Variation of normal tissue complication probability (NTCP) estimates of radiation-induced hypothyroidism in relation to changes in delineation of the thyroid gland

MARIANNE F. RØNJOM1,2, CARSTEN BRINK2,3, EBBE L. LORENZEN3, LASZLO HEGEDŰS2,4 & JØRGEN JOHANSEN1,2

1 Department of Oncology, Odense University Hospital, Odense, Denmark, 2 Institute of Clinical Research, University of Southern Denmark, Odense, Denmark, 3 Laboratory of Radiation Physics, Odense University Hospital, Odense, Denmark and 4 Department of Endocrinology and Metabolism, Odense University Hospital, Odense, Denmark

ABSTRACT

Background. To examine the variations of risk estimates of radiation-induced hypothyroidism (HT) from our previously developed normal tissue complication probability (NTCP) model in patients with head and neck squamous cell carcinoma (HNSCC) in relation to variability of delineation of the thyroid gland.

Patients and methods. In a previous study for development of an NTCP model for HT, the thyroid gland was delineated in 246 treatment plans of patients with HNSCC. Fifty of these plans were randomly chosen for re-delineation for a study of the intra- and inter-observer variability of thyroid volume, Dmean and estimated risk of HT. Bland-Altman plots were used for assessment of the systematic (mean) and random (standard deviation (SD)) variability of the three parameters, and a method for displaying the spatial variation in delineation differences was developed.

Results. Intra-observer variability resulted in a mean difference in thyroid volume and Dmean of 0.4 cm³ (SD 1.6) and -0.5 Gy (SD 1.0), respectively, and 0.3 cm³ (SD 1.8) and 0.0 Gy (SD 1.3) for inter-observer variability. The corresponding mean differences of NTCP values for radiation-induced HT due to intra- and inter-observer variations were insignificantly small, -0.4% (SD 6.0) and -0.7% (SD 4.8), respectively, but as the SDs show, for some patients the difference in estimated NTCP was large.

Conclusion. For the entire study population, the variation in predicted risk of radiation-induced HT in head and neck cancer was small and our NTCP model was robust against observer variations in delineation of the thyroid gland. However, for the individual patient, there may be large differences in estimated risk which calls for precise delineation of the thyroid gland to obtain correct dose and NTCP estimates for optimized treatment planning in the individual patient.

Over the last decades treatment planning in radiotherapy (RT) has experienced a considerable development both in the precision in treatment planning [including computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET)-CT simulation] and in delivery of the planned dose to the target area through image-guided therapy. This has made it possible to optimize target dose coverage and at the same time reduce radiation dose to organs at risk (OARs) [1]. However, the precision of most of these new technologies relies on the delineation of the target volumes as well as OARs. The variation in delineations is a limiting factor for the precision of treatment planning and consequently prediction of the risk of late effects in the OARs.

The tolerance estimates of OARs are based on normal tissue complication probability (NTCP) models, and variation in OAR delineation might affect the estimated risk of developing a specific late effect. As pointed out by the Qualitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) group [2], lack of consistency in contouring OARs among investigators is one of the challenges in developing NTCP models.

Radiation-induced hypothyroidism (HT) is a well known normal tissue late effect to RT in the neck
area due to radiation of the thyroid gland [3,4]. In a previous study [3], a NTCP-model for biochemical HT after definitive RT in patients with head and neck squamous cell carcinoma (HNSCC) was generated. The model is a mixture model, taking thyroid volume, thyroid mean dose ($D_{\text{mean}}$) and latency time of the normal-tissue reaction into account. Biochemical HT was defined as any elevated serum level of thyrotropin (TSH). The aims of this study were to evaluate variability of estimated NTCP values of our mixture NTCP model subsequent to assessing the intra- and inter-observer variability in delineation of the thyroid gland.

**Material and methods**

In the previous NTCP model study, the thyroid gland was delineated in the RT plans of 246 patients with HNSCC, 203 of whom were finally included in the NTCP study. All delineations were made by oncologist A. The patients were treated at the Department of Oncology, Odense University Hospital, Denmark, during 2002–2010. All patients were treated according to the Danish Head and Neck Cancer Groups (DAHANCA) guidelines [5] with definitive external beam RT. For further characteristics of patients included in the previous study, we refer to our publication [3]. The treatment planning CT-scans were performed on a Siemens (Siemens Healthcare, Germany) or Philips CT-scanner (Philips Healthcare, The Netherlands) with a 3-mm slice thickness (voxels $1 \times 1 \times 3$ mm$^3$). Treatment planning was carried out in Pinnacle$^3$ (Philips Healthcare, The Netherlands), using the collapsed cone algorithm for dose calculation.

The thyroid glands were delineated without guidelines but according to departmental clinical practice. No dose constraints were applied to the thyroid gland during dose planning, however, according to current DAHANCA guidelines, doses to the larynx should be kept below 50 Gy for 2/3 of the organ as a secondary priority compared to PTVs.

Delineations made for the previous study is referred to as A1. Fifty of the 246 RT plans were randomly chosen for blinded re-delineation of the thyroid gland by A (A2, intra-observer variability) and by an experienced oncologist, B (inter-observer variability). The first 50 patients delineated in the initial study were excluded in order to avoid the potential problem of a learning curve during the initial delineations. Patients with abnormal conditions or deformations in the neck area were excluded from analysis. Altogether, four patients were excluded due to: tracheostomia ($n = 2$), surgical implants in the cervical spine ($n = 1$), and severe deformation of the neck due to a previous accident ($n = 1$), leaving 46 treatment plans for evaluation. Of these 46 patients, 24 received an intravenous iodine-containing contrast injection, ioversol (Optiray), during the CT-scan for treatment planning. Patient characteristics are shown in Table I. The thyroid gland volume and $D_{\text{mean}}$ from all three delineation sets were extracted from the dose-planning system.

Intra-observer (A1-A2) and inter-observer (A1-B) variations in volume and $D_{\text{mean}}$ were evaluated using Bland-Altman (Ba) plots [6]. For all delineation pairs, the Ba-plots show the paired differences as a function of the averaged paired values; furthermore, the mean of all differences and associated 95% limits of agreement (95% confidence limits, 1.96 × SD) are illustrated by horizontal lines. A Wilcoxon-signed rank test was used to test for systematic differences between the delineations. The random variation was calculated as the SD of the paired differences. Spearman’s rank correlation was used to test for correlations between the paired differences and mean thyroid volume, $D_{\text{mean}}$ and nTCP, respectively. The correlation between the (random) variation and mean thyroid volume was tested by dividing the paired differences into five equally sized groups, depending on mean thyroid size, and test the SDs of these groups with Spearman’s rank correlation.

### Table I. Patient characteristics.

| Number (%) |
|------------|
| (N = 46)   |

| Sex       | Male 38 (83) | Female 8 (17) |
| Age (years) | Median (range) 56 (43–78) |
| Primary site | Oral cavity 3 (7) |
| Oropharynx 23 (50) |
| Larynx 18 (39) |
| Hypopharynx 2 (4) |
| Stage | I+ II 14 (30) |
| III+ IV 32 (70) |
| Ch-RT | None 39 (85) |
| Concomitant 7 (15) |
| Boost | Concomitant 25 (54) |
| Integrated 21 (46) |
| Contrast | None 22 (48) |
| Iodine 24 (52) |
| Total RT dose (Gy) 66 34 (74) |
| 68 11 (24) |
| 76 1 (2) |
| Number of fractions | 33 41 (89) |
| 34 4 (9) |
| 56 1 (2) |
For each patient the overlap between the intra- and inter-observer delineations was calculated and the Sørensen-Dice similarity index (DSI) was calculated. For two volumes (A and B) DSI is defined as:

\[
\text{DSI}(A,B) = \frac{2V_{A\cap B}}{V_A + V_B}
\]

For additional assessment of a potential spatial variation of differences between delineations, the spatial dependent minimum distance between two pairs of contours was calculated over the entire surface of the gland. Due to the anatomy, the thyroid gland was divided into the left and right lobe for the mapping. This was done manually from a clinical perception of the anatomy, and separation was done through the middle of the isthmus. The mapping was done separately for each lobe, and for both the intra- and inter-observer delineations, using the center of mass as origin and 100 measurement points in both the latitudinal and the longitudinal direction (10,000 surface points in total). The distance was measured, including the sign indicating whether the first contour was inside or outside the second contour. Based on these spatially resolved values, population averaged values (systematic deviations) as well as SD (random deviations) could be calculated and visualized as a color surface plot of the “mean thyroid gland” using the color to illustrate the local deviation.

Both thyroid volume and Dmean from the different sets of delineations were entered into our previously developed multivariate NTCP model. The estimated NTCP risks for each patient were calculated and evaluated using BA-plots.

Statistics were done using STATA 13 and MatLab version 2012b for the NTCP modeling. All tests were two-sided and a p-value < 0.05 was considered statistically significant.

### Results

Median thyroid volumes, Dmean and estimated NTCP of the three sets of delineations are shown in Table II. Table II. Median (range) volume, Dmean and NTCP for the three delineated datasets.

|                  | Observer A1 | Observer A2 | Observer B |
|------------------|-------------|-------------|------------|
| Volume (cm³)     | 17.7 (7.1–86.5) | 17.2 (7.6–82.8) | 17.7 (6.8–93.8) |
| Dmean (Gy)       | 39.9 (12.6–67.8) | 40.8 (12.5–67.9) | 41.7 (11.6–67.6) |
| NTCP (%)         | 11.9 (1.45×10⁻³–93.0) | 12.7 (3.9×10⁻³–93.8) | 10.6 (1.6×10⁻³–93.6) |

Figure 1. Bland-Altman plots demonstrating intra-observer (A) and inter-observer (B) variation in delineated volume.

Spearman’s rank correlation did not show a significant intra-observer relation between the variation and the size of the gland \((p = 0.188)\), but did show a significant inter-observer variation \((p = 0.005)\). When assessing the intra- and inter-observer BA-plots for the subgroups receiving contrast or no contrast no clear difference between mean and random variability between the subgroups was found (Supplementary Figure 1 and Supplementary Table I, to be found online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2014.1001034).

Mean DSI was 0.88 (SD ± 0.03) for the intra-observer delineations and 0.85 (SD ± 0.04) for inter-observer delineations. Systematic spatial variations of delineations were assessed from the color surface plots of the lobes of the “mean thyroid gland”. The plots displayed variability from the population mean values of ± 0.02 cm for the majority of the surfaces on both the intra- and inter-observer plots (Figure 2A). The most pronounced variability from the population mean values was found around the middle of the gland, especially on the anterior and lateral sides of the lobes and the medial parts of the gland (above the isthmus), with mean differences in delineations of ± 0.04–0.08 cm. Assessing the local random variation, the surface plots showed that the SD of the delineations were around 0.06–0.1 cm (intra-observer) and 0.06–0.12 cm (inter-observer) for the major areas of the lobes (Figure 2B). The highest random variations in
delineations were in the caudal and medial parts of the lobes (including the isthmus), as well as in small areas cranially, with an SD around 0.12–0.16 cm (intra-observer) and 0.16–0.22 cm (inter-observer).

Figure 3A and B show the variations in $D_{\text{mean}}$ derived from the different thyroid gland delineations. As demonstrated in Figure 3A and Table III, there is a small systematic difference in intra-observer $D_{\text{mean}}$; mean difference being 0.5 Gy lower in A1 compared with A2 ($p = 0.002$). However, intra-observer random variation was slightly smaller than the inter-observer, SD ± 1.0 and ± 1.3 Gy, respectively. Assessed by Spearman’s rank correlation no correlation was found between the differences and averaged values of $D_{\text{mean}}$.

The intra- and inter-observer variations in predicted risk for development of HT were assessed by estimating individual NTCP-values based on thyroid volumes and $D_{\text{mean}}$ from the three delineations sets. In Figure 4A and B the BA-plots demonstrate the intra- and inter-observer variations, respectively. We found that intra- and inter-observer systematic variations in the estimated risk of developing HT were small, below 1.0%, and that random variations were equal to or less than 6.0%. Assessed by Spearman’s rank correlation no correlation was found between the differences and averaged values on the risk of developing HT.

Discussion

Several studies of observer-variability delineation of OARs in HNSCC [7–9] have demonstrated the importance of precise delineation of organ structures in dose planning. One study, analyzing the impact of different guidelines in estimating NTCP-values for different late effects for swallowing organ structures [10], showed that differences in organ delineation may have major impact on NTCP estimates. We studied the impact of intra- and inter-observer variability in delineation of the thyroid gland on estimation of the risk of developing radiation-induced HT using our previously published NTCP model.

When predicting the risk of developing HT, by estimating individual NTCP-values based on thyroid volumes and $D_{\text{mean}}$ from the three delineations sets, only small and insignificant mean differences of NTCP values were found, i.e. -0.4% within observer, and -0.7% between observers. This shows that the applied NTCP-model is robust towards small systematic differences in volume and $D_{\text{mean}}$ caused by delineation inaccuracy and that NTCP-estimates are consistent. Thus, the variability in delineation of the thyroid gland had a modest effect on the predicted mean NTCP for HT after RT. This implies that our previously published tolerance levels for the thyroid gland in general, for dose planning, are consistent [3].
and in the third patient, there was no obvious explanation for the delineation differences. For the two first patients the differences might be explained by lack of intravenous contrast during the CT-scan, although we did not find any difference in delineation variation between the subgroups receiving contrast or not when assessing the Ba-plots of the subgroups.

The mean DSi of 0.88 and 0.85 in our study is relatively high compared to other structures in the head and neck region that have been analyzed for observer variability [7]. Although we cannot rule out an impact of local delineation traditions, the consistency in delineation of the thyroid gland may be explained by the fact that the thyroid gland is relatively well defined on CT-scans due to its well vascularized tissue [12] and high content of iodine [13,14]. This is also in accordance with Nygaard et al. [15] who found a good reproducibility of thyroid volume on CT scans when assessing intra- and inter-observer variability in moderate-sized goiters.

In this single center study, both intra- and inter-observer variability for thyroid volume delineation on CT scans were within the same range without using specific guidelines for delineation. Consistent thyroid volumes were observed in three sets of delineations of the thyroid gland by two observers. When visually assessing thyroid volume on the BA plots, we could not discern a particular structure and no systematic differences, nor were there significant correlations between the mean differences and volume (Figure 1A and B). For the inter-observer delineations, the random variation was positively correlated with thyroid volume, and this might be expected when the absolute volume of the gland increases. With one exception, all thyroid glands in our study were in or close to the normal size range of 10–28 ml [11].

Three patients were identified as outliers from the BA-plots due to a difference in delineated volume above 3.0 cm³. All three were found in the intra-observer set, two of whom were also outliers in the inter-observer set. When assessing the CT-scans of these patients, we found that in one of the patients, the scanning image was blurred and the thyroid difficult to visualize, the second had an irregular goiter and in the third patient, there was no obvious explanation for the delineation differences. For the two first patients the differences might be explained by lack of intravenous contrast during the CT-scan, although we did not find any difference in delineation variation between the subgroups receiving contrast or not when assessing the BA-plots of the subgroups.

The mean DSi of 0.88 and 0.85 in our study is relatively high compared to other structures in the head and neck region that have been analyzed for observer variability [7]. Although we cannot rule out an impact of local delineation traditions, the consistency in delineation of the thyroid gland may be explained by the fact that the thyroid gland is relatively well defined on CT-scans due to its well vascularized tissue [12] and high content of iodine [13,14]. This is also in accordance with Nygaard et al. [15] who found a good reproducibility of thyroid volume on CT scans when assessing intra- and inter-observer variability in moderate-sized goiters.

As stated, DSi showed good agreement in delineated volume. However, this study extended the assessment of spatial variation by assessing the size of the local variation and identifying problematic areas that might need particular attention in delineation by three dimensional (3D) mapping of the thyroid gland. The spatial variations in delineations, illustrated in Figure 2B, demonstrate that the random variation is greatest in the area around the isthmus (medial part), the caudal and cranial part of the thyroid gland. This is in agreement with Brouwer et al. [8] in their study of 3D variations of five other head and neck OARs. When we assessed the size of the variation around the gland (Figure 2B), we found that the areas with the largest variation (in red) had a range of around 0.12–0.2 cm, reflecting a difference in delineation corresponding to one slice of the CT-scan cranially or caudally (slice thickness 3 mm). In our opinion, this random variation is difficult to avoid and is not likely to be minimized with delineation guidelines. Variability might be reduced with a smaller slice thickness of the planning CT. Furthermore, delineations in multiple orientations and not

**Table III.** Mean differences, 95% limits of agreement and SD between the intra- and inter-observer delineations. Wilcoxon-signed rank test was used for the p-value calculation.

| Observer | Mean difference (95% limits of agreement) | ± SD | Is there a (systematic) difference? DSi (± SD) |
|----------|------------------------------------------|-----|--------------------------------------------|
| Volume (cm³) | A1-A2 0.4 (2.8–3.6) | 1.6 | No (p = 0.186) | 0.88 (0.03) |
| | A1-B 0.3 (3.3–3.8) | 1.8 | No (p = 0.104) | 0.85 (0.04) |
| D_mean (Gy) | A1-A2 −0.5 (2.6–1.6) | 1.0 | Yes (p = 0.006) | |
| | A1-B 0.0 (2.7–2.6) | 1.3 | No (p = 0.74) | |
| NTCP (%) | A1-A2 −0.4 (12.4–11.6) | 6.0 | No (p = 0.448) | |
| | A1-B −0.7 (10.3–8.9) | 4.8 | No (p = 0.154) | |

*aDSi, Sørensen-Dice similarity index.

![Figure 3. Bland-Altman plots demonstrating intra-observer (A) and inter-observer (B) variation in calculated thyroid D_mean.](image-url)
Variation of NTCP estimates of hypothyroidism regarding a set of OARs in patients with pharyngeal cancer they found that substantial inter session variability in volume was apparent, but that dose differences were small in accordance with our findings. However, the thyroid gland was not delineated in either of the two studies [7,21].

We have addressed the methodological limitation of the low numbers of observers by analyzing the impact on NTCP estimates caused by extreme systematic variations. This was done by calculating the standard error of the observed differences in the studied patient group and then applying the extreme value of the 95% confidence interval of the observed mean value. The percentage difference was then obtained by dividing the value from the extreme value of the confidence value with the median observed value. This resulted in a systematic difference of delineated volume of $\pm 5\%$ (both intra- and inter-observer) and a systematic difference in Dmean of $\pm 2\%$ (intra-observer) and $\pm 1\%$ (inter-observer), which in the most extreme combination ($-5\%$ volume and $-2\%$ Dmean) led to an estimated mean difference in NTCP of $3.4\%$ with a SD of $3\%$. This indicates that the estimated NTCP, compared to the intrinsic uncertainty of the model, is almost unchanged for the whole population, however, there might be large changes for the individual patient.

It must be emphasized that the current study did observe large differences in estimated risk of HT in some individuals (up to $26\%$), leading to a SD of the NTCP-values up to $\pm 6\%$. A small thyroid volume and high thyroid Dmean are highly significant and independent risk factors for development of radiation-induced HT [3,4]. Consequently, when estimating the variability in NTCP, this relies both on the differences in delineated volume and Dmean and also on the patients’ actual thyroid volume and Dmean. Furthermore, the variability is dependent on the patients’ risk of developing HT, i.e. the steepness of the dose-response curve giving a larger $\Delta$NTCP in the middle of the dose-response curve than in both ends for the same difference in volume or Dmean. This can be seen in Figure 4A and B, where the random variation is higher in the areas with a 20–80% risk of developing HT, compared to the areas with both lower and higher risk. When examining the three outliers according to thyroid volume mentioned earlier, one of these had a difference in NTCP of $26\%$ (due to a small thyroid volume and a NTCP in the steep part of the dose-response curve), whilst the other two outliers had a $<1\%$ difference in NTCP. Thus, despite the small variation in mean differences of NTCP values in the study population, variation in organ delineation does have an impact on NTCP estimates which still calls for a precise delineation of the thyroid gland to obtain correct dose and NTCP estimates for optimized treatment planning in the individual.
For the entire study population, the variation in predicted risk of radiation-induced HT in head and neck cancer was small and our NTCP model was robust towards observer variations in delineation of the thyroid gland. Therefore, our recommendations for tolerance dose levels for the thyroid gland to radiation treatment still hold true. However, for the individual patient there may be pronounced differences in the estimated risk due to variation in volume and organ delineation so that it is of utmost importance for the individual patient, to precisely define OARs to obtain correct dose and NTCP estimates for optimized treatment planning.

Acknowledgments
This study was supported by CiRRO – The Lundbeck Foundation Center for Interventional Research in Radiation Oncology and the Danish Cancer Society. Laszlo Hegedüs is supported by an unrestricted grant from the Novo Nordisk Foundation.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References
[1] Marta GN, Silva V, de Andrade Carvalho H, de Arruda FF, Hanna SA, Gadia R, et al. Intensity-modulated radiation therapy for head and neck cancer: Systematic review and meta-analysis. Radiother Oncol 2014;110:9–15.
[2] Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUanTEC): An introduction to the scientific issues. Int J Radiat Oncol Biol Phys 2010;76(3 Suppl):S3–9.
[3] Rønjom MF, Brink C, Bentzen SM, Hegedüs L, Overgaard J, Johansen J. Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: Normal tissue complication probability modeling with latent time correction. Radiother Oncol 2013;109:317–22.
[4] Boomsma MJ, Bijl HP, Christianen ME, Beetz I, Chouvalova O, Steenbakkers RJ, et al. A prospective cohort study on radiation-induced hypothyroidism: Development of an NTCP model. Int J Radiat Oncol Biol Phys 2012;84:e351–6.
[5] DAHANCA. Danish Head and Neck Cancer Group (DAHANCA) guidelines 2013 [updated 2013; cited 2014 Oct 28]. Available from: www.dahanca.dk/guidelines.
[6] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
[7] Nelms BE, Tome WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: Test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2012;82:368–78.
[8] Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32.
[9] Loo SW, Martin WM, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. Br J Radiol 2012;85:1070–7.
[10] Brouwer CL, Steenbakkers RJ, Gort E, Kamphuis ME, van der Laan HP, Van’t Veld AA, et al. Differences in delineation guidelines for head and neck cancer result in inconsistent reported dose and corresponding NTCP. Radiother Oncol 2014;11:148–52.
[11] Hegedüs L, Perrild H, Poulsen LR, Andersen JR, Holm B, Schnoor P, et al. The determination of thyroid volume by ultrasound and its relationship to body weight, age, and sex in normal subjects. J Clin Endocrinol Metab 1983;56:260–3.
[12] Policeni BA, Smoker WR, Reed DL. Anatomy and embryology of the thyroid and parathyroid glands. Semin Ultrasound CT MR 2012;33:104–14.
[13] Bashist B, Ellis K, Gold RP. Computed tomography of intrathoracic goiters. AJR Am J Roentgenol 1983;140:455–60.
[14] Kanelo T, Matsumoto M, Fukui K, Hori T, Katayama K. Clinical evaluation of thyroid CT values in various thyroid conditions. J Comput Tomogr 1979;3:1–4.
[15] Nygaard B, Nygaard T, Court-Payan M, Jensen L, Soe-Jensen P, Gerhard Nielsen K, et al. Thyroid volume measured by ultrasonography and CT. Acta Radiol 2002;43:269–74.
[16] Sims R, Isambert A, Gregoire V, Bidault F, Fresco L, Sage J, et al. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. Radiother Oncol 2009;93:474–8.
[17] Teguh DN, Levendag PG, Voet PW, Al-Mamgani A, Han X, Wolk TK, et al. Clinical validation of atlas-based autosegmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys 2011;81:950–7.
[18] Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiother Oncol 2014;9:173.
[19] Zhu M, Bzdusek K, Brink C, Eriksen JG, Hansen O, Jensen HA, et al. Multi-institutional quantitative evaluation and clinical validation of Smart Probabilistic Image Contouring Engine (SPICE) autosegmentation of target structures and normal tissues on computer tomography images in the head and neck, thorax, liver, and male pelvis areas. Int J Radiat Oncol Biol Phys 2013;87:254–8.
[20] Lorenzen EL, Taylor CW, Maraldo M, Nielsen MH, Offerens BV, Andersen MR, et al. Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: A multi-centre study from Denmark and the UK. Radiother Oncol 2013;108:254–8.
[21] Feng M, Demiroz C, Vineberg KA, Eisbruch A, Balter JM. Normal tissue anatomy for oropharyngeal cancer: Contouring variability and its impact on optimization. Int J Radiat Oncol Biol Phys 2012;84:e245–9.

Supplementary material available online
Supplementary Figure 1, Supplementary Table I to be found online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2014.1001034.