Treatment planning

Differences in delineation guidelines for head and neck cancer result in inconsistent reported dose and corresponding NTCP

Charlotte L. Brouwer *, Roel J.H.M. Steenbakkers, Elske Gort, Marije E. Kamphuis,
Hans Paul van der Laan, Aart A. van 't Veld, Nanna M. Sijtsema, Johannes A. Langendijk

University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, Groningen, The Netherlands

ABSTRACT

Purpose: To test the hypothesis that delineation of swallowing organs at risk (SWOARs) based on different guidelines results in differences in dose–volume parameters and subsequent normal tissue complication probability (NTCP) values for dysphagia-related endpoints.
Materials and methods: Nine different SWOARs were delineated according to five different delineation guidelines in 29 patients. Reference delineation was performed according to the guidelines and NTCP-models of Christianen et al. Concordance Index (CI), dosimetric consequences, as well as differences in the subsequent NTCPs were calculated.
Results: The median CI of the different delineation guidelines with the reference guidelines was 0.54 for the pharyngeal constrictor muscles, 0.56 for the laryngeal structures and 0.07 for the cricopharyngeal muscle and esophageal inlet muscle. The average difference in mean dose to the SWOARs between the guidelines with the largest difference (maxΔD) was 3.5 ± 3.2 Gy. A mean ΔNTCP of 2.3 ± 2.7% was found. For two patients, ΔNTCP exceeded 10%.
Conclusions: The majority of the patients showed little differences in NTCPs between the different delineation guidelines. However, large NTCP differences >10% were found in 7% of the patients. For correct use of NTCP models in individual patients, uniform delineation guidelines are of great importance.
© 2014 The Authors. Published by Elsevier Ireland Ltd. Radiotherapy and Oncology 111 (2014) 148–152
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

In head and neck radiotherapy, reducing the dose to healthy tissues is important, since radiation damage to organs at risk (OARs) may result in severe complications during and after completion of treatment. Some radiation-induced complications, in particular swallowing dysfunction, have a significant impact on health-related quality of life as reported by patients [1,2].

Several guidelines for OAR delineation have been published [3–11]. However, the definition, selection and delineation of OARs vary widely among the different publications and authors. This may lead to unjustified comparisons between institutes that apply different guidelines, jeopardizing the translation of results published into routine clinical practice.

Studies on the development of normal tissue complication probability (NTCP) models have identified numerous predictive factors for the development of radiation-induced dysphagia, such

as the radiation dose to anatomical structures involved in swallowing dysfunction (e.g. the superior pharyngeal constrictor muscle) [12]. NTCP models can be used to estimate the risk of a given complication. Moreover, the most important dose volume parameters included in these NTCP-models can be used for treatment plan optimization, and thus to compare different radiation treatment plans in order to select the most optimal treatment.

Radiation doses to specific swallowing organs at risk (SWOARs) are main parameters for the calculation of NTCPs of dysphagia. NTCPs directly result from specific dose parameters of the SWOARs. However, if the delineation of SWOARs markedly differs from the guidelines used for NTCP-model development, the translation of the results of such models into routine clinical practice may be incorrect.

Recently, Christianen et al. [12] published delineation guidelines for SWOARs in head and neck radiotherapy that differ at some points from the definitions of SWOARs and subsequent delineation guidelines used by other investigators [4–11]. So far, the magnitude of these differences is still unclear, and the possible clinical relevance regarding differences in corresponding NTCPs remains to be determined.

* Corresponding author. Address: Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9300 RB Groningen, The Netherlands.
E-mail address: c.l.brouwer@umcg.nl (C.L. Brouwer).

Therefore, the main objective of the present study was to test the hypothesis that SWOAR delineations based on different delineation guidelines lead to differences in dose–volume parameters and subsequent NTCPs for dysphagia.

Materials and methods

Delineation guidelines and patients

For the purpose of the present study, the guidelines as proposed by Christianen et al. were used as a reference [3]. We decided to use this publication as a reference as it was the only one dedicated to the description of SWOARs delineation guidelines and because these guidelines were actually used in a subsequent publication that reported on the development of multivariate NTCP-models for different endpoints related to dysphagia [12]. This publication also included an overview of eight other guidelines for delineation of SWOARs that were published between 2000 and 2010 [4–11]. The following SWOARs were included in this overview: the pharyngeal constrictor muscles (PCMs), cricopharyngeal muscle and 'esophageal inlet muscle' (EIM) (which was previously described as '1 cm of the muscular compartment of the esophageal inlet' (10) and 'upper esophageal sphincter' (9)) and the glottic and supraglottic larynx. For the purpose of the current study, we extracted the definitions from the original papers and defined different delineation groups (DGs) by clustering the structures into groups with corresponding definitions (Table S1). These groups with corresponding definitions will be referred to as 'DG1', 'DG2', etc..

The information in Table S1 was confined to the definitions of the cranial and caudal borders of the SWOARs, since the definitions of these borders showed the largest variation. A detailed description of all remaining borders can be found elsewhere [3].

SWOARs were delineated in Pinnacle^3 v9.0 (Philips, Madison) in 29 sample patients from our clinic according to the different guidelines of the DGs, resulting in a total number of 899 contoured SWOARs. Contouring was performed by one observer (EG) and checked by two others (MK and RS). The contours according to all DGs of the SWOARs which are input to the studied NTCP-models [12] are shown in Fig. 1.

Patients were randomly selected from our previous cohort [12]. The set comprised 6 laryngeal, 4 hypopharyngeal, 1 oral cavity, 15 oropharyngeal and 3 nasopharyngeal patients [13]. Planning computed tomography (CT)-scans were acquired in supine position with a 2 mm slice thickness.

Geometric comparison

Geometric differences between the DGs were expressed as the Concordance Index (CI) of different DGs with the reference DG (DG1). The CI provides information on volume as well as on positional differences.[14] The CI is the ratio of the intersection (Volume1∩Volume2) and union (Volume1∪Volume2) volume of two delineated volumes. A CI of 1.00 indicates perfect overlap (identical structures), whereas a CI of 0.00 indicates no overlap at all.

Dosimetric comparison

Standard clinically acceptable photon intensity modulated radiation therapy (IMRT) treatment plans were available for all patients. Plans were reviewed and/or replanned by a single experienced dosimetrist (HPL) for the purpose of plan consistency. When replanning (plan adjustment) was performed, this was done to make sure that: (1) Coverage of the planning target volumes (PTV) was adequate (exactly 98% of the PTV should receive 95% of the prescribed dose); (2) The mean dose in the parotid glands was as low as possible; (3) The dose outside the PTV was reduced as much as possible (optimized dose conformity). No efforts were taken to specifically reduce the dose to the SWOARs [13]. Thus, the IMRT treatment plans were not influenced by the SWOARs delineations.

We studied the differences in mean doses in the SWOARs between the different DGs. For each patient the two DGs that resulted in the largest difference in mean dose (maxΔD) for a
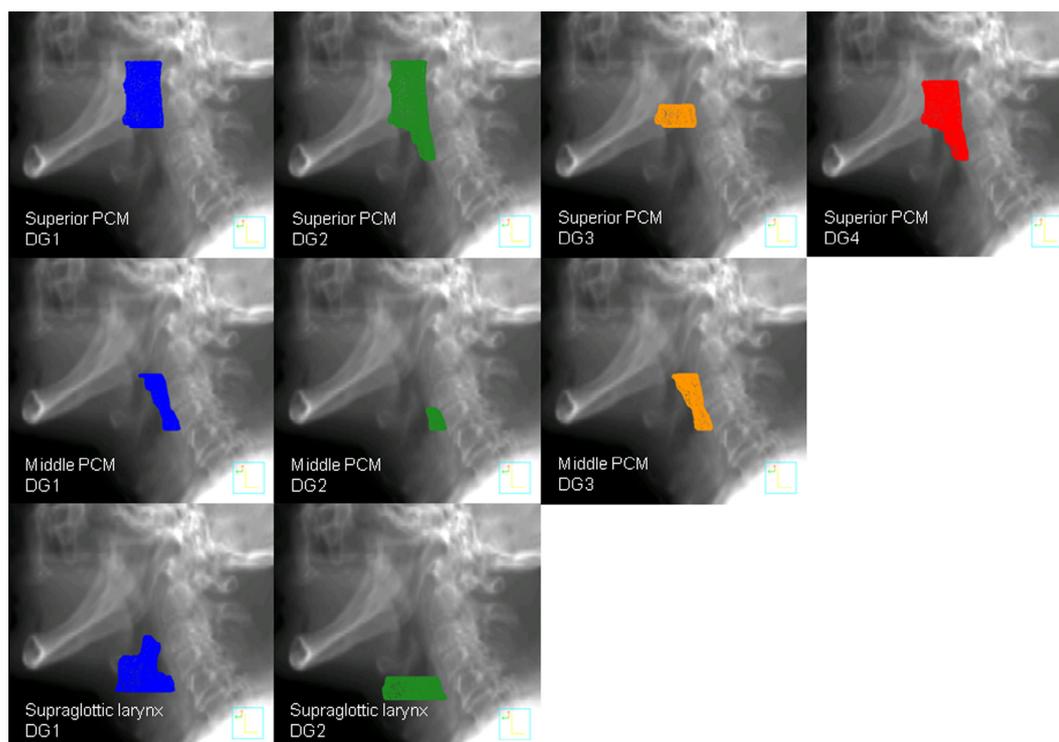

Fig. 1. Swallowing organs at risk (SWOARs) according to the different delineation groups (DGs) in one patient. Sagittal view. PCM = pharyngeal constrictor muscle.

particular SWOAR were selected. MaxΔD was averaged over all patients to obtain an average maxΔD per SWOAR. Estimates of the variability in this study are always reported as ±1 standard deviation (SD).

NTCP comparison

NTCPs were estimated for DG1 and DG2, in order to translate the differences in dose to differences in NTCPs. This will demonstrate the deviation from the model (ΔNTCP) in the situation of a clinical practice in which the contouring guidelines of DG2 are achieved, while the NTCP model belonging to DG1 is adopted. The analysis was confined to DG2 since it contained the most complete set of SWOARs' description in relation to DG1. Differences in the NTCPs between DG1 and DG2 (ΔNTCP) were calculated for each patient, based on four equations published by Christianen et al. [12]. The NTCP-models contained the endpoints:

– swallowing dysfunction grade 2–4 at 6 months after completion of radiotherapy, according to the RTOG Late Radiation Morbidity Scoring Criteria (1)
– patient-rated moderate-to-severe problems with swallowing solid (2), soft (3) and liquid (4) food

Table 1 lists the various parameters in the four different NTCP models. Radiation technique was IMRT for all patients in this study. Details on the NTCP calculation can be found in the Supplemental Material.

Results

Geometric Comparison

A statistically significant difference in SWOAR volume was observed between the different DGs ($p < 0.05$, two-way ANOVA, Table S2). Fig. S1 illustrates the CI of the different DGs reference to DG1 for each SWOAR. The average median CI value was 0.54 for the PCMs, 0.56 for the laryngeal structures and 0.07 for the cricopharyngeal muscle and EIM. For the cricopharyngeal muscle no overlap at all with DG1 was seen (CI = 0). CIs of a certain DG reference to DG1 varied between patients due to different anatomy and/or different flexion of the neck.

Dosimetric comparison

Differences in SWOAR mean dose between the DGs showed moderate to large variations (Fig. S2). Largest maxΔD was found for patient 11, for which the difference in mean dose to the PCM superior between DG1 and DG3 was 19.1 Gy. The average maxΔD of all SWOARs was 3.5 ± 3.2 Gy with the largest differences observed for the total PCM (6.0 ± 3.4 Gy), while differences for the glottic larynx (0.8 ± 0.9 Gy) remained limited.

NTCP comparison

Fig. 2 depicts ΔNTCP between DG1 and DG2 for the four NTCP-models studied. The mean absolute ΔNTCP over all patients

and complications was 2.3 ± 2.7%. Differences were related to patient's anatomy, posture and primary tumour site. Patients with tumours located in the oropharynx or nasopharynx showed higher NTCPs for the DG1-based SWOARs, while for patients with tumours located in the larynx and hypopharynx, the DG2-based SWOARs showed the highest NTCPs (grey vs. white bars in Fig. 2, respectively). This is mainly due to the larger overlap between the planning target volume (PTV) and the DG1-based SWOARs with respect to the DG2-based SWOARs for oropharynx/nasopharynx patients, and vice versa for larynx/hypopharynx patients. For two patients, the absolute ΔNTCP for at least one of the endpoints was larger than 10%. For patient 12 (primary tumour location in oropharynx), the mean dose to the supraglottic larynx according to DG1 was 70.5 Gy and for DG2 57.7 Gy (Fig. 3). The resulting ΔNTCP for RTOG grade 2–4 swallowing dysfunction was 11.6% (61.6 vs. 50.0%). For problems with swallowing solid food, ΔNTCP was 14.5% (47.3% vs. 32.8%). For the other patient (primary tumour located in oropharynx), ΔNTCP was 10.9% (35.0% vs. 24.1%) for the endpoint swallowing soft food (Fig. 2).

Discussion

This is the first study on the effect of variation in delineation guidelines on dose and subsequent NTCPs. We showed that dose parameters and corresponding NTCPs may vary widely depending on the definitions of the SWOARs. For the set of head and neck SWOARs included in the present study, the average maximal dose difference (maxΔD) was 3.5 ± 3.2 Gy. The translation of the dose variation to variation in NTCP for DG1 vs. DG2 resulted in a mean absolute ΔNTCP of 2.3 ± 2.7% (average over all patients and all four NTCP models studied). On average this seems a moderate difference, but it should be stressed that in individual cases ΔNTCP was much larger (>10%), which may lead to incorrect NTCP-predictions and possibly unjustified clinical decisions.

The magnitude of deviations from the reference volumes, dose, and subsequent NTCPs depended on patient's anatomy and posture, as well as on primary tumour site. The impact of the variation in patient anatomy was illustrated well in the box plots of the CI of Fig. S1 (large interquartile distances). This spread of CI values may be explained by the fact that for some patient anatomies and postures, the demarcations (e.g. certain bone and muscle structures) of different DGs may be more separated than for other cases. For example, patients with primary tumour sites located in the oropharynx or nasopharynx showed relatively large differences in NTCPs due to dose variation in the supraglottic larynx, while these differences were much smaller for laryngeal and hypopharyngeal cancers. According to DG1, the supraglottic larynx extends to the tip of the epiglottis, while according to DG2 the cranial border ends at the upper extension of the piriform sinus and aryepiglottic fold. Therefore, the overlap of the supraglottic larynx with the PTV in oropharyngeal cancer will generally be larger when using DG1 compared to DG2, resulting in higher dose values for DG1 compared to DG2 (Fig. 3). Therefore, the NTCP for patient-rated moderate-to-severe problems with solid, soft and liquid food (for which the model includes the mean dose to the supraglottic larynx)

Table 1
Parameters in the NTCP models for the four studied endpoints considering swallowing dysfunction [12].

NTCP model	Parameters						
	superior PCM	Middle PCM	Supraglottic larynx	EIM	Radiation technique	Age (y)	Tumour site
RTOG grade 2–4	Mean dose (Gy)		Mean dose (Gy)				
Solid food	Mean dose (Gy)		Mean dose (Gy)			18–65/65+	
Soft food		Mean dose (Gy)			3DCRT/IMRT	18–65/65+	Oropharynx, Nasopharynx/Hypopharynx, Larynx, Oral Cavity
Liquids			Mean dose (Gy)		3DCRT/IMRT		
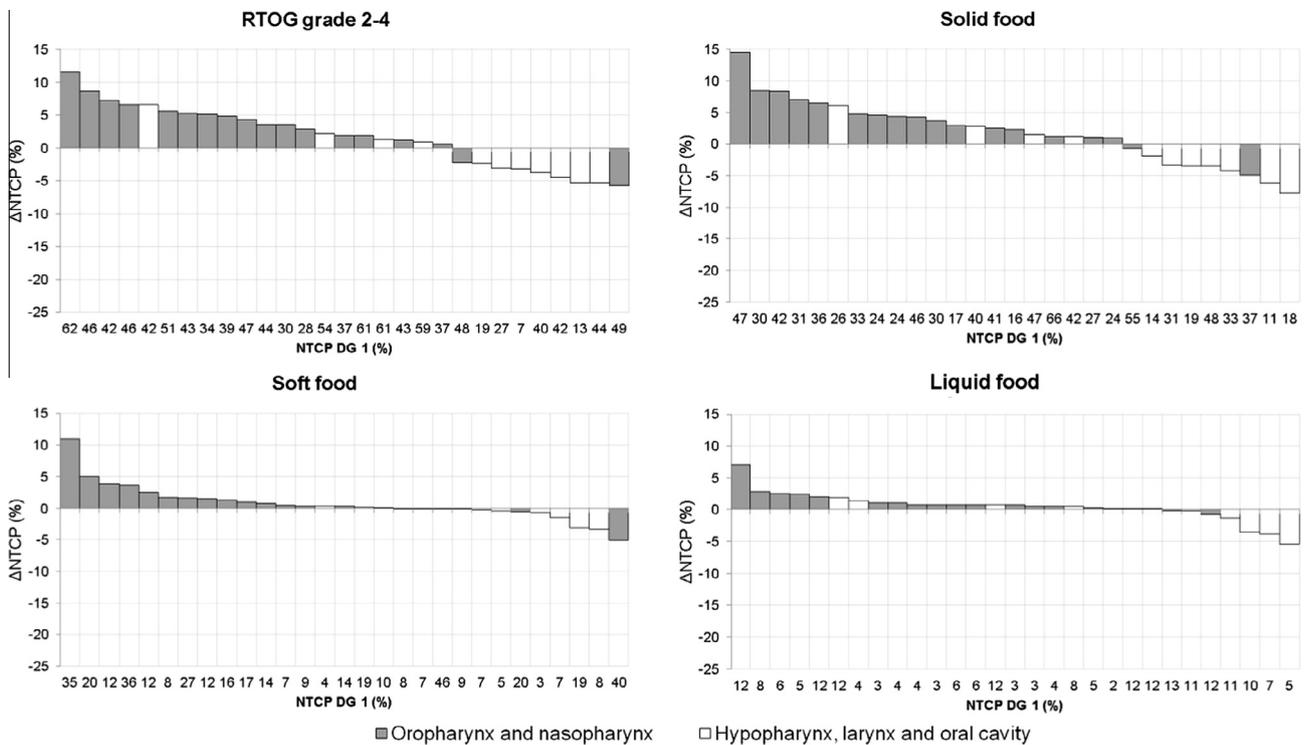
Fig. 2. Difference in normal tissue complication probabilities (ΔNTCP) between delineation group (DG) 1 and 2 for different complications [12]. ΔNTCP > 0 means underestimation and ΔNTCP < 0 means overestimation of the NTCP using DG2 in relation to DG1. Tumour location is indicated by grey/white filling of the bars.
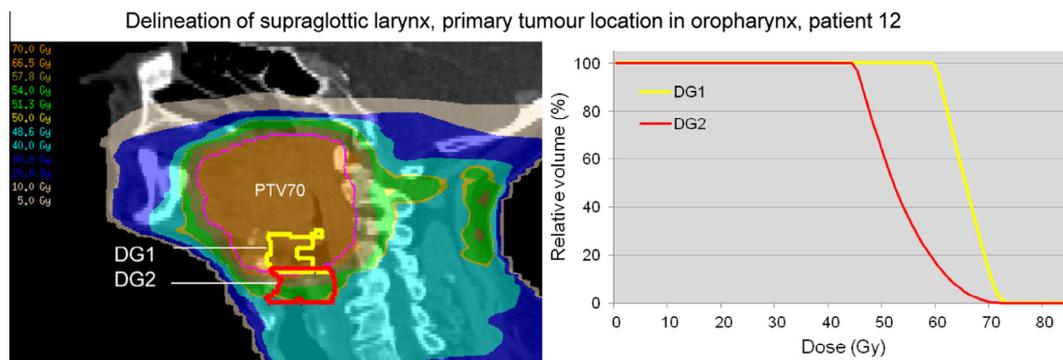
Fig. 3. Dose distribution (sagittal view) and dose–volume histogram of the supraglottic larynx for a patient showing large differences between DG (delineation group) 1 and 2.

was smaller for DG2 compared to DG1 (Fig. 2). For patient-rated moderate-to-severe problems with swallowing soft food, applying DG2 for contouring the middle PCM also resulted in underestimation of the NTCPs for patients with primary tumours located in the oropharynx in relation to DG1 due to less overlap of the PTV with the SWOAR using DG2.

The large differences in NTCPs in some individual patients emphasize the importance of uniform delineation guidelines. We propose to develop general consensus guidelines, which should be simple and unambiguously described. Probably, current delineation guidelines differ most because of different interpretation of anatomy, and different choices for (derived) structure borders. However, before we will be able to define a pragmatic set of simple delineation guidelines, we believe it is important to study dose–response relationships for swallowing problems more extensively and to understand the physiology of side effects, to be able to include the best predictive parameters in NTCP models. The (superior) pharyngeal constrictor muscles [15,16] and the supraglottic larynx [15] were, similar to our own research [12], recently associated with late radiation induced dysphagia. Besides,

De Ruyck et al. found that the rs3213245 (XRCC1) polymorphism was associated with radiation induced dysphagia [16]. Integrating biological and genetic (polymorphisms) information is promising to improve and individualize NTCP models.

Consensus meetings, multi-modality imaging, and the use of auto delineation tools could facilitate the introduction of uniform delineation guidelines [17,18]. The findings of this study may also have implications for the design of clinical trials, especially when radiation-induced dysphagia is a primary or secondary endpoint. In these cases (automated) review of delineations is recommended. Although there still may be differences resulting from interobserver variability, the concordance of head and neck OAR delineations within a guideline appears to be better than those between guidelines (results of this study) [19].

Feng and colleagues [20] reported on the effect of contouring variability and the resulting impact on IMRT treatment plan optimization in oropharyngeal cancer. A contouring variability up to 1.4 cm led to a 0.9 Gy mean difference between optimizations. We can, however, not compare the results of that study with our results, since these investigators studied variation in delineation

of repeated delineations by a group of experts, while the current study focussed on inter-guideline variation. Moreover, these investigators studied dose differences between optimizations (thus between different treatment plans) on different contours, while we studied the effect of using different guidelines for NTCP estimation within one treatment plan.

From a scientific point of view, it is important to externally validate NTCP-models developed in specific institutions, before they can be used in routine clinical practice. The results of the present study clearly illustrate that this external validation may be hampered by inconsistencies in delineation guidelines. This is particularly true for SWOARs with large dose variation and for NTCP-models for which the results are more sensitive to differences in contouring. Previous work has shown that the way we measure dysphagia (physician-rated, patient-reported, or objective measurements) is also of main importance for consistent NTCP modelling [21]. Therefore, clear definitions of organs at risk and endpoints are required to improve the external validity of NTCP-models.

The present study showed the consequences of not applying the matching input data to NTCP-models. In theory, all delineation guidelines would fit their own NTCP models. In practice however, multiple model versions should be constructed and validated, and this would also rule out pooling of dose–volume and follow-up data into large data sets to build a proper NTCP-model. We would therefore strongly advocate the use of uniform guidelines for NTCP-modelling studies as well as for studies on external validation and routine clinical practice.

In the current study, mean dose and corresponding NTCP differences between DGs were compared using IMRT plans that were not optimized based on the dose to the SWOARs, but particularly on the dose to the parotid glands. Therefore, the question arises what happens with the dose differences if the IMRT plans would be optimized for the different DGs. We expect the dose differences between the DGs to be similar or even larger when optimization on SWOARs would be performed, since dose gradients would be located closer to the SWOARs, resulting in larger dose differences between the different DGs. SWOAR optimization for different DGs was performed in two of our study patients, and results confirmed our presumption (see Supplemental Material II. for a case example).

Conclusion

The majority of the patients showed little differences in NTCPs for different delineation guidelines. However, large NTCP differences >10% were found in 7% of the patients. For correct use of NTCP models in individual patients uniform delineation guidelines are of great importance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2014.01.019.

References

[1] Nguyen NP, Frank C, Moltz CC, et al. Impact of dysphagia on quality of life after treatment of head-and-neck cancer. Int J Radiat Oncol Biol Phys 2005;61:772–8.

[2] Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, et al. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. J Clin Oncol 2008;26:3770–6.

[3] Christianen MEMC, Langendijk JA, Westerlaan HE, et al. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. Radiother Oncol 2011;101:394–402.

[4] Bhide SA, Gulliford S, Kazi R. Correlation between dose to the pharyngeal constrictors and patient quality of life and late dysphagia following chemo-IMRT for head and neck cancer. Radiother Oncol 2009;93:539–44.

[5] Caglar HB, Tishler RB, Othus M, et al. Dose to larynx predicts for swallowing complications after intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2008;72:1110–8.

[6] Caudell JJ, Schaner PE, Desmond RA, et al. Dosimetric factors associated with long-term dysphagia after definitive radiotherapy for squamous cell carcinoma of the head and neck. Int J Radiat Oncol Biol Phys 2010;76:403–9.

[7] Dirix P, Abbeel S, Vanstraelen B, et al. Dysphagia after chemoradiotherapy for head-and-neck squamous cell carcinoma: dose–effect relationships for the swallowing structures. Int J Radiat Oncol Biol Phys 2009;75:385–92.

[8] Feng FY, Kim HM, Lyden TH, et al. Intensity-modulated radiotherapy of head and neck cancer aiming to reduce dysphagia: early dose–effect relationships for the swallowing structures. Int J Radiat Oncol Biol Phys 2007;68:1289–98.

[9] Jensen K, Lambertsen K, Grau C. Late swallowing dysfunction and dysphagia after radiotherapy for pharynx cancer: frequency, intensity and correlation with dose and volume parameters. Radiother Oncol 2007;85:74–82.

[10] Levendag PC, Teguh DN, Voet P, et al. Dysphagia disorders in patients with cancer of the oropharynx are significantly affected by the radiation therapy dose to the superior and middle constrictor muscle: a dose–effect relationship. Radiother Oncol 2007;85:64–73.

[11] Li B, Li D, Lau DH, et al. Clinical-dosimetric analysis of measures of dysphagia including gastrostomy-tube dependence among head and neck cancer patients treated definitively by intensity-modulated radiotherapy with concurrent chemotherapy. Radiat Oncol 2009;4:52.

[12] Christianen MEMC, Schilstra C, Beetz I, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. Radiother Oncol 2012;105:107–14.

[13] Van der Laan HP, Christianen MEMC, Bijl HP, et al. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. Radiother Oncol 2012;103:76–81.

[14] Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin Oncol 2010;22:515–25.

[15] Mortensen HR, Jensen K, Aksglæde K, et al. Late dysphagia after IMRT for head and neck cancer and correlation with dose–volume parameters. Radiother Oncol 2013;107:288–94.

[16] De Ruyck K, Duprez F, Werbrouck J, et al. A predictive model for dysphagia following IMRT for head and neck cancer: introduction of the EMLasso technique. Radiother Oncol 2013;107:295–9.

[17] Steenbakkers RJHM, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. Int J Radiat Oncol Biol Phys 2006;64:435–48.

[18] Chao KSC, Bhide S, Chen H, 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.

[19] Brouwer C, Steenbakkers R, van den Heuvel E, et al. 3D variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32.

[20] Feng M, Demiroz C, Vineberg KA, et al. Normal tissue anatomy for oropharyngeal cancer: contouring variability and its impact on optimization. Int J Radiat Oncol Biol Phys 2012;84:e245–9.

[21] Eisbruch A, Kim HM, Feng FY, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. Int J Radiat Oncol Biol Phys 2011;81:e93–9.