Towards accurate target delineation for head and neck cancer

Lisanne Jager

❤

Layout by Maurits Fornier

'**Op weg naar nauwkeurige intekeningen voor hoofd-hals tumoren**'

ISBN 978-90-393-6778-0

The printing of this thesis was financially supported by: Elekta B.V., Bayer B.V., ChipSoft B.V., Cablon Medical B.V., Medical Consultancy at Work, the Michel Keijzer Fonds and the Department of Radiation Oncology of the University Medical Center Utrecht. The research project was financially supported by the KWF Cancer Foundation.

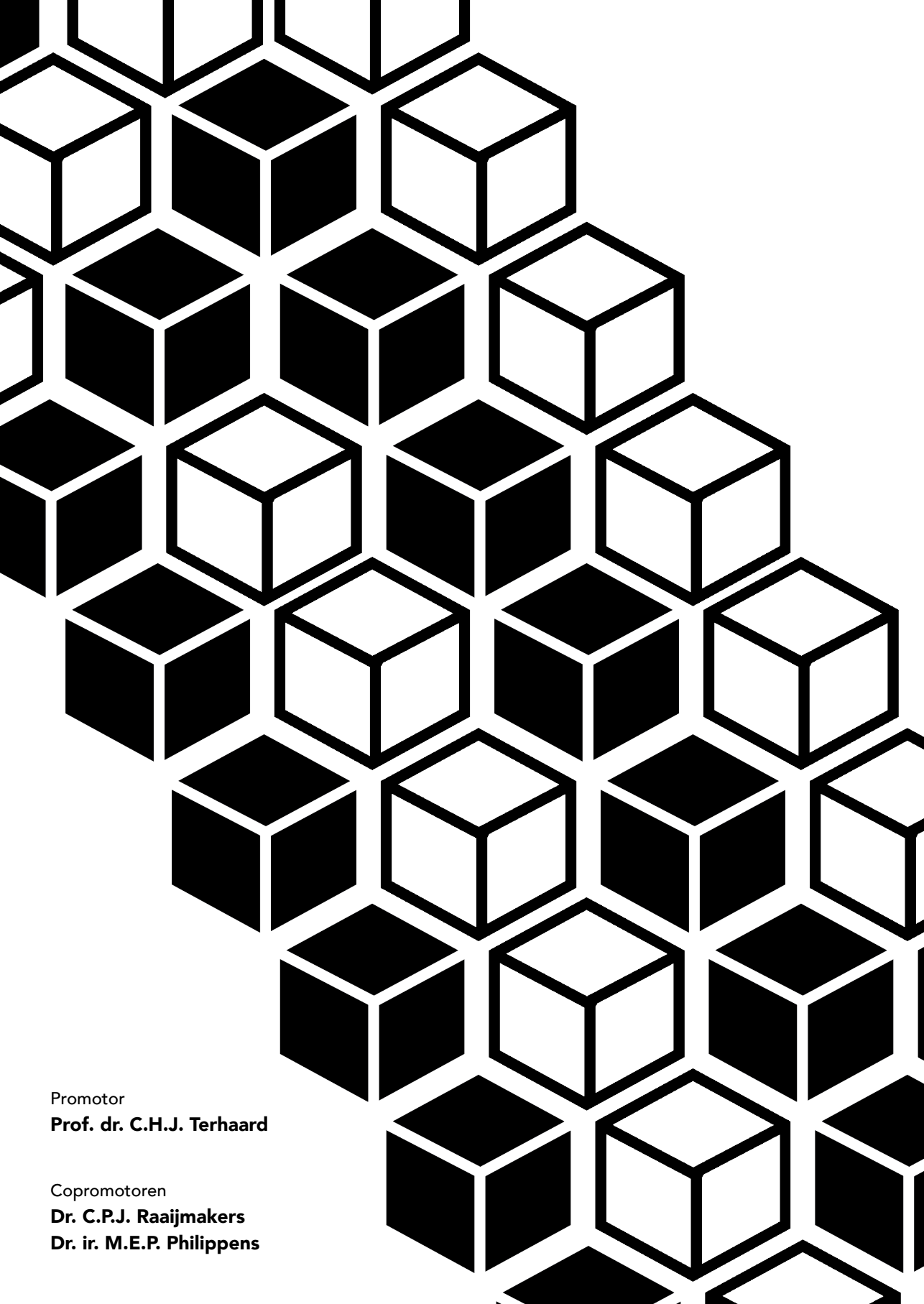
Proefschrift

Towards accurate target delineation for head and neck cancer

ter verkrijging van de graad van Doctor aan de Universiteit Utrecht
op gezag van de rector magnificus prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op donderdag 15 juni 2017 om 10.30 uur

Elise Anne Jager

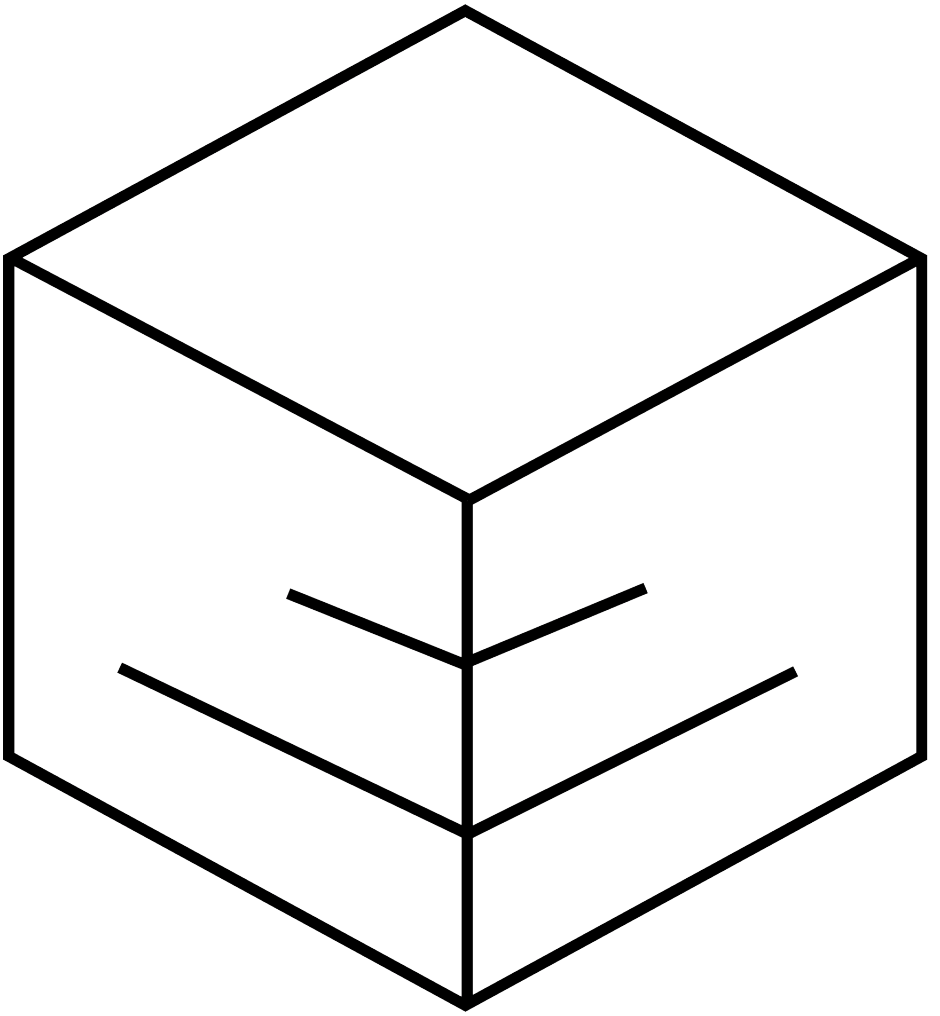
geboren op 22 april 1984 te Groningen



Promotor
Prof. dr. C.H.J. Terhaard

Copromotoren
Dr. C.P.J. Raaijmakers
Dr. ir. M.E.P. Philippens

Contents



Chapter 1
Introduction

Introduction

The past 30 years radiotherapy has developed in many ways to create the optimal treatment, delivering high dose to tumor cells while sparing healthy tissue. Major challenges in optimizing radiation treatment are: increasing visualization of the target with imaging, movement of the target, target planning, type of radiation/radiation techniques, and cell biology and radiation sensitivity. Geometrical uncertainties are introduced by equipment tolerances, setup error, organ motion, target volume delineation [1]. Especially the latter is a major source of error caused by variation between radiation oncologists who delineate the Gross Tumor Volume (GTV: Gross palpable or visible/demonstrable extent and location of malignant growth [2]) on images [3,4]. In clinical radiotherapy practice, this implies that the visible tumor volume on three-dimensional imaging is considered as the GTV. The volume that requires the highest radiation dose consists of the clinical target volume (CTV), which is the gross tumor volume (GTV) plus the inclusion of presumed microscopic spread. To extend the GTV to the CTV, a CTV margin is applied. This margin, which is highly variable in size, attempting to include the microscopic spread that is invisible on imaging because of its resolution below the limits of the resolution of modern radiology techniques. Most commonly used imaging modality for target delineation of head-and-neck cancer is Computed Tomography (CT) because it is widely available, fast, relatively cheap and can be used for treatment planning because CT images contain information on electron density, forming the basis for calculating three-dimensional radiation dose-distribution [5]. The past decennia various imaging modalities are developed and improved for their use in defining the target such as magnetic resonance imaging (MRI) and positron emission tomography (PET) [3,6-10]. In close relation to this, is the development of radiotherapy towards a precision technique, delivering high radiation doses to the tumor with tight treatment margins. Consequently, accurate three-dimensional target volume definition has become a crucial step. For an accurate and consistent definition of the target volume, validation of imaging modalities is fundamental, nevertheless this remains a major challenge in radiotherapy [1,11].

Target delineation for laryngeal and hypopharyngeal carcinoma

Medical imaging for radiotherapy purposes or diagnosis of cancer involves subjective interpretation of the various images. The visibility of the target, and consequently the GTV delineation, is dependent on the imaging modality and the data acquisition process i.e.: intermodality variation, whereas interpretation of the images, experience and appliance of delineation guidelines by the observer account for most of the interobserver variation [4,12,13]. The intraobserver variability is the consequence of day-to-day variations in subjective interpretation of images and clinical data of an observer and seems to be smaller than the interobserver variation [14,15].

Interobserver variation is high for head-and-neck cancer in comparison with other tumor sites [1]

and might have a greater impact on the accuracy of dose delivery than errors in treatment setup and organ motion [1,16]. Even with the introduction of new imaging techniques, the subjective interpretation is one of the main contributing factors to interobserver variation [4]. This subjective interpretation, leading to variations in target delineation, can have severe implications. Overestimation of the tumor will lead to damage to healthy tissue, whereas an underestimation may result in tumor recurrence [17].

The subjectivity is not only limited to defining the GTV, there are various interpretations concerning the extent of the microscopic involvement [14]. The estimate of the CTV margin can be hampered by a number of uncertainties. Microscopic extent can be highly variable depending on the tumor site and type, microscopic disease can be nodal, perineural, perivascular, intramuscular, intraparenchymal, perilymphatic, or bony extent. Although CTV margins would differ per tumor site, type and stage, the CTV margin that is clinically applied seems to be more of a historically determined margin that has developed with the clinical experience of an institution [13,18,19]. The size of the CTV margin differs between institutes and between observers within an institute or cooperative group [6,8-10,12,18,20] and usually varies between 0.5 mm and 2 cm [19,20] for CTV on CT images.

The CTV margin is an inconclusive definition of a barely validated margin which size is based on experience. Besides that, it does not take the variation in visibility of the target on various imaging modalities (intermodality variation) into account. Because GTV is subjected to these variations in visibility, modality based CTV margins are needed to correct for this modality dependent invisibility of the target.

One major attributing factor to the interobserver variation for GTV and CTV is claimed to be the lack of exact, unambiguous advice for outlining these volumes [7]. However, the influence of this factor could be reduced by the use of validated delineation guidelines [1,15,18,21,22].

Increasing visualization of the target with imaging

Development of imaging techniques and improving image quality of already existing techniques, are a field where expertise of nuclear medicine, radiology and radiotherapy join their forces. The choice of the adequate imaging modality and technique is of outstanding importance for the definition of target volumes. With CT currently being the imaging modality for GTV delineation and radiotherapy planning, other imaging modalities are gaining ground. MRI has a high contrast of soft tissue and might be better to detect cartilage invasion [23]. PET uses information of metabolic activity for identification of the tumor and might therefore be able to discriminate between tumorous and healthy tissue in the immediate vicinity. Because PET can be used with automatic delineation of the target, the effect of interobserver variation is absent. This implies more standardization and easier validation than manual GTV delineations. Conversely, automatic segmentation can very considerably with the acquisition and reconstruction method used. Besides, at this point are PET images often used in combination with CT images because PET lacks anatomical information that is needed for target delineation.

Target planning and radiation techniques for head-and-neck cancer

Target planning aims to give optimal dose to the tumor and minimize the dose given to healthy tissue. The past years, for head-and-neck cancer, three-dimensional conformal radiotherapy (3D-CRT) developed to intensity modulated radiotherapy (IMRT) with the development of a single beam to treatment from different positions to achieve the target as good as possible. IMRT uses non-uniform radiation beam intensities to maximize the delivery of radiation to the planning target volume by the use of volumetric

modulated arc therapy (VMAT) [24]. Proton therapy is an emerging treatment, which is still sparsely used for head-and-neck tumors. The mass of proton particles and their unique physical properties (ie, the Bragg peak) allow proton therapy to spare normal tissues distal to the tumor target from incidental irradiation. Initial observations show that proton therapy is particularly useful for treating tumors in challenging locations nearby non-target critical structures [25]. Image guided adaptive radiation therapy (IGRT) is designed for adjustment of the radiation treatment plan according to tumor size changes or normal organ shift during the course of treatment. Although treatment precision has developed incredibly throughout the last decennia, the treatment is only as good as the accuracy with which the target is known. Accurate radiation therapy involves knowing exactly where the tumor is at the time of treatment. As the famous Canadian medical physicist Harold Johns once said: "If you can't see it you can't hit it and if you can't hit it you can't cure it". This underwrites the importance of an accurate definition of the tumor.

'True tumor volume'

Hematoxylin and eosin (H&E) staining is one of the principal stains in histology. It is the most widely used stain for medical diagnosis and it is considered as the gold standard. The tissue of a specimen is transferred onto a glass slide and stained. In this thesis the glass slides with stained tissue is referred to as H&E-sections. For validation of images, histopathological results on H&E-sections are essential. However, this validation is a complex procedure and few studies have been performed in the research field of head-and-neck cancer [5,26,27]. Difficulties with processing the specimen and registering the histopathological results to in vivo images are serious obstacles that need to be overcome to draw conclusions.

Thesis outline

To investigate the variation between observers for GTV delineation using various imaging modalities, a delineation study on supraglottic laryngeal carcinoma has been performed where the use of CT images and CT images in combination with MR images were compared in chapter 2. This article is the basis of this thesis. It outlines the problem, determination of the variation between observers using various imaging modalities. Since interobserver variation remains large, even with combining imaging modalities, validation of these imaging modalities is a logical step. If the observers don't agree, the preciseness is low but what about the accurateness? For validation histopathological results are considered to be 'the truth'. In the third chapter, this gold standard has been investigated. How good is the gold standard? Variation between pathologists in tumor delineation on Hematoxylin and eosin (H&E) stained sections is determined in chapter 3. Since MRI is superior in showing contrast between soft tissues in comparison to CT, MRI can be of added value for the delineation of head-and-neck cancer, where various critical soft tissue structures are located close to each other. A next important step is to increase the preciseness and the accurateness of the delineation. Therefore, guidelines for delineation on MRI were developed and applied. GTV delineation of laryngeal and hypopharyngeal cancer on MRI with guidelines was compared to delineation according to current clinical practice in chapter 4. The 5th chapter of this thesis compares CT, MRI and PET for delineation of laryngeal and hypopharyngeal tumors and development of validated modality specific CTV margins. These CTVs are compared to the current clinically used CTV margin of 10 mm for GTV delineation on CT images.

References

1 Rasch C, Steenbakkers R, van Herk M. Target definition in prostate, head, and neck. Semin Radiat Oncol 2005;15:136-145.

2 Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT): Contents. J ICRU 2010;10:NP.

3 Anderson CM, Sun W, Buatti JM et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J Radiat Oncol 2014;1:006.

4 Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys 2008;33:136-140.

5 Daisne JF, Duprez T, Weynand B 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.

6 Daisne JF, Sibomana M, Bol A, Cosnard G, Lonneux M, Gregoire V. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 2003;69:237-245.

7 Khoo VS, Dearnaley DP, Finnigan DJ, Padhani A, Tanner SF, Leach MO. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother Oncol 1997;42:1-15.

8 Rasch CR, Steenbakkers RJ, Fitton I et al. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. Radiat Oncol 2010;5:21-717X-5-21.

9 Verduijn GM, Bartels LW, Raaijmakers CP, Terhaard CH, Pameijer FA, van den Berg CA. Magnetic resonance imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors. Int J Radiat Oncol Biol Phys 2009;74:630-636.

10 Thiagarajan A, Caria N, Schoder H et al. Target volume delineation in oropharyngeal cancer: impact of PET, MRI, and physical examination. Int J Radiat Oncol Biol Phys 2012;83:220-227.

11 Gregoire V, Coche E, Cosnard G, Hamoir M, Reychler H. Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience. Radiother Oncol 2000;56:135-150.

12 Jeanneret-Sozzi W, Moeckli R, Valley JF et al. The reasons for discrepancies in target volume delineation : a SASRO study on head-and-neck and prostate cancers. Strahlenther Onkol 2006;182:450-457.

13 Moeckli R, Sozzi WJ, Mirimanoff RO et al. Physical considerations on discrepancies in target volume delineation. Z Med Phys 2009;19:224-235.

14 Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlenther Onkol 2003;179:21-30.

15 Geets X, Daisne JF, Arcangeli S 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.

16 Hurkmans CW, Remeijer P, Lebesque JV, Mijnheer BJ. Set-up verification using portal imaging; review of current clinical practice. Radiother Oncol 2001;58:105-120.

17 Chen AM, Farwell DG, Luu Q, Chen LM, Vijayakumar S, Purdy JA. Marginal misses after postoperative intensity-modulated radiotherapy for head and neck cancer. Int J Radiat Oncol Biol Phys 2011;80:1423-1429.

18 Purdy JA. Current ICRU definitions of volumes: limitations and future directions. Semin Radiat Oncol 2004;14:27-40.

19 Caudell JJ, Meredith RF, Spencer SA, Keene KS, Dobelbower MC, Bonner JA. Margin on gross tumor volume and risk of local recurrence in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010;76:164-168.

20 Hong TS, Tome WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiother Oncol 2012;103:92-98.

21 Brouwer CL, Steenbakkers RJ, van den Heuvel E et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32-717X-7-32.

22 Brouwer CL, Steenbakkers RJ, Bourhis J et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015;117:83-90.

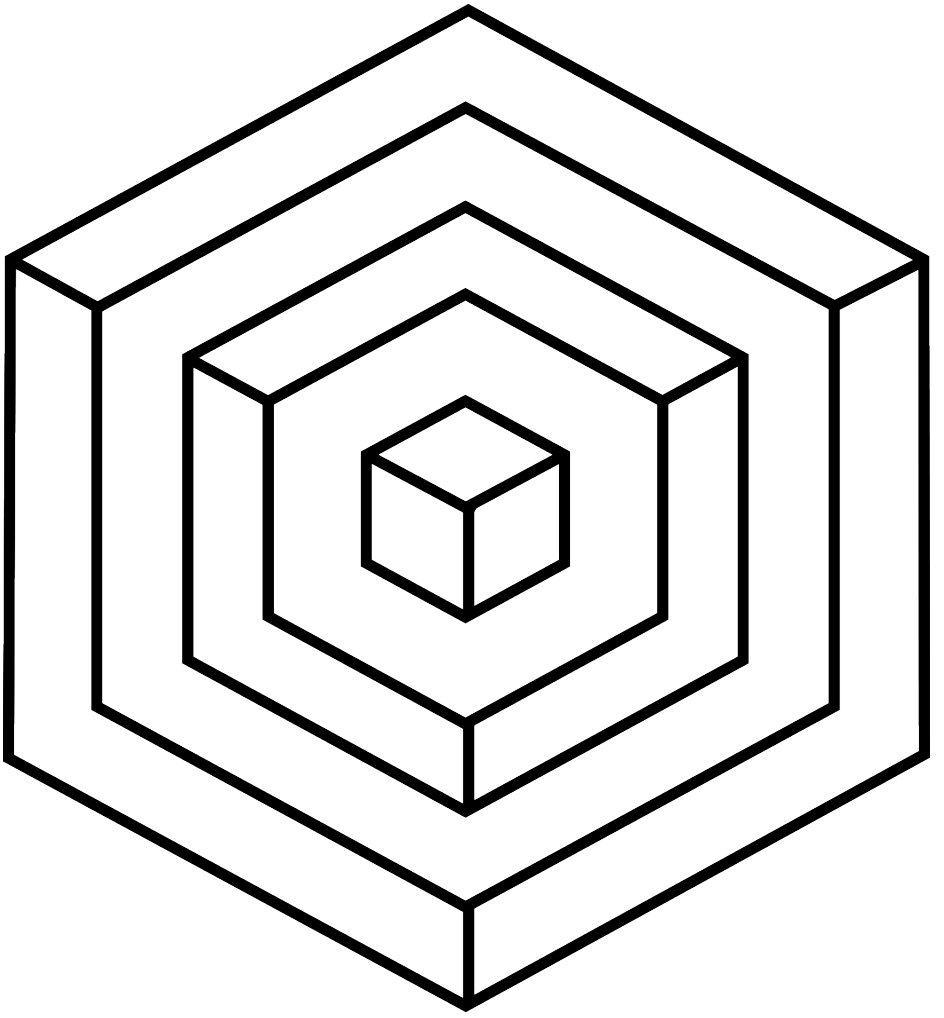
23 Becker M, Zbaren P, Casselman JW, Kohler R, Dulguerov P, Becker CD. Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis at MR imaging. Radiology 2008;249:551-559.

24 Lee N, Puri DR, Blanco AI, Chao KS. Intensity-modulated radiation therapy in head and neck cancers: an update. Head Neck 2007;29:387-400.

25 Holliday EB, Frank SJ. Proton radiation therapy for head and neck cancer: a review of the clinical experience to date. Int J Radiat Oncol Biol Phys 2014;89:292-302.

26 Caldas-Magalhaes J, Kasperts N, Kooij N et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 2012;82:e289-98.

27 Caldas-Magalhaes J, Kooij N, Ligtenberg H et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol 2015:1-7.



Chapter 2

GTV delineation in supraglotic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation

The aim of the study is to determine the interobserver agreement for GTV delineations of supraglottic laryngeal carcinoma on CT and on CT combined with MR images and to determine the effect of adding MR images to CT-based delineation on the delineated volume and the interobserver agreement.

Abstract

Background

GTV delineation is the first crucial step in radiotherapy and requires high accuracy, especially with the growing use of highly conformal and adaptive radiotherapy techniques. If GTV delineations of observers concord, they are considered to be of high accuracy.

Methods

Twenty patients with biopsy proven T1-T4 supraglottic laryngeal cancer, treated with curative intent were included. For all patients a contrast enhanced planning CT and a 1.5-T MRI with gadolinium were acquired in the same head-and-shoulder mask for fixation as used during treatment. For MRI, a two element surface coil was used as a receiver coil. Three dedicated observers independently delineated the GTV on CT. After an interval of 2 weeks, a set of co-registered CT and MR images was provided to delineate the GTV on CT. Common volumes (C) and encompassing volumes (E) were calculated and C/E ratios were determined for each pair of observers. The conformity index general (CIgen) was used to quantify the interobserver agreement.

Results

In general, a large variation in interobserver agreement was found for CT (range: 0.29-0.77) as well as for CT-MR delineations (range: 0.17-0.80). The mean CIgen for CT (0.61) was larger compared to CT-MR (0.57) ($p=0.032$). Mean GTV volume delineated on CT-MR (6.6 cm^3) was larger compared to CT (5.6 cm^3) ($p=0.002$).

Conclusion

Delineation on CT with co-registered MR images resulted in a larger mean GTV volume and in a decrease in interobserver agreement compared to CT only delineation for supraglottic laryngeal carcinoma.

Background

Radiotherapy for head-and-neck cancer can give rise to severe acute and late side effects [1-4]. To minimize damage to healthy tissues on one hand and eradicate macroscopic tumor on the other hand, the gross tumor volume (GTV) should be determined as accurate as possible. This is especially required when applying intensity-modulated radiation therapy (IMRT) and position verification, to maximize the benefits of high-precision radiation techniques using smaller radiation fields [5].

Various studies [6-10] have been performed, using different imaging modalities, to determine the agreement among observers when delineating the GTV in head-and-neck cancer. Interobserver agreement and interobserver variability (disagreement) are often used in the same context whereas these terms express the opposite. Generally, interobserver agreement is used to assess the quality of an image modality to visualize the tumor. Thus, when there is more agreement among the observers, the image modality is assumed to be more precise and even more accurate in visualizing the tumor, although high accuracy can only be assessed by pathology.

For delineating the GTV and treatment planning in head-and-neck cancer, Computed Tomography (CT) is the imaging modality of first choice in most cases [11,12]. The advantages of CT are that it is widely available, does not cause geometrical distortion and has intrinsic information on the relative electronical density of the various tissues used for dose calculation algorithms [11]. Where CT offers excellent bony detail, magnetic resonance imaging (MRI) uses various sequences to visualize soft tissue and bone contrasts. Especially the capability of MRI to visualize soft tissues is an improvement compared to CT, therefore permitting better definition of disease extent and organs at risk [12-14]. Because MRI does not carry intrinsic information on electronic density, it is currently precluded as sole imaging modality in clinical use for radiotherapy treatment planning in head-and-neck tumors [11,12]. Various studies demonstrated superior soft tissue contrast on MRI compared to CT [6,9,15,16]. Although there is agreement on the capacity of MRI to increase visibility of soft tissue structures in head-and-neck oncology, there is no agreement on the value of MRI for determination of the GTV and its influence on the interobserver agreement [7,8,10,11].

The aim of this study is to compare the interobserver agreement between delineations on CT and on CT with co-registered MR images in supraglottic laryngeal carcinoma and to determine the value of adding MR images to the 'gold standard' CT images.

Methods

Patient selection

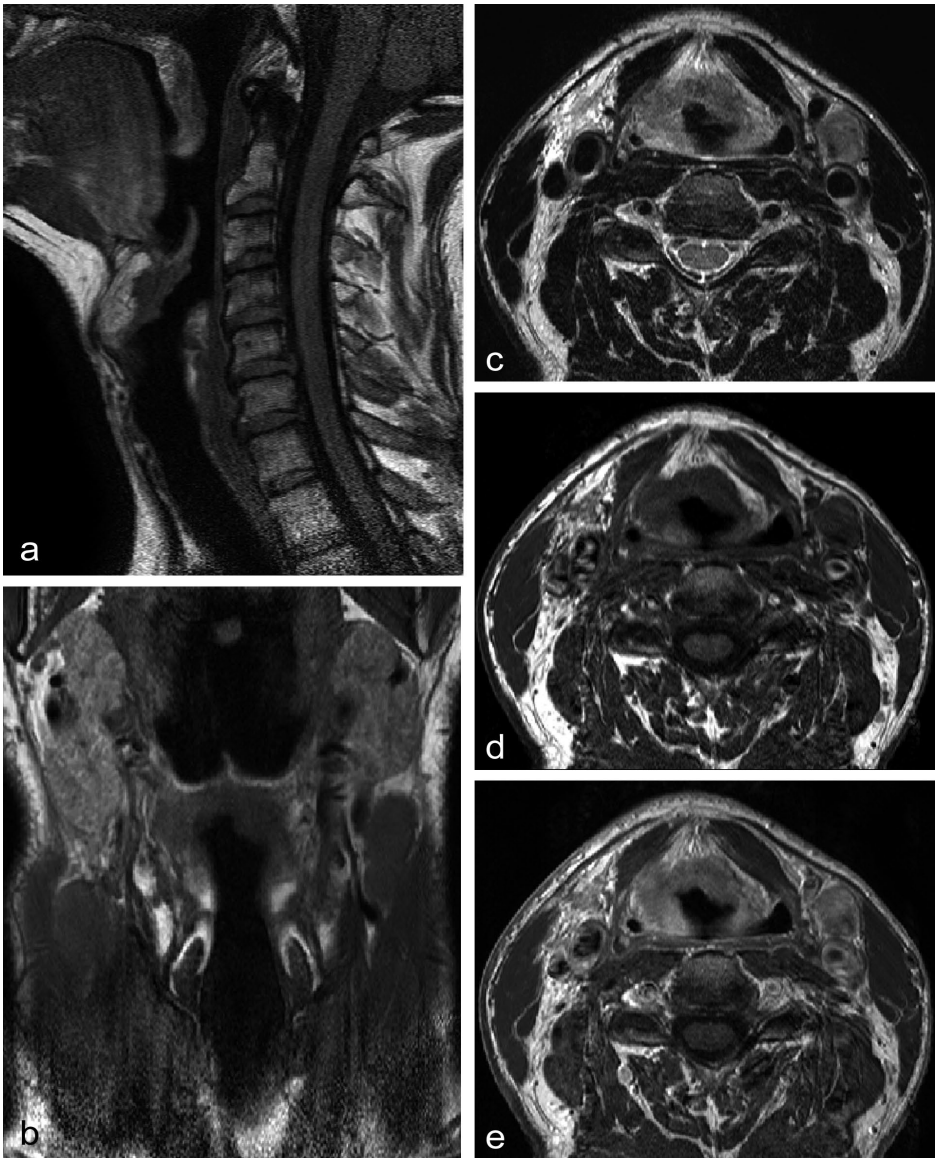
Twenty patients with biopsy proven T1-T4 supraglottic laryngeal cancer (squamous cell carcinoma, SCC) and treated with high-dose radiotherapy with curative intent at our institution between November 2005 and October 2009 were included in this study.

From a database of 120 patients with laryngeal and hypopharyngeal cancer, 39 patients fulfilled the criteria of inclusion. Which were; patients with a supraglottic tumor, the availability of a contrast enhanced CT scan and a MRI with gadolinium performed in a radiotherapy mask. Twenty patients were randomly selected from this group, Initial clinical assessment of tumor stage was performed based on triple-endoscopy under anesthesia, contrast enhanced CT scan, and indirect laryngoscopy to assess mobility of the vocal cords. The study group consisted of five female and 15 male patients with a mean age of 64 years (range: 40-80 yr).

Imaging technique and data acquisition

CT and MR imaging was performed prior to radiotherapy treatment and in radiotherapy position. Patients were immobilized in a radiotherapy mask (five-point head-and-shoulder mask,

Figure 1. MR images of the larynx acquired in a radiotherapy mask. a:T1-weighted sagittal view, b: T1-weighted coronal view. Transversal views: c: T2-weighted, d: T1-weighted, e: T1-weighted + gadolinium.



Posicast PR5; Civco, Reeuwijk, The Netherlands) and received a CT scan from the base of the skull to the carina after intravenous injection of iodinated contrast. The CT images were obtained by two different CT machines. Fifteen patients were scanned with a single slice Philips Aura machine and five patients with a Philips Big Bore Brilliance (multi-slice CT). Images were acquired with helical scans. A slice thickness of 2 mm and 3 mm, and a pitch of 1.0 (Philips Aura) and 0.7 (Philips Big Bore Brilliance) was used. Axial images were acquired using a matrix size of 512 × 512, with a pixel spacing of 0.95 × 0.95 mm^2 - 1.19 × 1.19 mm^2. After a mean interval of six days (range: 0-13 days) the patients underwent a 1.5 Tesla MRI scan (Achieva; Philips Medical System, Best, the Netherlands) in the same fixation device as used for CT scanning, combined with a two-element flexible surface coil [17,18]. For every patient T1-weighted images were obtained in transversal, sagittal and coronal directions as well as transversal T2-weighted and T1-weighted after injection of gadolinium according to our clinically used MR protocol for imaging the larynx and the hypopharynx. A 512 acquisition and reconstruction matrix was used. The field-of-view diameter was 210 mm and the slice thickness was 3 mm. An example of the acquired MR images is shown in *Figure 1*. The registration was performed by a medical physicist who defined a rectangular box containing the GTV and surrounding bony structures. The rigid registration was performed using the mutual information algorithm within this box and the registration was visually controlled. This procedure is according to clinical practice at our department. No approval of an ethics committee was needed according to Dutch law.

Delineation of GTV

Gross tumor volume (GTV) was defined as the macroscopic (gross) extent of the primary tumor that is demonstrable on the imaging modality e.g, MRI scan, CT scan. The following guidelines for delineation of the GTV were agreed upon by the three observers at a consensus meeting in advance of the delineations sessions. Areas of doubt had to be included in the GTV according to radiotherapy practice. Edema around the tumor had to be included and evident stasis of saliva had to be excluded in the delineation. Criteria for soft tissue infiltration were: left-right asymmetry, contrast enhancement and fatty space infiltration. For cartilage invasion on CT the following signs were used: osteolysis of dense mineralized areas (if in contact with the primary tumor), cortical erosions, abnormal increased asymmetrical density and presence of tumor on both sides of bony/cartilaginous structures. Sclerotic cartilage with an intact cortex was not to be included in the GTV. Guidelines for interpretation of neoplastic invasion of laryngeal cartilages, as defined by M. Becker et al. [19], were used during delineation on MRI.

Three dedicated and MRI-trained head-and-neck specialists (two radiation oncologists and one radiologist) respectively called observer a, b and c, independently delineated the GTV. They started with CT images at fixed window/level 350/50 with minor adjustments of 10-20 HU based on individual preferences.

After an interval of more than two weeks, to avoid possible bias due to recall of the previous delineation, the same CT (without previous contours) was delineated with the co-registered MR images (T1w, T1w+Gd, T2w) simultaneously visible.

The observers received an anonymised triple-endoscopy report and were instructed to record: delineation time, window/level and which anatomical parts of the larynx were involved by tumor, during delineating on CT and CT-MR. Observers were also asked to subjectively rate image quality (good, moderate, poor and not assessable) and tumor detectability [20]. The latter was scored as followed; 0, if tumor boundaries were not visible, 1: tumor is visible, boundaries not, 2: boundaries are visible but not clear, 3: tumors as well as boundaries are clearly visible.

Volumetric analysis and interobserver agreement

All GTVs were delineated in Volumetool [21], a software application which is capable of simultaneous visualization of multiple three-dimensional datasets. The volume of the GTV was determined by multiplying the number of voxels contained within a contour by the size of the voxel. The size of the voxel depends on the resolution of the image reconstruction and the slice thickness. If the center of the voxel is within the contour boundary, the voxel is regarded as being part of the volume.

For each pair of observers, the common volume (C; the volume that is part of both GTVs of one patient) and the encompassing volume (E; volume encompassing both GTVs of one patient) of the delineated GTVs for each patient, were automatically calculated. C/E ratios (Jaccard coefficient) were determined for each pair of observers (observer a&b, a&c, b&c). The Jaccard coefficient can only be used as a conformity index for a situation in which two delineated volumes are compared. Therefore we used the conformity index general (CIgen) to quantify the agreement between all observers. This index is independent of the number of observers or delineated volumes [22]. CIgen is defined as the sum of the common volumes of the various observer pairs divided by the sum of the encompassing volumes of these pairs and is written for the 3 observers as the following formula:

$$CI_{\text{gen}} = \frac{(a \cap b) + (a \cap c) + (b \cap c)}{(a \cup b) + (a \cup c) + (b \cup c)}$$

A CIgen of 1.00 indicates perfect overlap (identical delineations), whereas a CIgen of 0.00 indicates no overlap at all.

Clinical impact

Since the GTV is extended using margins to correct for several factors such as microscopic disease, movement and setup inaccuracies, the PTV is considerably larger than the GTV.

For each patient the GTV delineations were extended with a margin to create a planning target volume (PTV). Two scenarios were investigated according to the work of Vugts et al. [23]. In one scenario conventional margins were applied (PTVclinical). In the other scenario tight margins were investigated (PTVtight). A margin of 8 mm was used for PTVtight and 15 mm for PTVclinical. As a 'worst case scenario', the largest GTV was assumed to be the correct GTV. Subsequently, it was determined in how many cases this GTV was not covered by the PTVs.

Statistical analysis

Based on non-normality of the samples according to the Shapiro Wilk test, the Wilcoxon signed-rank test was applied for statistical comparison of the mean delineated volume (GTV) between CT and CT-MR.

The coefficient of variation (COV), defined as COV=standard deviation (SD)/mean volume, was determined for all delineated GTVs of the patients for each imaging modality. For each modality correlation between mean GTV volume and COV was measured using a Spearman rank correlation test.

A Student paired t-test was used for the comparison of the CIgen on CT and CIgen on CT-MR. These samples were both normally distributed according to the Shapiro Wilk test. For each modality the correlation between CIgen and GTV volume was measured using a Spearman rank correlation test. Statistical analyses were performed with SPSS 16.0 using a (alpha) level of significance of 0.05.

Results

Image quality was considered 'good' for the majority of the CT images as well as for MR images. For some patients the image quality of the CT scan was deteriorated by contrast insuf-

ficiency or due to swallowing. Movement due to swallowing had an adverse effect on MR image quality. However, the image quality was never considered to be 'not assessable', nor did the observers unanimously qualify the image quality as being 'poor'.

For the CT of 16 patients, at least one observer recorded that the tumor borders were 'visible but not clear' or worse (grade 2, 1 and 0). For MRI, this was the case for 11 patients. In eight patients at least one of the observers recorded explicit difficulties in the cranial and/or caudal direction on CT and in four patients referring to the MRI. These recorded difficulties in cranial and caudal direction were objectified by larger discrepancies (smaller common volumes) in the delineations in the cranial and caudal areas and in de region of the epiglottis for CT as well as for MRI in nearly all patients.

Volumetric analysis

The difference between GTV volume on CT and on CT-MR was considerably larger in two cases (patient 10 and 15) (*Table 1*) compared to the difference for the remaining patients.

The median GTV volume of the three observers on CT-MR (median: 6.6 cm^3, interquartile range: 2.9-9.2, 95% confidence interval: 4.8 - 11.8 cm^3) was significantly larger (p=0.002) compared to median of the GTV volume on CT (median: 5.6 cm^3, interquartile range: 2.5 - 7.9, 95% confidence interval: 4.2-9.8 cm^3) (*Table 1*).

A large variation in COV was found between the GTV per modality as well as between the modalities (*Table 1*). The mean COV on CT (0.17, SD 0.12) was comparable to CT-MR (0.19, SD 0.16) (*Table 1*).

No relation between COV and the mean GTV volume for any of the imaging modalities was observed (CT: rho=-0.28 p=0.23 CT-MR: rho=0.05 p=0.84).

Interobserver agreement

In general, a large variation in interobserver agreement was found for the delineated tumors on CT as well as for CT-MR delineations (*Table 1*, *Figure 2 and 3*).

The mean CIgen for CT was significantly larger (0.61, SD 0.12, range 0.29-0.77, 95% confidence interval: 0.56-0.67) compared to CT-MR (0.57, SD 0.15, range 0.17-0.80, 95% confidence interval: 0.50-0.64) (p=0.032).

Although the smallest CIgens were observed for the smallest tumors (*Table 1*), no relation between CIgen and the mean GTV volume for CT as well as for CT-MR delineations was observed (CT: rho=0.28 p=0.24, CT-MR: rho=0.31 p=0.18).

Clinical impact

When applying tight margins, for 12 of the 20 patients the largest GTV contour was not covered by all the PTVs. When using clinical margins this number decreased to two of the 20 patients. The anatomical sites where the GTV contour was not encompassed by the PTV contours were mostly in cranial and caudal direction.

Discussion

The present study on supraglottic laryngeal carcinoma demonstrates that adding MR images to CT resulted in a decrease in interobserver agreement compared to the interobserver agreement of the CT-only delineation-session. Furthermore, the median GTV volume was larger on CT-MR compared to CT although there was no relation found between the GTV volume and the CIgen. Subjectively, the observers reported an increased visibility of anatomical details on MRI.

According to other studies based on head-and-neck cancer where MRI was compared with CT, Ahmed et al. [6] demonstrated that the delineated GTV volume for base of tongue tumors on MRI was almost two times larger compared to CT. They also reported a superior subjective

visualization and delineation of base of tongue tumors on the MRI scans relative to CT. Several other studies concluded the same for tongue and floor of the mouth cancer [15,16] and nasopharyngeal carcinoma [9].

Although there is agreement on the capacity of MRI to increase visibility in head-and-neck oncology, there is no agreement on the value of MRI for determination of the GTV. Rasch et al. [10] showed better interobserver agreement with matched CT-MRI, for target delineation in nasopharynx cancer compared to CT alone. A large improving factor on the interobserver agreement was the decision to include entire anatomical structures invaded by tumor. A previous study done by Rasch et al. [7] reported that for six patients with advanced head-and-neck carcinoma, the delineated GTVs and interobserver agreement was better for delineations on MRI (with CT images available) than on CT (with MR images available). However, no difference between one single observers' mean GTV volume delineated on CT and on MRI for oropharyngeal, laryngeal and hypopharyngeal tumors was found by Daisne et al. [11]. Additionally, a study performed by Geets et al. [8] showed no clinical advantage of MRI over CT in terms of volume determination and interobserver agreement for pharyngo-laryngeal tumors. Concerning the design of the study, this study [8] was the only one that, to some extent, resembled ours. However, we were not able to adequately compare our findings with results from the mentioned study because MRI was used without CT for delineating. Furthermore, the use of a different metric to quantify the interobserver agreement, based on area of overlap between contours, hampers a detailed comparison. In general, a wide variety of metrics is used to quantify the interobserver agreement in delineation studies for example: Dice similarity coefficient, common to encompassing volume ratio and Jaccard index [22, 24, 25].

Since the GTV is extended by margins to correct for several factors such as microscopic disease, movement and setup inaccuracies, the PTV is larger than the GTV.

Our analysis indicates that large conventional margins partly compensate for the interobserver variation. However, when evidence-based tight margins are applied the interobserver variation for delineating the GTV might result in inadequate dose coverage of the GTV.

Tumor recurrence was diagnosed for two patients in this study. Due to the development in radiotherapy treatment schedules and tumor treatment planning between 2005 and 2009 we are not able to draw conclusions from this finding concerning treatment outcomes. Furthermore, the treatment plan was based on the delineation from the treating radiation oncologist while the delineations in this study were used for research purposes.

In our study a dedicated MR protocol for radiotherapy GTV delineation was applied. This protocol has been used at our department since 2005. Care was taken to optimize the MR image quality for radiotherapy purposes [17]. For the majority of the patients, MR image quality was considered 'good'. However, the introduction of 3.0 Tesla and recently 7.0 Tesla MRI scanners and the development of new fast scan protocols might further optimize MR image quality.

A shortcoming of the studies of Daisne [11] and Geets [8] was the use of a multipurpose body-coil as receiver coil for MR imaging. Rasch [7] used a head coil for MR imaging only without a mask or external markers, causing a decrease in image quality.

Although, in our study, the observers reported a subjectively increased visibility of anatomical details on MRI compared to CT, this did not improve agreement between observers. On the contrary, interobserver agreement was decreased. Apparently, additional MRI information resulted

in more options to interpret the imaging data, resulting in a greater variation in delineations and an increase in delineated volumes. The inclusion of areas of doubt in the GTV, as described in our delineation guidelines, further increased these variations and volumes. In our opinion, the increased visibility of anatomical details on MRI might be of value in radiotherapy practice when it is clear how to combine the information of different MR-sequences when delineating the GTV. To maximize the benefits of high-precision radiation techniques, the gross tumor volume (GTV) should be determined as accurate as possible. Clear guidelines for interpretation and GTV delineation of laryngeal carcinoma could therefore be very useful. To develop these guidelines, a validation-study with total laryngectomy specimens is currently being performed at our institution. In that study, tumor tissue is identified based on pathological findings and compared with GTV delineations on different image modalities [26].

The large variation in interobserver agreement for the GTVs delineated on CT as well as for CT-MR delineations (*Table 1*) suggests that for some tumors it was more difficult to delineate the GTV compared to others. In some cases, this might have been influenced by a mode rately decreased image quality. In our opinion, this variation was mostly caused by differences in location and characteristics of the tumor, and difficulties to distinguish tumor borders.

For the two T4 stage tumors, the differences between the delineated volumes on CT and CT-MR were the largest. This might be explained by the presence of edema that is increased in larger tumors and which could cause an increase in delineated volume on CT-MR since MRI is superior in visualizing soft tissues (e.g. edema) [12-14]. The observers also included more cartilage in their GTV on CT-MR compared to CT only. Besides the capacity of MRI to increase visibility of soft tissue, MRI might have an improved visibility for cartilage invasion compared to CT. Research performed by Becker et al. supports this presumption [14, 19]. Since there was no histopathological data available for this study we are not able to further investigate this finding.

The CT images used in this study were obtained on 2 different CT scanners and slice thickness varied between 2 and 3 mm. This did not influence the results since there were no remarkable differences between the CIgens comparing the two scanners. Besides, no difference in image quality and no specific matching related problems were reported.

Conclusions

The interobserver agreement was decreased in the CT-MR session compared to the CT only delineations and mean delineated volume on CT-MR was larger compared to CT. At this point MRI has no objective added value concerning the CIgen outcomes. The increased visualization of anatomical details on MRI might lead to an increased interobserver agreement and more accurate GTV estimation only when clear guidelines for interpretation and delineation of MR images of laryngeal tumors are present.

Authors

Elise Anne Jager, Nicolien Kasperts, Joana Caldas-Magalhaes, Mariëlle E.P. Philippens, Frank A. Pameijer, Chris H.J. Terhaard and Cornelis P.J. Raaijmakers.

Published

Radiation Oncology (2015) 10:26.

Table 1. Gross tumor volumes delineated by 3 observers on CT and CT-MR in supraglottic laryngeal carcinoma

Patient	Tumor stage	Mean GTV (cm^3)		CIgen		COV	
		CT	CT-MR	CT	CT-MR	CT	CT-MR
8	T1	1.2	1.3	0.52	0.54	0.23	0.11
6	T2	1.5	1.9	0.29	0.32	0.41	0.66
12	T3	1.6	1.5	0.62	0.67	0.18	0.03
16	T3	1.9	2.8	0.43	0.17	0.38	0.13
20	T2	2.3	2.8	0.61	0.41	0.14	0.38
14	T2	2.9	3.2	0.73	0.68	0.07	0.05
9	T2	5.0	5.1	0.65	0.61	0.17	0.25
4	T2	5.3	5.2	0.62	0.50	0.19	0.41
1	T3	5.5	6.7	0.76	0.66	0.04	0.06
19	T2	5.5	6.6	0.63	0.54	0.04	0.06
5	T2	5.7	6.7	0.49	0.49	0.32	0.30
11	T2	6.8	6.5	0.68	0.69	0.18	0.14
13	T2	7.0	6.8	0.77	0.80	0.10	0.03
7	T2	7.1	7.3	0.75	0.73	0.04	0.12
2	T3	7.9	9.6	0.70	0.59	0.09	0.27
18	T3	7.9	8.1	0.59	0.57	0.10	0.08
17	T3	11.1	11.8	0.67	0.62	0.05	0.15
10	T4	12.8	22.1	0.49	0.50	0.36	0.35
15	T4	14.7	20.6	0.69	0.68	0.09	0.16
3	T3	27.0	29.1	0.60	0.66	0.26	0.11

Tumor stage, mean delineated tumor volumes (GTV) in cm^3 of 3 observers ranked by mean GTV. COV (coefficient of variation) and CIgen (conformity index general) for GTV on CT and CT-MR.

Figure 2. Delineations of three observers on contrast enhanced CT and CT-MR. Delineations of three observers on contrast enhanced CT (a, c, e) and CT-MR (b, d, f) (T1-weighted). Transversal (a, b), sagittal (c, d) and coronal (e, f) views.

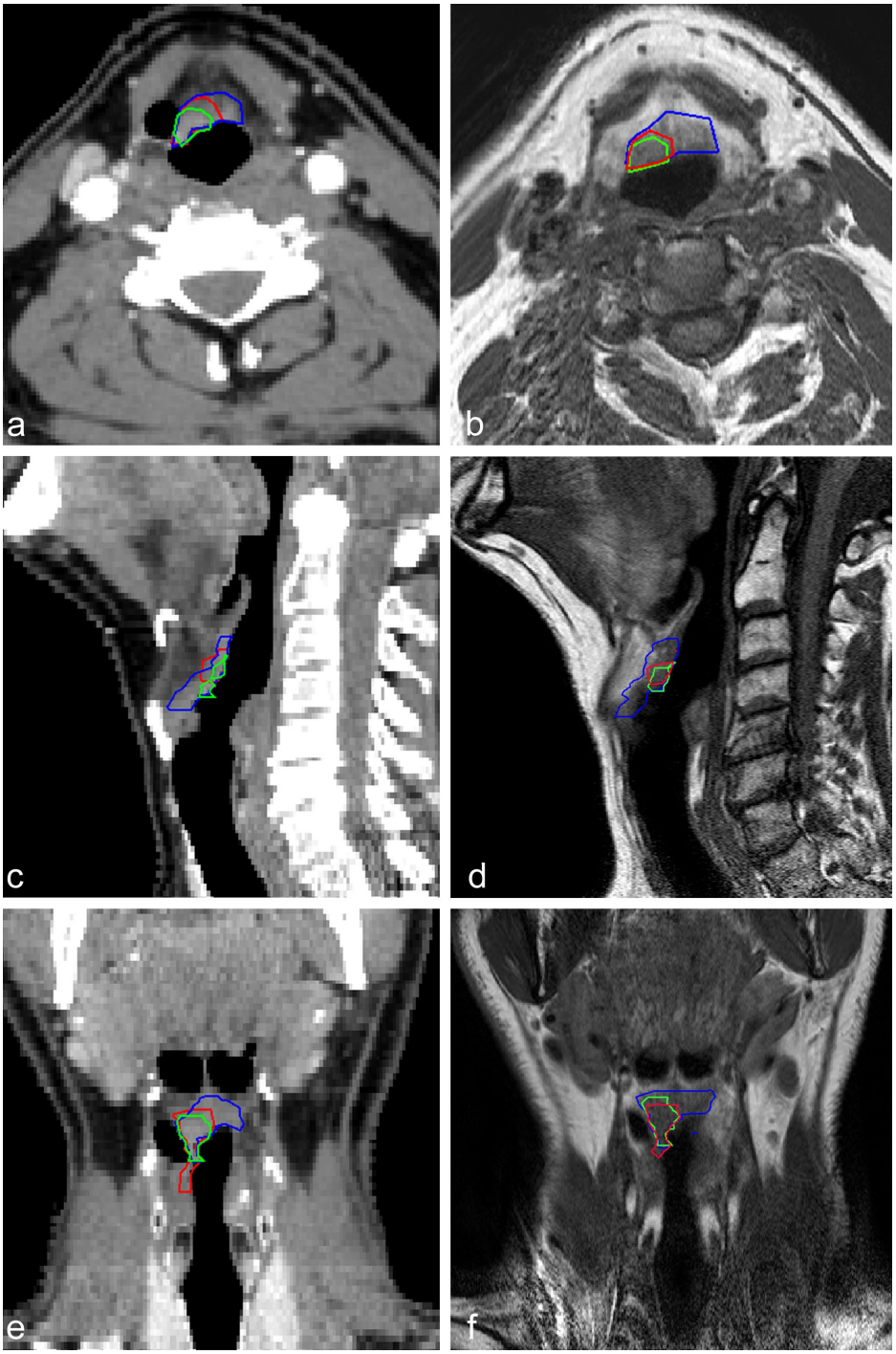
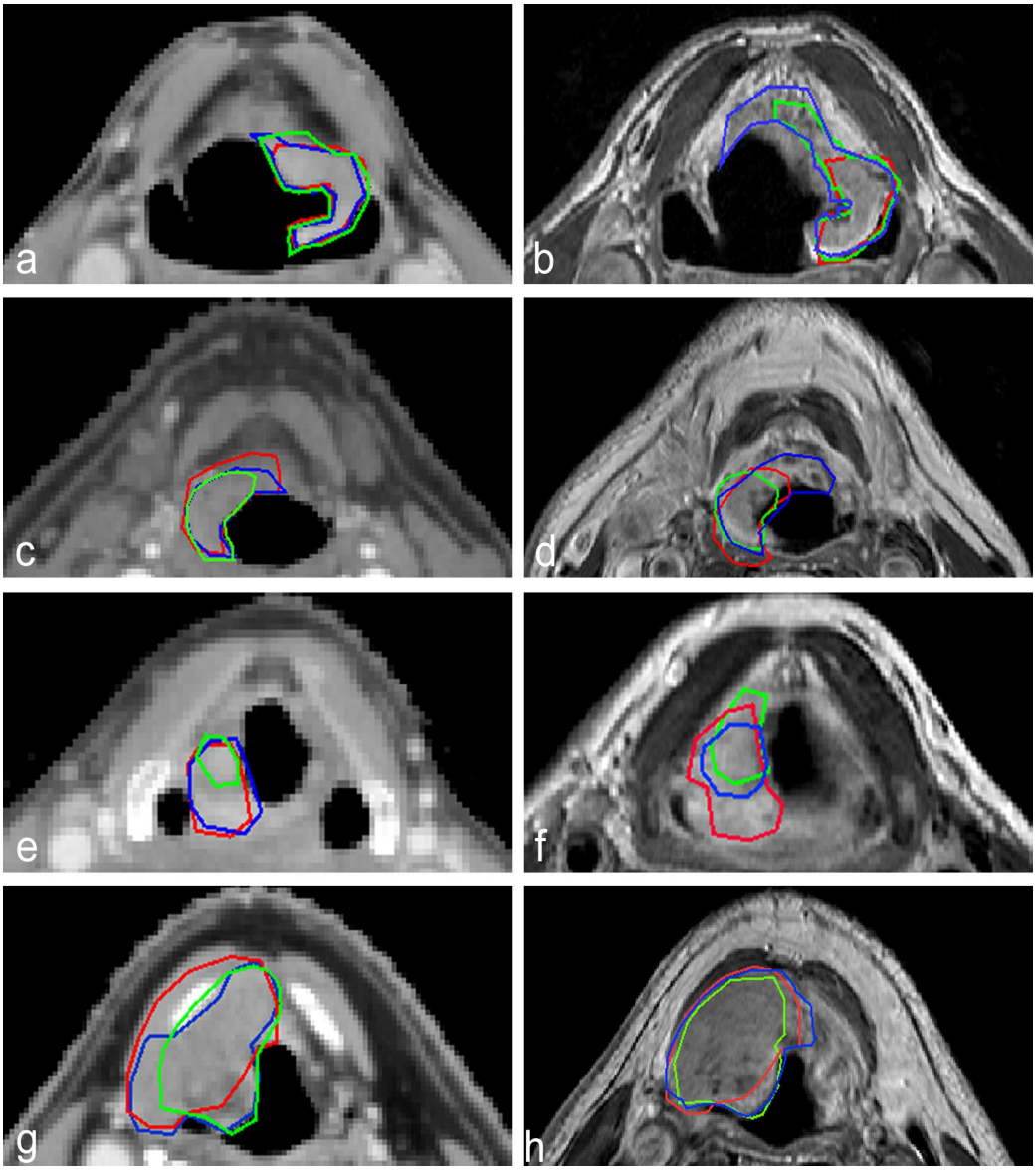


Figure 3. Delineations of three observers on contrast enhanced CT and CT-MR. Delineations of three observers on contrast enhanced CT (a, c, e, g) and on CT-MR (b, d, f, h) (T1-weighted + Gd) for four different patients.



References

1 Meyer F, Fortin A, Wang CS, Lui G, Bairati I: Predictors of severe acute and late toxicities in patients with localized head-and-neck cancer treated with radiation therapy. Int J Radiat Oncol Biol Phys 2012,82(Suppl 4):1454-1462.

2 Zackrisson B, Mercke C, Strander H, Wennerberg J, Cavallin-Ståhl E: A systematic overview of radiation therapy effects in head and neck cancer. Acta Oncol 2003,42(Suppl 5-6):443-461.

3 Caglar HB, Tishler RB, Othuis M, Burke E, Li Y, Goguen L, Wirth LJ, Haddad RI, Norris CM, Court LE, Aninno DJ, Posner MR, Allen AM: Dose to larynx predicts for swallowing complications after intensity-modulated radiotherapy. Radiat Oncol Biol Phys 2008,72(Suppl 4);1110-1118.

4 Dijkema T, Raaijmakers CPJ, Braam PM, Roesink JM, Monninkhof EM, Terhaard CH: Xerostomia:a day and night difference. Radiother Oncol 2012,104(Suppl 2):219-223.

5 Lambrecht M, Nevens D, Nuyts S. Intensity-modulated radiotherapy vs. parotid-sparing 3D conformal radiotherapy. Strahlenther Onkol 2013,189(Suppl 3):223-229.

6 Ahmed M, Schmidt M, Sohaib A Kong C, Burke K, Richardson C, Usher M, Brennan S, Riddell A, Davies M, Newbold K, Harrington KJ, Nutting CM: The value of magnetic resonance imaging in target volume delineation of base of tongue tumors – A study using flexible surface coils. Radiother Oncol 2010,94(Suppl 2):161-167.

7 Rasch C, Keus R, Pameijer FA Koops W, de Ru V, Muller S, Touw A, Bartelink H, van Herk M, Lebesque JV: The potential impact of CT-MRI Matching on tumor volume delineation in advanced head and neck cancer. Int J Radiat Oncol Biol Phys 1997,39(Suppl 4): 841-848.

8 Geets X, Daisne J-F, Arcangeli S Coche E, De Poel M, Duprez T, Nardella G, Grégoire V: 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(Suppl 1):25–31.

9 Chung NN, Ting LL, Hsu WC, Lui LT, Wang PM: Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck 2004,26(Suppl 3):241-246.

10 10] Rasch CR, Steenbakkers RJ, Fitton I Duppen JC, Nowak PJ, Pameijer FA, Eisbruch A, Kaanders JH, Paulsen F, van Herk M: Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. Radiat Oncol 2010,5;1-8.

11 Daisne J-F, Duprez T, Weynand B Lonneux M, Hamoir M, Reychler H, Grégoire V: Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR Imaging, and FDG Pet and validation with surgical specimen. Radiology 2004,233(Suppl 1):93-100.

12 Khoo VS, Dearnaley DP, Finnigan DJ, Padhani A, Tanner SF, Leach MO: Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother Oncol 1997,42(Suppl 1):1-15.

13 Castelijns JA, Hermans R, Brekel van den MWM, Mukherji SK: Imaging of laryngeal cancer. Semin Ultrasound CT MR 1998,19(Suppl 6):492-504.

14 Becker M, Zbären P, Laeng H, Stoupis C, Porcellini B, Vock P. Neoplastic invasion of the laryngeal cartilage. Comparison of MR imaging and CT with histopathologic correlation. Radiology. 1995;194 Suppl 3:661–9.

15 Sigal R, Zagdanski AM, Schwaab G, Bosq J, Auperin A, Laplanche A, et al. CT and MR imaging of squamous cell carcinoma of the tongue and floor of the mouth. Radiographics. 1996;16 Suppl 4:787–810.

16 Lam P, Au-Yeung KM, Cheng PW, Wei WI, Yuen AP, Trendell-Smith N, et al. Correlating MRI and histologic tumor thickness in the assessment of oral tongue cancer. AJR Am J Roentgenol. 2004;182 Suppl 3:803–8.

17 Verduijn GM, Bartels LW, Raaijmakers CP, Terhaard CH, Pameijer FA, van den Berg CA. Magnetic Resonance Imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors. Int J Radiat Oncol Biol Phys. 2009;74 Suppl 2:630–6.

18 Webster GJ, Kilgallon JE, Ho KF, Rowbottom CG, Slevin NJ, Mackay RI. A novel imaging technique for fusion high-quality immobilised MR imaged of head and neck with CT scans for radiotherapy target delineation. Br J Radiol. 2009;82 Suppl 978:497–503.

19 Becker M, Zbären P, Casselman JW, Kohler R, Dulguerov P, Becker CD. Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis of MR Imaging. Radiology. 2008;249 Suppl 2:551–9.

20 Murakami R, Baba Y, Furusawa M, Nishimura R, Nakaura T, Baba T, et al. Early glottic squamous cell carcinoma. Predictive value of MR imaging for the rate of 5-year control with radiation therapy. Acta Radiol. 2000;41 Suppl 1:38–44.

21 Bol GH, Kotte ANTJ, Van der Heide UA, Lagendijk JJ. Simultaneous multi-modality ROI delineation in clinical practice. Comput Methods Programs Biomed. 2009;96 Suppl 2:133–40.

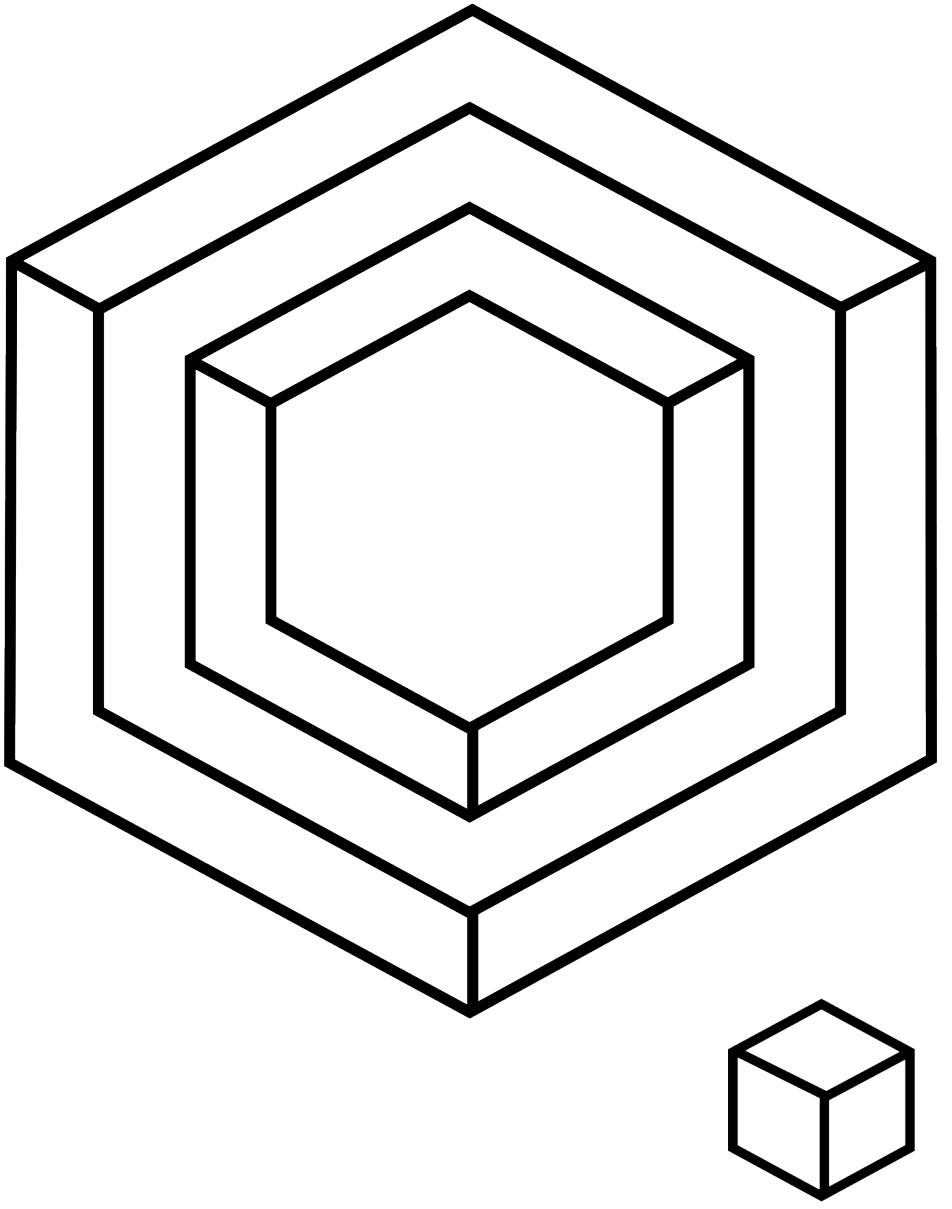
22 Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. Phys Med Biol. 2009;54 Suppl 9:2863–73.

23 Vugts CA, Terhaard CH, Philippens ME, Pameijer FA, Kasperts N, Raaijmakers CP. Consequences of tumor planning target volume reduction in treatment of T2-T4 laryngeal cancer. Radiat Oncol. 2014;9:195.

24 Fontina I, Lütendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol. 2012;188 Suppl 2:160–7.

25 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 Suppl 7:515–25.

26 Caldas-Magalhaes J, Kasperts N, van den Kooij N, Berg CA, Terhaard CH, Raaijmakers CP, et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys. 2012;82 Suppl 2:289–98.



Chapter 3
Interobserver variation among pathologists for delineation of tumor on H&E-sections of laryngeal and hypopharyngeal carcinoma. How good is the gold standard?

Cancer imaging has to deal with subjective interpretation of various imaging modalities. In radiotherapy, this subjective interpretation causes an interobserver variation when determining the gross tumor volume. Pathology is considered 'gold standard' for target delineation studies. However, there might also be a variation between pathologists. The aim of this study is to investigate the accuracy of the gold standard.

Introduction

In tumor delineation studies for radiotherapy, histopathology is used for validation purposes [1–7]. Validation of tumor delineation is a complex procedure and relatively few studies have been performed in the research field of head-and-neck cancer [3,5,8]. in these studies, whole mount sections of laryngectomy specimens were obtained.The tumor was delineated by a pathologist and used to validate various imaging modalities, e.g. computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) for their ability to distinguish tumor tissue. For the interpretation of validation studies, the variation of tumor outline on histopathology is crucial, as it is used as gold standard for tumor delineation in clinical imaging studies as currently performed by our institute [3,5,9]. However, a study on the reproducibility of tumor outline is missing. The aim of this study is to determine the variation of tumor delineation among pathologists on H&E-sections for laryngeal and hypopharyngeal carcinoma to quantify the uncertainties in the gold standard in the context of imaging validation studies for laryngeal and hypopharyngeal carcinoma.

Material and methods

Ten patients were randomly selected from a database of 22 patients enrolled in the imaging-validation study performed at our institute [3]. This study was approved by the Ethics committee of the University Medical Center Utrecht, the Netherlands and informed Consent was given by the patients included in this study.

All tumors from patients selected for this study were T3 or T4 squamous cell carcinoma of the larynx or hypopharynx, eligible for surgical resection (*Table 1*).

Three dedicated head-and-neck pathologists manually delineated carcinomatous tissue on the whole mount hematoxylin and eosin (H&E) stained sections. Overlap and distance analyses were performed.

Table 1. Tumor characteristics per patient.

Patient	Primary Site	T-Stage
1	hypopharyngeal	T4
2	hypopharyngeal	T3
3	supraglottic	T4
4	supraglottic	T3
5	supraglottic&glottic	T4
6	supraglottic	T4
7	transglottic	T4
8	glottic	T4
9	hypopharyngeal	T4
10	supraglottic	T4

Tumor site and tumor stage for each patient.

Delineation on H&E-sections

H&E-sections were obtained from whole mount sections of the laryngectomy-specimens according to the procedure used in our imaging-validation study. The sets of H&E-sections of the whole specimens of 10 patients consisted of in total 279 H&E-sections. The interval between the H&E-sections obtained from the specimens was 3 mm.Three head-and-neck pathologists from different institutions independently delineated tumor tissue on the H&E-sections using a permanent marker pen. The pathologists were blinded to each others' results. One pathologist used magnifying glasses and a light microscope for areas of doubt. The other two pathologists used a light microscope. No guidelines for specic magnications and settings were given.The observers were instructed to manually draw a line around the tumor tissue on the H&E-sections including cartilage invasion but excluding any positive lymph nodes. After delineation, the sections, including the tumor outlines, were scanned at 300 dpi resolution. For separate digitization of the lines drawn by the pathologist an in-house developed software package [10], used in clinical radiotherapy practice, was applied to manually trace the tumor outline by a researcher. The digitized lines were then projected on scanned H&E-sections without delineations. Re-evaluation of H&E-sec-

tions with remarkable discrepancies between delineations of the pathologists was performed. The delineations used in the analysis were not adjusted after re-evaluation.

Observation parameters

Volumetric analysis

The volumes of the delineated tumors were determined by multiplying the number of voxels contained within a contour by the size of the voxel. This value was multiplied by 3 mm which is the interval between the sections. Mean volumes and standard deviations were calculated.

Interobserver variation

In order to quantify the variation between the observers, the generalized conformity index (CIgen) was calculated [11]. CIgen is defined as the sum of the common volumes of the various observer pairs divided by the sum of the encompassing volumes of these pairs (observer A&B, A&C, B&C) and is defined for three observers as:

$$CI_{gen} = \frac{(a \cap b) + (a \cap c) + (b \cap c)}{(a \cup b) + (a \cup c) + (b \cup c)}$$

The common volume (CV) is the volume that is part of all individual delineations for one patient. The encompassing volume (EV) is the volume encompassing all individual delineations for one patient. A CIgen of 1.00 indicates perfect overlap (identical delineations, 100% agreement), whereas a CI of 0.00 indicates no overlap at all. This index is independent of the number of observers or delineated volumes [11].

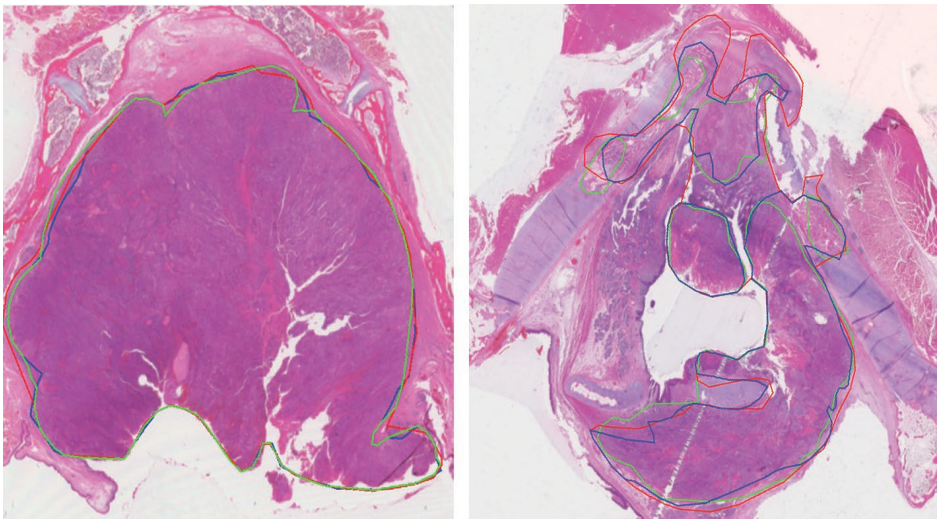
Analysis of variation in distance

From the three contours a common and an encompassing contour were derived. For each H&E-section the distance for each point on the common contour to the closest point on the encompassing contour was calculated. The root mean squares (RMS) of these distances were calculated.

Statistical analysis

The correlation between the mean size of the

Figure 1. H&E stained sections obtained from a laryngectomy-specimen with tumor delineations of the three pathologists. Good agreement between observers is perceived on the left section. The right H&E stained section of a tumor with irregular shaped areas showed larger interobserver variation.



tumor and variation between the delineated volumes for that tumor was analyzed with the Spearman's rank correlation test (two-tailed) which was also used to analyze the correlation between delineated tumor volumes and the CIgen. The comparison between the overlap of the various observer pairs was analyzed with related samples Friedman's two-way analyses of variance ranks. The Wilcoxon-signed ranks test was applied for analyzing the distribution of the delineated tumor volumes between observers.

Results

in total 124 of the 279 H&E-sections were delineated by the pathologists resulting in 372 delineations.

In general the agreement between observers appeared high. However, for several H&E-sections considerable variation between delineations was observed (*Figure 1*).

Volumetric analysis and interobserver variation

The mean delineated volume by the three observers was 12.95 cm^3 (SD 0.3, range 3.0–39.8). The variation between the delineated volumes per patient was approximately 2% and was not related to the size of the tumor (rho=-0.32, p=0.37) (*Table 2*).

One pathologist (observer A) delineated a larger volume in eight of the 10 cases compared to the other observers (*Table 2*). The distribution of the delineated tumor volumes was significantly different between observer A and B (p=0.022) and observer A and C (p=0.007). No difference between observer B and C was observed (p=0.74).

The mean interobserver agreement, expressed as the generalized conformity index, was 0.87 (SD 0.04, range 0.82–0.95) (*Table 2*).

There was no systematic difference in overlap (CV/EV) between the observer pairs (p=0.42) and no correlation between delineated tumor volumes and CIgen was observed (rho=0.37, p=0.30).

Re-evaluation H&E-sections with discrepant tumor delineations

Ten sections with remarkable discrepancies between pathologists were re-evaluated by all three observers in order to determine if the variation was due to interpretation, if tumor was overlooked and/or if the observers would come to the same conclusion after re-evaluation. The criteria for re-evaluation were: discrepant areas larger than 0.8 cm^2 and if one or more isolated delineated areas which were not delineated by the other observers. From this re-evaluation it was concluded that most discrepancies were based on the interpretation of the extension of cartilage invasion and the in- or exclusion of parts of the necrotic cartilage (*Figure 2*).

After re-evaluation all the observers agreed that these parts of the cartilage were necrotic as a consequence of tumor invasion. Difficulties to distinguish a lymph node with metastatic disease from the main tumor mass caused larger variation on two H&E-sections. For two cranial sections, small fragments of the tumor were clearly overlooked by two pathologists.

Analysis of variation in distance

Ninety-five percent of the measured distances between the encompassing and the common lines were smaller than 2.0 mm (SD 0.7) and in 90% smaller than 1.4 mm (SD 0.4) (*Table 3*).

The 5% of the calculated distances larger than 2.0 mm were mostly found in irregularly shaped tumor areas and in cartilage with tumor invasion (*Figure 1*). The RMS of the distances between the common and the encompassing contours amounted to 1.0 mm.

Figure 2. H&E stained sections obtained from a laryngectomy-specimen with tumor delineations of the three pathologists. For this specimen the largest distances between delineations were observed due to exclusion of thyroid cartilage by one of the observers. Distances up to 9.9 mm were measured.

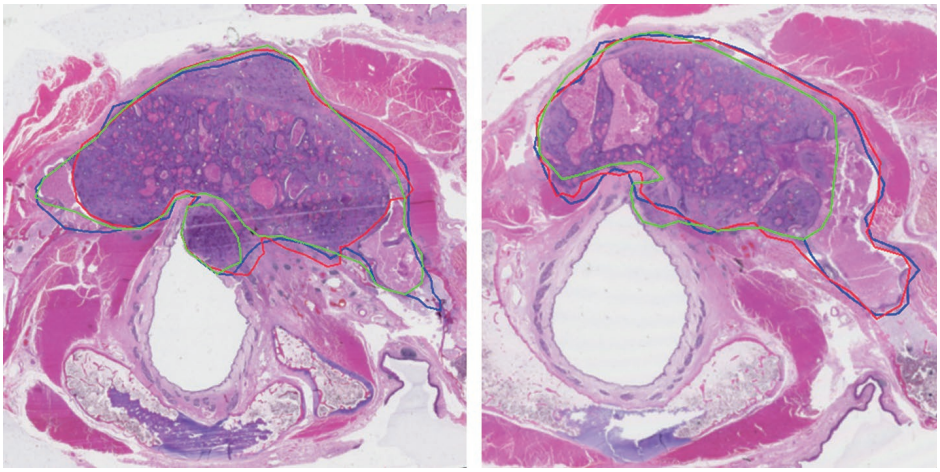


Table 2. Delineated tumor volumes per pathologist and interobserver variation according to the generalized conformity index.

Patient	Observer A (cm^3)	Observer B (cm^3)	Observer C (cm^3)	Mean (cm^3)	SD (cm^3)	RSD(%)	CIgen
2	3.10	2.93	2.91	2.98	0.10	3.50	0.88
4	3.37	3.47	3.07	3.30	0.21	6.30	0.85
6	5.15	4.92	4.79	4.95	0.18	3.68	0.82
5	5.36	5.13	4.97	5.15	0.20	3.80	0.86
3	6.30	6.25	5.65	6.07	0.36	5.96	0.85
8	9.78	9.87	9.82	9.82	0.05	0.46	0.85
1	13.17	12.17	11.71	12.35	0.75	6.04	0.85
7	14.30	13.95	13.22	13.82	0.55	3.99	0.85
10	31.57	31.37	30.81	31.25	0.39	1.26	0.95
9	40.52	38.48	40.45	39.82	1.16	2.91	0.91
mean	13.26	12.85	12.74	12.95	0.27	2.12	0.87

CIgen, generalized conformity index; RSD, relative standard deviation as percentage SD of mean; SD, standard deviation.

Discussion

To our knowledge this is the first study in which the variance of tumor delineation on H&E-sections by pathologists was investigated. The variation between tumor delineation on H&E-sections by the pathologists was relatively low and histopathology as the gold standard for imaging validation studies was highly reliable. The mean overlap between the delineations (expressed as CIgen) amounted to 0.87. This implies that on average the observers agreed on 87% of the total delineated volume. The distances between the delineations were in 95% of the measured distances smaller than 2 mm. Larger distances were found in irregularly shaped tumor areas and in the presence of cartilage invasion. The inclusion or exclusion of cartilage increased variation although after re-evaluation there was consensus about whether or not the cartilage was affected. The variation between the pathologists measured in this study was also caused by several other factors. The thickness of the used pencil (0.7 mm), the delineation style, e.g. the decision to delineate along the outside of the tumor border or along the inside; how precise the observer decided to delineate along the tumor border, and the in- or exclusion of necrotic tissue. Increasing accordance in delineation style by clear delineation guidelines would further decrease the variation between pathologists.

For the best case the overlap (CIgen) between the observers was 0.95.The variation of 0.05 was merely caused by the thickness of the pencil and the delineation style. This value may consequently be considered as the maximum overlap value. As this study is unique with regard to its purpose and method we were not able to adequately compare our findings with results from other studies. However, the results of this study can be compared to delineation studies performed on various imaging modalities. I
n an imaging-validation study [3] performed at our institution, the registration errors between various imaging modalities and histopathology using a threedimensional registration method, were determined. The calculated registration errors (RMSE CT: 1.5 mm, PET: 3.3 mm, MRI: 3.0 mm) exceeded the variation between the pathologists determined in this study. Therefore, in a study-setting in which pathology imaging registration is performed, the registration inaccuracy is larger than the variation of the gold standard. Much larger delineation inaccuracies varying from 3.1 to 16.1 mm were reported for delineation of the GTV by various observers delineating on CT images [5]. Rasch et al. [12] calculated the RMS of the standard deviation of distances between delineations of radiation oncologists and the median surface for tumor delineation for nasopharyngeal carcinoma. This resulted in 4.4 mm on CT and 3.3 mm on CT combined with MRI with an overlap agreement of 36% (CT) and 64% (CT+MRI). These values are considerably larger than the values reported in our study. Earlier work performed by our research group showed a CIgen of 0.61 for delineation of supraglottic laryngeal carcinoma by three radiation oncologists [13]. Therefore, it can be concluded that delineation inaccuracies on images are much larger than on histopathology.

Table 3. Distance analysis.

Patient	p90(mm)	p95(mm)
1	1.3	1.9
2	0.9	1.0
3	1.3	1.8
4	1.3	1.7
5	1.6	2.2
6	1.6	2.2
7	1.8	2.5
8	2.0	3.6
9	1.3	1.6
10	1.0	1.3
mean (SD)	1.4 (0.4)	2.0 (0.7)

Distances in mm measured between each point on the common and the encompassing delineation per patient. p90, p95: 90% respectively 95% of the measured distances is smaller than the values shown in the table.

Authors
Elise Anne Jager, Stefan M. Willems, Tim Schakel, Nina Kooij, Pieter J. Slootweg, Marielle E. P. Philippens, Joana Caldas-Magalhaes, Chris H. J. Terhaard & Cornelis P. J. Raaijmakers.

Published
Acta Oncologica, 2016; 55: 391–395.

References

1 ZhangY, Hu J, li J,Wang N, liW, ZhouY, et al. Comparison of imaging-based gross tumor volume and pathological volume determined by whole-mount serial sections in primary cervical cancer. Oncol Targets Ther 2013;6:917–23.

2 Stroom J, schlief A, Alderliesten T, Peterse H, Bartelink H, Gilhuijs k. Using histopathology breast cancer data to reduce clinical target volume margins at radiotherapy. int J Radiat Oncol Biol Phys 2009;74:898–905.

3 Caldas-Magalhaes J, kasperts N, kooij N, van den Berg CA, Terhaard CH, Raaijmakers CP, et al. Validation of imaging with pathology in laryngeal cancer: Accuracy of the registration methodology. int J Radiat Oncol Biol Phys 2012;82: e289–98.

4 Borren A, Moman MR, Groenendaal G, Boeken kruger AE, van Diest PJ, van der Groep P, et al. Why prostate tumour delineation based on apparent diffusion coef cient is challenging: An exploration of the tissue microanatomy. Acta Oncol 2013;52:1629–36.

5 Caldas-Magalhaes J, kooij N, ligtenberg H, Jager EA, schakel T, kasperts N, et al. The accuracy of target deline-
ation in laryngeal and hypopharyngeal cancer. Acta Oncol Epub 2015 Mar 3:1–7.

6 Campbell s, Poon i, Markel D, Vena D, Higgins k,
Enepekides D, et al. Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: implications for the management of head-and- neck cancers. int J Radiat Oncol Biol Phys 2012;82: 574–81.

7 Groenendaal G, Moman MR, korporaal JG, van Diest PJ, van Vulpen M, Philippens ME, et al. Validation of functional imaging with pathology for tumor delineation in the prostate. Radiother Oncol 2010;94:145–50.

8 Daisne JF, Duprez T, Weynand B, lonneux M, Hamoir M, Reychler 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.

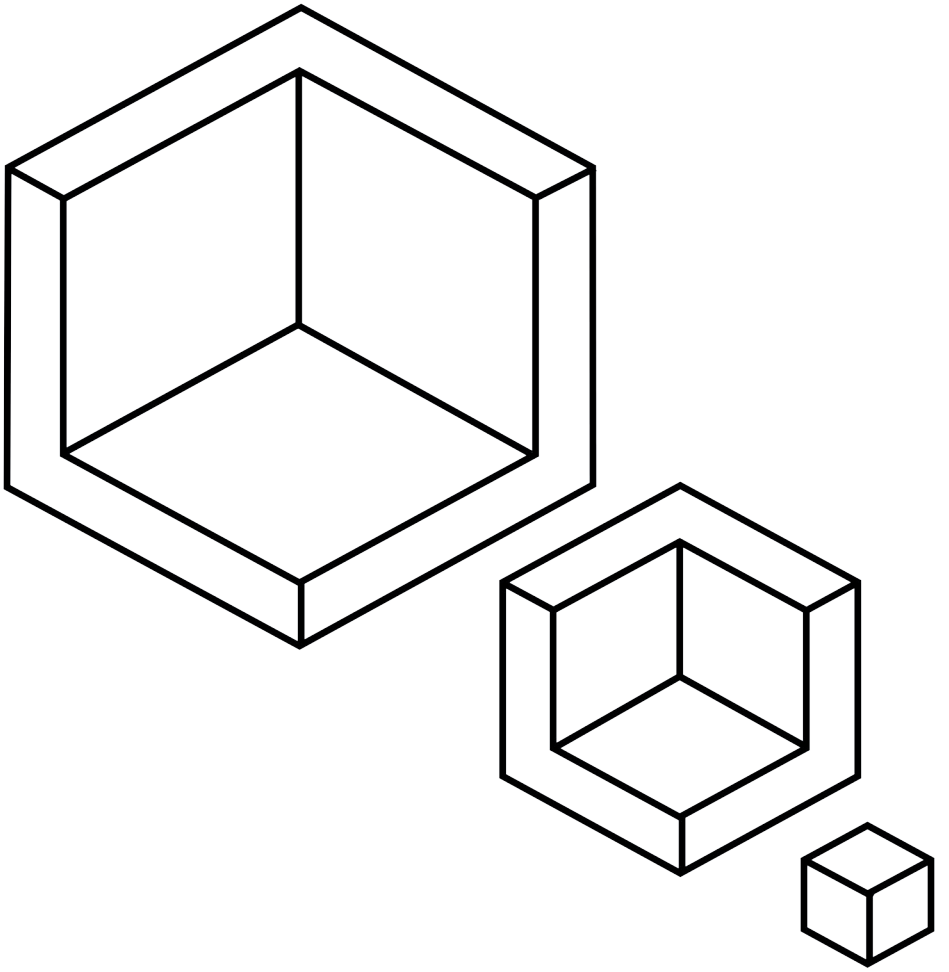
9 Driessen JP, Caldas-Magalhaes J, Janssen lM, Pameijer FA, kooij N, Terhaard CH, et al. Diffusion-weighted MR imaging in laryngeal and hypopharyngeal carcinoma: Association between apparent diffusion coeffcient and histologic findings. Radiology 2014;272:456–63.

10 Bol GH, kotte AN, van der Heide UA, lagendijk JJ. simultaneous multi-modality ROi delineation in clinical practice. Comput Methods Programs Biomed 2009;96: 133–40.

11 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.

12 Rasch CR, steenbakkers RJ, Fitton i, Duppen JC, Nowak PJ, Pameijer FA, et al. Decreased 3D observer variation with matched CT-MRi, for target delineation in Nasopharynx cancer. Radiat Oncol 2010;5:21–717X–5–21.

13 Jager E, kasperts N, Caldas-Magalhaes J, Philippens M, Pameijer FA, Terhaard C, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015;10:26.



Chapter 4
Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer

Validation of MRI for GTV delineation is of utmost importance to benefit from the visibility of anatomical details on MR images. Guidelines for GTV delineation of laryngeal and hypopharyngeal tumors on MR images were developed. The GTVs that were delineated by using guidelines and GTVs delineated according to clinical practice were both compared to histopathology, which is considered the be the gold standard.

Abstract

Background

Validation of magnetic resonance imaging (MRI) and development of guidelines for the delineation of the gross tumor volume (GTV) is of utmost importance to benefit from the visibility of anatomical details on MR images and to achieve an accurate GTV delineation. In the ideal situation, the GTV delineation corresponds to the histopathologically determined 'true tumor volume'. Consequently, we developed guidelines for GTV delineation of laryngeal and hypopharyngeal tumors on MRI and determined the accuracy of the resulting delineation of the tumor outline on histopathology as gold standard.

Material and Methods

Twenty-seven patients with T3 or T4 laryngeal/hypopharyngeal cancer underwent a MRI scan before laryngectomy. Hematoxylin and eosin sections were obtained from surgical specimens and tumor was delineated by one pathologist. GTV was delineated on MR images by three independent observers in two sessions. The first session (del1) was performed according to clinical practice. In the second session (del2) guidelines were used. The reconstructed specimen was registered to the MR images for comparison of the delineated GTVs to the tumor on histopathology. Volumes and overlap parameters were analyzed. A target margin needed to assure tumor coverage was determined.

Results

The median GTVs (del1: 19.4 cm^3, del2: 15.8 cm^3) were larger than the tumor volume on pathology (10.5 cm^3). Comparable target margins were needed for both delineation sessions to assure tumor coverage. By adding these margins to the GTVs, the target volumes for del1 (median: 81.3 cm^3) were significantly larger than for del2 (median: 64.2 cm^3) ($p \leq 0.0001$) with similar tumor coverage. Conclusions: In clinical radiotherapy practice, the delineated GTV on MRI is twice as large as the tumor volume. Validated delineation guidelines lead to a significant decrease in the overestimation of the tumor volume.

Introduction

Radiotherapy is developing towards a precision technique, delivering a high radiation dose to the tumor with tight treatment margins. Therefore, accurate three-dimensional target volume definition has become a crucial step. Target delineation, however, is one of the largest sources of uncertainty in head-and-neck cancer radiotherapy [1]. Even with the introduction of new imaging techniques the interobserver variability remains relatively large [2–5]. For an accurate definition of the target volume, validation of gross tumor volume (GTV) delineation is fundamental. For this validation, histopathology is the gold standard. However, this validation is a complex procedure and few studies have been performed for head-and-neck cancer [5–7]. For radiotherapy purposes, a detailed comparison of histopathology and imaging is required to validate the actual size, shape and location of the tumor. This involves three-dimensional reconstruction of the pathology specimen after slicing and matching of this specimen to the in vivo images.

Currently, computed tomography (CT) is the standard imaging modality for staging and delineation of laryngeal carcinoma. Magnetic resonance imaging (MRI) is gaining ground in radiation oncology as availability and image quality have been improved dramatically [7,8]. Several studies [2,3,5,9–11] have been performed, using various imaging modalities, to determine the agreement among observers when delineating the GTV in head-and-neck cancer. Studies in which CT was compared to MRI demonstrated contradictory results concerning the added value of MRI based on the interobserver variation [3,4,11]. However, a considerable number of studies reported an increased visualization of boundaries between different tissues and increased soft tissue contrast with MRI compared to CT [3,4,9–12]. Thus, in head-and-neck cancer, especially in the laryngeal region, there is no consensus on the added value of MRI for improving agreement between observers. However, there is agreement that visualization of soft tissue structures is better on MRI. A possible explanation for this contradiction is the increased visibility of anatomical details on MR images in combination with the lack of clear interpretation and delineation guidelines, what might result in an increased variability [4]. According to a study performed by Rasch et al. [3], the use of guidelines for delineation might be of value.

The aim of this study was to determine the accuracy of GTV delineation on MR images using delineation guidelines.

Material and methods

Patient selection

Thirty-six patients, treated with a total laryngectomy (TLE) for primary T3 or T4 laryngeal or hypopharyngeal cancer, were included in this study according to the inclusion criteria as further described in this section.

The first six patients were excluded for optimization of the pathology imaging registration procedure. Patient 12 was excluded because of a biopsy between preoperative imaging and surgery. The tumor of Patient 21 was too large for our whole mount standard analysis. The exclusion of Patient 30 was due to an incohesive tumor.

Resultantly, 27 patients (median age 62 years, range 49–79 years, two female and 25 male) with primary T3 (N=4) or T4 (N=23) laryngeal (supraglottic including the glottic: N=2, supraglottic: N=7, transglottic: N=4, glottic: N=2) or hypopharyngeal (N=12) carcinoma were used in this study for analysis. The patients underwent a TLE as a primary treatment in our institution between March 2009 and August 2014. The patients are numbered according to their study numbers.

This study is part of an extensive image validation project using positron emission tomography

Table 1. Volumes and overlap analysis.

Patient	$Tumor_{H\&E}$ (cm^3)	del1				del2			
		GTV (cm^3)	Sens	PPV	CI	GTV (cm^3)	Sens	PPV	CI
15	3.4	12.8	0.99	0.26	0.26	7.3	0.94	0.44	0.43
9	3.4	9.2	0.98	0.36	0.36	5.8	0.89	0.52	0.49
17	4.7	8.2	0.93	0.53	0.51	6.2	0.87	0.66	0.60
14	5.2	13.4	0.89	0.34	0.33	9.5	0.87	0.49	0.45
19	5.5	18.3	0.97	0.29	0.29	16.4	0.94	0.34	0.33
13	5.9	10.5	0.91	0.52	0.49	7.5	0.81	0.63	0.55
8	6.7	19.4	0.97	0.34	0.33	10.1	0.93	0.62	0.59
16	7.3	16.7	0.87	0.38	0.36	13.9	0.90	0.47	0.45
23	7.3	17.2	0.98	0.42	0.42	18.2	0.98	0.40	0.40
27	7.7	25.5	0.96	0.30	0.29	18.9	0.95	0.39	0.38
35	7.9	15.1	0.93	0.49	0.47	13.3	0.92	0.57	0.54
29	8.5	22.4	0.94	0.37	0.36	13.3	0.82	0.53	0.47
10	8.7	26.4	0.99	0.33	0.33	15.5	0.96	0.54	0.52
22	10.5	17.5	0.91	0.55	0.52	13.0	0.86	0.69	0.62
34	10.6	15.2	0.95	0.66	0.64	12.9	0.89	0.74	0.68
20	12.9	13.3	0.70	0.68	0.52	10.9	0.65	0.77	0.54
32	13.6	18.9	0.80	0.61	0.52	15.8	0.83	0.72	0.62
7	14.2	31.6	0.98	0.44	0.44	23.1	0.94	0.58	0.56
31	15.2	34.6	0.98	0.44	0.43	27.3	0.95	0.53	0.52
26	15.8	26.9	0.87	0.51	0.48	20.7	0.84	0.66	0.58
25	16.1	33.3	0.98	0.48	0.47	27.5	0.92	0.54	0.52
11	17.7	36.0	0.99	0.49	0.49	24.7	0.95	0.69	0.67
28	29.0	41.8	0.97	0.68	0.66	35.9	0.94	0.76	0.73
18	30.0	46.3	0.96	0.63	0.61	40.5	0.95	0.70	0.67
24	39.6	71.5	0.94	0.52	0.51	59.0	0.91	0.61	0.58
33	47.6	58.1	0.90	0.74	0.68	50.9	0.88	0.82	0.74
36	68.6	86.4	0.88	0.70	0.64	67.2	0.78	0.80	0.66
median	10.5	19.4	0.95	0.49	0.47	15.8	0.91	0.61	0.55

Volumes and overlap analysis for delineation on MRI without (del1) and with (del2) guidelines. Patients are listed by tumor size. $Tumor_{H\&E}$: tumor volume delineated by a pathologist on H&E-sections. CI: mean conformity index determined for the overlap between the GTV contour of the observers and the pathology contour; GTV: mean GTV volumes for three observers delineated on MRI. Overlap analysis; PPV: mean positive predictive value of the three observers; Sens: mean sensitivity of the three observers.

(PET), CT and MR images. Criteria of exclusion were: contraindications for MRI at 1.5 Tesla, contraindication for CT contrast administration as defined in the protocols of the radiology department and insulin dependent diabetes mellitus. The study was approved by our ethical review board.

Preoperative image acquisition

Before surgery, all patients underwent an MRI scan with T1- weighted (T1w) images in transverse, sagittal and coronal directions as well as transverse T2-weighted (T2w) and T1- weighted Gadolinium-enhanced images (T1w-Gd) while immobilized in a head and shoulder radiotherapy mask and using two small flexible surface coils (Intera, Philips Medical Systems) [12]. The MRI scan was made on a 1.5 Tesla spectrometer producing high resolution images. Imaging parameters can be found in *Supplementary Table 1*. The median time interval between MRI and surgery was 14 days (range 1–35 days).

Processing of the surgical specimen

Macroscopy

The pathology procedure was described in detail previously [6]. Briefly, the fresh larynx specimen was fixated in 10%-formaldehyde directly after surgery. After fixation, the specimen was embedded in an agarose block and transversely sliced in approximately 3 mm thick slices, which were subsequently photographed and digitized.

Microscopy

From each 3 mm thick slice, a 4 µm section was obtained and stained with hematoxylin and eosin (H&E). Histopathological analysis was performed by a dedicated head-and-neck pathologist who delineated all tumor tissue on the H&E-sections using a microscope. The delineated tumor is referred to as $tumor_{H\&E}$. The H&E-sections were subsequently digitized. These delineations were used as gold standard to validate the delineations on MRI.

Image registration and 3D reconstruction

First, the H&E-sections were registered to the corresponding thick-slice photos using cartilage landmarks to perform a point-based rigid registration with scaling. For three-dimensional reconstruction of the pathology specimen and registration to the MR images we used a previous published method [6]. To optimize the registration, the rigid registration between the pathology and MRI was visually verified and manually adjusted to correct for deformations of the pathology specimen. This was done by rigidly registering the $tumor_{H\&E}$ contour to the MRI and then manually adjusting the $tumor_{H\&E}$ contour to anatomical structures visible both on histopathology and MRI. The manual adjustments were done in consensus between two experts.

Delineation of GTV

Three dedicated and MRI-trained head-and-neck specialists independently delineated the GTV on MRI in Volumetool [13] an in-house developed clinical and research software application.

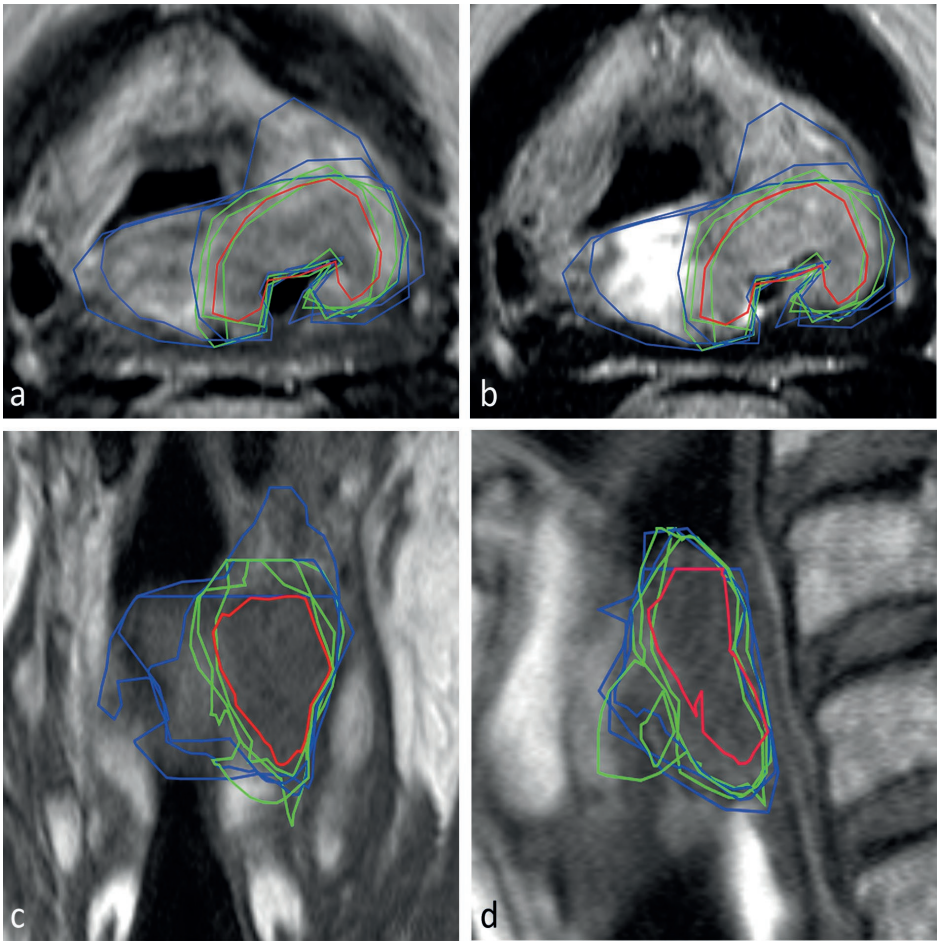
After a pilot study of five patients (not included in this study) followed by a meeting where the individual delineations were discussed, the approach was fine-tuned according to clinical practice and the experience of the observers. Consensus was reached on the following approach for the first delineation session (del1) using the various MRI sequences:

1. T1w with respect to the anatomy: delineation of all abnormal anatomy (i.e. presumed tumor).

2. T2w with respect to signal intensity: delineation of all hyper intense areas, with the exception of evident stasis of saliva. No differentiation between the primary tumor and surrounding edema and/or soft tissue swelling adjacent to the presumed tumor bulk.

3. T1w-Gd with respect to signal intensity:

Figure 1. Axial MR image at the level of the supraglottic larynx showing a T3 hypopharyngeal tumor centered at the left piriform sinus. GTV delineations of three observers according to clinical practice (del1: blue contours) and with guidelines (del2: green contours) and tumor delineations made by the pathologist ($tumor_{H\&E}$: red contour). T1w Gd-enhanced (a) and T2w (b) MR images showing a large asymmetric area of soft tissue on the right site of the tumor. This area was included in the first round (del1) of GTV delineations by two of the three observers. However, based on the increased signal intensity compared to the tumor, this area was excluded in the second round (del2) by all the three observers. (c,d) corresponding coronal and sagittal T1w reformations with contours permitting an overview of the delineations in non-axial planes. The delineated volumes of the second session (del2), where guidelines were used, show a large reduction in delineated volumes and reduction in overestimation of the tumor (red contour) compared to the delineated volumes in the first round (del1).



delineation of all enhancing areas that were suspect for tumor/tumor growth based on clinical experience of the observer.

T1w, T2w and T1w-Gd transverse and T1w sagittal and coronal images were available for image analysis.

The sagittal and coronal views permit an overview of the consistency of the delineations in non-axial planes and determination of the cranio-caudal tumor extension (*Figure 1*). The GTV was delineated on the T1w-Gd image and was modified according to the other MRI sequences. For all study patients, the observers were aware of the findings during endoscopy, which was performed before the MRI.

After an interval of one year, the tumors were re- delineated (del2) this time by using guidelines derived from criteria for diagnosis of neoplastic cartilage invasion on MR images [14] applied on soft tissue structures. The following delineation guidelines were used to analyze cartilage and soft tissue structures. The delineation guidelines were:

1. T2w or T1w-Gd with signal intensity higher than that of the adjacent tumor bulk was considered to indicate inflammation and was not included in the GTV (*Figure 2*).

2. Areas of strongly increased T2w signal intensity in the immediate surroundings of the tumor were considered to be stasis of saliva or trapped secretions and these were not included in the GTV (*Figure 3*).

From the GTV and the $tumor_{H\&E}$ contour, a common contour was derived for each patient per observer.

Analysis

Volumetric analysis and overlap analysis

The volume of the $tumor_{H\&E}$ was compared to the various GTVs. Three parameters were calculated to quantify the overlap. The sensitivity (1), which reflects the part of the $tumor_{H\&E}$ volume that was included in the GTV; the positive predictive value (PPV) (2), which reflects the part of the GTV that actually was tumor; and the conformity index (CI) (3) that quantifies the similarity between the two volumes (*Figure 4*).

$$(1)\ \textit{Sensitivity} = \frac{|\text{tumorH\&E} \cap \text{GTV}|}{|\text{tumorH\&E}|}$$

$$(2)\ \textit{Positive predictive value} = \frac{|\text{tumorH\&E} \cap \text{GTV}|}{|\text{GTV}|}$$

$$(3)\ \textit{Conformity index} = \frac{|\text{tumorH\&E} \cap \text{GTV}|}{|\text{tumorH\&E} \cup \text{GTV}|}$$

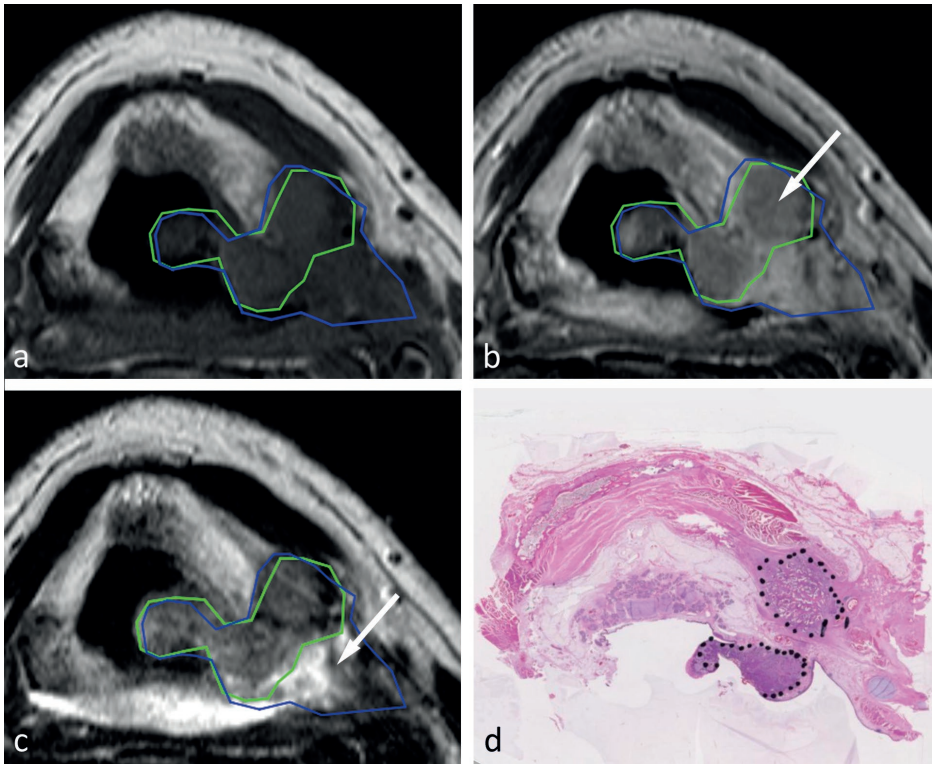
Distance analysis

From the GTV and the $tumor_{H\&E}$ contour, a common contour was derived for each patient per observer.

Two types of distances were calculated. Type I was the distance from each point of the contour of the $tumor_{H\&E}$ to the closest point on the common contour in 3D. This measure quantifies the distances resulting from missing tumor (underestimation). Type II was the distance from each point on the contour of the GTV to the closest point of the common contour and quantifies the overestimation (*Figure 4*).

The 95th percentiles of the distances per patient for each observer were determined. For each patient an average 95th percentile distance was calculated. The median value (p50) of the values (the average 95th percentile) of all patients was determined as the final type I and type II distance (p95 type I/II distances). The 95th percentile of the distances indicates that on average 95% of the outer contour (i.e. the $tumor_{H\&E}$ contour for

Figure 2. Patient with T4 hypopharyngeal tumor centered at the left aryepiglottic fold. GTV delineations of one observer according to clinical practice (del1: dark grey contour) and with guidelines (del2: light grey contour) and tumor delineations made by the pathologist ($tumor_{H\&E}$: black dotted contour). (a) axial T1w MR image at the level of the supraglottic larynx showing a tumor centered in the left aryepiglottic fold, (b) axial T1w Gd-enhanced MR image at the same level as (a), showing extensive involvement of the left paralaryngeal space (arrow). (c) T2w axial MR image at the same level as (a) and (b). In the first session of GTV delineation (del1) the observer included a large area on the left (arrow). In the second round (del2), this lateral extension was excluded because of high T2 signal intensity (considered as stasis of saliva or trapped secretions according to the guidelines). This separated the tumor from the left extralaryngeal soft tissues, which were not involved. (d) corresponding H&E-section with the $tumor_{H\&E}$ contour.



type I distance, and the GTV contour for type II distance) lies within these final type I and type II distances.

The 95th percentile as a cut-off point for the distance analysis was chosen to limit the influence of residual registration errors and deformations.

Derivation of target margin

From the type I distance, a margin that accounts for the underestimation was derived.

First, the 95th percentiles of the type I distances per patient for each observer were determined. Subsequently, a gamma distribution over these values was assumed per observer. The 95th percentile of this gamma distribution was determined per observer, resulting in three observer depended margins. Averaging these margins resulted in the margin accounting for the underestimation, i.e. the target margin.

For determination of the target margin, the 95th percentile of the gamma distribution was chosen to prevent extensive margins caused by the worst case and, consequently, overestimation of nearly all tumors.

Target volume determination

The target margin was added to the GTV contour of observer with the most average type I distances for delineation without and with guidelines. By adding the target margin around the GTV contour while correcting for anatomical boundaries such as air, pharyngeal constrictor muscles and vertebrae, target volumes were calculated.

Statistical analysis

Pairwise comparison of the volumes, overlap and distance parameters between the $tumor_{H\&E}$ and GTV were performed with a two-tailed Wilcoxon signed rank test.

Results

Volumetric analysis and overlap analysis

A large variation in tumor volume was observed (*Table 1*). In general a significant overestimation of the tumor was observed, which was reduced when guidelines were used for delineation ($p<.0001$) (*Table 1*, *Figure 1–3*). In some cases very good agreement between pathology and MRI delineations was observed and the delineations nearly completely covered the tumor (maximum sensitivity of 0.99). In other cases larger deviations were found and in the worst case 35% of the tumor was not included in the GTV (sensitivity of 0.65).

The median sensitivity of the three observers was larger for del1 (0.95) compared to del2 (0.91) ($p < .0001$). The amount of overestimation (PPV) varied largely although generally the overestimation for del2 was reduced compared to the del1 ($p < .0001$). The median CI increased by the use of guidelines (del1: 0.47, range 0.26–0.68, del2: 0.55, range 0.33–0.74) ($p < .0001$) (*Table 1*).

For Patients 20, 24, 26, 32, 33 and 36 larger tumor volumes were missed compared to other patients. The mean missed tumor volume of these patients was 4.0 cm^3 (SD 2.2) for del1 and 5.6 cm^3 (SD 4.7) for del2. The other patients, for whom less tumor volume was missed, the missed tumor volume was 0.45 cm^3 (SD 0.35) (del1) and 0.55 cm^3 (SD 0.48) (del2). Visual inspection of the tumors with larger missed tumor volumes showed that Patient 20, the irregular shape of the tumor for might have hampered an accurate delineation resulting in the lowest sensitivity of 0.70 (del1) and 0.65 (del2). For Patient 24 [sensitivity: 0.94 (del1), 0.91(del2)] and 36 [sensitivity: 0.88 (del1), 0.78 (del2)], the latter being an outlier with respect to tumor size (68.6 cm^3) and with on average the largest missed volume [7.9 (del1), 14.8 (del2) cm^3], the tumors encountered more deformations and registration problems. Furthermore, parts of tumor-invaded thyroid cartilage were not included in the GTV. For Patient 26 (sensitivity: 0.87 (del1), 0.84 (del2))

Figure 3. Patient with a T4 hypopharyngeal tumor centered in the left piriform sinus. GTV delineations of one observers according to clinical practice (del1: dark grey contour) and with guidelines (del2: light grey contour) and tumor delineations made by the pathologist ($tumor_{H\&E}$: black dotted contour). (a,b) T1w and T1w Gd enhanced axial MR image of the larynx at the level of the false vocal cord showing a tumor centered in the left piriform sinus. Note: enhancing area in the left thyroid laminae (arrow) adjacent to the tumor on the T1w Gd image. This area was included in the first round (del1) of GTV delineation by two of the three observers. The signal intensity of this area is higher compared to the tumor. Therefore in the second round (del2) considered as inflammation and excluded from the GTV. (c) T2w axial MR image of the larynx at the same level as (a) and (b). The same area (arrow) displays a higher T2 signal intensity compared to the tumor. Based on the guidelines this area was excluded in the second round (del2). (d) corresponding H&E-section with the $tumor_{H\&E}$ contour.

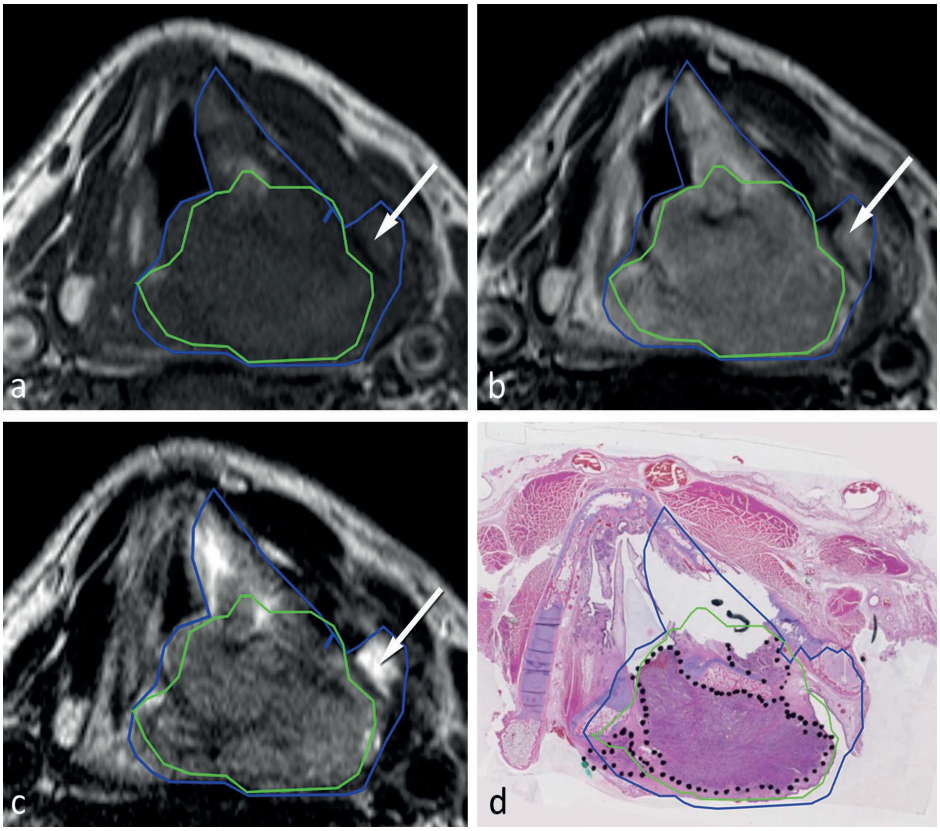


Figure 4. Schematic diagram of overlap and distance parameters. GTV: gross tumor volume (dark grey) and $tumor_{H\&E}$ (light grey). Blue dotted line reflects the encompassing contour. Arrow I: type I distance, arrow II: type II distance. The common volume is enclosed by a contour indicating the overlap between the GTV and $tumor_{H\&E}$ contours. Overestimation is delineation of tissue by the observers that is not tumor while underestimation is the part of the tumor that is not included in the GTV delineation, i.e. missed volume.

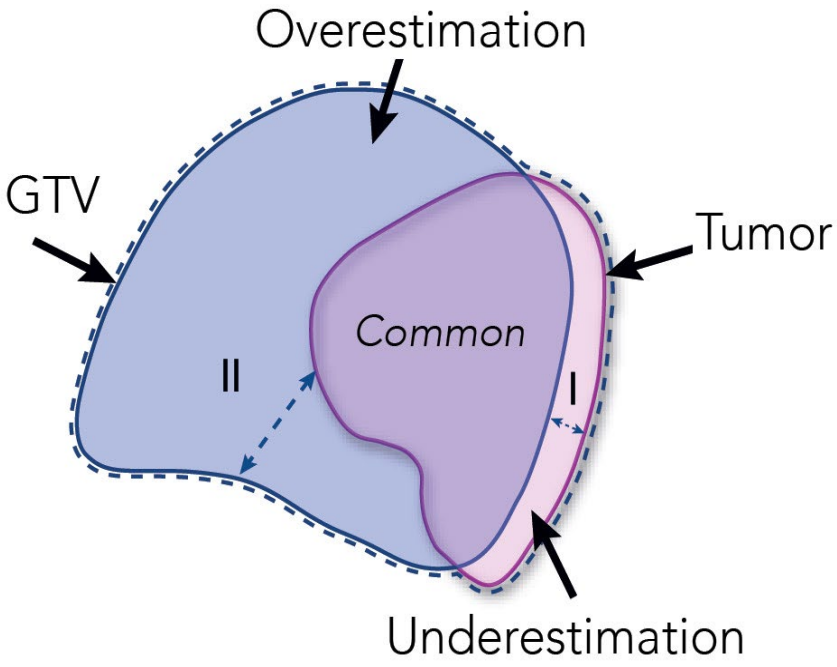


Table 2. Distance analysis.

Patient	Distance type I, p95 (mm)		Distance type II, p95 (mm)		Distance type I, max (mm)	
	del1	del2	del1	del2	del1	del2
15	0.1	0.8	10.6	6.8	0.9	2.4
9	0.4	1.5	8.9	4.9	1.8	4.1
17	1.5	1.9	6.2	4.2	3.8	3.3
14	2.0	2.7	8.9	6.0	4.9	5.7
19	1.0	2.0	11.8	12.9	4.3	4.3
13	2.0	3.1	9.6	6.6	5.4	5.3
8	0.6	1.2	20.5	5.1	2.8	3.4
16	2.9	2.1	12.9	8.7	4.8	4.3
23	0.6	0.3	11.8	12.5	3.3	2.8
27	0.6	1.1	12.0	9.4	2.4	3.1
35	1.6	1.4	7.5	6.1	3.8	3.5
29	1.5	4.7	11.4	9.1	3.9	8.5
10	0.1	1.0	12.6	6.2	3.3	4.4
22	2.7	3.9	9.4	5.3	7.2	7.9
34	1.3	2.2	6.9	6.1	3.3	4.8
20	6.5	6.1	6.9	3.4	12.0	10.3
32	7.7	6.4	8.1	4.8	14.7	12.3
7	0.3	1.7	9.6	6.7	4.1	5.9
31	0.4	1.2	13.3	8.9	3.7	4.8
26	8.0	7.2	11.9	7.0	13.2	12.2
25	0.7	2.0	7.9	6.5	3.7	5.4
11	0.4	1.0	10.6	5.5	1.9	3.3
28	1.0	1.3	6.1	4.5	3.2	4.3
18	1.9	1.8	14.1	11.1	5.0	5.3
24	2.3	3.7	13.8	9.2	9.1	11.4
33	3.3	3.6	6.3	5.1	8.7	9.3
36	3.5	6.3	10.3	8.1	10.0	13.0
median (p50)	1.5	2.0	10.3	6.5	3.9	4.8
p95	7.0	6.9			10.2	8.3

From three individual GTV contours and the $tumor_{H\&E}$ contour a common and an encompassing contour were derived. Type I distance: distance from each point of the common contour to the closest point on the contour of the $tumor_{H\&E}$. Type II distance: distance from each point of the common contour to the closest point on the contour of the GTV. Max: maximum distance measured for this patient, mean of the three observers. Distance type I and distance type II, p95 (columns): 95% of the mean distances of the three observers are smaller than the presented value. P95 (value in de last row): the 95th percentile of the 95th percentile of the mean distances of the three observers (distance type I, p95 and distance type II, p95). The target margins were based on the p95 of the distance type I, p95. P95 (last row) of distance type I, max: 95th percentile of mean of the maximum type I distances of the three observers. P50: the median values. The patients are listed by tumor size.

parts of the tumor were missed in retropharyngeal direction and for Patient 32 (sensitivity: 0.80 (del1), 0.83 (del2)) the tumor extended into the pharynx and the constrictor muscles crossing the midline. This extension was not clearly visible on the MR images and was missed by the observers. The tumor of Patient 33 (sensitivity: 0.90 (del1), 0.88 (del2)) showed extra laryngeal growth, which was missed.

Distance analysis

For type I distances, quantifying the distances resulting from missing tumor, the median (p50) of the 95% of the measured distances of the tumor extensions outside the GTV (distance type I, p95), were smaller than 1.5 mm (del1) and 2.0 mm (del2) (*Table 2*). The maximal distances from the GTV border where tumor extensions were found were 14.7 mm (del1) and 13.0 mm (del2). The median (p50) of the 95th percentile for type II distances, indicating the overestimation, was significantly smaller for delineation with guidelines (6.5 mm) compared to delineation without guidelines (10.3 mm) ($p < .0001$) (*Table 2*).

Target margin and target volume analysis

Required target margins were comparable for delineation according to clinical practice (del1: 7.0mm) and with guidelines (del2: 6.9 mm) (*Table 2*: p95 of the distance type I, p95). When applying these margins, significantly larger target volumes were calculated for del1 (median: 64.4 cm^3, range 33.5–194.7 cm^3) compared to del2 (median: 50.3 cm^3, range 26.2–167.5 cm^3) ($p < .0001$), with a decrease in overestimation (median del1: 53.9 cm^3, median del2: 39.8 cm^3) of more than 25% ($p<.0001$). By using this target margin, the maximal missed tumor volumes was 0.28 cm^3 (del1) and 0.32 cm^3 (del2) for Patient 20 and Patient 26 respectively. Only a small fraction of the total tumor volume (del1: 0.17%, del2: 0.23%) was missed after applying a target margin. These missed volumes were not significantly different ($p=0.53$) between del1 and del2 and were found in five of the 27 patients (number 20 (del1), 26 (del2), 32, 36 and 24).

Discussion

In clinical practice, the delineated GTV on MRI is twice as large as the tumor volume determined on pathology (*Figure 1*, *Table 1*). Clear delineation guidelines lead to a decrease of tumor volume overestimation and a smaller target volume.

For GTV delineation on MRI according to current clinical practice, the overestimation results in relatively large radiation fields. This overestimation might be explained by the inclusion of tissue with abnormal appearance other than tumor tissue, e.g. edema or inflamed tissue, in the GTV. Subsequently, delineation guidelines were used with the aim to differentiate between tumorous and non-tumorous tissue and, consequently, the exclusion of non-tumorous tissue from the GTV. The PPV implied that, on average, 49% of the delineated GTV on MRI consisted of tumor tissue while 5% of the tumor was missed (sensitivity). The delineation guidelines increased the percentage of tumor in the GTV with 12% at the expense of a slight decrease in sensitivity of 4% (0.95–0.91).

An explanation for this decrease in sensitivity is that by reduction of the GTV volume, the chance of missing tumor was increased for del2 compared to del1. By decreasing the delineated volumes, the effect of the remaining registration errors and deformations might become more apparent. Furthermore, by trying to delineate the tumor as accurately as possible, small delineation inconsistencies become apparent even though the correct tumor border has been delineated. The CI, however, increased from 0.47 to 0.55 indicating more resemblance of the GTV with the true tumor volume by the use of guidelines. This parameter takes into account the overestimation as well as the underestimation.

For some patients it was more difficult to include all tumor tissue in the GTV. In a previous study, in which the interobserver variation between pathologists for delineation of the tumor on H&E-sections was investigated, the largest vari-

ation was observed for Patient 20. In general, the interobserver variation between pathologists was small [15]. However, the irregular shape of this tumor, with a large amount of swelling around the tumor, hampered an accurate delineation on both the H&E-sections and the MR images. Another difficulty was tumor invading the cartilage structures leading to misses (Patients 24 and 36). In clinical radiotherapy practice, radiation oncologists sporadically delineate tumors with cartilage invasion because these tumors are seldom primary treated with radiation. Lack of experience in evaluating and delineating cartilage invasion on MRI might have caused this finding.

Important work on validation of imaging techniques for head-and-neck tumors has been performed by Daisne et al. [7]. A comparison between photographs of nine TLE specimens and delineated GTVs on MRI resulted in a lower sensitivity (0.83) and PPV (0.44) than what was found in our study. A shortcoming of this study was that microscopic evaluation was not performed. Furthermore, in our work a dedicated MRI protocol was used for GTV delineation in radiotherapy [12]. Overestimation of the tumor on MRI has also been reported for oropharyngeal and oral cavity cancer by Seitz et al., although this overestimation was smaller than observed in our study (tumor volume: 16.6 ± 18.6 cm^3, GTV MRI: 17.6 ± 19.1 cm^3) [16].

In our study, patients with surgery as primary treatment were included. For lower tumor stages (T1 and T2T1b-T3) radiotherapy usually is the treatment of choice and the lower stage tumors mostly concern smaller tumors. In this study the use of guidelines results in a reduction of overestimation for smaller tumors as well as for larger tumors, indicating that these guidelines could be of added value for reduction of overestimation for laryngeal and hypopharyngeal tumors that are treated with radiation therapy alone.

The median time interval between MRI and surgery was 14 days (range 1–35 days). For 24 patients surgery was performed within 17 days after MRI acquisition, whereas for three patients this time interval exceeded one month. This increased time interval might have enabled significant tumor growth between MRI and surgery, although no clear data on tumor growth rate are available. If the tumor on MRI was smaller than the tumor volume identified by histopathology, this might have caused a decrease in overestimation or an increase in tumor volume that was missed (decrease in sensitivity) by MRI. However, no significant correlation between the sensitivity and time interval between surgery and MRI (del1: Spearman's rho=-0.080, p=0.69, del2: Spearman's rho=-0.34, p=0.87) was found.

Due to flaws in one of the observers' delineation, two patients were re-delineated. These flaws were caused by an unfinished delineation and incorrect use of guidelines and were considered not to occur in clinical practice. These redelineations resulted in more reliable results and were used for the analysis.

Although we optimized the registration between pathology and the MR images and manually corrected deformations of the specimen according to the MR images, we could not fully prevent some registration mismatch. Furthermore, a variable amount of shrinkage of tumor specimens after fixation has been found for head-and-neck cancer depending on the fixation method, tumor site, and tissue of origin [17–20]. For oral tongue cancer, the tumors shrank on average 20.2% [17]. This is much more than what has been found by Caldas-Magalhaes et al. [6]. The shrinkage of larynx specimens due to fixation with formaldehyde was on average 3% within the cartilage skeleton, which included the tumor bulk. The larynx is a privileged site with a strong and rigid skeleton, which helps to maintain the shape and the size of the tissues inside it resulting in less shrinkage.

One of the results of this study is that overesti-

mation of the tumors on MRI was found. This overestimation is probably caused by the inclusion edema and/or inflamed tissue in the GTV. Histological investigation of these volumes is needed to confirm the presence of edema and inflamed tissue.

The target margin was added uniformly around the GTV and therefore, the direction of the extensions were not taken into account. It should be noted that the target margins are determined for research purposes. Detailed analysis of margins, which can be used to incorporate delineation uncertainty in clinical practice for various imaging modalities is subject of future research.

Conclusions

In clinical practice, the delineated GTV on MRI is twice as large as the tumor volume determined on pathology. Delineation guidelines can be used as a starting point for MRI delineation for laryngeal and hypopharyngeal carcinoma, leading to a decrease in tumor volume overestimation. After the addition of a target margin around the GTVs to include all tumor tissue, the use of guidelines resulted in a smaller target volume with similar tumor coverage.

Authors

Elise Anne Jager, Hans Ligtenberg, Joana Caldas-Magalhaes, Tim Schakel, Marielle E.P. Philippens, Frank A. Pameijer, Nicolien Kasperts, Stefan M. Willems, Christiaan H.J. Terhaard & Cornelis P.J. Raaijmakers.

Published

Acta Oncologica, 2016, 55:11, 1305-1312.

Supplementary table 1. MR imaging parameters.

MR Imaging parameters	MRI sequences			
	cT1w	**sT1w**	**tT1w**	**T2w**
TR (ms)	570	570	593	4200
TE (ms)	17	17	15	150
slice thickness (mm)	4	4	4	4
gap between slices (mm)	0,4	0,4	0,4	0,4
number of averages	2	2	2	2
acquisition matrix (mm^2)	512x307	512x307	512x307	512x307
reconstruction matrix ($mm^{2)}$	512x512	512x512	512x512	512x512
field of view (mm)	210x148	210x148	210x148	210x148
reconstructed resolution (mm^2)	0,41x0,41	0,41x0,41	0,41x0,41	0,41x0,41

MR imaging parameters and MRI sequences. cT1w: coronal T1 weighted image, sT1w: sagittal T1 weighted image, tT1w: transverse T1 weighted image, T2w: T2 weighted images. TE: echo time, TR: repetition time.

References

1 Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. J Med Phys 2008;33:136–40.

2 Geets X, Daisne JF, Arcangeli S, 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.

3 Rasch CR, Steenbakkers RJ, Fitton I, et al. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. Radiat Oncol 2010;5:21-717X-5-21.

4 Jager E, Kasperts N, Caldas-Magalhaes J, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015;10:26.

5 Caldas-Magalhaes J, Kooij N, Ligtenberg H, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol 2015;54:1181–1187.

6 Caldas-Magalhaes J, Kasperts N, Kooij N, et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 2012;82:e289–98.

7 Daisne JF, Duprez T, Weynand B, 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.

8 Khoo VS, Dearnaley DP, Finnigan DJ, et al. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother Oncol 1997;42:1–15.

9 Ahmed M, Schmidt M, Sohaib A, et al. The value of magnetic resonance imaging in target volume delineation of base of tongue tumors–a study using flexible surface coils. Radiother Oncol 2010;94:161–7.

10 Chung NN, Ting LL, Hsu WC, et al. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck 2004;26:241–6.

11 Rasch C, Keus R, Pameijer FA, et al. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. Int J Radiat Oncol Biol Phys 1997;39:841–8.

12 Verduijn GM, Bartels LW, Raaijmakers CP, et al. Magnetic resonance imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors. Int J Radiat Oncol Biol Phys 2009;74:630–6.

13 Bol GH, Kotte AN, van der Heide UA, et al. Simultaneous multimodality ROI delineation in clinical practice. Comput Methods Programs Biomed 2009;96:133–40.

14 Becker M, Zbaren P, Casselman JW, et al. Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis at MR imaging. Radiology 2008;249:551–9.

15 Jager EA, Willems SM, Schakel T, et al. Interobserver variation among pathologists for delineation of tumor on H&E-sections of laryngeal and hypopharyngeal carcinoma. How good is the gold standard? Acta Oncol 2016;55:391–395.

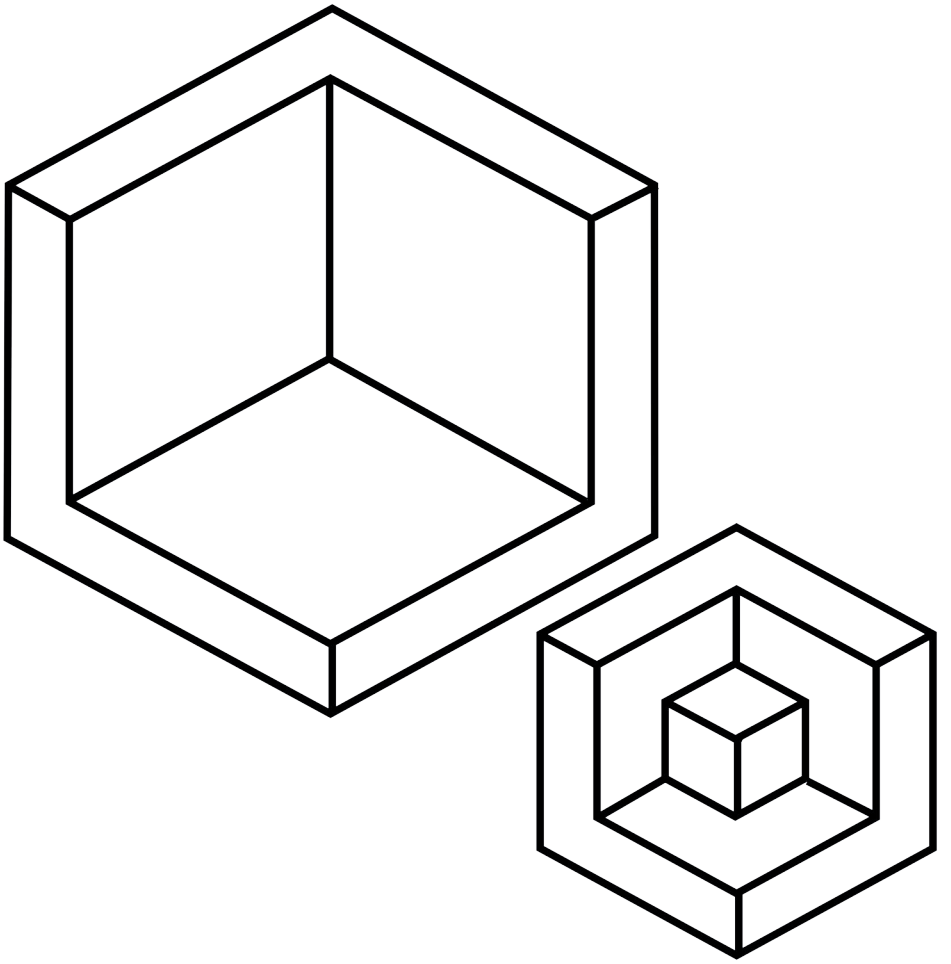
16 Seitz O, Chambron-Pinho N, Middendorp M, et al. 18F- Fluorodeoxyglucose-PET/CT to evaluate tumor, nodal disease, and gross tumor volume of oropharyngeal and oral cavity cancer: comparison with MR imaging and validation with surgical specimen. Neuroradiology 2009;51:677–86.

17 Brotherston D, Poon I, Peerani R, et al. Tumor shrinkage associated with whole-mount histopathologic techniques in oral tongue carcinoma. Pathol Res Pract 2015;211:398–403.

18 Campbell S, Poon I, Markel D, et al. Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: implications for the management of head- and-neck cancers. Int J Radiat Oncol Biol Phys 2012;82:574–81.

19 Chen CH, Hsu MY, Jiang RS, et al. Shrinkage of head and neck cancer specimens after formalin fixation. J Chin Med Assoc 2012;75:109–13.

20 Cheng A, Cox D, Schmidt BL. Oral squamous cell carcinoma margin discrepancy after resection and pathologic processing. J Oral Maxillofac Surg 2008;66:523–9.



Chapter 5
Modality-specific target definition for laryngeal and hypopharyngeal cancer on FDG-PET, CT and MRI

Correct GTV delineation is the basis for accurate radiotherapy treatment. The goal of this study was to improve target definition by deriving modality-specific margins for CTV for laryngeal and hypopharyngeal cancer on CT, MRI and 18-FDG-PET.

Abstract

Background and Purpose

The goal of this study was to improve target definition by deriving modality-specific margins for clinical target volumes (CTV) for laryngeal and hypopharyngeal cancer on CT, MRI and 18-FDG-PET.

Material and Methods

Twenty-five patients with T3/T4 laryngeal/hypopharyngeal cancer underwent CT, MRI and 18-FDG-PET scans before laryngectomy. H&E-sections were obtained from the surgical specimen and tumor was delineated ($tumor_{H\&E}$). The GTVs on CT and MRI were delineated in consensus. PET-based GTVs were automatically segmented. The three-dimensionally reconstructed specimen was registered to the various images. Modality-specific CTV margins were derived and added to the GTVs to achieve adequate tumor coverage. The resulting CTVs were compared with each other, to $tumor_{H\&E}$, and to CTVCT10 constructed on CT with the clinical margin of 10 mm.

Results

CTV margins of 4.3 mm (CT), 6.1 mm (MRI) and 5.2 mm (PET) were needed to achieve adequate tumor coverage. The median volumes of the resulting modality-specific CTVs were 44 ml (CT), 48 ml (MRI) and 39 ml (PET), while the CTV10mm was 80 ml.

Conclusion

For laryngopharyngeal tumors, 45% - 52% target volume reduction compared with CTV10mm is achievable when modality-specific CTV margins are used. PET-based CTVs were smaller compared to CT- and MRI-based CTVs.

Introduction

The past two decades, radiotherapy has experienced impressive technological innovations with increased precision in dose delivered to the tumor [1–3]. Accurate target definition of the target volumes is essential to exploit these advances. In radiotherapy practice, disease extent is derived from imaging data, since pathological data is usually absent. Incorrect target definition can lead to unnecessary damage to healthy tissues or undertreatment near the tumor boundary, particularly for head-and-neck cancer, where the target is surrounded by critical anatomical structures.[1,2]. Despite the significance, accurate and consistent definition of the target volume remains a major challenge in radiotherapy [3,4].

The gross tumor volume (GTV) is delineated on an image as a first estimate for the tumor. This volume is expanded with a margin to include microscopic tumor spread that is invisible on the images to create the CTV. In the ICRU83 report, the CTV is defined as "a volume of tissue that contains a demonstrable gross tumor volume (GTV) and/or subclinical malignant disease…" [3]. While the GTV represents the tumor visible in imaging, the CTV is more closely related to the histopathological tumor extent.

In spite of this, comparison of GTV delineation with (histo)pathology is used to determine the accuracy of imaging modalities for target definition in literature [5]. To overcome the discrepancy between GTV in imaging and tumor volume in (histo)pathology, research groups made different choices to come to a surrogate GTV in macroscopic pathology [5] or microscopic histology [6] for head-and-neck tumors.

Another issue is that tumor visibility and, consequently, GTV delineation depends on the imaging modality and the data acquisition parameters. Further, interpretation of the images, differences in experience level and delineation guidelines result in interobserver variation [7–9]. Assessment of the necessary CTV margin is hampered by these limiting factors.

In addition to variations between imaging modality and between observers, microscopic extent can be highly variable depending on tumor site and type. As a consequence, "delineation of the CTV is currently based on clinical experience", as the ICRU 83 report states [3], which is reflected by a large variation in CTV margins between institutes [10,11] and between observers [7,10,12], ranging from 5 mm to 2 cm [6]. In our institute, in general a 10 mm CTV margin for head-and-neck cancer is added to the GTV determined on computed tomography (CT), as suggested by Caudell [11].

CT has traditionally been used for tumor contouring. However, with the development of magnetic resonance imaging (MRI) and positron emission tomography (PET), research focus for target definition has advanced accordingly to exploit the additional information available from these imaging modalities. MRI is known to have a range of excellent soft tissue contrasts, which could enhance tumor visibility [13]. Fluorodeoxyglucose (FDG)-PET provides information on the uptake of the radiopharmaceutical glucose analog, reflecting metabolic differences between tumor and surrounding tissues. The difference in metabolic properties can result in a high contrast image, although with low spatial resolution. These modality specific qualities lead to modality-specific variations in GTV delineation [5,9,13], which might influence the CTV margin needed for adequate tumor coverage.

Here, we have compared the gross tumor volumes defined on MRI, CT and FDG-PET datasets with the corresponding tumor delineations on hematoxylin and eosin (H&E) stained histopathology, which was considered to be the gold standard, in order to investigate the impact of different modalities on target definition of primary tumor. The main goal of this study was to derive modality-specific (MRI, CT and FDG-PET) clinical target definitions for laryngeal and hypopharyngeal tumors. The margins for these CTVs were optimized to ensure coverage of microscopic tumor extent defined by the histopathological 'true tumor' volumes.

Materials and Methods

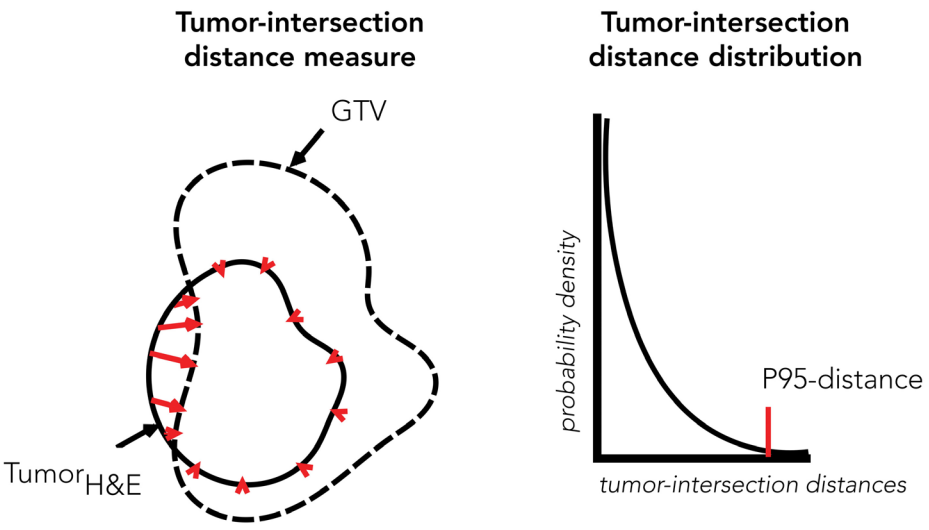
Patient selection

Thirty-six patients, treated with a total laryngectomy (TLE) for primary T3/T4 laryngeal or hypopharyngeal cancer were included in this study according to the inclusion criteria as further described in this section. In total nine patients were excluded. The first six patients were used for optimization of the pathology-imaging registration procedure. Patient 12 was excluded because of a biopsy between preoperative imaging and surgery. The tumor of Patient 21 was too large for our standard whole-mount procedure. The exclusion of Patient 30 was due to a tumor which fragmented during surgery.

As a result, 27 patients (median age: 62 years, range: 49-79 years, two female and 25 male) with primary T3 (N=4) or T4 (N=23) laryngeal (supra-&glottic:2, supraglottic:7, transglottic:4, glottic:2) or hypopharyngeal (12) squamous cell carcinoma were included in the study. The patients underwent TLE as primary treatment between March 2009 and August 2014.

Furthermore, if the time interval between imaging with a particular modality and surgery was ≥20 days, images from that modality were excluded from further analysis. For two patients (11 and 32) all imaging was performed ≥20 days prior to surgery. Therefore, in total 25 patients were suitable for analysis with 25, 23 and 22 for CT, MRI and PET respectively.

Figure 1. P95-distance derivation scheme. For all patients, GTV and tumor$_{H\&E}$ were geographically compared and the intersection of both contours was determined. The distance from the tumor$_{H\&E}$ to the contour of the intersection of the tumor$_{H\&E}$ with the GTV was measured (red arrows). For each patient, the 95th percentile of these distances is calculated (red bar in 'Tumor-intersection distance measure' figure).



Criteria of exclusion were contraindications for MRI or for CT contrast administration as defined in the protocols of the department of Radiology, and insulin-dependent diabetes mellitus. Patients are numbered according to their study numbers. The study was approved by our ethical review board.

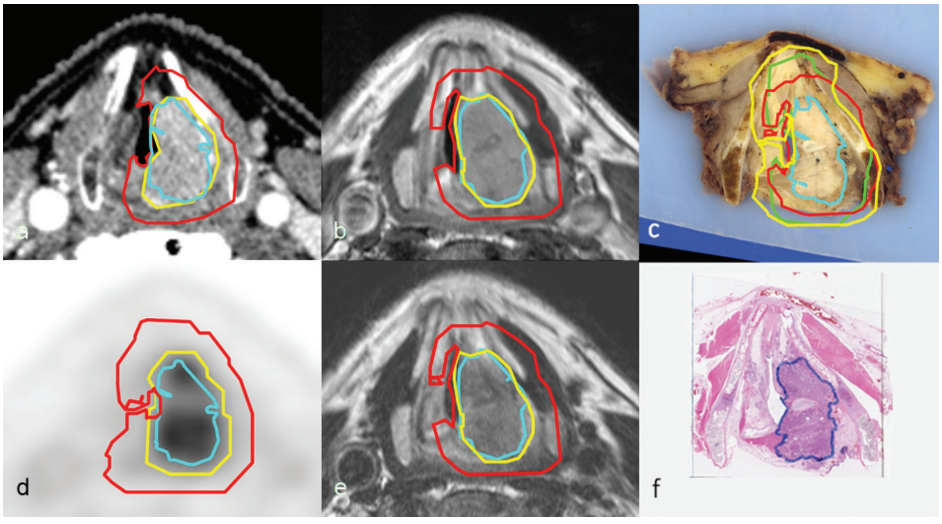
Preoperative image acquisition

Before surgery, all patients underwent CT, MRI and 18-FDG-PET scans while immobilized in a head-and-shoulder radiotherapy mask with a head-and-neck support. The CT scan was obtained after intravenous contrast agent administration. Two small flexible surface coils (Flex S coils, Philips Medical Systems)[14] were used for the MRI scan. Details on the imaging parameters are shown in *Supplementary tables 1-3*.

Surgical specimen processing

The pathology procedure was described in detail previously [15]. Briefly, the fresh larynx specimen was fixated in 10%-formaldehyde directly after surgery. After fixation, the specimen was scanned on the CT for registration purposes, as described in the next section. Subsequently, the specimen was embedded in an agarose block and transversely sliced in approximately 3 mm thick slices, which were then photographed and digitized. The thick slices were reconstructed in 3D into a digitized specimen using the contour of the agarose block to steer the registration as described by Caldas-Magalhaes et al [15]. From each slice, a 4 µm section was obtained and stained with H&E. Histopathological analysis was performed by a dedicated head-and-neck pathologist who delineated all tumor tissue on

Figure 2. Laryngeal tumor delineated on (a) contrast-enhanced CT, (b, e) MRI (T1-weighted Gd-enhanced, T2 weighted, respectively) and (d) FDG-PET with corresponding (c) macro- and (f) histo-pathology of patient 34 with a T4 stage glottic laryngeal carcinoma. For a, b, d and e the contours of $tumor_{H\&E}$ (blue), GTV (yellow) and CTV (red) are shown. For c, the contours of $tumor_{H\&E}$ (blue) and CTVCT (green), CTVMRI (red) and CTVPET (yellow) are given. In f, the manually delineated pathology contour (dark blue) is depicted. The $tumor_{H\&E}$ contour shown in a–e is the non-rigidly transferred delineation of the pathology delineation shown in f. The difference of the pathology based tumor delineation between the imaging modalities is caused by slice thickness differences between MRI and CT/PET resulting in sampling differences.

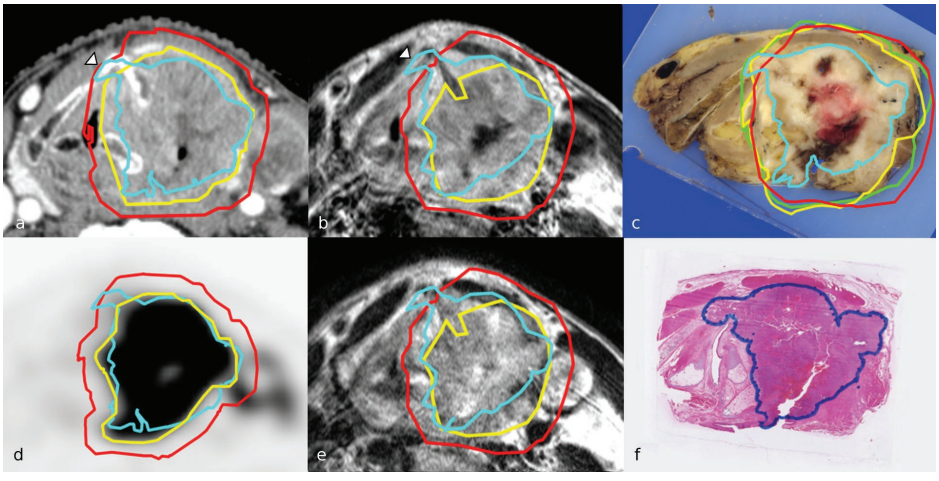


the H&E-sections using a microscope, generating a three-dimensional structure referred to as tumor$_{H\&E}$. Subsequently, the H&E-sections were digitized and registered to the 3D-reconstructed specimen. The structure tumor$_{H\&E}$ was used as the gold standard to validate tumor definition from the various imaging modalities. The H&E-sections were registered to the corresponding thick-slice using cartilage landmarks to perform point-based rigid registration with scaling. Shrinkage of the H&E-sections, which amounted to 12%, was taken into account. Shrinkage of the thick-slices was limited and therefore not taken into account [15].

Image registration

Registration between the 3D-reconstructed histopathology block and the three imaging data sets was needed to compare the GTV delineations with the pathological tumor contour. The registration was performed in three main steps (*Supplementary figure 1*).

Figure 3. Comparison of GTV delineation in MRI, CT and PET with tumor on histopathology illustrated by patient 24 with a T4-hypopharyngeal tumor with thyroid cartilage invasion: (a) contrast-enhanced CT, (b, e) MRI (T1w Gd-enhanced, T2w) and (e) FDG-PET with corresponding (c) macro- and (f) histo-pathology. For a, b, d and e the contours of tumor$_{H\&E}$ (blue), GTV (yellow) and CTV (red) are shown. For c, the contours of tumor$_{H\&E}$ (blue) and CTVCT (green), CTVMRI (red) and CTVPET (yellow) are given. In f the manually delineated pathology contour (dark blue) is depicted. The tumor$_{H\&E}$ contour shown in a-e is the non-rigidly transferred delineation of the pathology delineation shown in f. (a) For CT, the tumor extension in the thyroid cartilage was missed by the GTV although just included in the CTV (indicated by arrow). (b) On MRI, a part of the thyroid cartilage invaded by tumor tissue and some other small parts of the tumor volume were not included in the GTV. The cartilage was not fully enclosed by the CTV margin (indicated by arrow point). (c) On PET, small parts of the tumor volume were missed in the GTV as well as a part of the thyroid cartilage that was invaded. The whole tumor was covered after adding a CTV margin to the GTV. The difference in depiction of the pathology based tumor delineation between the various imaging modalities is caused by differences in sampling due to differences in slice thickness between MRI and CT/PET. The analyses are performed on resampled images.



The first step was an automated rigid registration of the CT scans to the specimen using the CT of the specimen (specimenCT) as an intermediate step [15]. Firstly, the contour of the pathology specimen and the contour of specimenCT where rigidly registered. Secondly, the specimenCT was registered to the CT of the patient. To guide this registration step a region of interest (ROI) was drawn around the thyroid and cricoid cartilage. The MRI and PET were rigidly registered to the CT based on mutual information. Additionally, the PET image was translated by overlaying the center of gravity of the PET-based GTV with the pathology based GTV, which corrects for larynx movement but might give a positive bias to the final results.

Secondly, the registration was optimized by manual rigid registration based on tumor and anatomical structures visible both on the specimen and on the MR and/or CT images. The high soft-tissue contrast on MRI and the high contrast between cartilage and soft-tissue on CT was exploited here. The single transformation derived from this registration step was applied to all imaging.

Lastly, due to deformations of the specimen further non-rigid transference of the tumor contour from pathology to the MRI was performed if necessary. The procedure was based on anatomical structures visible on both histopathology and MRI.

Table 1. $Tumor_{H\&E}$ volume, GTV, distances and CTV for all modalities.

Parameters	Modality	Median	Range	p-values		
				$Tumor_{H\&E}$	MRI	PET
$Tumor_{H\&E}$ volume (ml)	Pathology	8.7	3.4-68.6	-	-	-
GTV (ml)	CT	17.5	5.9-88.7	**<0.0001**	**<0.0001**	**<0.0001**
	MRI	15.5	4.9-66.3	**<0.0001**	-	0.44
	PET	14 .5	6.1-82.7	**<0.0001**	0.44	-
P95-distance (mm)	CT	1.6	0.4-5.6	-	**0.015**	0.12
	MRI	1.7	0.4-6.3	-	-	0.74
	PET	2.3	0.7- 6.2	-	0.74	-
Maximum distance (mm)	CT	4.6	1.7-10.7	-	0.50	0.72
	MRI	4.3	1.9-12.3	-	-	0.48
	PET	5.3	2.0-12.1	-	0.48	-
CTV (ml)	CT	43.9 (55%)	18.9-152.0	-	0.052	**0.0022**
	MRI	47.7 (60%)	24.4-152.3	-	-	**0.0005**
	PET	38.7 (48%)	19.8-159.0	-	**0.0005**	-
	CT10mm	79.9 (100%)	39.2-226.3	-	-	-
Tumor volume not included in CTV (ml)	CT	0.01	0.00-0.39	-	-	-
	MRI	0.01	0.00-0.34	-	-	-
	PET	0.03	0.00-0.53	-	-	-
	CT10mm	>0.01	0.00-0.24	-	-	-

$Tumor_{H\&E}$ volume: histopathological tumor volume based on hematoxylin-eosin staining; p95-distance: 95th percentile of the distances measured form the $tumor_{H\&E}$ to the $tumor_{H\&E}$-GTV intersection. Numbers in bold are significantly different (p-values < 0.05; Wilcoxon-signed-ranked test).

Figure 4. Determination of the margin required to cover the tumor for the GTV delineation of the different imaging modalities and the resulting missed volume after application of this margin in all patients. (a) p95 distances for CT (circle), MRI (square) and PET (triangle) plotted against the tumor$_{H\&E}$ volume. Target margins are represented for CT (dotted line), MRI (dashed-dotted line) and PET (dashed line). (b) Missed tumor$_{H\&E}$ volume after adding the target margin ranked according to the tumor$_{H\&E}$ volume for CT (white), MRI (black) and PET (gray). # MRI excluded due to time interval ≥20 days. *PET excluded due to time interval ≥20 days.

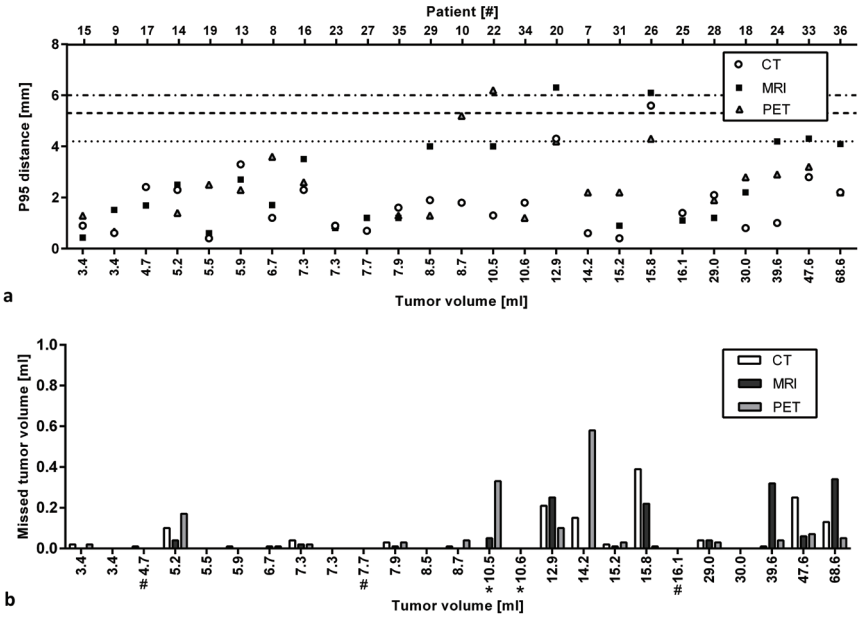
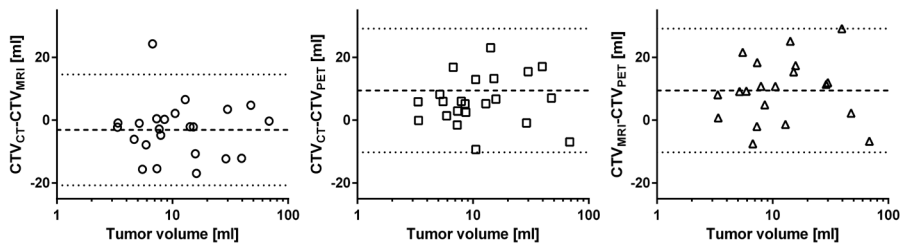


Figure 5. Difference of target volumes between CT and MRI (a), CT and PET (b) and MRI and PET (c) versus tumor$_{H\&E}$ volume. Each point represents a patient. Mean difference between target volume and tumor$_{H\&E}$ volume (thick dashed line) and 95% confidence interval (thin dotted lines) are depicted.



GTV delineation

Three dedicated head-and-neck specialists (two radiation oncologists and one radiologist) manually delineated the GTV in consensus on CT as well as on MRI datasets. An endoscopic report was available during the delineation sessions. For MRI, the GTV was delineated on the T1-weighted gadolinium(Gd)-enhanced transverse MRI while applying dedicated delineation guidelines that have been previously validated [16]. All tumor delineation on CT and MRI was performed in Volumetool [17].

The GTV on PET was automatically segmented using a Gaussian mixture model [18], which was adapted for laryngeal and hypopharyngeal tumors [19]. The Gaussian mixture model determines a patient-specific threshold based on the signal intensity histogram of a volume of interest by fitting Gaussian distributions to the histogram using expectation maximization. These distributions, representing regions with different intensity levels in the image, are classified to tumor or background.

The sensitivity and positive predictive value (PPV) were derived for overlap analysis between the $tumor_{H\&E}$ volume and the modality-specific GTV. The sensitivity is defined as the percentage coverage of the $tumor_{H\&E}$ volume by the GTV, and the PPV is defined as the percentage of the GTV that covers the $tumor_{H\&E}$ volume.

Modality-specific derivation of the CTV margin

From the GTV and $tumor_{H\&E}$ contours, a common contour was derived for each patient per modality. The common contour is the delineation enclosing the overlapping volume of the GTV and $tumor_{H\&E}$. The three-dimensional distances from each point of the $tumor_{H\&E}$ to the closest point of the common contour were determined (*Figure 1*) for each patient per modality, yielding a distribution of distances per patient per modality. These measures are referred to as 'tumor-intersection distances' and quantify the amount of tumor that is not included in the GTV.

We aimed to derive margins for each imaging modality that would ensure coverage of at least 95% of the surface of $tumor_{H\&E}$ that is not already included in the GTV identified on images for 95% of patients. Thus, the 95th percentile of the tumor-intersection distances was calculated (p95-distances) for each patient per modality. This 95th percentile cut-off point was chosen to limit the influence of residual registration errors and deformations.

The modality specific margin was then defined as the second highest p95-distance for each modality. This value was chosen as the largest value, representing the patient dataset with the greatest discrepancy between $tumor_{H\&E}$ and the GTV, would result in overestimation of the microscopic tumor extent for nearly all tumors.

To accurately compare the various imaging modalities on a high resolution, all imaging was resampled to a resolution of 0.43x0.43x0.67 mm.

CTV analysis

Each CTV was constructed by isotropically expanding the GTV using the derived modality-specific margin. Subsequently, the CTV for CT and MRI was corrected for anatomical boundaries such as air, pharyngeal constrictor muscles and vertebrae (*Figure 2*, *Figure 3*). As PET imaging does not give anatomical information, correction for air and vertebrae of PET CTVs was based on CT data.

These various CTVs were compared to the CTVs derived via our current clinical practice of expanding the CT-based GTVs is 10 mm margin (CTVCT10).

Statistical analysis

Statistical analysis was performed in GraphPad Prism Version6.07. The non-parametric Wilcoxon-signed-ranked test was used because there was no justification to assume a particular distri-

bution. Differences were assumed to be significant for p-values <0.05.

Results

Large variation in $tumor_{H\&E}$ volume was observed (*Table 1* and *Supplementary table 5*). Nearly all GTVs overestimated the tumor volume. However, the $tumor_{H\&E}$ was not completely covered by the GTV contoured on each patient, which is reflected by a sensitivity less than one for all modalities and patients. The sensitivity was highest for CT compared to MRI and PET. However, the CT images also had the lowest positive predictive value (PPV) (*Supplementary Table 4*). Therefore, from these parameters, the optimal image modality for clinical practice could not be derived.

The median value of the p95-distances, quantifying the margin needed for adequate tumor coverage, was smaller for CT than for PET and MRI (*Table 1*). The modality-specific CTV margins are 4.3 mm (CT), 5.2 mm (PET) and 6.1 mm (MRI) (Figure. 4a).

The CTVs derived using modality-specific margins resulted in similar tumor coverage (*Figure 2 and 3*, *Supplementary table 5*). These CTVs were compared with the $tumor_{H\&E}$ volume and the CTVCT10. Modality-specific CTVs were all smaller than the clinically used CTVCT10 ($p<0.0001$) with a reduction of the median volume of 45% (CT), 40% (MRI) and 52% (PET) (*Table 1*). PET-based CTVs were smaller than CT- and MRI-based CTVs. No significant differences were found between CT- and MRI-based CTVs no differences were found (*Table 1*). No correlation was found between the $tumor_{H\&E}$ volume and the volume differences of the CTVs for the different modalities (*Figure 5*).

As a result of our CTV definition, tumor tissue was located outside the CTVs for several patients. Less tumor tissue was missed for CTVs of smaller tumors (<10 ml) than of larger tumors (*Figure 4b* and *Table 1*). The volume located outside the CTV is a small fraction of the tumor volume with median values of 0.1% for all modalities and maximum fractions of 2.5% (CT), 1.9% (MRI) and 4.1% (PET). Cartilage invasion was not entirely covered in 3 (CT), 2 (MRI) and 4 (PET) patients (*Figure 3*).

Discussion

In this study, clinical target definition was improved by introducing the first evidence-based margins for CTV construction that are optimized for each imaging for laryngeal and hypopharyngeal squamous cell carcinoma. Expansion of the GTV by these margins resulted in a reduction of the average volume of the CTV by 45% - 52% compared with CTVs constructed via our current clinical practice CTVCT10 by expanding the GTV identified on CT images with a margin of 10 mm. This reduction improves normal tissue complication probability (NTCP) of organs at risk [20,21]. The CTV margins were designed to achieve adequate tumor coverage while minimizing the inclusion of healthy tissue and amounted to 4.3 mm (CT), 5.2 mm (PET) and 6.1 mm (MRI). Since all modality-specific CTVs had equal tumor coverage and PET-based CTVs were smallest, i.e. the lowest overestimation, PET-based CTVs were considered to be most accurate. Nevertheless, the differences between the various imaging modalities were small.

For target definition it was remarkable that PET-based CTVs were most accurate. High contrast between tumor and non-tumorous tissue and an accurate automatic segmentation using the Gaussian mixture model may improve the exclusion of non-tumorous tissue in GTVPET compared to CT and MRI. Consequently, the image analysis resulted in a small GTV and a moderate CTV margin. The low resolution of PET imaging appeared to be sufficient for target definition. PET-based target definition has been used for head-and-neck tumors, resulting in smaller target volumes compared with CT [5,15,22].

CT-based CTVs were larger than PET-based CTVs, although the CTV margin was smallest for CT. Low contrast between tumor and surrounding soft tissue on contrast-enhanced CT makes differentiation between tumorous and non-tumorous tissue difficult. This challenge may have encouraged the observers to enlarge the GTV in order to be convinced that all tumorous tissue was included [23].

Target definition on MRI was comparable with CT and was less accurate than PET, despite the advantage offered by superior soft tissue contrast and anatomical detail. The CTV margin for MRI was larger compared to the other imaging modalities. Apparently, the signal intensity of the tumor and the surrounding tissue is heterogeneous, and peritumoral inflammation may mimic neoplastic invasion [24], leading to difficulties in interpretation of the large variety of signal intensities [16]. New MRI contrasts, such as diffusion weighted imaging (DWI), may improve GTV delineation and target definition [25].

The tumor tissue missed outside the GTV was mostly due to partial volume effects. Cartilage invasion was also not adequately included in GTVs delineated on all modalities. This is inconvenient as cartilage invasion is used for TNM-stage classification and influences the choice of treatment. The preferred treatment of patients with observed invasion through the outer cortex of the thyroid is surgery as radiation treatment leads to higher recurrence rates [26].

The main purpose of our study is the translation of these histological findings into a validated CTV concept, which has been shown to decrease target volumes and thus is expected to reduce complications to organs at risk. In the literature, no consensus on how to determine a CTV margin exists, but some valuable remarks are made in ICRU report 83. It states that the CTV contains "subclinical malignant disease with a certain probability of occurrence considered relevant for therapy. There is no general consensus on what probability is considered relevant for therapy, but typically a probability of occult disease higher than from 5% to 10% is assumed to require treatment." In other words, the CTV is not considered to cover all tumor tissue for all patients. Here, we aimed to derive target margins that would ensure tumor coverage of at least 95% of the histopathologically-identified tumor for at least 95% of the patients. These margins resulted in a maximum volume miss of 2% (0.4 ml) with a median of 0.1%. A margin that ensured 100% tumor coverage for all patients would overestimate the volume relevant for therapy for the majority of patients.

The general concept is that the CTV includes microscopic tumor islands, which are separated from the gross tumor bulk. In our study, however, isolated tumor was not observed on approximately 800 H&E-sections. Future research with cytokeratin staining of histopathological sections, which has a high sensitivity for detection of squamous cell carcinoma, will be performed to confirm the absence of isolated tumor islands.

In this study, isotropic CTV margins were used and afterwards corrected for clear anatomical boundaries such as air cavity, vertebrae and pharyngeal constrictor muscles. Further reduction of the CTV might have been possible when anisotropic margins were used. However, due to the limited number of patients involved in this study derivation of guidelines for anisotropic margins was not possible.

A limitation of our study is that tumors eligible for radiotherapy are typically lower staged (T1b-T3) than tumors in the present study (T4-tumors). Higher stage tumors are often larger than lower stage tumors, what could have impact on the CTV margin necessary to ensure tumor coverage. In our study, less $tumor_{H\&E}$ volume was missed on the modality-specific CTVs for smaller tumors (<10 ml) compared to larger tumors (*Figure 4b*). This result indicates that smaller tumors have less microscopic tumor extent or that observers

delineate smaller tumors more accurately. Based on this finding, we expect that the CTV margins derived for the higher staged tumors are sufficient for lower staged tumors.

Although image registration was optimized, registration mismatch could not fully be prevented. We manually improved the automatic rigid registration process which was described in a previous publication [15]. Despite this additional non-rigid registration, some deformations remained and resulted in a suboptimal registration between histology and the imaging modalities. To account for this, less than 5% tumor volume outside the CTV is accepted.

It is important to note that for CT and MRI the CTV margin is based on manual GTV delineations, which also include observer delineations errors, i.e. visible gross tumor that is not delineated. In the determination of the CTV margin, both human errors and misses due to lack of visibility are jointly analyzed, although delineation errors are by definition not included on the CTV margin. To minimize human errors, we used consensus delineation by three observers. It is strongly advised to review by a second observer and compare imaging of various modalities during GTV delineation to reduce delineation errors. We observed that in case of delineation errors, the delineations on one of the other did include the missed tumor part.

Another issue is that inter- and intra-observer variation can be large for manual GTV delineations on CT and MRI [9,23], whereas automatic GTV segmentation on PET does not suffer from these variations. This implies more standardization and easier validation than manual GTV delineations. Conversely, automatic segmentation can vary considerably with acquisition and reconstruction method. Using other PET segmentation methods, for example 50% SUVmax or manual delineation, might result in suboptimal GTV delineations, increasing the CTV margin needed. Furthermore, PET cannot be used as a stand-alone modality because of the lack of anatomical information which is needed to reduce the CTV by excluding anatomical structures like uninvolved muscle and cartilage. It should be noted that all observers involved in this study were experienced in delineating tumors in head-and-neck.

Generally, a fixed margin for CTV is used independent of imaging modality. In our study, this would have led to large differences in CTV following the differences in GTV. These differences would be partly compensated with the introduction of our new CTV approach using modality-specific margins which also substantially reduce the CTV. The final CT- and MRI-based CTVs were comparable, but were larger than PET-based CTVs. Whether these differences are clinically significant should be analyzed in a broader perspective. For example, costs and availability of the various imaging modalities should be taken into account. MRI and CT scans are less expensive than PET scans, while MRI and PET are generally less easily accessible than CT. When PET is available, we suggest to use automatic segmentation for primary target definition to decrease inter- and intraobserver variability. However, CT or MRI based target definitions are good alternatives when proper, i.e. modality specific, margins are used.

A combination of the various imaging modalities might further improve target definition, because the modalities provide complementary information. CT can provide anatomical information, MRI can add soft-tissue contrast and PET shows the metabolic activity. However, further research is needed to decide which combination is optimal for target definition.

In conclusion, although in general GTV overestimated the tumor volume in all modalities, CTV margins were needed to achieve adequate tumor coverage. In our study, target definition was improved by developing modality-specific

CTV margins of 4.3 mm (CT), 6.1 mm (MRI) and 5.2 mm (PET). This resulted in CTVs with 45%-52% volume reduction compared with a clinical CTV based on GTV delineation on CT with a margin of 10 mm. PET-based CTVs were smallest, and CT and MRI-based CTVs were not significantly different from each other.

Authors

Elise Anne Jager (first author), Hans Ligtenberg (first author), Joana Caldas-Magalhaes, Tim Schakel, Frank A. Pameijer, Nicolien Kasperts, Stefan M. Willems, Chris H. J. Terhaard, Cornelis P. J. Raaijmakers, Marielle E. P. Philippens.

Published

Radiother Oncol. 2017 Apr;123(1):63-70.

Supplementary table 1. CT imaging parameters.

Parameters	CT	FDG-PET/CT
Number of patients	21	6
Median time interval between imaging and surgery	1 days (range: 1-13)	8 days (range: 1-16)
Scanner model	Brilliance iCT 256	mCT-Biograph
Manufacturer	Philips Medical Systems	Siemens Medical Solutions
Injected Iodine contrast	Total volume: 90 ml; rate: 2 ml/s; delay: 65 s	Total volume: 90 ml; rate: 4 ml/s
Tube potential	120 kVp	120 kVp
Tube current-time product	220 mAs with dose reduction	100 mAs with dose reduction
Reconstructed voxel size	(0.43-0.98)x(0.43-0.98)x(0.67-2.0) mm	(0.49-0.57)x(0.49-0.57)x2.0 mm

Supplementary table 2. MRI scanner and sequence parameters.

Median time interval between imaging and surgery		10 days (range: 1-17)
Scanner model		Achieva 1.5 T
Manufacturer		Philips Medical Systems
MR Imaging parameters	**MRI sequences**	
	tT1w	**T2w**
TR (ms)	593	4200
TE (ms)	15	130
Slice thickness (mm)	4*	4-5
Gap between slices (mm)	0.4*	0.4-0.5
Number of averages	2	2
Acquisition matrix (mm^2)	512x307	512x299
FOV (mm^2)	210x148	210x148
Reconstruction matrix (mm^2)	512x512	512x512
Reconstructed resolution (mm^2)	(0.41-1.00) x (0.41-1.00)	(0.41-0.50x0.41-0.50)

tT1w: transverse T1 weighted image, T2w: T2 weighted images. TR: repetition time,TE: echo time, FOV: field of view *patient had a slice thickness of 2.2 mm and spacing between slices of 1.0 mm.

Supplementary table 3. CT imaging parameters.

Parameters	FDG-PET	FDG-PET/CT
Number of patients	13	14
Median time interval between imaging and surgery	2 days (range: 1-13)	8 days (range: 1-17)
Scanner model	Allegro PET System	mCT-Biograph
Manufacturer	Philips Medical Systems	Siemens Medical Solutions
Injected FDG dose	3.7 MBq/kg	2.0 MBq/kg
Acquisition time per bed positions	2-3	2
Reconstruction method	OSEM with 3D row-action maximum-likelihood algorithm	OSEM with PSF and TOF
Post-reconstruction filter	4.0 mm FWHM Gaussian filter	5.0 mm FWHM Gaussian filter
Reconstructed voxel size	4.0x4.0x4.0 mm	2.04x2.04x1.5 mm

FDG: Fluorodeoxyglucose; OSEM: Ordered subset expectation maximization; PSF: Point-spread function; TOF: Time-of-flight; FWHM: Full width at half maximum

Supplementary table 4. GTV analysis parameters.

Patient	Tumor$_{H\&E}$ volume (ml)	GTV (ml)			Sensitivity*			PPV			p95 distance (mm)			max. distance (mm)		
		CT	MRI	PET	CT	MRI	PET	CT	MRI	PET	CT	MRI	PET	CT	MRI	PET
15	3.4	9.1	6.8	6.1	0.93	0.96	0.91	0.34	0.48	0.51	0.9	0.4	1.3	2.3	2.2	2.6
9	3.4	8.1	4.9	8.4	0.97	0.86	0.98	0.40	0.59	0.39	0.6	1.5	0.7	1.7	4.3	2.0
17	4.7	5.9	6.0	†6.2	0.83	0.90	†0.74	0.65	0.70	†0.56	2.4	1.7	†5.1	4.6	3.6	†7.1
14	5.2	10.9	8.2	8.1	0.86	0.85	0.92	0.41	0.54	0.59	2.3	2.5	1.4	6.9	5.1	3.9
19	5.5	13.6	15.5	9.9	0.98	0.97	0.84	0.40	0.35	0.47	0.4	0.6	2.5	3.4	1.9	4.0
13	5.9	7.7	7.9	6.7	0.81	0.84	0.85	0.63	0.63	0.75	3.3	2.7	2.3	6.4	5.2	4.7
8	6.7	23.0	9.0	13.9	0.94	0.89	0.82	0.28	0.66	0.40	1.2	1.7	3.6	4.1	3.4	6.2
16	7.3	15.4	12.3	15.5	0.85	0.85	0.87	0.40	0.50	0.41	2.3	3.5	2.6	4.1	5.3	4.9
23	7.3	14.3	16.2	11.3	0.96	0.97	0.96	0.50	0.44	0.63	0.9	0.8	0.9	3.6	2.9	3.4
27	7.7	19.4	15.5	†8.5	0.98	0.96	†0.83	0.39	0.48	†0.75	0.7	1.2	†3.4	2.1	2.8	†5.0
35	7.9	15.6	12.6	12.7	0.93	0.93	0.93	0.47	0.59	0.58	1.6	1.2	1.3	5.2	3.4	3.8
29	8.5	17.5	13.2	14.2	0.90	0.82	0.93	0.44	0.53	0.56	1.9	4.0	1.3	3.6	7.8	3.2
10	8.7	17.2	†15.2	14.8	0.91	†0.96	0.75	0.46	†0.55	0.44	1.8	†0.9	5.2	4.9	†3.1	7.0
22	10.5	16.5	12.6	7.5	0.95	0.84	0.59	0.60	0.70	0.84	1.3	4.0	6.2	2.7	8.7	12.1
34	10.6	14.0	†13.2	17.7	0.90	†0.85	0.96	0.68	†0.68	0.58	1.8	†4.1	1.2	3.8	†6.6	2.5
20	12.9	18.3	10.6	15.3	0.81	0.65	0.80	0.57	0.80	0.67	4.3	6.3	4.2	9.2	12.3	8.8
32†	†13.6	†17.3	†14.2	†16.1	†0.81	†0.77	†0.83	†0.64	†0.74	†0.71	†4.3	†8.7	†3.8	†9.5	†14.7	†8.4
7	14.2	32.7	25.5	18.5	0.97	0.97	0.89	0.42	0.54	0.68	0.6	0.6	2.2	2.8	3.8	6.6
31	15.2	34.1	26.4	24.6	0.98	0.96	0.90	0.44	0.55	0.56	0.4	0.9	2.2	4.8	4.1	5.7
26	15.8	22.6	21.8	11.3	0.90	0.90	0.57	0.63	0.65	0.79	5.6	6.1	4.3	10.7	10.7	6.1
25	16.1	28.7	27.0	†19.3	0.96	0.97	†0.85	0.54	0.58	†0.71	1.4	1.1	†3.0	5.3	4.0	†5.7
11†	†17.7	†27.9	†22.6	†15.6	†0.93	†0.97	†0.74	†0.59	†0.76	†0.84	†1.4	†0.8	†4.6	†3.3	†2.5	†10.3
28	29.0	33.2	34.8	32.6	0.92	0.95	0.91	0.80	0.79	0.80	2.1	1.2	1.9	5.4	2.6	3.9
18	30.0	46.2	36.6	33.3	0.98	0.93	0.89	0.64	0.77	0.80	0.8	2.2	2.8	3.1	6.1	5.8
24	39.6	63.1	58.1	47.4	0.98	0.89	0.90	0.61	0.61	0.76	1.0	4.2	2.9	4.9	12.2	6.7
33	47.6	64.9	44.7	52.5	0.95	0.83	0.85	0.70	0.88	0.77	2.8	4.3	3.2	10.6	8.5	6.2
36	68.6	88.7	66.3	82.7	0.94	0.82	0.93	0.72	0.84	0.77	2.2	4.1	2.2	7.6	10.4	8.8
Median	**8.7**	**17.5**	**15.5**	**14.5**	**0.94**	**0.90**	**0.90**	**0.50**	**0.59**	**0.61**	**1.6**	**1.7**	**2.3**	**4.6**	**4.3**	**5.3**
Minimum	**3.4**	**5.9**	**4.9**	**6.1**	**0.81**	**0.65**	**0.57**	**0.28**	**0.35**	**0.39**	**0.4**	**0.4**	**0.7**	**1.7**	**1.9**	**2.0**
Maximum	**68.6**	**88.7**	**66.3**	**82.7**	**0.98**	**0.97**	**0.98**	**0.80**	**0.88**	**0.84**	**5.6**	**6.3**	**6.2**	**10.7**	**12.3**	**12.1**

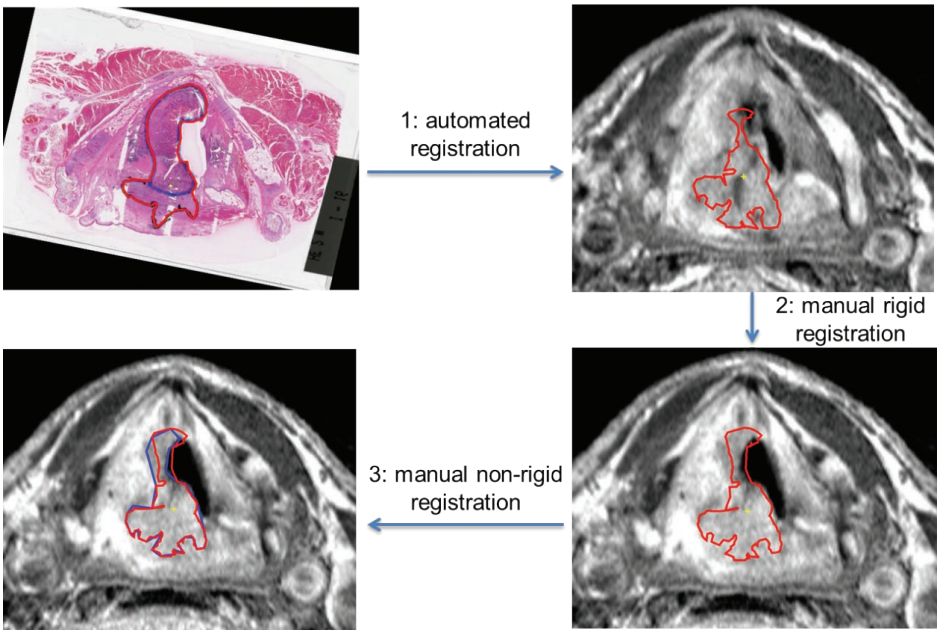
GTV analysis parameters. Overlap analyses for GTV and tumor volume (tumor$_{H\&E}$) reflected as sensitivity and positive predictive value (PPV). Distance analysis: p95-distance: 95th percentile of the tumor-intersection distance per patient for each modality. Patients are listed by tumor size. †timeinterval between imaging and surgery ≥20 days; these data are excluded from further analysis, the median, minimum and maximum. * Sensitivity is part of tumor$_{H\&E}$ covered by GTV; PPV (Positive Predictive Value) is part of GTV covering tumor$_{H\&E}$.

Supplementary table 5. GTV analysis parameters

Patient	$Tumor_{H\&E}$ volume (ml)	CTV (ml)			
		CT_{10mm}	$CT_{4.3mm}$	$MRI_{6.1mm}$	$PET_{5.2mm}$
15	3.4	44	25.65	27.89	20
9	3.4	44	23.5	24.37	24
17	4.7	39	18.88	24.96	†19
14	5.2	63	30.48	31.46	22
19	5.5	59	35.44	51.08	29
13	5.9	44	22.14	29.97	21
8	6.7	93	53.82	29.47	37
16	7.3	69	36.12	35.67	38
23	7.3	60	34.57	47.65	32
27	7.7	81	45.08	47.89	†28
35	7.9	65	38.95	43.75	33
29	8.5	80	43.91	43.62	39
10	8.7	72	41.21	†43.29	39
22	10.5	62	37.38	35.19	24
34	10.6	68	36.93	†37.66	46
20	12.9	80	44.14	37.56	39
32†	†13.6	†76	†43.81	†46.56	†41
7	14.2	105	67.4	69.5	44
31	15.2	108	69.65	71.78	56
26	15.8	85	49.94	60.62	43
25	16.1	106	62.53	79.47	†52
11†	†17.7	†88	†58.27	†61.24	†39
28	29.0	121	72.04	84.32	73
18	30.0	142	88.85	85.38	73
24	39.6	180	117.31	129.41	100
33	47.6	170	114.81	110.02	108
36	68.6	226	152.04	152.3	159
Median	**8.7**	**80**	**44**	**48**	**39**
Minimum	**3.4**	**39**	**19**	**24**	**20**
Maximum	**68.6**	**226**	**152**	**152**	**159**

CTV after expansion of GTVs by modality dependent CTV margins. CTV10mm: addition of a clinically used CTV margin of 10 mm to the GTVs. Patients are listed by tumor size. †time-interval between imaging and surgery ≥20 days; these data are excluded from further analysis, the median and the range.

Supplementary figure 1. Example of the three main registration steps applied to the imaging and tumor$_{H\&E}$ delineation. Shown is an H&E-section and transverse view of the corresponding MRI scan. The first step is an automatic rigid registration as described in [15], which is mainly controlled by. The second manual rigid registration step improves the registration near the air cavity. In the third step de contour is manually reshaped to further improve the delineation with the air cavity and the visible anatomical boundaries.



References

1 Rasch C, Steenbakkers R, van Herk M. Target Definition in Prostate, Head, and Neck. Semin Radiat Oncol 2005;15:136–45.

2 Grégoire V, Coche E, Cosnard G, Hamoir M, Reychler H. Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience. Radiother Oncol 2000;56:135–50.

3 Report 83: Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). J ICRU 2010;10:41–53.

4 Burnet NG, Thomas SJ, Burton KE, Jefferies SJ. Defining the tumour and target volumes for radiotherapy. Cancer Imaging 2004;4:153–61.

5 Daisne J-F, Duprez T, Weynand B, Lonneux M, Hamoir M, Reychler 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.

6 Campbell S, Poon I, Markel D, Vena D, Higgins K, Enepekides D, et al. Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: Implications for the management of head-and-neck cancers. Int J Radiat Oncol Biol Phys 2012;82:574–81.

7 Purdy JA. Current ICRU definitions of volumes: limitations and future directions. Semin Radiat Oncol 2004;14:27–40.

8 Daisne J-F, Sibomana M, Bol A, Cosnard G, Lonneux M, Grégoire V. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 2003;69:237–45.

9 Jager EA, Kasperts N, Caldas-Magalhaes J, Philippens MEP, Pameijer FA, Terhaard CHJ, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015;10:1–9.

10 Moeckli R, Jeanneret Sozzi W, Mirimanoff R-O, Ozsahin M, Zouhair A, Valley JF, et al. Physical considerations on discrepancies in target volume delineation. Z Med Phys 2009;19:224–35.

11 Caudell JJ, Meredith RF, Spencer SA, Keene KS, Dobelbower MC, Bonner JA. Margin on Gross Tumor Volume and Risk of Local Recurrence in Head-and-Neck Cancer. Int J Radiat Oncol Biol Phys 2010;76:164–8.

12 Hong TS, Tome WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiother Oncol 2012;103:92–8.

13 Bird D, Scarsbrook AF, Sykes J, Ramasamy S, Subesinghe M, Carey B, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. BMC Cancer 2015;15:844.

14 Verduijn GM, Bartels LW, Raaijmakers CPJ, Terhaard CHJ, Pameijer FA, van den Berg CAT. Magnetic Resonance Imaging Protocol Optimization for Delineation of Gross Tumor Volume in Hypopharyngeal and Laryngeal Tumors. Int J Radiat Oncol Biol Phys 2009;74:630–6.

15 Caldas-Magalhaes J, Kasperts N, Kooij N, Van Den Berg C a T, Terhaard CHJ, Raaijmakers CPJ, et al. Validation of imaging with pathology in laryngeal cancer: Accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 2012;82:289–98.

16 Jager EA, Ligtenberg H, Caldas-Magalhaes J, Schakel T, Philippens MEP, Pameijer FA, et al. Validated guidelines for tumor delineation on MR-imaging for laryngeal and hypopharyngeal cancer. Acta Oncol 2016.

17 Bol GH, Kotte ANTJ, van der Heide UA, Lagendijk JJW. Simultaneous multi-modality ROI delineation in clinical practice. Comput Methods Programs Biomed 2009;96:133–40.

18 Aristophanous M, Penney BC, Martel MK, Pelizzari CA. A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. Med Phys 2007;34:4223–35.

19 Caldas Magalhaes J, Raaijmakers CP, Aristophanous M, Lee JA, Kasperts N, Jager EA, et al. FDG-PET Semi automatic Segmentation Methods for GTV Delineation in Laryngeal and Hypopharyngeal Cancer. Int J Radiat Oncol 2014;90:S536.

20 Vugts CAJM, Terhaard CHJ, Philippens MEP, Pameijer FA, Kasperts N, Raaijmakers CPJ. Consequences of tumor planning target volume reduction in treatment of T2-T4 laryngeal cancer. Radiat Oncol 2014;9:195.

21 Chen AM, Yu Y, Daly ME, Farwell DG, H. Benedict S, Purdy JA. Long-term experience with reduced planning target volume margins and intensity-modulated radiotherapy with daily image-guidance for head and neck cancer. Head Neck 2014;36:1766–72.

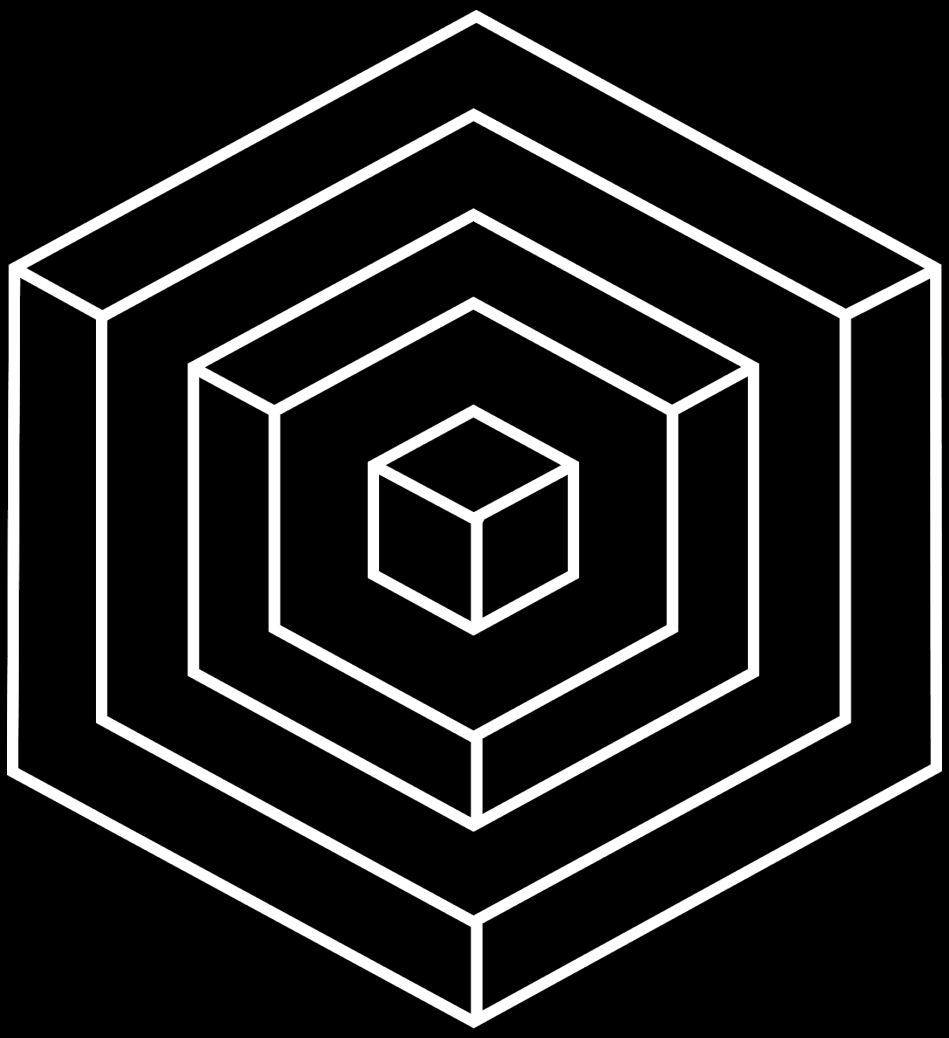
22 Leclerc M, Lartigau E, Lacornerie T, Daisne J-F, Kramar A, Grégoire V. Primary tumor delineation based on 18FDG PET for locally advanced head and neck cancer treated by chemo-radiotherapy. Radiother Oncol 2015;116:87–93.

23 Caldas-Magalhaes J, Kooij N, Ligtenberg H, Jager E a., Schakel T, Kasperts N, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol (Madr) 2015;54:1181–7.

24 Becker M, Zbaren P, Casselman JW, Kohler R, Dulguerov P, Becker CD. Neoplastic Invasion of Laryngeal Cartilage : Reassessment of Criteria for Diagnosis at MR Imaging. Radiology 2008;249:551–9.

25 Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother Oncol 2013;109:394–7.

26 Castelijns JA, Becker M, Hermans R. Impact of cartilage invasion on treatment and prognosis of laryngeal cancer. Eur Radiol 1996;6:156–69.



Chapter 6
Summary

During the past decennia, radiation of the target has become more precise (1–5). In order to benefit from these high precision techniques, an accurate delineation of the target is essential. In this thesis, imaging modalities for Gross Tumor Volume (GTV) delineation are validated by using histopathology as a gold standard. The need for this research stems from a general problem in radiotherapy compromising the accuracy of the radiation treatment: variation among observers in delineation of the target. This variation is mainly due to several factors: visibility of the target including its extensions, disagreement on target extensions, and interpretation or lack of delineation protocols (6, 7). However, efforts undertaken in radiotherapy have not yet been able to reduce the influence of these factors in such way that variation is no longer the major source of errors (8). The variation remains large for head-and-neck cancer, even with the use of various imaging modalities (8–12). In chapter 2 this problem was outlined by an interobserver study, which resulted in high variation between observers for delineation of the GTV for supraglottic laryngeal carcinoma. By combining CT images and MR images for delineation, this variation was even higher than for CT only delineations (11). Our conclusion was endorsed by research performed by Anderson et al (9), however, the opposite was also asserted (6, 10). Because of higher soft tissue contrast on MRI, the visibility of soft tissue structures is enhanced (13, 14). However, combining MRI and CT did not result in better agreement. In my opinion this increased visibility contributes to a wider variety of choices that need to be made during GTV delineation. This consequently leads to an increase in variation between observers instead of the desired decrease in variation. Clear interpretation and delineation guidelines might reduce this interobserver variation (11, 15–18). When interobserver variation is high, precision i.e.: the resemblance of the delineation when performed by various observers, is low. What about the accuracy i.e.: trueness of the delineation, and how to increase the accuracy for GTV delineation? Validation studies, necessary to establish the accuracy, have been conducted although these are scarce and complex to perform (19, 20). It is this complexity that also had to be dealt with in this thesis. In this thesis, images on various imaging modalities were compared to histopathology of on average 800 H&E-sections of 27 surgical laryngeal specimens. The specimens and H&E-sections were reconstructed in 3D to enable comparison to the in vivo images accomplished before surgery. Several steps were undertaken to optimize the method used for validation (21), where as a first step the precision of the gold standard was investigated (chapter 3). Due to the enhanced soft tissue contrast of MRI compared to CT (6, 11, 13, 15, 22–24), MRI was investigated for GTV delineation. Guidelines for GTV delineation based on guidelines for cartilage invasion on MRI (25) were developed to investigate whether these guidelines would increase the accuracy of the delineation (chapter 4). Several studies have suggested that delineation guidelines could increase the agreement between observers and therefore increase the precision. By developing GTV delineation guidelines, we therefore aimed to increase the accuracy of the delineation. In this section, overestimation and underestimation of tumor tissue were determined before and after the appliance of the MRI specific guide-

lines for GTV delineation. Even more difficult and clinically relevant is the determination of the CTV, which is the volume of tissue that contains a demonstrable GTV and/or subclinical malignant disease with a certain probability of occurrence considered relevant for therapy (26). This microscopic spread is not visible on images. However, the degree of visibility, and therefore invisibility, varies between imaging modalities depending on the specific qualities of the imaging modality. The size of the CTV margin, which is needed to include the presumed microscopic spread in the target, varies considerably between observers and institutes and is merely historically determined (27).

A distance analysis was conducted based on underestimation of the histopathologically determined tumors for CT, MRI and FDG-PET. These distances were used to develop modality specific CTV margins (chapter 5). GTVs with addition of these margins resulted in CTVs that are halved in volume compared to CTVs that were derived from GTVs with the clinically used CTV margin of 10 mm. The addition of modality specific CTV margins would lead to smaller volumes to be treated with high dose radiation and consequently to a decrease in radiation to non-tumorous tissues resulting in less damage to healthy tissue (chapter 5). Combining validated imaging modalities in order to create an optimal combination for visualization of the target could be a future goal in radiotherapy. This would enable radiation treatment to benefit from high precision radiation techniques (2). In order to benefit from systems that allow imaging of the tumor during radiation treatment, such as: MR-linac (28) and MRIdian (29), validated guidelines for interpretation and delineation of the target are essential. Additionally, further development of isotropic CTV margins to anisotropic CTV margins based on the prevalence of tumor misses geographically should be a future research item which would lead further reduction of the overestimation.

Chapter 6
General Discussion

Precision and interobserver variation

Precision is a term that is used to indicate the consistency of results when repeated. In terms of interobserver variation, the precision is the resemblance of the delineation when performed by various observers. In delineation studies for radiotherapy, precision is frequently confused with accuracy. Consensus among observers is considered to be an accurate GTV delineation. Although, when determining the accuracy of the delineation, the exact location and the size of the tumor has to be known. Variation between observers for target delineation studies is high (8–12), especially in head-and-neck cancer where various critical tissue structures are located very close to each other. To express interobserver variation, several measures are used (30–32). These measures are predominantly based on volume comparison, the distances measured between points on contours, and the volume of the overlap between contours. Moreover, the indices used to express the measures vary. This creates difficulties when study results are compared and thus only general conclusions from such comparisons can be drawn. To increase the agreement between observers, various steps have been undertaken i.e.: the development of delineation guidelines and atlases (33), improving image quality and combining various imaging modalities for improvement of tumor visualization. Despite efforts to increase agreement between observers, interobserver variation remains high. In head-and-neck cancer more variation between observers is detected compared to other tumor sites (34). Furthermore, the variation between observers for GTV delineation is greater than variation between pathologists in tumor delineation for the gold standard (chapter 3). The gold standard and its precision will be further explained later in the discussion. More important is the effort that has been undertaken in radiotherapy to increase the precision in radiation of the target. Three sources of geometrical uncertainty may hamper the exact delivery of radiation to the target: patient- set up variation, organ motion and deformation and machine related set up errors (35). Patient set-up errors are attributable to variations in the daily positioning of the patient on the treatment couch. Set-up errors are a source of uncertainty that have been well investigated for all tumor sites. Although this error is limited for head-and-neck radiation when thermoplastic fixation masks are used (36), it is estimated to vary within a range of 3 mm. It has been known that set-up errors are much smaller than the variation in delineation. Therefore this delineation variation undermines the benefits of high precision radiation and the developments that have been undertaken to improve the patient set-up (37, 38).

In chapter 2, CT images were combined with MR images for GTV delineation of supraglottic laryngeal carcinoma and were compared to CT only GTV delineations. CT in combination with MRI showed no improvement concerning the interobserver variation, the variation even increased. This in contrast to what some researchers have claimed (6, 15). A wide range of factors (e.g.: imaging protocols, quality of images, experience of observers, use of guidelines and interpretation in the basic definitions of the Commission on Radiation Units and Measurements (ICRU)) influencing the variation between observers may lead to different conclusions (27).

Our conclusion was that it was more difficult to delineate an imaging modality such as MRI where many different structures are visualized and when no clear guidelines are available. As John Lennon already appeared to know in 1980, he sang: "The more I see, the less I know for sure."(39)

The development and use of guidelines for delineation have been suggested in various studies with the aim of improving the precision of the delineation (11, 15–18). I assert that it would be logical and necessary to first investigate the accuracy of the delineations and improve accuracy before increasing the precision. Validation studies for head-and-neck cancer are both scarce (20) and complex to perform robustly, but are, however, essential for the optimization of the effect of the precise radiation treatment.

Accuracy for delineation of the target

Accuracy for contouring the GTV is the resemblance of a delineation compared to the outline of the 'true tumor'. In intensity modulated radiotherapy, which is increasingly used as a treatment for head-and-neck cancer, there is a steep dose gradient outside the target volume. It is, therefore crucial to both accurately determine and delineate the target volume thereby avoiding geographical misses or excessive radiation doses to non-tumorous tissue (18, 40, 41). The development of imaging and radiation techniques to monitor the target (i.e: Image Guided Radiotherapy) addresses the problem: not knowing the exact location of the tumor is at the time of treatment, hence although the treatment can be precise, the accuracy is low (8). "Proper contouring of the target improves accuracy whereas image guidance improves precision. In other words, you can consistently hit the wrong target (high precision but poor accuracy), but what is in fact required is to hit the right target consistently (high precision, high accuracy)"(8). By delineating the target as accurately and precisely as possible, one would, in an ideal situation, neither miss any tumor tissue nor harm healthy tissue because there would be no overestimation. This ideal situation is currently utopian. 'Accurate' in this thesis refers to a situation where almost no tumor would be missed and overestimation greatly diminished. At this point, as shown in chapter 4 and 5 of this thesis, missing no tumor in any of the delineations for laryngeal and hypopharyngeal carcinoma seems impossible. For tumor delineation in radiotherapy definitions as 'GTV' and 'CTV' are used. These definitions are not used for pathological analysis. To enable comparison between delineated tumor on imaging and histopathology, the tumor delineated on the H&E-sections was considered as the CTV while the tumor delineated on imaging was regarded as the GTV. The tumor tissue located outside the GTV was determined by comparing the imaging based GTV with the CTV contour on histopathology. In other words, the histopathology contour extensions outside the imaging based GTV contour, were the result of tumor tissue that was missed by the radiation oncologists when delineating the GTV. This missed tumor tissue was regarded as subclinical malignant disease, which was not visible on images and should yet be encompassed by a CTV margin. The extensions of the missed tumor tissue differed between imaging modalities. Consequently, a modality specific margin was added around the GTV to compensate for these misses.

Margins in radiotherapy

In clinical radiotherapy, several margins are designed to compensate for movement of the tumor, patient set-up variations and microscopic spread of the tumor (26). The 'types' of margins used and their size differ per tumor site and even differ between institutions and cooperative groups (18, 27, 40, 41). Where GTV is defined as the gross palpable or demonstrable extent and location of malignant growth as defined by the ICRU (International Commission on Radiation Units & Measurements) (26). In clinical radiotherapy practice this is applied as GTV being

the part of tumor that is visible with the use of 3D imaging in combination with findings based on endoscopy and physical examination. Consequently, the delineated tumor volume is mostly dependent on the imaging modality utilized and the data acquisition process (8). The clinical target volume (CTV) is the volume that includes the GTV as well as subclinical and microscopic anatomical spread patterns (26). The subclinical and microscopic anatomical spread is not visualized by images because they are below the resolution limits of modern imaging techniques. The GTV volume requires the highest radiation dose to eradicate high density of cancer cells whereas the lower density of cells found in the CTV should be irradiated with a lower radiation dose according to radiological principles (42) and control probability models (43). The estimate of a CTV margin will be influenced by a number of uncertainties. Microscopic extent can be extremely variable depending on the tumor site and type. Microscopic disease can be nodal, perineural, perivascular, intramuscular, intraparenchymal, perilymphatic, or extent in bone. Although CTV margins could differ per tumor site, type and stage, the CTV margin that is clinically applied seems to be more of a historical determined margin that has developed with the clinical experience of an institution (27, 44, 45) ranging from 5 mm to 2 cm (46). The CTV definition itself is, unless it's clear definition given by the ICRU, particularly prone to variation between observers and institutions. This is due to variable knowledge and/or interpretation in the basic ICRU definitions, for example whether or not lymph nodes should be included in the CTV. In other words, the CTV definition given by the ICRU is interpreted differently by observers and institutes. The important discrepancies observed in the GTV and CTV delineations are 'moderated' by the PTV (Planning Target Volume) delineations (27). Moreover, "the estimate of the microscopic extent, leading to the magnitude of the CTV margin, depends much on previous experience, based on a few pathological studies of surgical specimens or on patterns of failure after surgery or radiotherapy", according to Jeanneret-Sozzi et al (40). Although the policy within an institution may well have changed since the development of the GTV and CTV definitions in 1993 (47) as a result of experience and research, in my experience there has been no pioneering scientific research that has led to consensus on the appropriate size of the CTV margin. In a study of Campbell (46) based on ten specimens, investigating microscopic disease (MD) outside the GTV for oral tongue cancer, it was concluded that 95% of the microscopic disease was found within 3.95 mm of the GTV. Additionally, the amount of the MD fell rapidly as the distance from the GTV increased. GTV was identified by naked eye resolution while MD was detected on microscopic level. GTV and MD were determined by a pathologist on H&E-sections. Imaging was not involved in the study of Campbell (46) and consequently intermodality variation and interobserver variation, the latter being one of the main sources of errors, were not determined or taken into account.

With the aim of advancing the consensus of delineation of the pathological lymph nodes at risk, various nodal atlases for head-and-neck cancer are published (33, 48) where visibility of malignant nodes are validated with pathology. Unfortunately, robustly performed research for delineation of a margin for microscopic spread (CTV) around the GTV on various imaging modalities for laryngeal and hypopharyngeal tumors is lacking.

Although there is no agreement on the optimal size of the CTV margin between institutes and observers within institutes, questioning the adequateness of a CTV margin in the current form has led to controversy. To quote Moeckli et al (27): "In this study we confirm the well-known issue that there exist wide interobserver variations in the GTV, CTV and PTV delineations. This thesis raises the question of the usefulness of GTV and CTV delineation in the present situation." As soon as one questions the histor-

ically determined usefulness of this margin in its current form, a good deal of resistance will be the response. It is, in my opinion that science is about questioning things that we held to be true, and good scientists are always aware that they might adjust their point of view. "Who is more humble? The scientist who looks at the universe with an open mind and accepts whatever the universe has to teach us, or somebody who says everything in this book must be considered the literal truth and never mind the fallibility of all the human beings involved?" (Carl Sagan, astronomer, scientist). I am aware that this thesis is not about astronomy, nevertheless, in my opinion, an open-mindedness, curiosity and a certain degree of humility allows us to learn, improve and develop whether in the realm of medical science, astronomy or in personal life. The resistance previously mentioned in this section, can be understood from the perspective of doctors who are unwilling to miss tumor tissue because of the risk of tumor recurrence. They are unwilling to change what they regard as a 'winning team' and less treatment seams counterintuitive. However, treating less does not necessarily lead to higher risk of recurrence if the treatment is more accurate. The difficulty arises in balancing the risk of recurrence on one hand and toxicity on the other. I maintain that due to a lack of validation studies, the current use of a CTV margin is chiefly to cover up interobserver variations and delineation inaccuracies more than microscopic spread that is not visible on images.

How accurate is the gold standard?

During the conduct of the research for this thesis, a question that often arose was: "How accurate is our gold standard?" In chapter 3 in this thesis, it was concluded that interobserver variation between pathologists is low and the precision is therefore high. Even though pathologist do not delineate tumor on H&E-sections in clinical practice, a variation of 2% in volume and 2.0 mm between common and encompassing contours was found. The variation was caused by the in- or exclusion of necrotic tissue, more disagreement concerning cartilage invasion, occasional misses, pencil thickness, and delineation style. The question whether the most commonly used H&E staining is sufficient to detect isolated tumor cells at a distance from the main tumor has been asked. Isolated tumor cells would eventually grow out to islands and consequently would be detected by H&E staining. However, these isolated islands are not present on the H&E-sections manufactured from the specimens. In our experience, no isolated tumor cells on a distance from the delineated tumor on the H&E-sections are missed with H&E staining. To confirm this, keratin stained sections were manufactured from the specimens. This staining is more specific for detection of isolated tumor cells. This research (not in this thesis) showed no tumor cells at a distance more than from the GTV delineated on H&E-sections, and therefore less variation than the interobserver variation between pathologists for delineation of the GTV on H&E-sections. Furthermore, the distance between H&E-sections was 3 mm, in accordance with the slice thickness of the images. Tissue of the specimen between the H&E-sections was not analyzed. However, in total, a large number of approximately 800 whole-mount H&E-sections was obtained and analyzed.

Registration

Much effort has been undertaken to optimize the registration between in vivo images and histopathology. This is one of the most challenging areas in this research. Within this project the discrepancy between the in vivo imaging and the ex vivo specimen that has been processes in multiple steps, causes registration difficulties and consequently registration errors (21). To reach an accurate conclusion, precise comparison of histopathological tumor and GTV outlines is essential. We decided, therefore to improve and optimize the rigid image registration, a time consuming but very valuable step. The last step in the optimization was shaping the tumor contour by visual information on pathology and MRI. This was necessary in order to

overcome remaining deformations that had not been resolved by optimization of the automatic rigid registration with a non-rigid registration. After this more subjective step, results were more comparable. All adjustments were made by two competent experts, individually and in consensus. In this situation subjective means in addition to expert knowledge was not replaceable by the automatic performance of a computer.

Modality specific target margins and tumor coverage

Modality specific CTV margins were based on tumor found outside the GTV in comparison to histopathology. They were designed to adequately cover tumor tissue on PET, CT and MRI while minimizing the inclusion of healthy tissue and amounted to 4.3 mm (CT), 5.2 mm (PET) and 6.1 mm (MRI). These margins were based on a distance analysis on underestimation of the tumor. In chapter 4 of this thesis, this distance was referred to as type I distance to discriminate between this distance and type II distance. In the 5th chapter of this thesis the definition 'tumor-intersection distance' was used with similar meaning as type I distance.

Since the modality dependent CTV margins were designed to include for 95 percent of the patients at least 95 percent of the tumor surface in the CTV, these CTVs had a comparable tumor overlap for all modalities. Consequently, the various modalities for delineation could be compared by the amount of overestimation of the CTV. The method for development of the modality specific CTV margins will be discussed in detail later in this section. The CTV margins derived from this research could be considered as a recommendation for CTV margins used in clinical practice, dependent on which imaging modality is preferred for GTV delineation. It takes subclinical malignant disease and GTV delineation inaccuracies into account. Although the CTV margin is originally designed to compensate for microscopic spread, its current clinical application covers up delineation inaccuracies as well, this is also described in the IRCU report 83 "uncertainties in the GTV delineation generally propagate to CTV delineation" (26).

The modality specific CTV margin was based on the second highest p95 tumor-intersection distance for each modality. By using this method, the modality specific CTV margins included for 95 percent of the patients at least 95 percent of the tumor surface. One could debate whether to choose a margin that guarantees in 100% of the cases 95% tumor coverage (highest p95 distance) or even in 100% of the cases 100% coverage (highest value of maximal distances) to diminish or minimize the chance of missing any tumor in all situations. This would not be in line with the definition of the CTV, which refers to a volume that includes "subclinical disease with a certain probability of occurrence considered relevant for therapy" according to the ICRU report 83 (26). Furthermore this report describes that "there is no general consensus on what probability is considered relevant for therapy, but typically a probability of occult disease higher than from 5% to 10% is assumed to require treatment". It has not been stated that a CTV margin should include 100% of the subclinical tissue in 100% of the situations. Choosing the tumor extension (that was not visible on imaging and therefore missed) on the largest distance from the GTV (maximum tumor-intersection distance) as a reflection of the overall situation, would lead to a large overestimation of all other patients in all directions. Even for the tumor for which this maximum tumor-intersection distance was observed and this large CTV margin was based upon, the large CTV margin would lead to overestimation in all other directions because of isotropy of the margin. The CTV volumes based on the maximum tumor-intersection distances, would even be larger than the CTV volumes with addition of the clinically used 10 mm CTV margin. Also, by using the cut-off point of p95 of the tumor-intersection distance for determination of the margin, the effect of remaining registration errors, which are the consequence

of the research setting, in the development of the CTV margins is limited.

As a result of our CTV definition, tumor tissue was located outside the CTVs for several patients. The volume located outside the CTV was a small fraction of the tumor volume with median values of 0.1% for all modalities and maximum fractions of 2.5% (CT), 1.9% (MRI) and 4.1% (PET). Even with the clinical 10 mm CTV margin the maximum fraction of tumor outside the CTV was 1.7%.

On inspection of the areas where tumor was missed, most of the tumor volume was missed due to remaining registration errors between imaging and the H&E-sections generally concerning areas adjacent to the lumen of the larynx/hypopharynx. The $tumor_{H\&E}$ contour (pathology contour) extended in the air when projected on the in-vivo images and was consequently excluded by adjustment of the target delineation to anatomical boundaries. The remaining missed tumor volumes, despite the addition of CTV margins and a clinical CTV 10 mm margin, were observed for tumors extending in retropharyngeal direction into the posterior pharyngeal wall and post cricoid area and for laryngeal cartilage invasion. Concerning the maximum distances, the largest maximum distances were found for larger tumors for CT, MRI as well as for PET. In certain areas it is apparently more difficult to detect tumor tissue independently of the imaging modality.

The margins for the comparison of the PET, CT and MRI were isotropic although adjusted to anatomical boundaries as performed in clinical practice for the CTV margin (49). Because PET lacks anatomical information, CT was used for adjustment of the contour to anatomical boundaries.

By designing anisotropic margins, with larger margins in the direction where the chance of missing tumor is increased and reduction of the margin where smaller margins are sufficient to cover tumor, optimal margins could be achieved (49). Although the number of patients included in this study is larger compared to other validation studies, the number of patients is not sufficient for performance of geographical analyses. This should be a topic for future research when more patients are included.

CTV margins in this thesis are based on T3/T4 tumors and do not include pathological lymph nodes. However, for lower stage tumors, infiltration and tumor extent might be less than for the tumors included in this thesis and a smaller margin could be sufficient. Applying the target margins developed in this thesis could cause overestimation rather than underestimation of the tumor.

Variation between tumors

Some tumors seemed "more difficult to delineate". In other words, the delineations encountered relatively low sensitivity compared to the sensitivity of the delineations of other tumors. This low sensitivity was generally seen in cranial and caudal direction, this direction was also more prone to interobserver variation. The variation might be due to less contrast uptake in caudal and cranial directions, which makes it harder to discriminate between tissue structures, partial volume effect and registration errors. Cranial and caudal directions of the specimen are more prone to deformation and therefore more difficult to register accurately to the various imaging modalities. In addition, it seemed more difficult to accurately determine the GTV in the posterior pharyngeal wall and the postcricoid area. More tumor was missed in that region than in other areas by one or more observers. This finding might be explained by the fact that it is more difficult to differentiate between tumor and swallow muscles because of similar signal intensities or repression of these muscles due to the space occupying mass. An irregular tumor shape was associated with a lower sensitivity compared to compact shaped tumors. For compact shaped

tumor it seemed easier to determine the tumor border. Cartilage was more often missed or more variation between observers was seen. Various studies (25, 50, 51) found MRI to be more sensitive in detection cartilage invasion than CT, although over-diagnosis of the tumor invasion of the thyroid cartilage was seen in MRI due to underlying pathological processes. This has been attributed to reactive inflammation, edema and fibrosis secondary to the underlying pathological process, which cause similar radiological appearances to those seen when thyroid cartilage is infiltrated by tumor. CT can miss minor or early cartilage invasion, as a result of the normal irregular patterns of calcification and ossification of thyroid cartilage (52, 53). The low spatial resolution of PET imaging might not be sufficient to detect areas with minor cartilage invasion or very small tumors and for this reason it is suggested that PET images be combined with MRI or CT images. Unfortunately, have not performed a quantitative analysis for cartilage invasion and calculate its sensitivity and specificity in this thesis. Moreover, cartilage invasion is apparently more difficult to detect (50, 54), unfamiliarity of radiation oncologists with diagnosing and delineation of cartilage invasion might play a role.

In clinical practice, cartilage invasion is often an indication for primary treatment with a total laryngectomy. Therefore, in clinical practice this 'unfamiliarity' or "difficulty to delineate cartilage invasion" may have limited radiotherapeutic consequences. Radiological diagnosis of cartilage invasion is difficult. However, it is essential to determine the stage of the tumor at which the treatment depends (50, 54–57). Currently, treatment for cartilage invaded tumors is shifting increasingly towards primary treatment with radiation. I content therefore, that it is essential to increase the delineation accuracy of cartilage pathology and to develop anisotropic CTV margins adjusted to the extend of cartilage pathology. This in order to deal with various situations that can arise in cartilage pathology. Invasion, penetration, erosion of the cartilage can be observed or the scenario in which tumor is adjacent to the cartilage without changes of the cartilage (54). These situations may all require a different approach concerning the treatment.

FDG-PET, CT and MRI comparison or combination

Expansion of the GTV by modality specific margins resulted in a large CTV reduction of 45%-52% compared with the CTV margin of 10 mm, which is currently used in clinical practice. Since all modality specific CTVs had equal tumor coverage and PET-based CTVs were smallest, i.e. the lowest overestimation, PET-based CTVs was considered to be most accurate. Nevertheless, the differences between the various imaging modalities were small and whether this difference is clinically relevant would be a point of discussion.

The PET images in this thesis were delineated using an automatic segmentation for delineation based on a Gaussian mixture model (58). This makes delineations on PET images not prone to interobserver variation in contrast to CT or MRI. Besides, the automatic delineation can save radiation oncologists' time and diminish the workload.

The imaging modalities investigated in this thesis have their specific qualities and could well be used in combination to create a setting with optimal visualization of the tumor (6, 15). Where CT Is superior is visualizing bony structures, MRI is superior to CT in showing contrast between soft tissues. PET benefits from the ability of radiopharmacological uptake for the visualization of tissues and therefore discriminates between tissues. With the development of various MRI parameters, MRI allows physiological (e.g., dynamic contrast enhanced MRI, metabolic (e.g., MR spectroscopy), and molecular (e.g., diffusion weighted imaging) phenomena to be observed. PET is currently not used as a sole imaging modality because it lacks anatomical information. PET should therefore be used in

combination with CT of MRI. Several studies have investigated the value of PET for the detection of recurrences, nodal metastasis and the primary tumor in head-and-neck cancer (59–61). A large review by Spick et al (62) showed that in general PET combined with MRI and PET combined with CT provide comparable diagnostic information (staging and restaging of the tumor) when MRI is used simply to provide the anatomic framework. Spick's review used research based on the diagnostic use of PET in combination with CT or MRI to provide anatomical data. Its use for determination of the GTV was not studied. By exploiting the qualities of MRI in addition to its use to provide an anatomical framework, MRI might be of additional value in combination with PET compared to PET-CT, CT or MRI as shown by Huang et al (63). Besides, PET-MRI is possibly superior in head-and-neck cancer and for diagnosing cancer recurrence (64).

Consensus delineation or individual delineation

The delineations in this thesis were performed by three observers (chapter 2,3,4) and for the study where imaging modalities were compared (chapter 5), a consensus delineation of those three observers was performed. One single consensus delineation was found to be more practical for comparison of imaging modalities than three individual delineations. Is a consensus delineation also more accurate than an individual delineation? Would one recommend radiation oncologists to delineate in consensus or, which might be more achievable in a clinical setting, an individual delineation with supervision of one or more colleagues? On average, consensus delineations did not show a higher sensitivity than individual delineations for MRI. The sensitivity of one observer for GTV delineation on MRI was higher than the sensitivity of the consensus delineation (Wilcoxon signed rank test p=0.001). The PPV was lower for two observers than for the consensus delineation (Wilcoxon signed rank test, p= 0.055 and p=0.001). The consensus GTV delineations on CT were compared to the individual delineations. Since individual delineations on CT were performed for the first 14 patients, the mean sensitivity (0.84 ± 0.09) and the PPV (0.52 ± 0.16) of the delineations of these patients was compared to the mean sensitivity (0.85 ± 0.08) and PPV (0.49 ± 0.13) of the consensus delineation of these patients. The sensitivity of the consensus en individual delineations did not differ significantly (Wilcoxon signed rank test, sensitivity: p=0.20, PPV: p=0.055). To conclude, it seems that having delineations controlled by a colleague would only be beneficial in a situations when flaws in the delineation are made (i.e.: 'stupid mistakes').

Clinical impact

Which clinical advice could be given based on this thesis? For a clinician to decide which imaging modality to use is complex and multifactorial. Elements include the body part under evaluation, the histology of the cancer, the stage of disease, and the anticipated treatment. Proper modality selection aids the choice of operative approach, potential organ and functional preservation of affected or adjacent tissue, as well as helping to determine the role of surgical, radiation, and medical oncology in treatment. Imaging enhances clinical information beyond examination by assessing the submucosal extent of head-and-neck tumors and the presence of cervical nodal metastases (14).

In this thesis CT, MRI and PET were investigated, leaving other imaging modalities out of the scope.

In the University Medical Center Utrecht, we are fortunate to have various high quality imaging modalities available. The majority of the radiotherapy departments mostly use CT for imaging and delineation of cancer. To create an optimal combination of imaging modalities for specific situations, the special benefits of certain imaging modalities in visualizing tumorous and non-tumorous tissue should be investigated. PET has

shown to be superior for visualization of metastases, pathological lymph nodes, and possible tumor recurrence (61). Besides, an automatic delineation saves time and decreases workload. To benefit from the qualities of MRI one should be instructed how to interpret MRI with various sequences and delineate the GTV according to the guidelines developed in chapter 4. Further development of interpretation of MRI, for example distinction between cartilage invasion and cartilage penetration which can have clinical implications (54), would in my personal opinion lead to an increased benefit from MRI. A study performed by Huang et al. showed that MRI in combination with PET increases the sensitivity, specificity and provides superior positive predictive ratio and negative predictive ratio compared to PET-CT, MRI and CT imaging. The tumor diameter in this research showed greater resemblance to the pathological specimen compared to the other imaging modalities, which overestimated the tumor (63). Even the co-registration of FDG-PET CT and MRI (STIR and DWI) provides more accurate results for trans-compartemental extensions in T and N staging than for head-and-neck cancer according to a study performed by Stecco et al (65).

General limitations

CTV margins developed in this thesis were based on large tumors with cartilage invasion (stage T3/T4 tumors) and extension outside the laryngeal cartilage. Laryngeal and hypopharyngeal tumors primarily treated with radiation do not frequently have extensive cartilage involvement. Cartilage invasion was more often missed. The target margins were also designed to compensate for these misses. For lower stage tumors infiltration and tumor extent is less than the tumors used in this thesis and a smaller margin could be sufficient. The modality dependent margins designed in this thesis could lead to overestimation of lower stage tumors but even then, the margins are smaller than those used in current clinical practice.

A more subjective step was necessary to overcome remaining deformations that had not been resolved by optimizing the automatic rigid registration. It would have been ideal to get to an automatically optimized registration. The shaping of the tumor contour by visual information on pathology and MRI has been well-considered performed and has increased the accuracy of the registration.

For some patients the time between the various scans or time between surgery and imaging was relatively long. For two patients this time was longer than 20 days and for the remaining patients this was less than 20 days. These two patients were therefore excluded in the analyses in chapter 5 where modality dependent CTV margins were obtained. For an accurate comparison of imaging and the specimen, time between the imaging and surgery should be as short as possible. The imaging was accomplished not only for diagnostics and clinical purposes but also for research purposes. For logistical reasons it was not always possible to schedule the scans shortly before surgery and shorten the time between scans.

One of the limitations of this study was that imaging quality for some patients was not optimal. Movement of the patient due to obstruction and irritation during scanning of patients with large larynx and hypopharynx tumors is a factor that might influence the image quality. In some cases, a different imaging protocol was used to scan the patient, which resulted in suboptimal images. In recent years image quality has been improved and imaging protocols have been developed in order to increase the quality of the images. In an ideal setting, one could evaluate the image quality of a scan directly after scanning and decide to re-scan the patient directly or to postpone the scan. Unfortunately this is not current practice for various reasons such as: logistics, infusion of contrast and limited scan time, and patient burden/ethical reasons.

References

1. Bucci MK, Bevan A, Roach M. Advances in radiation therapy: conventional to 3D, to IMRT, to 4D, and beyond. CA. Cancer J. Clin. 55:117–34.
2. Grégoire V, Langendijk JA, Nuyts S. Advances in Radiotherapy for Head and Neck Cancer. J. Clin. Oncol. 2015;33:3277–84.
3. Hurkmans CW, Remeijer P, Lebesque J V, et al. Set-up verification using portal imaging; review of current clinical practice. Radiother. Oncol. 2001;58:105–20.
4. Lukens JN, Lin A, Hahn SM. Proton therapy for head and neck cancer. Curr. Opin. Oncol. 2015;27:165–71.
5. Perez CA, Mutic S. Advances and future of Radiation Oncology. Reports Pract. Oncol. Radiother. J. Gt. Cancer Cent. Pozn. Polish Soc. Radiat. Oncol. 2013;18:329–32.
6. Rasch C, Keus R, Pameijer FA, et al. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. Int. J. Radiat. Oncol. Biol. Phys. 1997;39:841–8.
7. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlenther. Onkol. 2003;179:21–30.
8. Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J. Med. Phys. 2008;33:136–40.
9. Anderson CM, Sun W, Buatti JM, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J. Radiat. Oncol. 2014;1:6.
10. Bird D, Scarsbrook AF, Sykes J, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. BMC Cancer. 2015;15:844.
11. Jager EA, Kasperts N, Caldas-Magalhaes J, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat. Oncol. 2015;10:26.
12. Geets X, Daisne J-F, Arcangeli S, 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.
13. Ahmed M, Schmidt M, Sohaib A, et al. The value of magnetic resonance imaging in target volume delineation of base of tongue tumours--a study using flexible surface coils. Radiother. Oncol. 2010;94:161–7.
14. Alberico RA, Husain SHS, Sirotkin I. Imaging in head and neck oncology. Surg. Oncol. Clin. N. Am. 2004;13:13–35.
15. Rasch CRN, Steenbakkers RJHM, Fitton I, et al. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. Radiat. Oncol. 2010;5:21.
16. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother. Oncol. 2015;117:83–90.
17. Brouwer CL, Steenbakkers RJHM, van den Heuvel E, et al. 3D Variation in delineation of head and neck organs at risk. Radiat. Oncol. 2012;7:32.
18. Hong TS, Tomé WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiother. Oncol. 2012;103:92–8.
19. Caldas-Magalhaes J, Kooij N, Ligtenberg H, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol. 2015;54:1181–7.
20. Daisne J-F, Duprez T, Weynand B, 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.

21. Caldas-Magalhaes J, Kasperts N, Kooij N, et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int. J. Radiat. Oncol. Biol. Phys. 2012;82:e289-98.

22. Chung N-N, Ting L-L, Hsu W-C, et al. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck. 2004;26:241–6.

23. Khoo VS, Dearnaley DP, Finnigan DJ, et al. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother. Oncol. 1997;42:1–15.

24. Sigal R, Zagdanski AM, Schwaab G, et al. CT and MR imaging of squamous cell carcinoma of the tongue and floor of the mouth. Radiographics. 1996;16:787–810.

25. Becker M, Zbären P, Casselman JW, et al. Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis at MR imaging. Radiology. 2008;249:551–9.

26. Anon. Definition of Volumes. Int. Comm. Radiat. Units Meas. 2010:41–53.

27. Moeckli R, Sozzi WJ, Mirimanoff R-O, et al. Physical considerations on discrepancies in target volume delineation. Z. Med. Phys. 2009;19:224–35.

28. Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, et al. MRI/linac integration. Radiother. Oncol. 2008;86:25–9.

29. Wooten HO, Green O, Yang M, et al. Quality of Intensity Modulated Radiation Therapy Treatment Plans Using a 60Co Magnetic Resonance Image Guidance Radiation Therapy System. Int. J. Radiat. Oncol. 2015;92:771–778.

30. 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.

31. Fotina I, Lütgendorf-Caucig C, Stock M, et al. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther. Onkol. 2012;188:160–7.

32. Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin. Oncol. (R. Coll. Radiol). 2010;22:515–25.

33. Grégoire V, Ang K, Budach W, et al. Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. Radiother. Oncol. 2014;110:172–81.

34. Rasch C, Steenbakkers R, van Herk M. Target definition in prostate, head, and neck. Semin. Radiat. Oncol. 2005;15:136–45.

35. Stroom JC, Heijmen BJM. Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report. Radiother. Oncol. 2002;64:75–83.

36. Weltens C, Kesteloot K, Vandevelde G, et al. Comparison of plastic and Orfit masks for patient head fixation during radiotherapy: precision and costs. Int. J. Radiat. Oncol. Biol. Phys. 1995;33:499–507.

37. Kanakavelu N, Jebaseelan Samuel J. Determination of patient set-up error and optimal treatment margin for intensity modulated radiotherapy using image guidance system. J. BUON. 21:505–11.

38. Cacicedo J, Perez JF, Ortiz de Zarate R, et al. A prospective analysis of inter- and intrafractional errors to calculate CTV to PTV margins in head and neck patients. Clin. Transl. Oncol. 2015;17:113–20.

39. Lennon J. Song: Borrowed time. Album Milk Honey. 1884.

40. Jeanneret-Sozzi W, Moeckli R, Valley J-F, et al. The reasons for discrepancies in target volume delineation : a SASRO study on head-and-neck and prostate cancers. Strahlenther. Onkol. 2006;182:450–7.

41. Caudell JJ, Meredith RF, Spencer SA, et al. Margin on gross tumor volume and risk of local recurrence in head-and-neck cancer. Int. J. Radiat. Oncol. Biol. Phys. 2010;76:164–8.

42. Hall EJ. Radiobiology for the radiologist. 6th ed. Philidelphia: Lippincott Williams & Wilkins; 2006.

43. Brenner DJ. Dose, volume, and tumor-control predictions in radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 1993;26:171–9.

44. Burnet NG, Thomas SJ, Burton KE, et al. Defining the tumour and target volumes for radiotherapy. Cancer Imaging. 2004;4:153–61.

45. Purdy JA. Current ICRU definitions of volumes: limitations and future directions. Semin. Radiat. Oncol. 2004;14:27–40.

46. Campbell S, Poon I, Markel D, et al. Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: implications for the management of head-and-neck cancers. Int. J. Radiat. Oncol. Biol. Phys. 2012;82:574–81.

47. Bethesda M. ICRU Report 50. Prescribing, Recording, and Reporting Photon Beam Therapy. Int. Comm. Radiat. Units Meas. 1993.

48. Sjöberg C, Lundmark M, Granberg C, et al. Clinical evaluation of multi-atlas based segmentation of lymph node regions in head and neck and prostate cancer patients. Radiat. Oncol. 2013;8:229.

49. Eisbruch A, Gregoire V. Balancing risk and reward in target delineation for highly conformal radiotherapy in head and neck cancer. Semin. Radiat. Oncol. 2009;19:43–52.

50. Becker M, Burkhardt K, Dulguerov P, et al. Imaging of the larynx and hypopharynx. Eur. J. Radiol. 2008;66:460–79.

51. Becker M, Zbären P, Laeng H, et al. Neoplastic invasion of the laryngeal cartilage: comparison of MR imaging and CT with histopathologic correlation. Radiology. 1995;194:661–9.

52. Kinshuck AJ, Goodyear PWA, Lancaster J, et al. Accuracy of magnetic resonance imaging in diagnosing thyroid cartilage and thyroid gland invasion by squamous cell carcinoma in laryngectomy patients. J. Laryngol. Otol. 2012;126:302–6.

53. Kazkayasi M, Onder T, Ozkaptan Y, et al. Comparison of preoperative computed tomographic findings with postoperative histopathological findings in laryngeal cancers. Eur. Arch. Otorhinolaryngol. 1995;252:325–31.

54. Kendi AT, Corey A, Magliocca KR, et al. Is there a role for PET/CT parameters to differentiate thyroid cartilage invasion from penetration? Eur. J. Radiol. 2016;85:319–23.

55. Chen SA, Muller S, Chen AY, et al. Patterns of extralaryngeal spread of laryngeal cancer: Thyroid cartilage penetration occurs in a minority of patients with extralaryngeal spread of laryngeal squamous cell cancers. Cancer. 2011;117:5047–5051.

56. Kats SS, Muller S, Aiken A, et al. Laryngeal tumor volume as a predictor for thyroid cartilage penetration. Head Neck. 2013;35:426–430.

57. Ryu IS, Lee JH, Roh J-L, et al. Clinical implication of computed tomography findings in patients with locally advanced squamous cell carcinoma of the larynx and hypopharynx. Eur. Arch. Otorhinolaryngol. 2014:2939–2945.

58. Aristophanous M, Penney BC, Martel MK, et al. A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. Med. Phys. 2007;34:4223–35.

59. Schwartz DL, Ford E, Rajendran J, et al. FDG-PET/CT imaging for preradiotherapy staging of head-and-neck squamous cell carcinoma. Int. J. Radiat. Oncol. Biol. Phys. 2005;61:129–36.

60. Park JT, Roh J-L, Kim JS, et al. (18)F FDG PET/CT versus CT/MR Imaging and the Prognostic Value of Contralateral Neck Metastases in Patients with Head and Neck Squamous Cell Carcinoma. Radiology. 2016;279:481–91.

61. Mak D, Corry J, Lau E, et al. Role of FDG-PET/CT in staging and follow-up of head and neck squamous cell carcinoma. Q. J. Nucl. Med. Mol. Imaging. 2011;55:487–99.

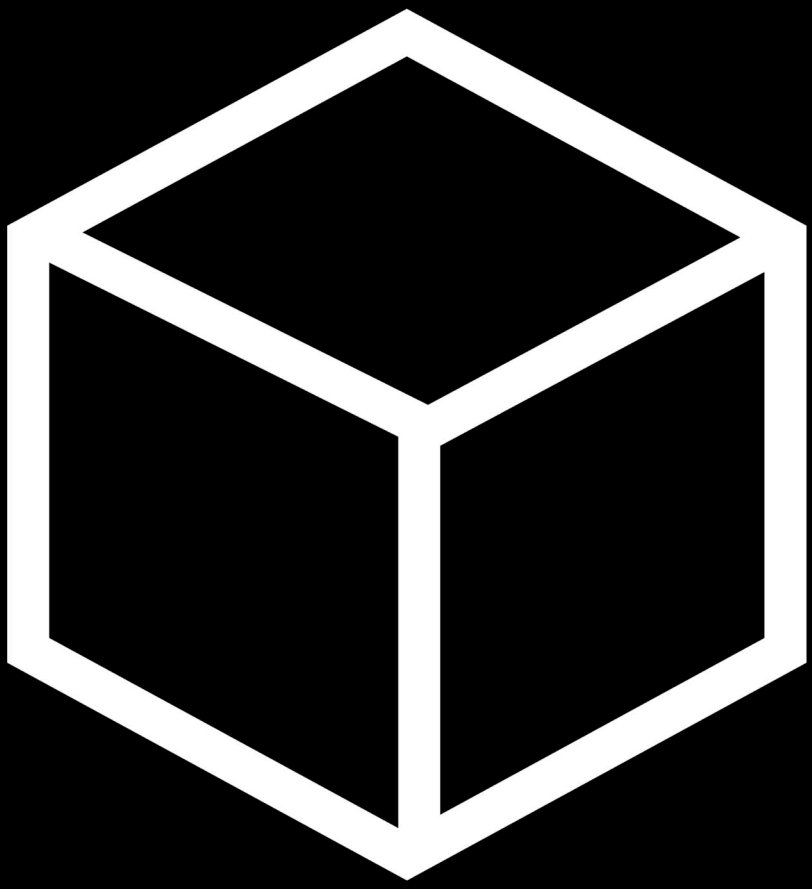
62. Spick C, Herrmann K, Czernin J. 18F-FDG PET/CT and PET/MRI Perform Equally Well in Cancer: Evidence from Studies on More Than 2,300 Patients. J. Nucl. Med. 2016;57:420–30.

63. Huang S-H, Chien C-Y, Lin W-C, et al. A comparative study of fused FDG PET/MRI, PET/CT, MRI, and CT imaging for assessing surrounding tissue invasion of advanced buccal squamous cell carcinoma. Clin. Nucl. Med. 2011;36:518–25.

64. Bashir U, Mallia A, Stirling J, et al. PET/MRI in Oncological Imaging: State of the Art. Diagnostics (Basel, Switzerland). 2015;5:333–57.

65. Stecco A, Ciolfi S, Buemi F, et al. Combined multimodal co-registration of PET/CT and MRI images increases diagnostic accuracy in squamous cell carcinoma staging. Radiol. Med. 2016;121:502–9.

66. Schakel T, Hoogduin JM, Terhaard CHJ, et al. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother. Oncol. 2013;109:394–7.



Hoofdstuk 7
Samenvatting

Het doel van radiotherapie is het vernietigen van maligne cellen. Hierbij is het van belang dat zoveel mogelijk omliggend gezond weefsel gespaard blijft. Wanneer gezond weefsel een grote hoeveelheid bestraling krijgt kan dit op korte en lange termijn tot ernstige complicaties leiden. In de afgelopen 30 jaar zijn er technieken binnen de radiotherapie ontwikkeld die de precisie van de bestraling vergroten. Andere uitdagingen bij het optimaliseren van de behandeling zijn onder andere het vergroten van de zichtbaarheid van de tumor (target) met behulp van beeldvorming en het daarmee verbonden intekenen (omlijnen) van het Gross Tumor Volume (GTV: Gross palpable or visible/demonstrable extent and location of malignant growth [1]) [2-4]. Voor het definiëren van de tumor zijn radiotherapeuten voornamelijk afhankelijk van de zichtbaarheid van de tumor op beelden gemaakt met een beeldvormende modaliteit. Deze zichtbaarheid wordt grotendeels bepaald door de eigenschappen van de toegepaste beeldvormende modaliteit. De gebruikte beeldvormende modaliteit is bepalend voor de intekening; de intekeningen van dezelfde tumor gedaan op beelden van verschillende modaliteiten verschillen van elkaar (intermodaliteit variatie) [5-11]. Ook tekent elke radiotherapeut de tumor net wat anders in. Hierdoor ontstaat er een verschil tussen de intekeningen gedaan op basis van dezelfde beelden van een bepaalde tumor (interobserver variatie). Het is dus niet altijd even makkelijk de tumorgrenzen te bepalen. Dit leidt tot een aanzienlijke interobserver variatie voor hoofd-hals tumoren.

Ook voor dit proefschrift (hoofdstuk 2) werd de interobserver variatie bepaald voor supraglottische larynxtumoren weergegeven op CT beelden, de standaard beeldvormende modaliteit binnen de radiotherapie [12]. Vervolgens werden MR beelden toegevoegd aan deze CT beelden. Dit omdat MRI een voordeel heeft ten opzichte van CT: het contrast tussen weke delen op MR beelden is hoger waardoor verschillende structuren beter van elkaar te onderscheiden zijn [13]. Op basis van de gecombineerde beelden werden dezelfde tumoren opnieuw ingetekend. Dit resulteerde in een toename van de interobserver variatie. Onze conclusie was dat de toegenomen zichtbaarheid van verschillende structuren het lastiger maakte de omvang en locatie van de tumor te bepalen. Hoe meer er te zien is, hoe meer afwegingen er gemaakt moeten worden met betrekking tot het in- of excluderen van bepaalde structuren in de intekening. Wanneer het niet duidelijk is hoe de beelden geïnterpreteerd dienen te worden en wanneer duidelijke intekenrichtlijnen afwezig zijn, zou dit kunnen leiden tot een toename van de interobserver variatie [7,14]. Er wordt een vaste marge om het GTV heen geplaatst om microscopische uitbreiding – die niet zichtbaar is op de beelden – te includeren in het doelgebied. Door middel van deze vaste marge wordt het Clinical Target Volume (CTV) bepaald. Dit volume krijgt de hoogste stralingsdosis. De CTV marge is bepaald op basis van historische gegevens en de klinische ervaring van radiotherapeuten. De CTV marge is niet nauwkeurig bepaald: de marge verschilt aanzienlijk tussen instituten en radiotherapeuten [2,3,15,16]. Als de GTV intekeningen en de daaruit voortkomende CTVs in grote mate verschillen tussen radiotherapeuten, is de precisie van de targetbepaling laag. Het is echter belangrijker om de nauwkeurigheid van de intekenin-

gen te bepalen. Hierbij wordt gekeken naar hoe correct de daadwerkelijke tumor wordt ingetekend. Om de nauwkeurigheid te kunnen bepalen zijn validatiestudies van essentieel belang, al zijn deze uiterst schaars voor hoofd-hals kanker en moeilijk uit te voeren [17,18]. Bij validatiestudies worden intekeningen (gedaan op beelden) vergeleken met de afgrenzing van de tumor op histologische coupes. De tumor op histologische coupes wordt beschouwd als weergave van de daadwerkelijke tumor. Bij het bepalen van nauwkeurigheid is niet alleen het vaststellen van gemist tumorweefsel van belang, er wordt ook gekeken naar de overschatting, het onterecht aanmerken van gezond weefsel als tumor.

Histopathologie wordt gezien als gouden standaard voor validatiestudies. Maar hoe precies is deze 'gouden standaard' eigenlijk? Uit dit proefschrift (hoofdstuk 4) blijkt dat er weinig verschil is tussen de intekeningen van tumorweefsel op histologische coupes. De intekeningen van 3 pathologen werden, voor elke tumor afzonderlijk, over elkaar heen geprojecteerd om ze te kunnen vergelijken. Vervolgens werden de afstanden tussen deze intekeningen bepaald. Hieruit bleek dat 95% van de afstanden minder was dan 2.0 mm [19]. Om de nauwkeurigheid van de GTV intekeningen te bepalen werd een validatiestudie uitgevoerd. Van 27 patiënten met T3/T4 larynx of hypofarynxtumoren werd, alvorens een totale laryngectomie plaatsvond, beeldvorming (FDG-PET, MRI en CT) vervaardigd. Om de beelden goed te kunnen vergelijken met het larynxpreparaat werd een methode gebruikt die beschreven is door Caldas-Magalhaes et al [20]. Kort gezegd werden de larynxpreparaten verwerkt tot H&E-coupes (in totaal ongeveer 800 stuks), waarop de tumor werd ingetekend door een patholoog. De H&E-coupes werden geregistreerd aan de beelden die voor de operatie gemaakt waren. De registratie werd op verschillende manieren geoptimaliseerd. Met de gebruikte methode was het mogelijk de 3D gereconstrueerde tumoren te vergelijken met de gemaakte beelden.

Vervolgens tekenden drie radiotherapeuten het GTV in op MR beelden (hoofdstuk 4). Dit gebeurde in twee sessies. Tijdens de eerste sessie werd het GTV ingetekend zoals radiotherapeuten dit doen in de klinische praktijk. In de tweede sessie gebruikten de radiotherapeuten duidelijke, nieuw ontwikkelde intekenrichtlijnen gebaseerd op informatie over de diagnose van kraakbeeninvasie op MRI [21]. Van beide intekensessies werd de nauwkeurigheid van deze GTV intekeningen bepaald. De conclusie was dat het GTV ingetekend volgens de klinische praktijk (de eerste sessie) tweemaal zo groot was als het histologisch bepaalde tumorvolume. Met het gebruik van de nieuwe intekenrichtlijnen (tweede sessie) nam deze overschatting significant af.

In een andere studie (hoofdstuk 5) werden meerdere frequent gebruikte beeldvormende modaliteiten (FDG-PET, CT en MRI), waarop het GTV was ingetekend, vergeleken met histopathologie. Vervolgens werden CTV marges berekend. Aangezien de zichtbaarheid van de tumor per beeldmodaliteit kan verschillen, is de marge die nodig is om alle niet zichtbare tumorcellen te omsluiten verschillend. Voor elke beeldvormende modaliteit werd een CTV marge bepaald aan de hand van een afstandsanalyse. Deze analyse werd gedaan door de H&E-contour en de GTV intekening over elkaar heen te leggen. Hieruit bleek dat de H&E-contour niet altijd omsloten werd door de GTV intekening. Vervolgens werden de afstanden vanaf de GTV contour waarbuiten nog histopathologische tumor aanwezig was, bepaald. Deze afstanden werden gebruikt om de CTV marge die nodig is om wel al het tumorweefsel te omsluiten, te bepalen. De CTV marges compenseren dus de onderschatting van de tumor door de radiotherapeut. De CTV marges werden berekend (PET:5,5mm, CT:4,4mm, MRI:6,7mm) en vervolgens werden deze CTV marges toegevoegd aan de GTVs. Vervolgens werden de CTV volumes bepaald. De mediane CTV op FDG-PET (39 ml) was kleiner dan de CTV op MRI (48 ml) en CT (44 ml). Aangezien de CTV

marges alle tumorcellen omsloten betekende dit dat het FGD-PET CTV het minst gezonde weefsel bevatte.

De nieuwe CTV marges zorgden voor een volumereductie van 45%-52% ten opzichte van het CTV volume gebaseerd op de huidige klinische CTV marge van 10 mm. Dit betekent dat met de ontwikkelde modaliteitsspecifieke CTV marges veel minder gezond weefsel met de hoogste dosis zou worden bestraald dan met de huidige marge. Een kleiner bestralingsvolume draagt bij aan het verminderen van de toxiciteit.

Dit proefschrift is uniek omdat er een robuuste validatiemethode gebruikt is om de nauwkeurigheid te bepalen in plaats van de precisie. Je hebt niets aan precisie als je niet nauwkeurig bent. Kortom: "Je kunt consequent de verkeerde target raken (grote precisie maar lage nauwkeurigheid), maar belangrijker is het consequent raken van de juiste target (grote precisie en hoge nauwkeurigheid)"[4].

Voetnoten

1 Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT): Contents. J ICRU 2010;10:NP.

2 Jeanneret-Sozzi W, Moeckli R, Valley JF et al. The reasons for discrepancies in target volume delineation : a SASRO study on head-and-neck and prostate cancers. Strahlenther Onkol 2006;182:450-457.

3 Moeckli R, Sozzi WJ, Mirimanoff RO et al. Physical considerations on discrepancies in target volume delineation. Z Med Phys 2009;19:224-235.

4 Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys 2008;33:136-140.

5 Anderson CM, Sun W, Buatti JM et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J Radiat Oncol 2014;1:006.

6 Bird D, Scarsbrook AF, Sykes J et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. BMC Cancer 2015;15:844-015-1867-8.

7 Brouwer CL, Steenbakkers RJ, van den Heuvel E et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32-717X-7-32.

8 Chung NN, Ting LL, Hsu WC, Lui LT, Wang PM. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck 2004;26:241-246.

9 Daisne JF, Sibomana M, Bol A, Cosnard G, Lonneux M, Gregoire V. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 2003;69:237-245.

10 Geets X, Daisne JF, Arcangeli S 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.

11 Rasch C, Keus R, Pameijer FA et al. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. Int J Radiat Oncol Biol Phys 1997;39:841-848.

12 Jager E, Kasperts N, Caldas-Magalhaes J et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat Oncol 2015;10:26.

13 Metcalfe P, Liney GP, Holloway L et al. The potential for an enhanced role for MRI in radiation-therapy treatment planning. Technol Cancer Res Treat 2013;12:429-446.

14 Brouwer CL, Steenbakkers RJ, Bourhis J et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015;117:83-90.

15 Burnet NG, Thomas SJ, Burton KE, Jefferies SJ. Defining the tumour and target volumes for radiotherapy. Cancer Imaging 2004;4:153-161.

16 Fotina I, Lutgendorf-Caucig C, Stock M, Potter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 2012;188:160-167.

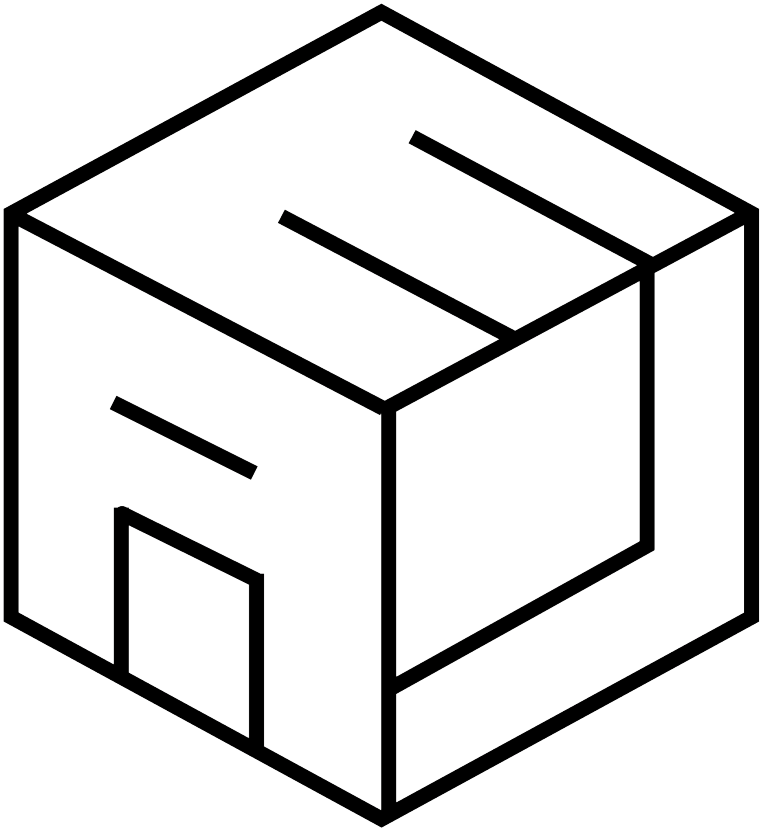
17 Caldas-Magalhaes J, Kooij N, Ligtenberg H et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol 2015:1-7.

18 Daisne JF, Duprez T, Weynand B 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.

19 Jager EA, Willems SM, Schakel T et al. Interobserver variation among pathologists for delineation of tumor on H&E-sections of laryngeal and hypopharyngeal carcinoma. How good is the gold standard? Acta Oncol 2015:1-5.

20 Caldas-Magalhaes J, Kasperts N, Kooij N et al. Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 2012;82:e289-98.

21 Becker M, Zbaren P, Casselman JW, Kohler R, Dulguerov P, Becker CD. Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis at MR imaging. Radiology 2008;249:551-559.



Curriculum Vitae
Elise Anne Jager...

....was born on 22 April 1984 in Groningen, the Netherlands. After finishing the VWO she moved to the University of Leiden to study medicine in 2002. Combining study, extracurricular activities such as sports, student life and medical jobs for a couple of years, she moved to The Hague during her internships. Extracurricular research and activities were mostly directed to surgical specialties. After graduating from medical school in 2011 she decided to develop in Radiation Oncology and started her PhD at the University Medical Center Utrecht at the department of Radiation Oncology. She moved to Amsterdam where she currently lives. During the past 4,5 years she performed research in validation of imaging for GTV delineation in laryngeal and hypopharyngeal cancer. In April 2016 she started her training as a radiation oncologist at the VU University Medical Center.