5/28 positive nodes and extra capsular spread (positive surgical margin at masseter muscle)	T2N2bMx
2	Left parotidectomy and level 2 dissection	18-mm acinic cell carcinoma with 0/2 nodes positive (closest margin 0.5 mm)	T1N0Mx
3	Right parotidectomy	20-mm carcinoma in pleomorphic adenoma (closest margin 1 mm)	T1NxMx
4	Right parotidectomy and level 2 dissection	23-mm adenocarcinoma with 0/2 nodes positive (close margins)	T2N0Mx
5	Left parotidectomy and level 2 dissection	30-mm acinic cell tumour with 1/1 node positive (tumour excised from the upper facial tract)	T2N1Mx

point dose and are independent of the length of OARs outlined. We therefore also calculated the CL for brain stem and spinal cord after restricting the analysis to the CT slices on which OARs were segmented by all four clinicians; this we denote CL (axial).

Results

A total of 40 CTV outlines and 20 OAR outlines each for spinal cord, brain stem and contralateral parotid gland were compared using the outline analysis method.

Interobserver variation in clinical target volume (baseline)

There was a significant variation in the total volume of the CTV among clinicians, with observer B consistently delineating the largest and observer C consistently delineating the smallest CTV. The mean volume and standard deviation (SD) for each case is presented in Table 2. A significant difference in CTV morphology was observed between the four observers, with the mean CL 30% (range 25–39%; Table 2).

Interobserver variation in clinical target volume (using segmentation guidelines)

Using the segmentation guidelines, the mean CL improved from 30% to 54% ($p=0.043$ using the Wilcoxon paired analysis; Table 2). This improvement in concordance

between clinicians was also confirmed by reduced variation in CTV volume (reduced SD; Table 2). Using the segmentation guidelines, observer B's volumes were largest in four patients, with no observer consistently responsible for the smallest volume. Overall, there was a significant increase in the mean CTV volume calculated for each patient after the use of the COSTAR segmentation protocol from 170 to 273 cm³ ($p=0.043$ using the Wilcoxon paired analysis).

Interobserver variation in segmentation of organs at risk

There was a significant variation in the volume of contralateral parotid gland between clinicians (Table 3) with a mean CL of 60% (range 53–68%). The concordance for brain stem and spinal cord was poor, with mean CLs of 23% (range 13–27%) and 25% (range 22–31%), respectively (Tables 4 and 5). The interobserver variation for brain stem and spinal cord was predominantly related to the segmentation of different lengths of the OARs; however, significant interobserver variation was also noted in the axial direction. The mean CLs (axial) for brain stem and spinal cord were 45% (range 42–51%) and 60% (58–64%), respectively.

Discussion

This study demonstrates variability in target segmentation between experienced HNC oncologists with a mean CL for the CTV of 30%. This degree of uncertainty

Table 2. Clinical target volume conformity between observers at baseline and after using the COSTAR segmentation guidelines; use of segmentation guidelines improves conformity level, increases the mean clinical target volume and reduces the standard deviation

Case	Baseline		Using COSTAR segmentation protocol	
	Conformity level	Mean volume in ml (SD)	Conformity level	Mean volume in ml (SD)
1	38.7%	207.9 (101.7)	54.8%	329.1 (51.0)
2	25.5%	118.0 (94.1)	65.4%	270.8 (14.0)
3	27.3%	157.1 (115.1)	50%	243.3 (59.1)
4	26.3%	197.3 (142.1)	51.2%	287.7 (35.8)
5	32.7%	169.0 (85.7)	50%	236.2 (53.5)
Mean	30.1%	169.8	54.2%	273.4

COSTAR, cochlear sparing therapy and conventional radiation; CTV, clinical target volume; SD, standard deviation.

Table 3. Conformity level, mean total volume and standard deviation (SD) for contralateral parotid gland between the four observers

Case	Conformity level	Mean total volume in ml (SD)
1	60.9%	16.9 (3.4)
2	68.1%	25.5 (0.7)
3	59.5%	32.4 (3.3)
4	57.9%	45.6 (10.6)
5	52.7%	23.4 (1.8)
Mean	59.8%	28.7

is not accounted for in the CTV-to-PTV margin calculation (typically 5 mm for HNC). It also confirms considerable variation among radiation oncologists in segmenting OARs. The uncertainty in segmentation of OARs is not accounted for in the currently used planning OAR volume (PRV) margin calculation (3 mm in our unit). Use of the COSTAR segmentation protocol improved the mean CTV CL (from 30% to 54%), and this was associated with an increase in the mean CTV volume for all five patients.

This study has a number of limitations. The number of clinicians involved was small, and, though all clinicians had more than 5 years of experience in HNC IMRT, each would see fewer than five patients each year with malignant salivary tumours. Clinicians had no access to the pre-operative radiology and surgical notes. The CT images used were non-contrast; use of contrast agent to highlight vascular structures may facilitate nodal delineation. However, this was standard practice in our region at the time, and it is unlikely that use of contrast would have significantly altered the results, as all clinicians had extensive experience of delineating without contrast. Finally, it should be remarked that this study was concerned with interobserver variation; there was no attempt to analyse intraobserver variation.

Interobserver variation in clinical target volume

The potential reasons for significant interobserver variations in CTV outlines in the absence of a segmentation protocol (Figure 1, left image) include:

- (1) difference in interpretation of CT-based cross-sectional anatomy among different observers
- (2) lack of consensus guidelines on the selection of cervical lymph node treatment volumes in parotid tumours

Table 4. Conformity level for brain stem between the four observers

Case	Conformity level	Conformity level (axial)
1	27.7%	41.7%
2	27.4%	46.6%
3	13.2%	42.1%
4	25.5%	51.1%
5	22.3%	42.6%
Mean	23.2%	44.8%

The conformity in axial direction is shown as conformity level (axial).

Table 5. Conformity level for spinal cord between the four observers

Case	Conformity level	Conformity level (axial)
1	31.1%	63.6%
2	22.1%	61.5%
3	25.3%	57.9%
4	21.7%	60.2%
5	24.5%	58.3%
Mean	24.9%	60.3%

The conformity in axial direction is shown as Conformity level (axial).

- (3) use of different window settings by observers during segmentation
- (4) differences in clinical judgement in defining extent of microscopic disease
- (5) editing of the CTV from the ramus and body of the mandible by some observers.

The use of the COSTAR segmentation protocol improved the CL in target segmentation from 30% to 54%. There remained significant variation in CTV segmentation anteriorly and in the infratemporal region between different observers after using the COSTAR protocol, possibly due to lack of bony reference points in the anterior part of the neck and in the infratemporal region. These results are consistent with previous publications for other tumour sites, including prostate cancer [30], oesophageal cancer [12] and HNC [22]. Mitchell et al [30] showed significant interobserver variation in CTV segmentation for post-prostatectomy radiotherapy, which was reduced by the use of a contouring protocol. Tai et al [12] reported an improved consistency in cervical oesophagus target segmentation after one-to-one training. Three cases were used to represent different clinical situations and the use of a segmentation protocol reduced the variation of the longitudinal positions of GTV, CTV and PTV for all cases. Berson et al [22] asked two neuroradiologists and two radiation oncologists to delineate the GTV on PET-CT fusion images for 16 HNC patients after a short tutorial and using a contouring protocol. They reported reduced interobserver variation after use of the contouring protocol. The observed increase in mean CTV volume with use of a segmentation protocol is also consistent with previously published studies [12, 30].

Interobserver variation in organs at risk

For the contralateral parotid gland, the mean CL was 60% (range 53–68%). This variation appeared to be due mainly to difficulty in defining the deep lobe of the parotid gland on CT, lack of bony landmarks anteriorly and the use of different CT windowing during segmentation. The total volume outlined by the four observers differed significantly.

We found significant variation in the length of brain stem and spinal cord outlined. This was partly related to the variation in the length of the CTV; however, these large variations between experienced clinicians highlight that any contouring protocol used needs to be meticulous about defining the extent of OARs by agreed

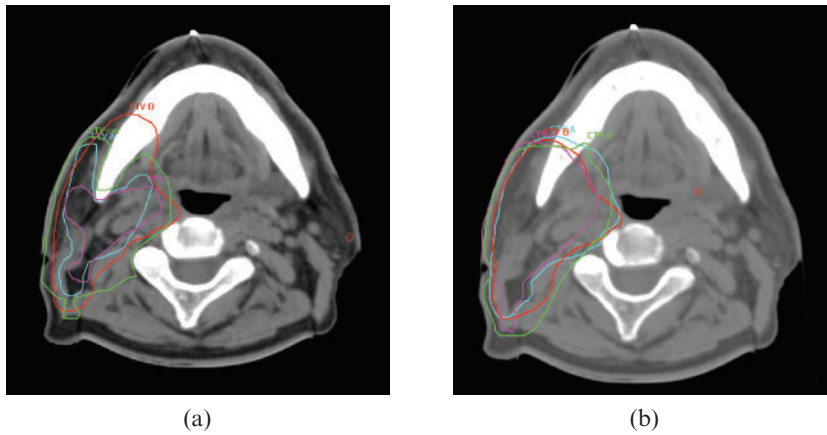


Figure 1. (a) Clinical target volume (CTV) outlined as baseline based on clinician's routine practice shows marked difference between observers A–D. (b) CTV outlined after using the cochlear sparing therapy and conventional radiation segmentation protocol shows improved concordance between observers A–D.

anatomical boundaries. We also found significant variation between outlines in the axial direction (Figure 2), which may have an impact on radiotherapy plan optimisation. We suspect that poor soft tissue characterisation on CT and use of different CT windowing contributed to the observed axial variation. We also found a significant variation among observers in defining the junction between brain stem and spinal cord. The inferior border of brain stem is at the level of the foramen magnum [31], and as the spinal cord is more radio-sensitive than brain stem, we propose that the lower end of foramen magnum (posterior lip of the foramen magnum) is used as the junction between spinal cord and brain stem.

To our knowledge, this study is the first to look at interobserver variation in CTV segmentation for parotid tumours that has included nodal segmentation and assessed the impact of segmentation guidelines. We found that the CL improves after using the COSTAR protocol, but the results were still below 80%. Complete agreement between two or more observers (CL of 100%) is rare and a conformity of >80% is generally accepted as highly concordant [32]. These levels of concordance are typically observed in the segmentation of intact gross tumour with non-infiltrative margins and high-quality soft tissue imaging. CTV segmentation in the post-operative setting is more difficult, as the clinician must

make allowances for altered anatomy post surgery and estimate the extent of potential microscopic disease that cannot be imaged using CT or MRI. The use of contouring protocols can reduce interobserver variation but cannot completely eliminate it. Discussion with the head and neck radiologist in interpretation of CT-based cross-sectional anatomy and defining the optimal window setting for segmentation may be useful.

This study highlights the need for international segmentation guidelines for OARs. The advantage of high-dose sculpting with IMRT to reduce dose to OARs is neutralised if the segmentation is inaccurate or inconsistent. The low CLs for OARs observed in our study are similar to previous published results [25, 33]. For example, Geets et al [25] showed poor reproducibility between five observers in segmenting the parotid gland and spinal cord on both CT and MRI. The MRI-based parotid volumes were however smaller than CT and more accurate in delineation of the deep lobe. The interobserver variation for OARs needs to be minimised in order to accurately compare and interpret treatment-related morbidity from different radiotherapy centres. To our knowledge, there are no studies looking at the effect of segmentation guidelines for OARs for HNC. van de Water et al [34] have defined segmentation guidelines for OARs involved with salivary function, which we anticipate will reduce interobserver variation.

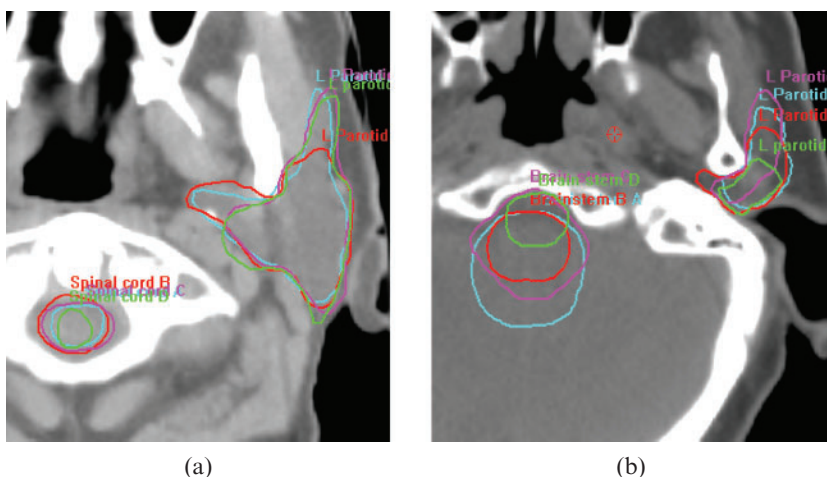


Figure 2. Variation in organs-at-risk segmentation. (a) Spinal cord and contralateral parotid gland segmentation by observers A–D. (b) Brain stem and contralateral parotid gland segmentation by observers A–D.

Conclusion

This study demonstrates significant interobserver variation in CTV and OAR segmentation between experienced HNC radiation oncologists. The CL for CTV delineation improved significantly after using the COSTAR segmentation protocol, but delineations were still not highly concordant. Successful implementation of newer radiotherapy techniques such as IMRT and IGRT in HNC depends crucially on minimising this systematic treatment planning error. We strongly recommend that contouring protocols are used in the segmentation of the target volume in the post-parotidectomy radiotherapy setting and segmentation guidelines are developed for OARs. Outcome data for patients treated with IMRT should include assessment of local failure in relation to target delineation. More efforts are still needed to minimise systematic errors in radiotherapy planning.

Acknowledgments

We would like to acknowledge Dr Christopher Nutting, Principal Investigator for the COSTAR trial, and Dr Teresa Provost, Medical Statistician at the Centre of Medical Applied Statistics, Cambridge.

References

1. Chen AM, Granchi PJ, Garcia J, Bucci MK, Fu KK, Eisele DW. Local-regional recurrence after surgery without post-operative irradiation for carcinomas of the major salivary glands: implications for adjuvant therapy. *Int J Radiat Oncol Biol Phys* 2007;67:982–7.
2. Pow EH, Kwong DL, McMillan AS, Wong MC, Sham JS, Leung LH, et al. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial. *Int J Radiat Oncol Biol Phys* 2006;66:981–91.
3. Stroom JC, Heijmen BJ. Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report. *Radiation Oncol* 2002;64:75–83.
4. Van de Steene J, Linthout N, de Mey J, Vinh-Hung V, Claassens C, Noppen M, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiation Oncol* 2002;62:37–49.
5. Weiss E, Richter S, Krauss T, Metzethin SI, Hille A, Pradier O, et al. Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume. A comparison between gynaecologic and radiation oncologists. *Radiation Oncol* 2003;67:87–95.
6. Hurkmans CW, Borger JH, Pieters BR, Russell NS, Jansen EP, Mijnheer BJ. Variability in target volume delineation on CT scans of the breast. *Int J Radiat Oncol Biol Phys* 2001;50:1366–72.
7. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiation Oncol* 1998;47:285–92.
8. Tai P, Van Dyk J, Yu E, Battista J, Stitt L, Coad T. Variability of target volume delineation in cervical esophageal cancer. *Int J Radiat Oncol Biol Phys* 1998;42:277–88.
9. Rasch C, Eisbruch A, Remeijer P, Bos L, Hoogeman M, van Herk M, et al. Irradiation of paranasal sinus tumors, a delineation and dose comparison study. *Int J Radiat Oncol Biol Phys* 2002;52:120–7.
10. Riegel AC, Berson AM, Destian S, Ng T, Tena LB, Mitnick RJ, et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int J Radiat Oncol Biol Phys* 2006;65:726–32.
11. Yamamoto M, Nagata Y, Okajima K, Ishigaki T, Murata R, Mizowaki T, et al. Differences in target outline delineation from CT scans of brain tumours using different methods and different observers. *Radiation Oncol* 1999;50:151–6.
12. Tai P, Van Dyk J, Battista J, Yu E, Stitt L, Tonita J, et al. Improving the consistency in cervical esophageal target volume definition by special training. *Int J Radiat Oncol Biol Phys* 2002;53:766–74.
13. Cooper JS, Mukherji SK, Toledano AY, Beldon C, Schmalfluss IM, Amdur R, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). *Int J Radiat Oncol Biol Phys* 2007;67:972–5.
14. Hermans R, Feron M, Bellon E, Dupont P, Van den Bogaert W, Baert AL. Laryngeal tumor volume measurements determined with CT: a study on intra- and interobserver variability. *Int J Radiat Oncol Biol Phys* 1998;40:553–7.
15. McKenzie A, Coffey M, Greener T, Hall C, Van Herk M, Mijnheer B, et al. Technical overview of geometric uncertainties in radiotherapy. In: BIR Working Party, ed. *Geometric uncertainties in radiotherapy*. London, UK: The British Institute of Radiology; 2003:11–27.
16. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys* 2005;62:70–5.
17. Daisne JF, Duprez T, Weynand B, Lonneux M, Hamoir M, Reyckler H, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 2004;233:93–100.
18. Paulino AC, Koshy M, Howell R, Schuster D, Davis LW. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2005;61:1385–92.
19. El-Bassiouni M, Ciernik IF, Davis JB, El-Attar I, Reiner B, Burger C, et al. [18FDG] PET-CT-based intensity-modulated radiotherapy treatment planning of head and neck cancer. *Int J Radiat Oncol Biol Phys* 2007;69:286–93.
20. Wang D, Schultz CJ, Jursinic PA, Bialkowski M, Zhu XR, Brown WD, et al. Initial experience of FDG-PET/CT guided IMRT of head-and-neck carcinoma. *Int J Radiat Oncol Biol Phys* 2006;65:143–51.
21. Giraud P, Elles S, Helfre S, De Rycke Y, Servois V, Carette MF, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiation Oncol* 2002; 62:27–36.
22. Berson AM, Stein NF, Riegel AC, Destian S, Ng T, Tena LB, et al. Variability of gross tumor volume delineation in head-and-neck cancer using PET/CT fusion, part II: the impact of a contouring protocol. *Med Dosim* 2009;34:30–5.
23. van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys* 2007;68:771–8.
24. Chao KS, Bhide S, Chen H, Asper J, Bush S, Franklin G, et al. Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach. *Int J Radiat Oncol Biol Phys* 2007;68:1512–21.

25. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. *Radiother Oncol* 2005;77:25–31.
26. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *Int J Radiat Oncol Biol Phys* 2007;68:763–70.
27. Gregoire V, Eisbruch A, Hamoir M, Levendag P. Proposal for the delineation of the nodal CTV in the node-positive and the post-operative neck. *Radiother Oncol* 2006;79:15–20.
28. ISRCTN Register. COSTAR: multicentre randomised study of cochlear sparing intensity modulated radiotherapy versus conventional radiotherapy in patients with parotid tumours (ISRCTN81772291). [Cited 28 June 2010.] Available from: <http://www.controlled-trials.com/ISRCTN81772291>.
29. Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. *Phys Med Biol* 2009;54:2863–73.
30. Mitchell DM, Perry L, Smith S, Elliott T, Wylie JP, Cowan RA, et al. Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation. *Int J Radiat Oncol Biol Phys* 2009;75:990–3.
31. Mayo C, Yorke E, Merchant TE. Radiation associated brainstem injury. *Int J Radiat Oncol Biol Phys* 2010;3(Suppl. 76):S36–41.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
33. Saarnak AE, Boersma M, van Bunningen BN, Wolterink R, Steggerda MJ. Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. *Radiother Oncol* 2000;56:37–42.
34. van de Water TA, Bijl HP, Westerlaan HE, Langendijk JA. Delineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia. *Radiother Oncol* 2009;93:545–52.