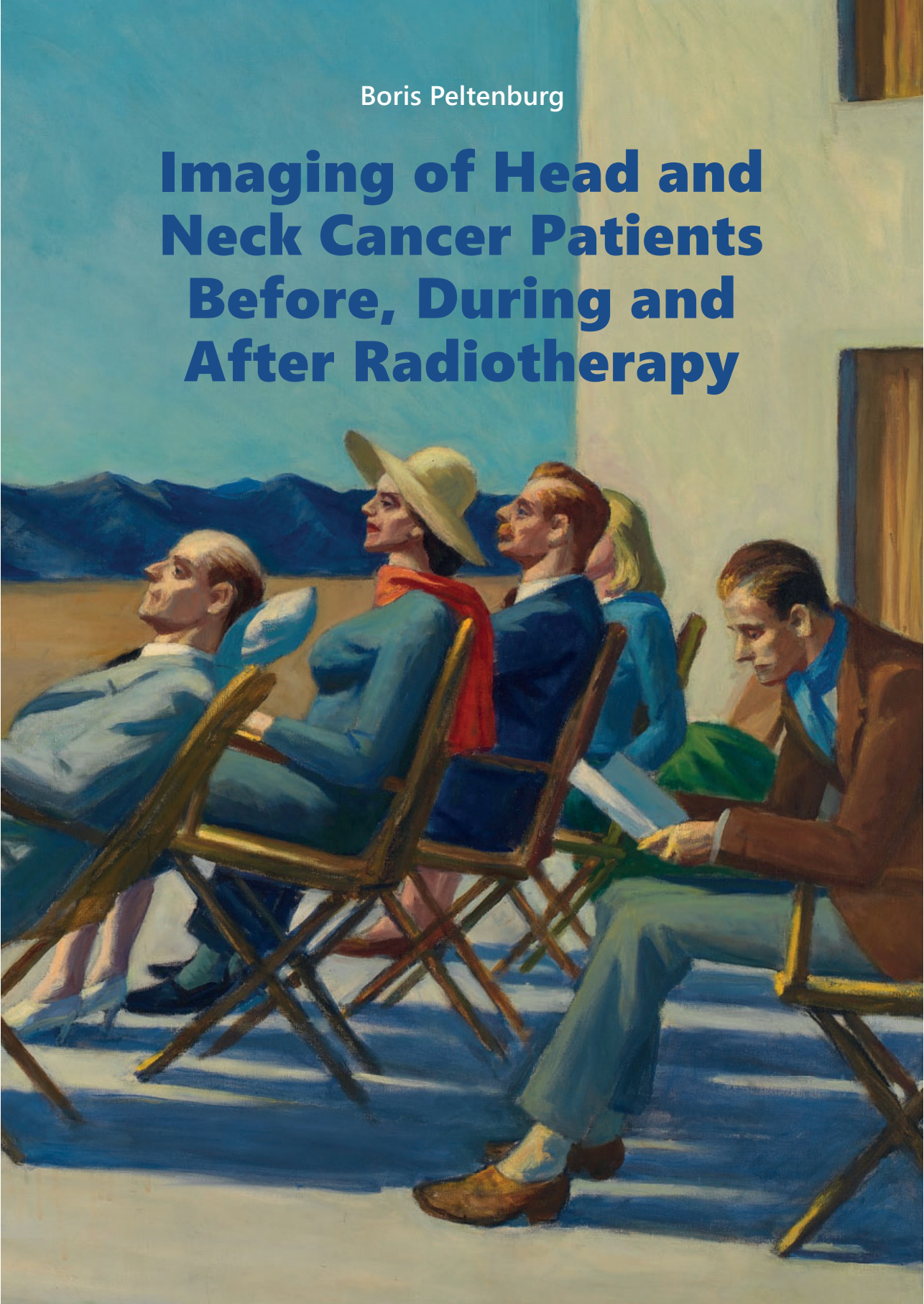Boris Peltenburg

Imaging of Head and Neck Cancer Patients Before, During and After Radiotherapy

Imaging of Head and Neck Cancer Patients Before, During and After Radiotherapy

Boris Peltenburg

Imaging of Head and Neck Cancer Patients Before, During and After Radiotherapy

PhD thesis, Utrecht University, the Netherlands

Imaging of Head and Neck Cancer Patients Before, During and After Radiotherapy

Medische beeldvorming bij patiënten met hoofd-halskanker voor, tijdens en na radiotherapie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 4 februari 2020 des middags te 4.15 uur

door

Boris Peltenburg

geboren op 13 december 1988
te Utrecht

Promotoren
Prof. dr. R. de Bree
Prof. dr. C.H.J. Terhaard

Copromotor
Dr. ir. M.E.P. Philippens

Contents

"Words can be like X-rays, if you use them properly--they'll go through anything. You read and you're pierced."

Helmholtz Watson - Brave New World

Written by Aldous Huxley who died of laryngeal cancer in 1963

Chapter 1

Introduction

The inside of the human mouth, throat and airway duct is lined with a protective network of epithelial cells. The outermost layer of this network resembles fish scales, we therefore call the cells of this epithelium 'squamous cells' after the Latin for scale, *squama*. Our body is constantly replacing these cells as abrasion from food and outside air causes cells to be lost. This is a natural process. However, sometimes errors occur during replication that lead to mutated cells. These mutated cells may start to divide uncontrollably and invade surrounding tissues. This is the start of a head and neck squamous cell carcinoma (HNSCC) [1].

Worldwide an estimated 900,000 new cases of HNSCC developed in 2018. This is approximately 5% of all new cancer cases [2]. Symptoms of head and neck cancer can vary greatly depending on the primary site. Patients might notice a growing mass in the mouth or neck, or they might experience dysphagia or otalgia. If the tumor grows on the vocal cords, hoarseness may be one of the first noticeable symptoms [3]. Incidence of HNSCC increases with age, mostly not occurring before the age of 40-50 [4]. Men are more likely to be diagnosed with HNSCC compared to women. However, this difference is diminishing as smoking, one of the major risk factors for HNSCC, is now almost equally common in men and women [5]. Next to tobacco, alcohol use is an important risk factor for HNSCC, with amplification of the harmful effect if alcohol and tobacco are used concomitantly. In recent years, the human papilloma virus (HPV) has been identified as another risk factor for oropharyngeal carcinomas. It appears that so-called HPV positive oropharyngeal tumors are a different entity than HPV negative tumors. For example, patients with HPV positive tumors are often younger and in many cases have no history of excessive tobacco or alcohol use [6]. Furthermore, HPV positive tumors seem to respond better to treatment than HPV negative tumors [7,8]. The prevalence of HPV differs around the world although the number of HPV positive oropharyngeal cancer cases has steadily risen in recent decades [9]. A possible reason cited for this increase are changes in sexual patterns, mainly the increased prevalence of oral sex and increasing numbers of sexual partners [10].

Anatomy and staging

The area of the head and neck where HNSCC most often occurs includes the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx. Less frequently HNSCC is localized in the paranasal sinuses and salivary glands. The structure of head and neck area is very heterogeneous: the mouth, nose and pharynx together with the paranasal sinus are filled with air and surrounded by layers of muscle, bone and cartilage. The area is furthermore responsible for a great variety of functions. Breathing, swallowing, speaking,

taste and smell are all heavily reliant on the proper functioning of head and neck organs. In addition, several cranial nerves run through the area. Damage to these nerves can cause loss of functionality leading for example to numbness or muscle paralysis.

HNSCC tumors are staged according to the 8th edition AJCC [11]. The size or extend of the primary tumor, involvement of lymph nodes and distant metastases are recorded using T, N and M stage respectively. Numbers are used for each field to indicate the severity of disease with higher numbers representing more extensive disease. This staging is important as patients with higher stage tumors will likely experience progressively worse outcomes. With similar treatment, patients with higher stages of disease are more likely to experience a recurrence and have a lower overall and disease-free survival. This is why the TNM stage of a tumor is considered to be a prognostic factor for disease outcome for individual patients.

Treatment

Treatment of HNSCC consists of surgery, radiotherapy, systemic treatment or a combination of these different treatment modalities. Based on the size and position of the tumor the most fitting treatment is selected. For higher stages of HNSCC a multi-modality approach to treatment is often necessary. In these cases, the toxicity of the treatment plays an important role in the final decision. Surgical removal of a tumor or high radiation dose can impact the head and neck organs surrounding the tumor, with transient or permanent loss of function as result. Evidently the preference of the patient is of major importance when selecting the definitive treatment as some patients accept a lower chance of curation if this means that functions can be preserved.

Radiotherapy with or without concomitant chemotherapy – with salvage surgery in reserve for residual or recurrent disease – is increasingly being used as it often has comparable curative rates with greater retention of functionality compared to surgery. In some cases the (chemo)radiotherapy ((C)RT) treatment fails and salvage surgery becomes the only option for a curative treatment. Due to tissue changes induced by the earlier (C)RT such as inflammation or fibrosis, salvage surgery is more difficult and less successful than primary surgery [12].

Radiotherapy uses ionizing radiation to destroy malignant cells. The radiation dose is delivered by using radiation fields, often these fields also contain healthy tissue. Generally, by increasing the dose the probability of malignant cells dying increases as well. In order to spare the healthy tissue the dose to these fields has to be limited. This creates a balancing act between maximizing the dose in order to destroy the tumor,

while keeping the dose to the healthy tissues and organs at risk (OARs) to acceptable levels [13]. Radiotherapy for HNSCC is furthermore divided into several fractions spread out over typically five to seven weeks. Between each fraction the healthy tissue is given time to recuperate from the radiation dose while the malignant cells, being less able to repair themselves, do not experience a recovering effect.

In order to increase the effectiveness of (C)RT many improvements have been made to the treatment over the years. Initially the radiation dose used to be delivered to the patient in one or two large rectangular fields. Eventually, improvements in radiotherapy machines (linear accelerators) allowed for more precise targeting of tumors. It is now possible to create a dose field that is specifically molded to the shape of the tumor of an individual patient [14]. A steep dose fall-off at the edges of the tumor spares the surrounding tissue while still delivering maximum dose to the tumor. This concept places a new focus on accurate definition of the tumor borders.

Target volume definition

In order to obtain information about the location, shape and size of a tumor, head and neck surgical oncologists, radiologists, nuclear medicine physicians, pathologists and radiation oncologists combine their strengths. Results from physical examination and medical imaging are discussed multidisciplinary to define the size and extent of a tumor. Furthermore, the number and distribution of metastatic lymph nodes in the neck is recorded.

When radiotherapy is chosen as the preferred treatment, the radiation oncologist is responsible for defining the exact borders between tumor and healthy tissue. The Gross Tumor Volume (GTV) is defined as the area demonstrably containing tumor cells. It is often not easy to discriminate between the borders of a tumor and surrounding healthy tissue. The physician has to make an interpretation using experience, training, guidelines and the available clinical information. This subjectivity can lead to interobserver variation in the delineation of tumors [15–17].

The GTV forms the basis for the Clinical Target Volume (CTV). The CTV is several millimeters larger than the GTV, accounting for the fact that microscopic extension of malignant cells may exist outside of the visible tumor. In turn, the CTV is used to determine the Planned Target Volume (PTV). Again slightly larger than the CTV, the PTV is the volume that will eventually receive the highest dose. It is constructed in order to minimize the effects of patient set-up inaccuracies or intrafraction motion. Typically the radius of the PTV is

several millimeters to centimeters larger than the GTV. The total volume of the PTV is therefore significantly larger than the volume of the GTV. This means that the PTV has a higher probability of containing all of the malignant cells. However, it also has a higher probability of containing healthy tissue.

During the course of treatment some tumors display a rapid decline in size [18,19]. This is experienced by the patients as a reduction of swelling and a return of function such as improved swallowing or less restricted mouth opening. Up until now, radiotherapy treatment for these patients often remains unchanged even though tumor volume has visibly been decreased. Adapting plans once or multiple times during the treatment could theoretically reduce the areas of healthy tissue receiving a high dose [20].

Imaging

Traditionally, computed tomography (CT) has been the preferred imaging modality for planning the radiotherapy treatment. Aside from being widely available, CT provides information on the electron density of the treatment area, which is essential for creating radiotherapy treatment plans. In spite of this, other imaging modalities might be better suited for target volume definition. Magnetic resonance imaging (MRI) uses strong magnetic fields and radio waves to generate images with a higher soft tissue contrast than CT [21,22]. Moreover, MRI can be used to generate images that provide information not just on the anatomy of the region but on the functionality as well [23]. An example of this is diffusion weighted MRI (DW-MRI). Images constructed via this technique provide information on the ability of water to move through the tissues of the body. Normally, water molecules can diffuse freely through a process called Brownian motion. However, in tissues movement is restricted when water reaches the cell membranes or large molecules. As tumors are often hypercellular, diffusion is more restricted in tumors than in surrounding tissue. This creates a high contrast between tumor and surrounding tissue on diffusion weighted images, aiding the target volume definition process. DW-MRI has an additional benefit; it is possible to quantify the diffusion restriction in a certain area by calculating the apparent diffusion coefficient (ADC). By adjusting the magnetic gradients, the sensitivity of diffusion weighted imaging for motion restriction can be increased. A new image can be obtained, this time with a higher so-called b-value. The higher b-value image experiences a decay of the signal in all tissues but this decay is lower in tissues with diffusion restriction (i.e. tumors). The signal decay is exponential and defined by the ADC. The ADC can be calculated as follows:

$S_b = S_0 e^{(-bADC)}$

S_b is the signal intensity at diffusion gradient b and S_0 is the signal intensity without diffusion weighting [24].

Quantification of tissue characteristics is important as it might lead to objective delineation of tumors. This could reduce subjectivity and interobserver variability during the delineation process.

Another method for quantifying tissue functionality can be obtained by performing positron-emission tomography (PET) [25–27]. PET uses a radioactive tracer designed to accumulate in specific tissues. The tracer emits positrons which annihilate after collision with an electron and produce two opposing photons, which can be detected by the scanner. The tracer most often used in HNSCC is fluorodeoxyglucose (FDG) which is labelled with radioactive fluorine-18. In the human body FDG behaves like glucose, but is only bound in the tissue. It accumulates in HNSCC tumors, metastatic lymph nodes and distant metastases due to their high glucose consumption.

Posttreatment period

Imaging remains important after the primary treatment is finished. Short-term radiotherapy effects such as mucositis disappear within 3 months after treatment. An MRI can be performed at this time point in order to evaluate treatment response. However, assessing imaging after radiotherapy treatment is often much more difficult than assessing imaging made prior to treatment. The radiation has caused damage to the area, changing normal anatomy and creating areas with edema or abnormal contrast and tracer uptake [28,29]. At this 3 month time point, it may be very difficult to differentiate between postradiation changes and recurrent or residual disease [30]. For some patients the radiotherapy treatment is not successful. Depending on the primary stage 10-50% of HNSCC patients will experience a recurrence of disease [31,32]. A local recurrence is clinically defined as the growth of malignant cells at the site of the primary tumor within three years after an initial treatment that appeared successful. Similar to the response assessment 3 months after treatment, detection of recurrent HNSCC can be extremely difficult due to posttreatment effects. Inflammation and the formation of fibrosis can obscure recurring tumors or in other cases be mistaken for recurrent disease.

Thesis outline

Patients with HNSCC can greatly benefit from medical imaging. Especially for patients receiving radiotherapy, imaging plays a major part in the treatment process. Imaging allows for adequate diagnosis and staging of tumors, it facilitates targeted radiation treatment and it is important for treatment response evaluation and recurrence detection. Furthermore, it may be possible to accurately predict treatment outcome using imaging. This thesis aims at evaluating and improving the application of medical imaging for HNSCC patients treated with radiotherapy.

In **part 1**, the role of imaging prior to the start of treatment is evaluated. **Chapter 2** discusses the use of FGD PET-CT for the detection of lymph node metastases. This is compared to the arguably more invasive and less patient-friendly ultrasound fine needle aspiration cytology. **Chapter 3** continues with a comparison of FDG PET-CT, this time to DW-MRI. Here the delineations made using DW-MRI of three observers are compared to each other and to automatically segmented PET-CT images in order to determine interobserver agreement. Finally, **chapter 4** examines if ADC values obtained by DW-MRI prior to treatment are a prognostic factor that can be used to optimize treatment selection.

Part 2 of this thesis evaluates imaging during the radiotherapy treatment. **Chapter 5** contains the protocol for a prospective study (PREDICT) with the ultimate aim to determine if a change in ADC attributed to radiotherapy treatment can be an early predictor of treatment outcome. **Chapter 6** uses data obtained in the PREDICT study to evaluate the usefulness of plan adaptation during treatment.

Part 3 handles the imaging patients receive after their radiotherapy treatment has ended. **Chapter 7** and **chapter 8** have DW-MRI for recurrence detection as a common theme. First, the effectiveness of DW-MRI for recurrence is compared to PET-CT. Secondly, the interobserver agreement in quantitative analysis of the DW-MRI images is analyzed.

Finally, **chapter 9** focuses on the optimal use of MRI in the radiotherapy setting. It contains a small recap of parts of this thesis and provides information on how to use MRI as a monitoring device during the entire treatment period of a patient. It also presents the current application of MRI before, after and during radiotherapy.

References

1. Young B, Lowe JS, Stevens A, Heath JW. Wheater's Functional Histology. Fifth Edit. Elsevier Ltd; 2006. 82-87 p.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov;68(6):394–424.

3. Lambert R, Sauvaget C, de Camargo Cancela M, Sankaranarayanan R. Epidemiology of cancer from the oral cavity and oropharynx. Eur J Gastroenterol Hepatol. 2011 Aug;23(8):633–41.

4. Döbrossy L. Epidemiology of head and neck cancer: magnitude of the problem. Cancer Metastasis Rev. 2005 Jan;24(1):9–17.

5. Centers for Disease Control and Prevention (CDC). State-specific prevalence of cigarette smoking among adults and quitting among persons aged 18-35 years--United States, 2006. MMWR Morb Mortal Wkly Rep. 2007 Sep 28;56(38):993–6.

6. Vokes EE, Agrawal N, Seiwert TY. HPV-Associated Head and Neck Cancer. J Natl Cancer Inst. 2015 Dec;107(12):djv344.

7. Sturgis EM, Ang KK. The epidemic of HPV-associated oropharyngeal cancer is here: is it time to change our treatment paradigms? J Natl Compr Canc Netw. 2011 Jun 1;9(6):665–73.

8. O'Sullivan B, Huang SH, Su J, Garden AS, Sturgis EM, Dahlstrom K, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. Lancet Oncol. 2016;17(4):440–51.

9. Van Dyne EA, Henley SJ, Saraiya M, Thomas CC, Markowitz LE, Benard VB. Trends in Human Papillomavirus-Associated Cancers - United States, 1999-2015. MMWR Morb Mortal Wkly Rep. 2018 Aug 24;67(33):918–24.

10. Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst. 2008 Mar 19;100(6):407–20.

11. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin. 2017 Mar;67(2):93–9.

12. Wong LY, Wei WI, Lam LK, Yuen APW. Salvage of recurrent head and neck squamous cell carcinoma after primary curative surgery. Head Neck. 2003;25(11):953–9.

13. Grégoire V, Langendijk JA, Nuyts S. Advances in Radiotherapy for Head and Neck Cancer. J Clin Oncol. 2015 Oct 10;33(29):3277–84.

14. Grégoire V, Maingon P. Intensity-modulated radiation therapy in head and neck squamous cell carcinoma: an adaptation of 2-dimensional concepts or a reconsideration of current clinical practice. Semin Radiat Oncol. 2004 Apr;14(2):110–20.

15. Jager EA, Ligtenberg H, Caldas-Magalhaes J, Schakel T, Philippens ME, Pameijer FA, et al. Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer. Acta Oncol. 2016 Nov;55(11):1305–12.

16. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. Int J Radiat Oncol Biol Phys. 2007 Jul 1;68(3):763–70.

17. Cardoso M, Min M, Jameson M, Tang S, Rumley C, Fowler A, et al. Evaluating diffusion-weighted magnetic resonance imaging for target volume delineation in head and neck radiotherapy. J Med Imaging Radiat Oncol. 2019 Jun;63(3):399–407.

18. Paudyal R, Oh JH, Riaz N, Venigalla P, Li J, Hatzoglou V, et al. Intravoxel incoherent motion diffusion-weighted MRI during chemoradiation therapy to characterize and monitor treatment response in human papillomavirus head and neck squamous cell carcinoma. J Magn Reson Imaging. 2017 Apr;45(4):1013–23.

19. Barker JL, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int J Radiat Oncol Biol Phys. 2004 Jul 15;59(4):960–70.

20. Castelli J, Simon A, Lafond C, Perichon N, Rigaud B, Chajon E, et al. Adaptive radiotherapy for head and neck cancer. Acta Oncol. 2018 Oct;57(10):1284–92.

21. 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(2):630–6.

22. Gage KL, Thomas K, Jeong D, Stallworth DG, Arrington JA. Multimodal Imaging of Head and Neck Squamous Cell Carcinoma. Cancer Control. 2017 Apr;24(2):172–9.

23. Nooij RP, Hof JJ, van Laar PJ, van der Hoorn A. Functional MRI for Treatment Evaluation in Patients with Head and Neck Squamous Cell Carcinoma: A Review of the Literature from a Radiologist Perspective. Curr Radiol Rep. 2018;6(1):2.

24. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology. 1986;161(2):401–7.

25. Evangelista L, Cervino AR, Chondrogiannis S, Marzola MC, Maffione AM, Colletti PM, et al. Comparison between anatomical cross-sectional imaging and 18F-FDG PET/CT in the staging, restaging, treatment response, and long-term surveillance of squamous cell head and neck cancer: a systematic literature overview. Nucl Med Commun. 2014 Feb;35(2):123–34.

26. Xu G, Li J, Zuo X, Li C. Comparison of whole body positron emission tomography (PET)/PET-computed tomography and conventional anatomic imaging for detecting distant malignancies in patients with head and neck cancer: a meta-analysis. Laryngoscope. 2012 Sep;122(9):1974–8.

27. Newbold K, Powell C. PET/CT in Radiotherapy Planning for Head and Neck Cancer. Front Oncol. 2012;2.

28. Weissman JL, Akindele R. Current imaging techniques for head and neck tumors. Oncol. 1999;13(5):697–709; discussion 713.

29. Yao M, Smith RB, Graham MM, Hoffman HT, Tan H, Funk GF, et al. The role of FDG PET in management of neck metastasis from head-and-neck cancer after definitive radiation treatment. Int J Radiat Oncol Biol Phys. 2005 Nov;63(4):991–9.

30. Wang SJ. Surveillance radiologic imaging after treatment of oropharyngeal cancer: a review. World J Surg Oncol. 2015;13:94.

31. Forastiere AA, Goepfert H, Maor M, Pajak TF, Weber R, Morrison W, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. N Engl J Med. 2003 Nov 27;349(22):2091–8.

32. Spector GJ, Sessions DG, Lenox J, Newland D, Simpson J, Haughey BH. Management of stage IV glottic carcinoma: therapeutic outcomes. Laryngoscope. 2004 Aug;114(8):1438–46.

Part 1

Imaging prior to the start of therapy

Chapter 2

Prediction of ultrasound guided fine needle aspiration cytology results by FDG PET-CT for lymph node metastases in head and neck squamous cell carcinoma patients

Boris Peltenburg, Bart de Keizer, Jan Willem Dankbaar, Mirthe de Boer, Stefan M. Willems, Marielle E. P. Philippens, Chris H. J. Terhaard & Remco de Bree

Abstract

Introduction

Accurate assessment of cervical lymph node status is essential in patients with head and neck squamous cell carcinoma (HNSCC) as it influences prognosis and treatment decisions. During patient work-up lymph node status is often examined by ultrasound guided fine needle cytology (USgFNAC). ^{18}F-Fluorodeoxyglucose PET-CT (FDG PET-CT) is frequently used to assess primary tumor and distant metastases but provides information on lymph node status as well. It is possible that FDG PET-CT (if already made for abovementioned indications) can predict the results of USgFNAC in subgroups of lymph nodes based on FDG-uptake and size. The objective of this study to identify maximum standardized uptake (SUVmax) and short axis diameter cutoff values of lymph nodes at which FDG PET-CT can reliably predict USgFNAC results.

Methods

117 patients with HNSCC were retrospectively analyzed. Patients were included when a FDG PET-CT and USgFNAC were available. SUVmax measurements were performed and compared to the USgFNAC results.

Results

Using USgFNAC as a reference standard, the area under the curve of the receiver operating curve was 0.91. At a SUVmax cutoff value of 4.9 the accuracy of FDG PET-CT was the highest (85%). Lymph nodes with short axis diameter ≥1.0 cm and SUVmax ≥4.9 were in 91% positive on USgFNAC. If SUVmax was below 2.2 no nodes were positive on USgFNAC. Of all lymph nodes 52% either had a short axis diameter ≥1.0 cm and SUVmax ≥4.9 or a SUVmax <2.2. FDG PET-CT and USgFNAC results were very similar in these nodes.

Conclusion

By measuring SUVmax values and minimal axial diameters of lymph nodes and using appropriate cutoff values, FDG PET-CT can predict the results of USgFNAC examinations in half of the examined lymph nodes. This information may lead to a reduction of USgFNAC examinations in HNSCC patients if FDG PET-CT is already performed for other indications.

Introduction

The presence of cervical lymph node metastases is an important prognostic factor in patients with head and neck squamous cell carcinoma (HNSCC). It is associated with reduced survival and affects treatment decision-making [1–4]. Accurate determination of the lymph node status is therefore of great importance in order to reduce over- and under-treatment.

Ultrasound guided fine needle aspiration cytology (USgFNAC) is frequently used to identify or confirm tumor-positive lymph nodes. Other objectives during diagnostic workup of HNSCC patients are identification of occult distant metastases and preparation for radiation treatment planning if indicated. For this, ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography (FDG PET-CT) is often used [5–7]. However, FDG PET-CT provides information on lymph node status as well and might be able to differentiate benign from malignant nodes based on quantitative measurements [8,9].

We hypothesize that FDG PET-CT criteria can be developed to accurately identify lymph nodes which are positive or negative on USgFNAC. These criteria may reduce the need to perform USgFNAC for the detection of lymph node metastases in subgroups of patients who already underwent FDG PET-CT for (whole body) staging or in preparation for primary (chemo)radiotherapy. Reducing the number of USgFNAC examinations could result in shortened diagnostic work-up time, reduced costs and decreased patient discomfort.

Methods

The database of the radiology department of a tertiary care center was retrospectively screened for HNSCC patients who had received USgFNAC within 1 month before or after a FDG PET-CT from January 2013 until August 2015. Only patients with a diagnosis of HNSCC confirmed by biopsy and histopathological examination were included in the final analysis. PET-CT and USgFNAC were performed as part of routine diagnostic work-up. The institution review board approved this study with a waiver of informed consent.

FDG PET-CT
PET of the head and neck area was acquired using a TruePoint Biograph mCT40 scanner (Siemens, Erlangen, Germany). After a fasting period of at least 6 hours, patients received intravenous injection of 2 MBq/kg ^{18}F-FDG. Approximately 60 min after the

administration of the tracer, head and neck PET images were acquired according to the European Association of Nuclear Medicine (EANM) recommendations [10]. The following parameters were used during imaging: A low dose CT scan was performed using Care Dose 4D and Care kV, reference parameters: 40 mAs, 120 kV. Subsequently, two 4 minute bed positions PET with time-of-flight and point spread function (TrueX) reconstruction, 4 iterations, 21 subsets, with a filter of 5 mm full width at half maximum, slice thickness 3mm were acquired. SUV calculations were performed using a lean body mass corrected formula.

USgFNAC

The examination was conducted by sonography radiologists using a IU 22 ultrasound machine (Philips Healthcare, Best, The Netherlands) with a 12-5 MHz linear-array transducer. Lymph nodes with a rounded shape, absence of a visible fatty hilum or suggested necrosis were considered abnormal and were targeted for aspiration cytology. Furthermore, cytology was obtained from all lymph nodes with a short axis diameter greater than 5 mm with the exception of level II lymph nodes where a 7 mm short axis diameter cutoff was used. Lastly, nodes considered positive on FDG PET-CT were targeted for aspiration irrespective of short axis diameter. A 21 gauge needle was used to obtain the cytology samples. For each targeted node two needle passes were performed with application of suction. The obtained aspirates were directly smeared onto separate glass microscope slides. If by visual inspection the amount of material was deemed insufficient or too bloody, a third pass was performed.

Image analysis

All lymph nodes from which cytology was obtained were retrospectively analyzed on FDG PET-CT. Localization of individual lymph nodes was based on the anatomical location described in the original ultrasound report. In some cases ultrasound images were compared to FDG PET-CT in order to select the correct node. If the node could not be localized, for example due to multiple pathological lymph nodes in the same level or incomplete reporting, a second observer compared the ultrasound and FDG PET-CT images. If this did not obtain a clear consensus, the node was excluded from analysis. SUVmax was measured by drawing a region of interest around the lymph node in the imaging software Syngo.via (Siemens Healthcare GmbH, Erlangen). The short axis diameter measurements were performed on the CT images. Both the first and second observer had access to the ultrasound report but were blinded to the outcome as described by the cytology report.

Statistical analyses

Histograms and receiver operating characteristic (ROC) curves were used to determine the correlation between SUV and cytological examination. USgFNAC was used as the reference test in order to calculate the positive predictive value, negative predictive values and other test characteristics of FDG PET-CT. An independent samples t-test was used to determine the difference in SUVmax between USgFNAC positive and negative lymph nodes. All statistical analyses were performed with SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp)

Sensitivity analysis

Robustness of the results was tested after the final analysis. First, all cytopathology reports were retrospectively reviewed by a cytopathologist (M.B.) and for all lymph nodes with a SUVmax ≥ 4.0 and a negative USgFNAC cytology, slides were revised. During this process lymph nodes with an unreliable conclusion due to low cell counts of the cytology sample were recorded. Secondly, the study's results were recalculated with the omission of these lymph nodes. The results were also recalculated by considering all of these lymph nodes as positive and finally as all negative regardless of their original outcome.

Results

Patients

A total of 142 patients matched the inclusion criteria. Thirty-five patients had received their USgFNAC before the FDG PET-CT with four patients having both examinations less than 7 days apart. Twenty-five patients were excluded for analysis: in 22 patients it was not possible to match any nodes examined by USgFNAC to the FDG PET-CT images. One patient could not be analyzed because the FDG PET-CT was not performed according to EANM guidelines. In two patients it was not possible to reliably determine the lymph node status of the specific lymph nodes from the cytology report due to incomplete reporting.

For the final analysis 170 lymph nodes in 117 patients were included. On USgFNAC 96 nodes were diagnosed positive and 71 nodes negative for squamous cell carcinoma (SCC). For three lymph nodes the diagnosis of the initial cytological examination was inconclusive due to low cell counts. Patients characteristics are shown in table 1 and an example of the imaging and cytology is shown in figure 1 .

Table 1. Baseline patient characteristics. TNM: AJCC 7th edition.

	Patients (n=117)	Percentage (%)
Gender		
Male	87	74
Female	30	26
Tumor location		
Oral cavity	12	10
Oropharynx	53	45
Nasopharynx	13	11
Hypopharynx	29	25
Larynx	10	9
T-stage		
T1-T2	40	34
T3-T4	77	66
N-stage		
N0	20	17
N1	17	15
N2	5	4
N2a	4	3
N2b	40	34
N2c	29	25
N3	2	2
M-stage		
M0	108	92
M1	9	8
HPV		
Positive	14	12
Negative	45	38
Unknown	58	50

For the nodes that could not be matched (n=29) to the FDG PET-CT images, the USgFNAC was positive for SCC in 9 cases, nonmalignant in 19 and inconclusive in 1.

SUVmax

The average SUVmax of the examined lymph nodes was 7.1, ranging from 1.5 to 25.7. The distribution of SUVmax values is shown in figure 2.

The SUVmax of USgFNAC proven lymph nodes metastases was higher than in USgFNAC negative lymph nodes. Positive lymph nodes had an average SUVmax of 9.5 (SD 5.1). Nonmalignant lymph nodes had a significantly different average SUVmax of 3.9 (SD 2.5; $p < 0.001$) (Figure 3).
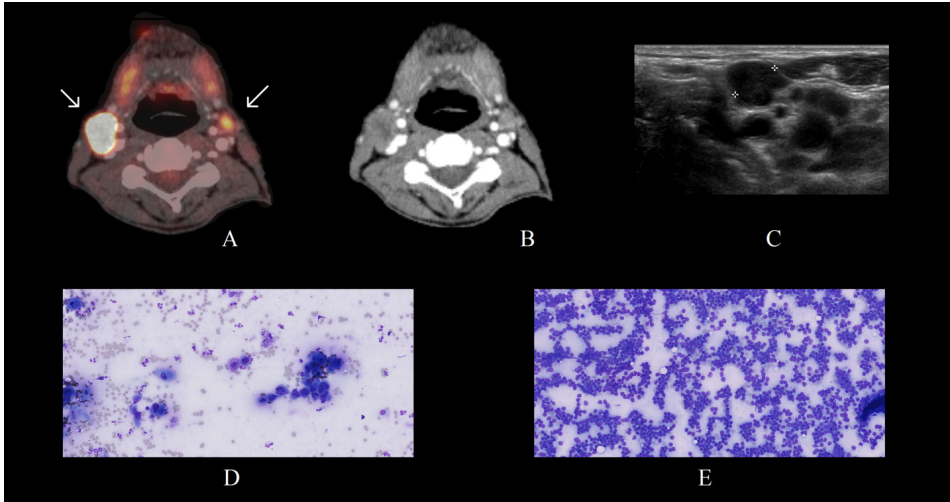
Figure 1. Examples of imaging and cytology: A. FDG PET-CT image, the white arrows point to lymph nodes. The large lymph node on the right side of the patient (left on the image) had a SUVmax of 12.13, the smaller node on the left had a SUVmax of 3.25. B. The corresponding CT image. C. The ultrasound image of the smaller node in the left neck, the distance between both crosses on the image is 0.8cm. D. Aspiration cytology of the right lymph node showing metastases (40x magnification). E. Aspiration cytology of the left lymph node without metastases (40x magnification).
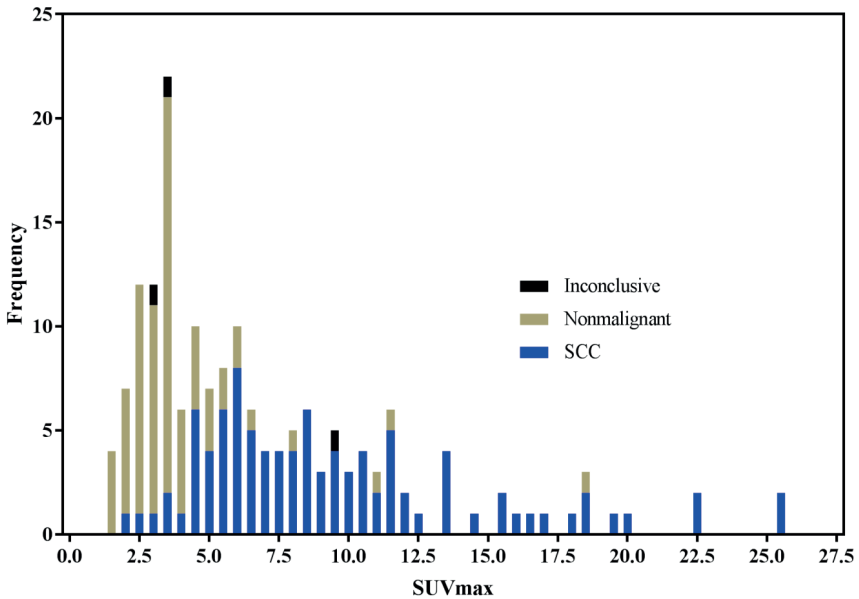
Figure 2. Histogram of all lymph node SUVmax values in bins of 0.5 with their corresponding USgFNAC results.
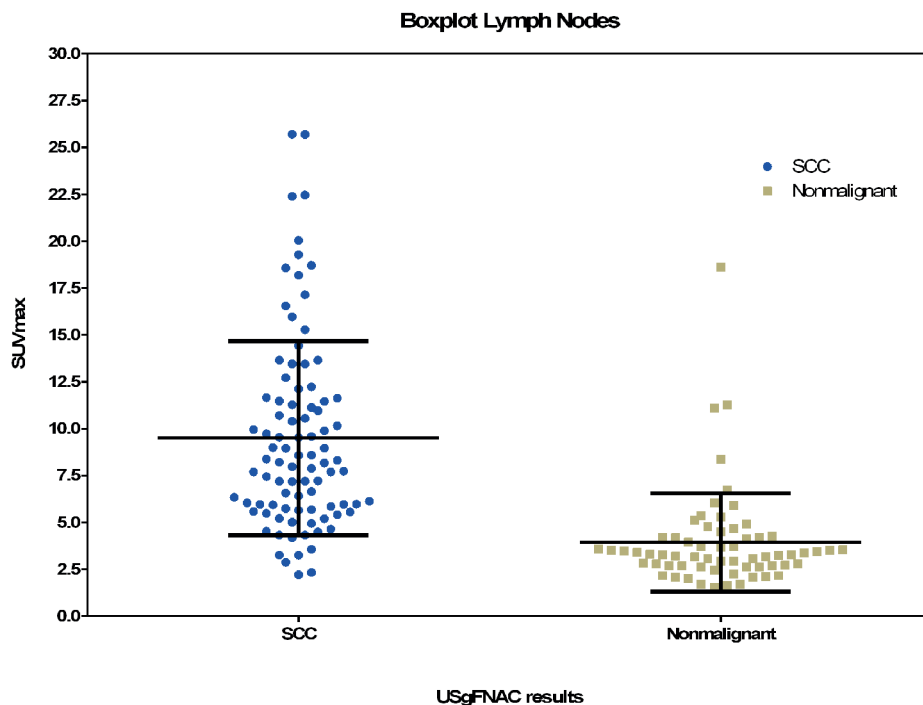
Figure 3. Boxplot of lymph node SUVmax values grouped by USgFNAC results. The black bars represent mean and standard deviation.

When comparing SUVmax to the USgFNAC results (reference standard) an area under the ROC curve of 0.91 was found (Figure 4). The cutoff value with the highest accuracy (86%) was at SUVmax 4.9. Corresponding test characteristics at this point are: sensitivity 88%, specificity 85%, PPV 88%, NPV 83%. Table 2 shows the 2x2 contingency table of FDG PET-CT using this cutoff value. The three lymph nodes with an initial inconclusive cytology report were excluded for this analysis.

Table 2. 2x2 contingency table at the cutoff of SUVmax 4.9

| | | **USgFNAC** | | |
		SCC	Other	Total
SUVmax	≥4.9	84	11	95
	<4.9	12	60	72
	Total	96	71	167
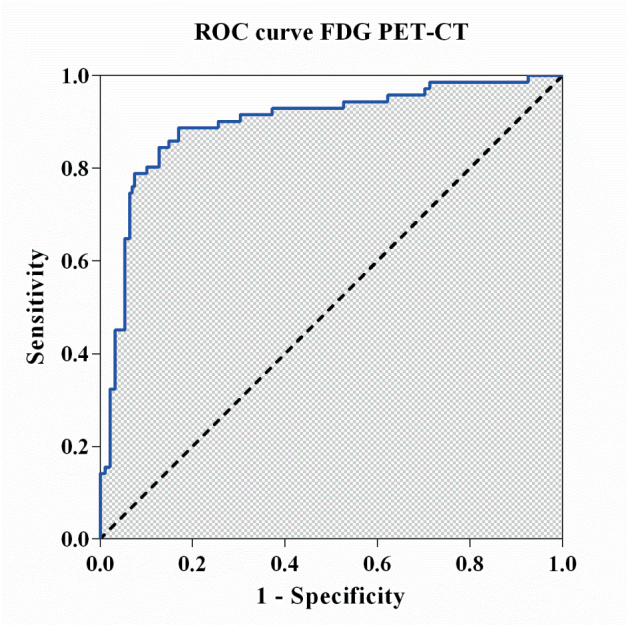
ROC curve FDG PET-CT

Figure 4. ROC curve FDG PET-CT vs USgFNAC

All nodes (n=10) with SUVmax values lower than 2.2 had a corresponding negative USgFNAC. Table 3 shows the test characteristics of FDG PET-CT using different cutoff values.

Table 3. Test characteristics at several SUVmax cutoff values. The underlined values are the highest in that category from among this group.

SUVmax cutoff	Sensitivity	Specificity	PPV	NPV	Accuracy
	Test characteristics				
2.0	<u>100%</u>	6%	59%	<u>100%</u>	60%
4.0	95%	72%	82%	91%	85%
5.0	86%	86%	89%	82%	<u>86%</u>
6.0	71%	92%	92%	70%	80%
8.0	53%	94%	93%	60%	71%
10.0	35%	<u>96%</u>	<u>92%</u>	53%	61%

Lymph node size

The average size of all examined lymph nodes, as measured on PET-CT, was 1.2 cm (range 0.3 – 3.8). Lymph nodes with positive cytology results had on average a short

axis diameter of 1.4 cm (SD 0.6) compared to 0.9 cm (SD 0.3) for nodes with a negative cytology. This difference was statistically significant (p<0.001).

Figure 5 shows the short axis diameter and SUVmax value of all individual lymph nodes sorted by the outcome of the USgFNAC. The cytology result for the nodes in the upper right quadrant (SUVmax ≥4.9 and short axis diameter ≥1.0 cm) was positive in 71, negative in 6 of the cases and inconclusive in 1 case resulting in a positive predictive value for USgFNAC results of 91% (71/78) if lymph nodes with SUVmax ≥4.9 and short axis diameter ≥1.0 cm were considered test positive.

The group with a high (≥4.9) SUVmax but a short axis diameter of <1.0 cm consisted of 17 lymph nodes out of which 12 were positive for SCC and 5 nonmalignant according to USgFNAC examination (PPV=71%).

Sensitivity analysis

Retrospective revision of the cytology report and slides yielded 5 lymph nodes with an uncertain negative conclusion and 4 lymph nodes with an uncertain positive conclusion
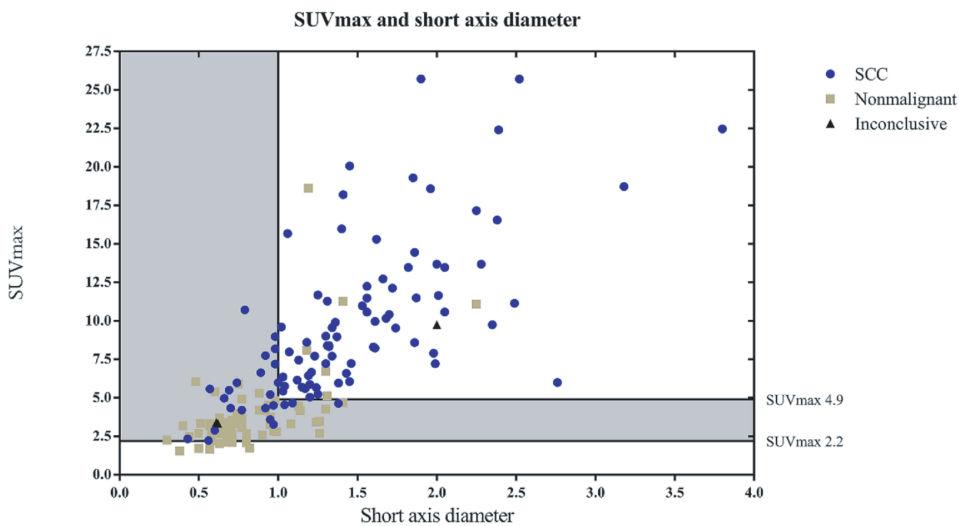
Figure 5. The horizontal line represents the SUVmax cutoff point with the highest accuracy (4.9). The vertical line separates nodes with a short axis diameter of >1.0 cm from those with a diameter of <1.0 cm. Lymph nodes with short axis diameter ≥1.0 cm and SUVmax ≥4.9 were in 69/76 positive on USgFNAC. Lymph nodes with SUVmax <2.2 were in 10/10 negative on USgFNAC. In all other lymph nodes (short axis diameter < 1.0 cm with SUVmax ≥2.2 and short axis diameter ≥1.0 cm and SUVmax ≥2.2 but <4.9) no pattern in USgFNAC results was observed.

due to low cell counts. Additionally, in 1 lymph node revision revealed suspicion of a B-cell lymphoma (SUVmax 4.2). These lymph nodes represented 6% (10/170) of all lymph nodes. Omission of these nodes or considering all these nodes positive or negative did not significantly influence the study result: The sensitivity of FDG PET-CT using a SUVmax 4.9 cutoff changed from 88% to 83% and the PPV changed from 88% to 92% by considering all these nodes positive. When omitting the nodes from analyses PPV changed to 91%. In all other cases PPV and sensitivity remained the same when rounded to the nearest whole number. In conclusion revision of the original cytology report changed the outcome of the USgFNAC in 10 cases, this had a very limited effect on the study results.

Discussion

This study shows that quantitative analysis of PET-CT images can be used to identify groups of lymph nodes in which FDG-PET can accurately predict the USgFNAC results. Of all lymph nodes in this study, 52% were selected by using a SUVmax <2.2 or ≥4.9 combined with a short axis diameter of ≥1.0 cm. This indicates that in around half of all nodes, USgFNAC performed after FDG PET-CT only confirms but does not provide additional information concerning the presence of lymph node metastases. Different sensitivity analyses showed the robustness of these findings.

It is important to note that 6 (4%) lymph nodes which were designated positive by the above criteria had a nonmalignant or inconclusive USgFNAC. This result could either be considered as false positive FDG PET-CT, or as false negative USgFNAC either due to sampling error or insufficient material [11]. Unfortunately, no neck dissection was performed and no comparison with the gold standard, i.e. histopathological examination of the resected lymph nodes, was possible.

This comparison between quantitative FDG PET-CT using SUVmax and histopathology has been done by others. Kitajima et al. [12] report an optimal SUVmax cutoff value of 3.5 for differentiating between malignant and benign lymph node metastases in 36 oral squamous cell carcinoma patients, while Matsubara et al. [13] examined 498 lymph nodes and found all nodes with a SUVmax of 4.5 or greater to contain lymph node metastases. These values are in line with the optimal cutoff found in our study of 4.9. Although the methodologies of these studies are not directly comparable to ours, they do provide some evidence that the FDG PET-CT criteria reported in our study not only determine when USgFNAC is unnecessary but also discriminate positive from negative nodes. The results might not be completely similar due to differences in imaging protocols. Our

study tried to find a cutoff value using standardized EANM directives in order to improve comparability between centers. Furthermore, these studies solve the difficulty of adequately matching a lymph node on FDG PET-CT with the corresponding lymph node in the neck dissection specimen by correlating lymph node levels or neck sides, scoring a true positive if a lymph node is positive on FDG PET-CT and any metastatic node is found by histopathological examination in the corresponding neck level or side. This method is reasonably effective when determining the presence of positive lymph nodes, but it lacks the accurate correlation needed for quantitative analysis. This might be possible by describing lymph node location based on surrounding anatomical structures during the neck dissection procedure. In contrast, one of the strengths of this study is that it correlates individual lymph nodes identified on FDG PET-CT with the pathology results obtained with USgFNAC of the same lymph node. This allows analysis of SUVmax of individual positive and negative lymph nodes.

FDG PET-CT is often performed during HNSCC workup in order to detect or delineate the primary tumor or to detect distant metastases. Additionally, in clinical practice FDG PET-CT examination will generally occur before USgFNAC as the aspiration of a lymph node might induce an inflammatory reaction. Inflammation causes accumulation of the FDG tracer resulting in a 'hot' lymph node on FDG PET-CT images [14]. This hot node might be erroneously considered metastatic. Therefore, FDG PET-CT images will be available in many HNSCC patients prior to USgFNAC results even though it is generally more expensive. As this study shows, these images can then be used to select lymph nodes of HNSCC patients which would not benefit from additional USgFNAC examinations.

It should be taken into consideration that cytological analysis of lymph nodes have an extremely high specificity and can confirm the presence of a lymph node metastasis in patients with HNSCC [15]. Also cytomorphological evaluation can confirm the squamous nature of the metastasis and exclude or diagnose metastasis of other tumors. This is particularly important in patients with a known history of other malignancies. Finally, due to the cellular material obtained through USgFNAC molecular examinations like HPV analysis and tumor DNA sequencing can be performed [16,17].

Limitations of this study are the retrospective design of the study which might have hindered accurate correlation between ultrasound images and FDG PET-CT images. Even though great care was taken to match lymph nodes from which cytology was obtained and SUVmax measurements, it is possible that in some cases the measured node was a different node than the node from which cytology was obtained. Secondly, the conclusions are based on a single center analysis. External validation of the reported cutoff values may be necessary to determine generalizability of results to other centers.

In conclusion, this study describes a pragmatic approach to predict the results of USgFNAC examinations of cervical lymph nodes in HNSCC patients by quantitative FDG PET-CT parameters and reduce the need for USgFNAC after FDG PET-CT already performed for other indications.

References

1. Johnson JT. A surgeon looks at cervical lymph nodes. Radiology. 1990 Jun;175(3):607–10.

2. Layland MK, Sessions DG, Lenox J. The influence of lymph node metastasis in the treatment of squamous cell carcinoma of the oral cavity, oropharynx, larynx, and hypopharynx: N0 versus N+. Laryngoscope. 2005 Apr;115(4):629–39.

3. Xing Y, Zhang J, Lin H, Gold KA, Sturgis EM, Garden AS, et al. Relation between the level of lymph node metastasis and survival in locally advanced head and neck squamous cell carcinoma. Cancer. 2016 Feb 15;122(4):534–45.

4. Dünne AA, Müller HH, Eisele DW, Kessel K, Moll R, Werner JA. Meta-analysis of the prognostic significance of perinodal spread in head and neck squamous cell carcinomas (HNSCC) patients. Eur J Cancer. 2006 Aug;42(12):1863–8.

5. Evangelista L, Cervino AR, Chondrogiannis S, Marzola MC, Maffione AM, Colletti PM, et al. Comparison between anatomical cross-sectional imaging and 18F-FDG PET/CT in the staging, restaging, treatment response, and long-term surveillance of squamous cell head and neck cancer: a systematic literature overview. Nucl Med Commun. 2014 Feb;35(2):123–34.

6. Xu G, Li J, Zuo X, Li C. Comparison of whole body positron emission tomography (PET)/PET-computed tomography and conventional anatomic imaging for detecting distant malignancies in patients with head and neck cancer: a meta-analysis. Laryngoscope. 2012 Sep;122(9):1974–8.

7. Newbold K, Powell C. PET/CT in Radiotherapy Planning for Head and Neck Cancer. Front Oncol. 2012;2.

8. Ozer E, Naiboglu B, Meacham R, Ryoo C, Agrawal A, Schuller DE. The value of PET/CT to assess clinically negative necks. Eur Arch Otorhinolaryngol. 2012 Nov;269(11):2411–4.

9. Manca G, Vanzi E, Rubello D, Giammarile F, Grassetto G, Wong KK, et al. (18)F-FDG PET/CT quantification in head and neck squamous cell cancer: principles, technical issues and clinical applications. Eur J Nucl Med Mol Imaging. 2016 Jul;43(7):1360–75.

10. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015 Feb;42(2):328–54.

11. Amedee RG, Dhurandhar NR. Fine-needle aspiration biopsy. Laryngoscope. 2001 Sep;111(9):1551–7.

12. Kitajima K, Suenaga Y, Minamikawa T, Komori T, Otsuki N, Nibu K-I, et al. Clinical significance of SUVmax in (18)F-FDG PET/CT scan for detecting nodal metastases in patients with oral squamous cell carcinoma. Springerplus. 2015;4:718.

13. Matsubara R, Kawano S, Chikui T, Kiyosue T, Goto Y, Hirano M, et al. Clinical Significance of Combined Assessment of the Maximum Standardized Uptake Value of F-18 FDG PET with Nodal Size in the Diagnosis of Cervical Lymph Node Metastasis of Oral Squamous Cell Carcinoma. Acad Radiol. 2012 Jun;19(6):708–17.

14. Malik AI, Akhtar N, Loya A, Yusuf MA. Endoscopic ultrasound - Fine needle aspiration of 2-deoxy-2-[^{18}F] fluoro-D-glucose avid lymph nodes seen on positron emission tomography- computed tomography -what looks like cancer may not always be so. Cancer Imaging. 2014;14(1):1–6.

15. de Bondt RBJ, Nelemans PJ, Hofman PAM, Casselman JW, Kremer B, van Engelshoven JMA, et al. Detection of lymph node metastases in head and neck cancer: a meta-analysis comparing US, USgFNAC, CT and MR imaging. Eur J Radiol. 2007 Nov;64(2):266–72.

16. Krane JF. Role of cytology in the diagnosis and management of HPV-associated head and neck carcinoma. Acta Cytol. 2013;57(2):117–26.

17. Rizzo G, Black M, Mymryk JS, Barrett JW, Nichols AC. Defining the genomic landscape of head and neck cancers through next-generation sequencing. Oral Dis. 2015 Jan;21(1):e11-24.

Chapter 3

Evaluation of diffusion weighted imaging for tumor delineation in head-and-neck radiotherapy by comparison with automatically segmented 18F-fluorodeoxyglucose positron emission tomography

Tim Schakel, Boris Peltenburg, Jan-Willem Dankbaar, Carlos E. Cardenas, Michalis Aristophanous, Chris H.J. Terhaard, Johannes M. Hoogduin, Marielle E.P. Philippens

Abstract

Background and purpose
Diffusion weighted (DW) MRI may facilitate target volume delineation for head-and-neck (HN) radiation treatment planning. In this study, we assessed the use of a dedicated, geometrically accurate, DW-MRI sequence for target volume delineation. The delineations were compared with semi-automatic segmentations on ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) images and evaluated for interobserver variation.

Methods and materials
Fifteen HN cancer patients underwent both DW-MRI and FDG-PET for RT treatment planning. Target delineation on DW-MRI was performed by three observers, while for PET a semi-automatic segmentation was performed using a Gaussian mixture model. For interobserver variation and intermodality variation, volumes, overlap metrics and Hausdorff distances were calculated from the delineations.

Results
The median volumes delineated by the three observers on DW-MRI were 10.8, 10.5 and 9.0 cm^3 respectively, and was larger than the median PET volume (8.0 cm^3). The median conformity index of DW-MRI for interobserver variation was 0.73 (range 0.38–0.80). Compared to PET, the delineations on DW-MRI by the three observers showed a median dice similarity coefficient of 0.71, 0.69 and 0.72 respectively. The mean Hausdorff distance was small with median (range) distances between PET and DW-MRI of 2.3 (1.5–6.8), 2.5 (1.6–6.9) and 2.0 (1.35–7.6) mm respectively. Over all patients, the median 95th percentile distances were 6.0 (3.0–13.4), 6.6 (4.0–24.0) and 5.3 (3.4–26.0) mm.

Conclusion
Using a dedicated DW-MRI sequence, target volumes could be defined with good interobserver agreement and a good overlap with PET. Target volume delineation using DW-MRI is promising in head-and-neck radiotherapy, combined with other modalities, it can lead to more precise target volume delineation.

Introduction

Intensity modulated radiotherapy (RT) has allowed for conformal dose distributions, aiming to maximize the dose to the gross tumor volume (GTV) and minimize the dose to normal tissue. Accurate determination of the GTV is a key issue in radiotherapy and forms the basis of treatment planning [1]. However, it is also one of the major sources of uncertainty [2–4]. The accurate delineation of the GTV is a time consuming step, prone to errors. Delineations of the GTV using computed tomography (CT) and magnetic resonance imaging (MRI) show large variations between observers and an overestimation of the actual pathological tumor volume [5–7]. Additionally, FDG-PET has shown to result in more conformal segmentations than CT or MRI [8,9].

To address the interobserver variation, automatic segmentation of the GTV would be ideal. Automatic segmentation is also preferable in online and/or adaptive MRI-guided treatments. Several authors have proposed ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) as imaging modality for automatic segmentation due to its high signal to background ratio [10–14]. However, the use of PET for segmentation is challenging. The segmentations can depend on the contrast and noise characteristics of the PET images, which, apart from tumor characteristics, originate from different acquisition and reconstruction protocols [15,16]. Additionally, a variety of segmentation algorithms have been proposed which result in large variations in target volumes [17,18]. Therefore, successful automatic segmentation requires a careful choice of acquisition and reconstruction parameters and segmentation algorithms. Other drawbacks are the increased radiation burden, resolution limitations and relatively high costs.

Diffusion weighted (DW) MRI might be an alternative imaging modality for FDG-PET as DW-MRI generates images with high contrast between tumor and surrounding normal tissue making it a candidate to improve MRI-based delineations and suitable for automatic segmentation. The contrast in DW-MRI is based on differences in the restriction of water diffusion on a microscopic level. The resulting high contrast images, along with the derived quantitative apparent diffusion coefficient (ADC) maps, could be used for delineation purposes [19,20]. However, the geometrically accurate acquisition of DW-MRI in the head and neck region is difficult due to the presence of air cavities causing large magnetic field inhomogeneities. Using conventional DW-MRI, acquired with echo planar imaging (EPI), image distortions up to centimeters can arise [21]. Recently, we have proposed the use of an alternative DW-MRI sequence; split acquisition of fast spin-echo signal for diffusion imaging (SPLICE). Here, EPI is replaced with turbo spin echo for data acquisition, which leads to less distortions as it is relatively insensitive to magnetic field inhomogeneities introduced by the patient. The DW-SPLICE sequence

has excellent geometric performance, allowing the DW-MRI images to be used for target delineation [22].

As the overestimation of the tumor volume results in a high burden on normal tissues surrounding the tumor due to the large treatment volume, the tumor has to be segmented as close (conformal) as possible. By definition the GTV only includes macroscopic tumor tissue, and requires a clinical target volume (CTV) margin to encompass microscopic spread. To avoid normal tissue burden, the CTV needs to be minimized. Recently, Ligtenberg et al. showed in an image validation study with pathology that GTVs automatically segmented from PET showed a tumor volume closest to the tumor volume on histology with a high coverage and the lowest overestimation [8]. Unfortunately, such validation studies cannot be performed in patients treated with primary radiotherapy. However, since the PET-based volume resembled the tumor volume best, and with the smallest overestimation, we chose to use PET-based segmentations as a reference to evaluate GTVs created using DW-SPLICE.

The aim of the present study was to assess the use of the dedicated DW-MRI sequence in target volume delineation. The GTVs were evaluated in terms of interobserver agreement, volume and spatial concordance with PET.

Materials & Methods

Patients
A total of 15 head and neck patients were included in this study, patient details are given in supplementary table 4.1. The institutional review board provided a waiver of consent. These patients, scheduled for RT treatment, underwent a clinical FDG-PET/CT and an MRI exam, both in treatment position using a personalized head support and 5-point head and shoulder immobilization mask [23].

FDG-PET/CT imaging
Patients fasted 6 hours before imaging on a mCT-Biograph PET/CT scanner (Siemens Medical Solutions, USA). An FDG dose of 2.0 MBq/kg was injected 60 minutes before image acquisition. Image acquisition was done at 2 bed positions, at 4 minutes per position. Image reconstruction was done using an ordered subset expectation maximization algorithm, utilizing point-spread function and time-of-flight information, at a reconstructed voxel size of 2.04 x 2.04 x 1.5 mm^3 (AP x RL x FH) with body weight corrected standardized uptake values.

MR imaging

MR imaging was performed on a 3.0~T Ingenia wide bore MR system (Philips Healthcare, The Netherlands) using 2-element Flex-M surface coils. The DW-SPLICE sequence was prototyped and loaded on the system, generating geometrically accurate diffusion weighted images [22]. Diffusion weighting was applied in an isotropic fashion using three orthogonal directions with b-values of 0 and 800 s/mm^2 (b0 and b800 respectively). ADC maps were generated on the MR system. Fat suppression was executed using spectral presaturation with inversion recovery. Detailed sequence parameters are given in table 1. The exam also contained T1 (with and without contrast) and T2 weighted TSE scans.

Table 1. Sequence parameters for the DW-SPLICE sequence. The protocol was adapted for laryngeal imaging (denoted in round brackets).

Scan	DW-SPLICE
FOV (RL×AP×FH) [mm3]	230×280×120 (200x200x120)
Acquired voxel size [mm2]	1.8×1.8
Slice thickness [mm]	4
TE [ms]	52
TR [ms]	16,366 (15,844)
SENSE acceleration factor	2
(Echo) train length {dummies}	64 {1} (55 {1})
Bandwidth [Hz]	900.3 (858.5)
b-values [s/mm2] {averages}	0 {2}, 800 {5}
Total acquisition time	4m38s (4m29s)

Target volume delineations

For the acquired PET scans, the GTV was segmented semi-automatically using a Gaussian mixture model (GMM). With this method, voxels are classified based on statistical differences in their intensity distributions [24–26]. The GMM assumes independent observations and models the image intensities with 10 Gaussian distributions. Following an initialization step by placing a selection box around the lesion, Gaussian distribution parameters are initialized automatically. The method was adapted for head and neck by fully automating the initialization of the Gaussian distribution means as described in detail in previous publications [24–26]. Following initialization, the Gaussian distributions are fitted to the image intensities using maximum likelihood estimation method. The algorithm then classifies regions to either tumor or background, generating binary masks.

The GTV was delineated in the diffusion weighted images by three observers. First, the hyperintense region on the b800 image was identified, this was aided by a semi-automatic segmentation. On the b800 image, a seed point was placed centrally within the tumor and using a threshold of 50% of the maximum signal intensity, a region was segmented. Following this initial segmentation, the contours were manually adapted using the contrast of the ADC map, where the individual observers excluded regions of high ADC typically at the edge of the tumors. While doing this, the b0 and b800 images were also available to the observers for reference. All contours (referred to as DW_1, DW_2, DW_3) were converted to binary masks.

Additionally, as all of the patients received radiation treatment following imaging, the target volumes were retrieved from the treatment plans for these patients (GTV_{RT}) as reference. The radiation oncologists mostly relied on CT and anatomical MRI (T1 weighted, with and without contrast, and T2 weighted), but PET and DW-MRI were also available.

Data analysis

All scans from the MRI exam were matched with the PET/CT using rigid registration. The resulting transformations were used to resample both the PET and DW-SPLICE masks on the grid of the CT ($0.98 \times 0.98 \times 2.0$ mm^3), in order to compare the segmentations. While resampling the masks, the volume was kept the same on the new grid.

Using the resampled masks the delineated volumes for PET and DW-SPLICE were determined. For each observer the DW-MRI delineation was further compared to the PET segmentation by means of Dice similarity coefficients (DSC) and Hausdorff distance (HD) analysis. The DSC is a measure of spatial overlap and is defined as:

$$DSC\ (A,B) = \frac{2(A \cap B)}{(A + B)}$$

where A and B are the respective delineated regions and $\cap$ is the intersection. A DSC of 0 indicates no spatial overlap at all and a DSC of 1 indicates complete overlap. The distance between two delineated contours is represented using the Hausdorff distance, which measures, for each point on a delineated contour, the Euclidean distance to the nearest point on the other contour. Per contour pair, the mean Hausdorff distance (HD_{mean}) and the 95th percentile distance (HD_{95}) were calculated. The results for each metric are given as the median (range) over all patients. We tested for significant differences ($p < 0.05$) in metrics using a Wilcoxon signed rank test.

For the interobserver analysis of the DW-SPLICE delineations, a generalized conformity index

$$CIgen = \frac{\sum_{pairs ji} A_i A_j \vee}{\sum_{pairs ji} A_i A_j \vee}$$

with A_i and A_j the respective delineated regions and U the union over all observers was calculated per patient, because this metric is independent of the number of observers [27]. Additionally, DSC were calculated per observer pair. Additionally, the HD analysis was also performed for the observer pairs as this metric is independent of the volume of the tumor and has clinical relevance as it is related to uncertainty margins.
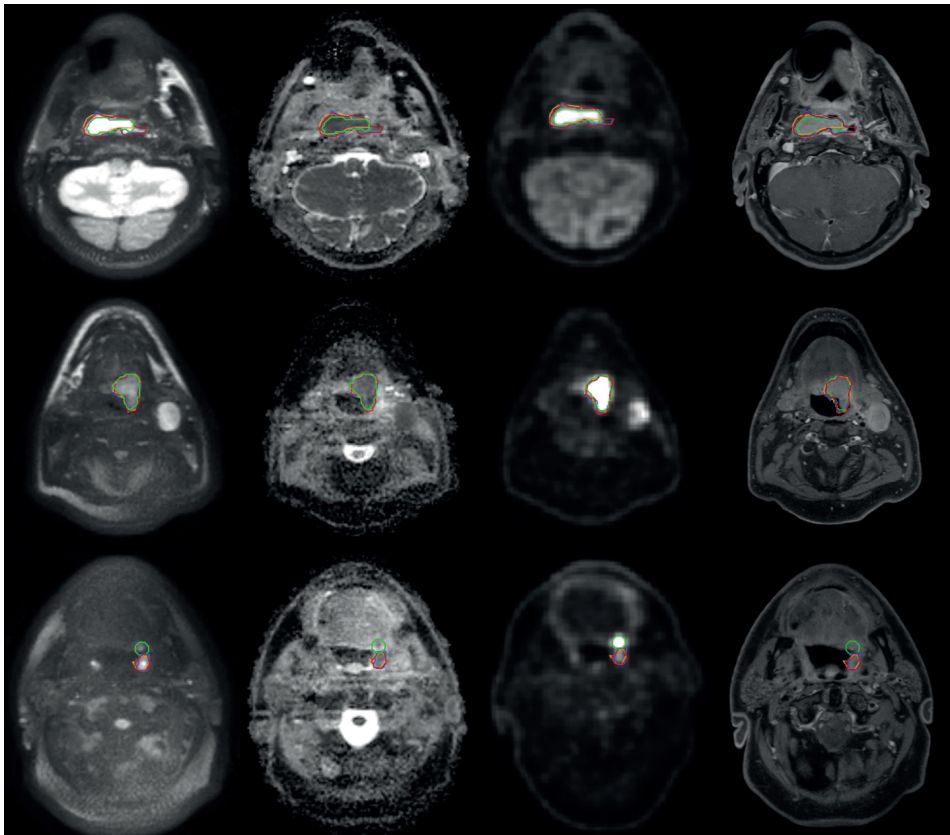
Figure 1. Patient examples showing the correspondence of the delineated target volumes. Each row of images is from a single patient and shows transverse slices with, from left to right: DW b800; ADC map; FDG-PET; T1 weighted TSE mDixon water reconstruction after gadolinium contrast agent injection. The images are taken from patient 9, 12 and 1 respectively. The delineations from FDG-PET (green), DW_1 (red), DW_2 (blue) and DW_3 (orange) are shown on all imaging.

Results

An example of the imaging data is given in figure 1, showing the correspondence between the different modalities as well as between the different observers on the diffusion weighted imaging. Additionally, anatomical T1 weighted images after gadolinium contrast agent injection are shown for reference. Initial semi-automatic segmentation of the b800 images proved difficult in three out of 15 patients, where the pixel intensity gradient was low around the tumor in images with limited SNR. For these patients, the tumor region was cropped manually to avoid other structures, such as unsuppressed fat, to be included in the region of interest.

Volumes

The volumes found on the GMM segmentation of PET had a median volume of 8.0 cm^3 (1.2 - 38.9), while the DW-SPLICE volumes were larger for all three observers: DW_1 10.8 (1.8 - 44.3), DW_2 10.5 (1.8 - 38.1), DW_3 9.0 (1.4 - 37.8) cm^3 (Figure 2). The difference with PET was significant for DW_1 (p=0.0001) and DW_2 (p=0.0043), but not for DW_3 (p=0.0637). The median GTV_{RT} was 15.63 cm^3, a significant difference compared to both PET and all DW-MRI volumes. An overview of all delineated volumes is given in supplementary table 1.
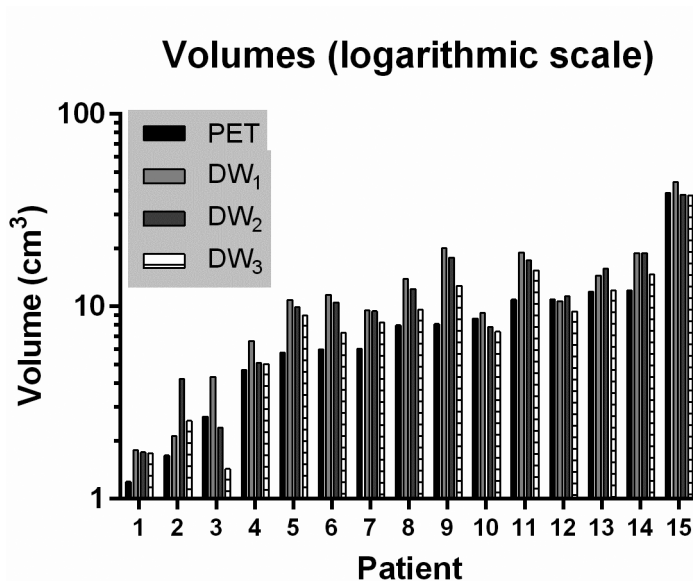

Figure 2. The delineated volumes per patient for the different observers and modalities. Patients were ranked in ascending order of the volume determined on FDG-PET.

Intermodality delineation overlap and distance analysis

The median (range) DSC of the PET segmentation with each DW-MRI delineation was very similar for the three observers: 0.71 (0.03 - 0.80), 0.69 (0.03 - 0.79) and 0.72 (0.03 - 0.82) respectively (figure 3a). An overview of correspondence measures is given in supplementary table 1. The DSC was reasonable to good (0.44 - 0.82) in all but 2 patients; these were the patients with the smallest volumes. Patient 1 had practically no overlap between PET and DW-MRI. This patient showed two neighboring hotspots on PET, with the anterior one showing most tracer uptake so this region was segmented by the GMM method. On DW-MRI however, the posterior hotspot showed most diffusion restriction and low ADC (figure 1, bottom row). The other patient (number 2) showing poor correspondence, had a tumor located at the base of the tongue. The mismatch was mainly caused by tongue movement between the PET scan and MRI exam.
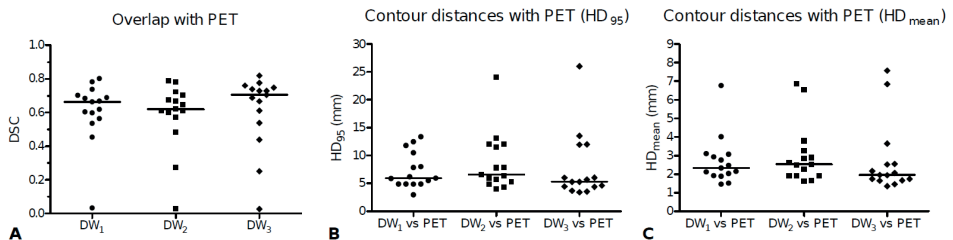
Figure 3. The agreement between DW-MRI and FDG-PET as measured with Dice Similarity Coefficient (a), the 95th percentile contour distance (b) and the mean contour distance (c). The horizontal bars represent the median.

Using the same delineation pairs, distance analysis was performed (figure 3b and 3c) and both the HD_{95} and HD_{mean} were calculated. For all three observers, the HD_{mean} was small with median (range) distances between PET and DW-MRI of 2.3 (1.5 - 6.8), 2.5 (1.6 - 6.9) and 2.0 (1.35 - 7.6) mm, respectively. Over all patients, the median 95th percentile distances were 6.0 (3.0 - 13.4), 6.6 (4.0 - 24.0) and 5.3 (3.4 - 26.0) mm. Similar to the DSC, patients 1 and 2 showed the largest distances between the PET and DW-MRI contours.

Interobserver agreement

The median CI_{gen} over all patients was 0.73 (0.38 - 0.80) indicating good agreement among the different observers (figure 4a). This is also seen in the three pairwise DSCs, median (range): observer 1 and 2, 0.87 (0.52 - 0.90); observer 1 and 3, 0.83 (0.43 - 0.91); observer 2 and 3, 0.82 (0.59 - 0.87). Patient 2 and 3 showed the least interobserver agreement. These were patients with small tumor volumes, located in the oropharynx

Figure 4. Interobserver agreement for all patients on diffusion weighted imaging, expressed as generalized conformity index (a), contour distances HD_{95} (b) and HD_{mean} (c). The horizontal bars are the medians.
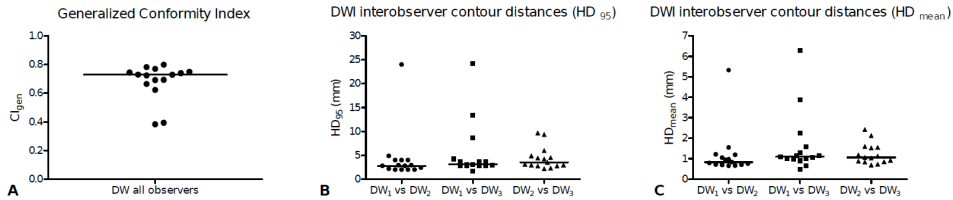
Discussion

In this study, we assessed semi-automated target volume delineation in head and neck cancer patients using a non-EPI diffusion weighted MRI sequence, DW-SPLICE. Furthermore, we investigated the variation between different observers on DW-MRI as well as the agreement between DW-MRI and [18]FDG-PET generated GTV delineations. The agreement between observers was good with a median CI_{gen} of 0.73 and small distances between the individual contours (median HD_{95} <3.5 mm). The delineated volumes on DW-MRI were found to be significantly larger (median volumes 10.8 cm^3, 10.5 cm^3 and 9.0 cm^3) than the PET segmentations (median volume 8.0 cm^3), nevertheless there was a substantial overlap between PET and DW-MRI with median DSCs of 0.71, 0.69 and 0.72 for the three different observers respectively. Both the PET and DW-MRI volumes were substantially smaller than the clinically used target volume GTV_{RT} (median volume 15.63 cm^3), which is known to overestimate the GTV [8,9].

High b-value DW-MRI images have a similar appearance as PET images, which makes them relatively easy to interpret by the observers. This was demonstrated by the good agreement between the different observers in this study. The median CI_{gen} over all patients was 0.73, where 0.7 is considered to be an indicator of good overlap [28]. The two cases showing a low CI_{gen} concerned patients with small tumors located in the oropharynx, adjacent to lymphatic tissue. Lymphatic tissue shows diffusion restriction and thus has diffusion properties similar to that of tumors, i.e. showing high signal on high b-value DW-MRI images and a low ADC. Therefore, DW-MRI should always be used as an addition to high quality anatomical images. Other delineation studies in the head-and-neck area show lower or similar CI_{gen} values [29,30]. Jager et al. [5] showed 0.61 for laryngeal cancer using CT and 0.57 using CT/MR; Geets et al. reported 0.41 and 0.42 for laryngeal and oropharyngeal GTVs respectively [31]; Mukesh et al. reported 0.54 for CTV delineations on CT [32].

Despite the different biological background and acquisition method of DW-MRI and FDG-PET, there was a large overlap between PET and DW-MRI. For a large part, both techniques identified the same target for treatment. Differences mainly occur at the edges of the delineated volumes, but the distances between the different contours were small in general. Over all patients, the median of the average distance between DW-MRI and PET contours was between 2.0 and 2.5 mm for the three different observers. For comparison, the acquired in-plane voxel sizes in DW-MRI and PET were 1.8 and 2.0 mm respectively. Furthermore, the PET GTVs, derived by the GMM segmentation, still require an appropriate CTV margin to encompass all microscopic tumor tissue. According to Ligtenberg et al. [8], this CTV margin is 5.2 mm in laryngeal and hypopharyngeal cancer. In a similar fashion, the DW-MRI GTV delineations will also require an appropriate CTV margin.

Generally, the difficulty of incorporating DW-MRI in the practice of target volume delineation in head-and-neck is the large geometrical distortion. To overcome this issue we used a dedicated sequence, DW-SPLICE, for the acquisition of diffusion images. Although this dealt with the geometrical distortions, some mismatch with other imaging modalities, like PET-CT, can appear. Despite the fact that patients were fixated in an immobilization mask, some misalignment may remain since patients were scanned at two different time points. Additionally, internal motion of mobile structures due to breathing, tongue movement or swallowing can still occur. This could explain some of the differences in the delineations between DW-MRI and PET, especially for patient 2.

The largest differences between DW-MRI and PET were found in small tumors, adjacent to lymphatic tissue. Both DW-MRI and FDG-PET are prone to false positives in these tissues, as it can show restricted diffusion behavior and increased FDG uptake. PET and DW-MRI visualize different biological processes and previous studies have investigated the relation between PET and DW-MRI in head-and-neck cancer [33–39]. However, these studies mainly looked at the relationship between SUV and ADC characteristics within the tumor and relating these to staging [38], histopathological parameters [35], or response prediction [34] and evaluation [34,39]. Houweling et al. [33] investigated a radiotherapy application of PET and DW-EPI after deformable registration. They specifically looked into targets for dose painting within a previously defined GTV. The spatial resolution is a limitation of both techniques. The detection limit and point spread function in PET could account for some of the differences found in the smaller lesions (i.e. patients 1 and 2). In PET the spatial resolution is mainly limited by the detector width, where in DW-MRI the resolution is limited due to the usage of a single-shot readout sequence. Advances in MRI sequence design (i.e. robust multi-shot DW-MRI sequences, faster imaging etc.), could still provide additional gains in spatial resolution.

A limitation of the study is that no histopathology was available to compare the target volume with. In order to fully assess the value of DW-MRI for target volume delineation, it would be ideal to use pathology specimens in image validation studies [9,25]. Unfortunately, this is very challenging for some anatomical sites such as the oro- and nasopharynx. As we know from previous studies, GTV on CT and MRI overestimates the tumor largely and shows large interobserver variation [5,7,8]. Therefore, we compared the target volume with an automatic segmentation on PET, as the PET volume approaches the true tumor volume closest and the automatic segmentation method is validated versus pathology. The acquisition of only 2 b-values, b=0 s/mm^2 and b=800 s/mm^2 in the diffusion sequence is another limitation. This could result in an overestimation of the ADC value in areas with a large blood plasma volume [40]. Finally, the number of patients in this study is limited.

The target volumes derived from DW-MRI were generally smaller than the GTV defined in current clinical practice (DW$_1$, DW$_2$, DW$_3$ vs GTV$_{RT}$). These smaller target volumes already lead to smaller treatment volumes. The addition of DW-SPLICE to the current practice of target volume delineation can help reduce variation among observers, since it allows for semi-automatic segmentations. The region identified by both DW-MRI and PET could be used as a first estimate for the tumor outline. Subsequently, using conventional T1 and T2 weighted MRI and CT, the GTV delineation can be refined. The combination of all these imaging modalities can further improve target volume delineation.

In conclusion, using an optimized DW-MRI sequence, target volumes were defined with good interobserver agreement and a good overlap with PET. Target volume delineation using undistorted DW-MRI is promising in head-and-neck radiotherapy.

Acknowledgements

Supplementary material

Supplementary table 1: Comparison between PET and DWI. Volumes of the different modalities and delineators. Generalized conformity indices Dice similarity coefficients and contour distances between PET and DWI. GTV_{RT} is the treatment volume retrieved from the treatment plan. Patient 11 was ultimately treated at a different institution, therefore, no volume was available.

Patient #	Volumes (cm³)					CI_{gen}	DSC vs PET			HD_{mean} vs PET (mm)			HD_{95} vs PET (mm)		
	PET	DW_1	DW_2	DW_3	GTV_{RT}	DW_{all}	DW_1	DW_2	DW_3	DW_1	DW_2	DW_3	DW_1	DW_2	DW_3
1	1.23	1.79	1.75	1.72	12.17	0.73	0.03	0.03	0.03	6.77	6.54	6.84	11.78	12.06	11.95
2	1.68	2.12	4.21	2.55	6.32	0.38	0.45	0.28	0.25	2.12	6.85	7.57	4.88	24.04	26.00
3	2.68	4.31	2.34	1.43	3.16	0.39	0.68	0.61	0.44	1.51	1.65	3.64	4.88	4.29	13.54
4	4.69	6.63	5.07	5.02	7.75	0.69	0.54	0.48	0.54	2.93	2.85	2.52	8.00	6.39	6.00
5	5.74	10.80	9.90	9.00	18.59	0.77	0.62	0.62	0.67	3.07	2.89	2.55	13.35	12.04	12.00
6	6.00	11.5	10.50	7.30	9.46	0.62	0.60	0.57	0.74	3.10	3.26	1.66	10.49	11.55	5.69
7	6.01	9.52	9.46	8.26	19.05	0.74	0.70	0.70	0.76	1.89	1.90	1.46	4.87	5.68	3.68
8	7.99	13.9	12.32	9.61	2.56	0.72	0.60	0.61	0.61	2.47	2.26	2.16	5.93	5.28	5.28
9	8.12	20.16	17.93	12.83	17.77	0.66	0.56	0.60	0.73	4.01	3.80	1.74	12.46	13.11	4.39
10	8.67	9.27	7.80	7.39	13.49	0.74	0.74	0.72	0.75	1.92	1.90	1.74	5.86	4.88	4.64
11	10.86	19.11	17.34	15.39	-	0.75	0.66	0.67	0.73	2.76	2.49	2.06	6.19	6.61	5.28
12	10.87	10.64	11.32	9.41	20.19	0.80	0.80	0.79	0.82	1.46	1.62	1.35	2.96	4.00	3.41
13	11.91	14.47	15.70	12.12	38.14	0.73	0.69	0.67	0.69	2.03	2.61	1.94	4.88	7.81	4.45
14	12.15	18.97	18.97	14.68	52.36	0.69	0.67	0.65	0.70	2.15	2.53	1.65	5.52	7.92	3.55
15	38.89	44.31	38.1	37.77	63.42	0.78	0.78	0.78	0.78	2.33	1.92	1.95	7.87	6.00	6.04
Median	7.99	10.80	10.50	9.00	15.63	0.73	0.66	0.62	0.70	2.33	2.53	1.95	5.93	6.61	5.28
Minimum	1.23	1.79	1.75	1.43	2.56	0.38	0.03	0.03	0.03	1.46	1.62	1.35	2.96	4.00	3.41
Maximum	38.89	44.31	38.10	37.77	63.42	0.80	0.80	0.79	0.82	6.77	6.85	7.57	13.35	24.04	26.00

Supplementary table 2: Patient characteristics: tumor, lymph node and distant metastasis staging (TNM), tumor location and patient sex.

Patient	T	N	M	Location	Sex
1	2	0	0	Tonsillar Region	M
2	2	3	0	Base of Tongue	M
3	2	1	0	Tonsillar Region	M
4	2	0	0	Tonsillar Region	F
5	2	2b	0	Tonsillar Region	M
6	1	2	0	Nasopharynx	M
7	3	2c	0	Hypopharynx	M
8	1	0	0	Base of Tongue	F
9	1	0	0	Nasopharynx	M
10	3	2b	0	Tonsillar Region	M
11	1	0	0	Nasopharynx	M
12	2	2b	0	Base of Tongue	M
13	3	2b	0	Tonsillar Region	M
14	4a	2b	0	Oropharynx	M
15	4a	2b	0	Oropharynx	F

References

1. Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer. 2016 Apr;16(4):234–49.

2. Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys. 2008 Oct;33(4):136–40.

3. 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 Jan;179(1):21–30.

4. Chen AM, Chin R, Beron P, Yoshizaki T, Mikaeilian AG, Cao M. Inadequate target volume delineation and local–regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer. Radiother Oncol. 2017 Jun;123(3):412–8.

5. 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;10:26.

6. Caldas-Magalhaes J, Kooij N, Ligtenberg H, Jager EA, Schakel T, Kasperts N, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol. 2015;54(8):1181–7.

7. Cooper JS, Mukherji SK, Toledano AY, Beldon C, Schmalfuss IM, Amdur R, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). Int J Radiat Oncol Biol Phys. 2007 Mar 15;67(4):972–5.

8. Ligtenberg H, Jager EA, Caldas-Magalhaes J, Schakel T, Pameijer FA, Kasperts N, et al. Modality-specific target definition for laryngeal and hypopharyngeal cancer on FDG-PET, CT and MRI. Radiother Oncol. 2017 Apr;123(1):63–70.

9. 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 Oct;233(1):93–100.

10. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging. 2010 Nov 25;37(11):2165–87.

11. Cheebsumon P, Boellaard R, de Ruysscher D, van Elmpt W, van Baardwijk A, Yaqub M, et al. Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. EJNMMI Res. 2012;2(1):56.

12. Daisne J-F, Sibomana M, Bol A, Doumont T, Lonneux M, Grégoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol. 2003 Dec;69(3):247–50.

13. Zaidi H, Abdoli M, Fuentes CL, El Naqa IM. Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging. 2012 May 31;39(5):881–91.

14. Geets X, Lee JA, Bol A, Lonneux M, Grégoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging. 2007 Aug 20;34(9):1427–38.

15. Brambilla M, Matheoud R, Secco C, Loi G, Krengli M, Inglese E. Threshold segmentation for PET target volume delineation in radiation treatment planning: The role of target-to-background ratio and target size. Med Phys. 2008 Mar 7;35(4):1207–13.

16. Matheoud R, Monica P Della, Loi G, Vigna L, Krengli M, Inglese E, et al. Influence of reconstruction settings on the performance of adaptive thresholding algorithms for FDG-PET image segmentation in radiotherapy planning. J Appl Clin Med Phys. 2011 Mar;12(2):115–32.

17. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging. 2010 Nov;37(11):2165–87.

18. Schinagl DAX, Vogel W V., Hoffmann AL, van Dalen JA, Oyen WJ, Kaanders JHAM. Comparison of Five Segmentation Tools for 18F-Fluoro-Deoxy-Glucose–Positron Emission Tomography–Based Target Volume Definition in Head and Neck Cancer. Int J Radiat Oncol. 2007 Nov;69(4):1282–9.

19. Burbach JPM, Kleijnen J-PJ, Reerink O, Seravalli E, Philippens MEP, Schakel T, et al. Inter-observer agreement of MRI-based tumor delineation for preoperative radiotherapy boost in locally advanced rectal cancer. Radiother Oncol. 2016 Feb;118(2):399–407.

20. van Rossum PSN, van Lier ALHMW, van Vulpen M, Reerink O, Lagendijk JJW, Lin SH, et al. Diffusion-weighted magnetic resonance imaging for the prediction of pathologic response to neoadjuvant chemoradiotherapy in esophageal cancer. Radiother Oncol. 2015 May;115(2):163–70.

21. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother Oncol. 2013 Dec;109(3):394–7.

22. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Technical Note: Diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method. Med Phys. 2017 Aug;44(8):4188–93.

23. 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(2):630–6.

24. 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 Oct 17;34(11):4223–35.

25. 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 Sep;90(1):S536.

26. Aristophanous M, Pelizzari CA. The evaluation of a highly automated mixture model based technique for PET tumor volume segmentation. In: Reinhardt JM, Pluim JPW, editors. 2008. p. 69141M.

27. Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. Phys Med Biol. 2009 May 7;54(9):2863–73.

28. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. Acad Radiol. 2004 Feb;11(2):178–89.

29. Loo SW, Martin WMC, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: a study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. Br J Radiol. 2012 Aug;85(1016):1070–7.

30. Thiagarajan A, Caria N, Schöder H, Iyer NG, Wolden S, Wong RJ, et al. Target Volume Delineation in Oropharyngeal Cancer: Impact of PET, MRI, and Physical Examination. Int J Radiat Oncol. 2012 May;83(1):220–7.

31. Geets X, Daisne J-F, Arcangeli S, Coche E, Poel M De, Duprez T, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. Radiother Oncol. 2005 Oct;77(1):25–31.

32. Mukesh M, Benson R, Jena R, Hoole A, Roques T, Scrase C, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? Br J Radiol. 2012 Aug;85(1016):e530–6.

33. Houweling AC, Wolf AL, Vogel W V, Hamming-Vrieze O, van Vliet-Vroegindeweij C, van de Kamer JB, et al. FDG-PET and diffusion-weighted MRI in head-and-neck cancer patients: implications for dose painting. Radiother Oncol. 2013 Feb;106(2):250–4.

34. Schouten CS, Hakim S, Boellaard R, Bloemena E, Doornaert PA, Witte BI, et al. Interaction of quantitative (18) F-FDG-PET-CT imaging parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. Head Neck. 2014 Oct 29 [cited 2016 Jan 20];

35. Schouten CS, de Graaf P, Alberts FM, Hoekstra OS, Comans EFI, Bloemena E, et al. Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT. Oral Oncol. 2015 May;51(5):541–7.

36. Surov A, Stumpp P, Meyer HJ, Gawlitza M, Höhn A-K, Boehm A, et al. Simultaneous 18F-FDG-PET/MRI: Associations between diffusion, glucose metabolism and histopathological parameters in patients with head and neck squamous cell carcinoma. Oral Oncol. 2016 Jul;58:14–20.

37. Varoquaux A, Rager O, Lovblad K-O, Masterson K, Dulguerov P, Ratib O, et al. Functional imaging of head and neck squamous cell carcinoma with diffusion-weighted MRI and FDG PET/CT: quantitative analysis of ADC and SUV. Eur J Nucl Med Mol Imaging. 2013 Jun 22;40(6):842–52.

38. Fruehwald-Pallamar J, Czerny C, Mayerhoefer ME, Halpern BS, Eder-Czembirek C, Brunner M, et al. Functional imaging in head and neck squamous cell carcinoma: correlation of PET/CT and diffusion-weighted imaging at 3 Tesla. Eur J Nucl Med Mol Imaging. 2011 Jun 5;38(6):1009–19.

39. Subesinghe M, Scarsbrook AF, Sourbron S, Wilson DJ, McDermott G, Speight R, et al. Alterations in anatomic and functional imaging parameters with repeated FDG PET-CT and MRI during radiotherapy for head and neck cancer: a pilot study. BMC Cancer. 2015;15:137.

40. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988 Aug;168(2):497–505.

Chapter 4

The added value of apparent diffusion coefficient to clinical T-stage as a prognostic factor for local recurrence in patients with head and neck squamous cell carcinoma treated with (chemo)radiotherapy

Boris Peltenburg, Juliette P. Driessen, Jeanine E. Vasmel, Frank A. Pameijer, Luuk M. Janssen, Chris H.J. Terhaard, Remco de Bree, Marielle E.P. Philippens

Abstract

Objectives

For patients with head and neck squamous cell carcinoma (HNSCC), pretreatment identification of radio-insensitivity of tumors would affect treatment planning. The apparent diffusion coefficient (ADC) of a tumor has been reported to be a predictor of local recurrence. However, this has rarely been examined in combination with known prognostic factors such as clinical T-stage.

The aim of this study is to determine the added value of pretreatment ADC to clinical T-stage as a prognostic factor for local recurrence.

Methods

This retrospective cohort study included 217 patients with HNSCC treated with (chemo) radiotherapy. Included patients were treated between April 2009 and December 2015. All patients underwent diffusion weighted MRI prior to treatment. Median ADC values of all tumors were obtained using a semi-automatic delineation method. A sensitivity analysis was performed in 188 patients using a reduced delineated volume. Univariate models containing ADC and T-stage were compared to a multivariable model containing both variables.

Results

Univariate analysis showed no significant association between tumor ADC and local recurrence within three years after (chemo)radiotherapy (p=0.09). However, ADC and local recurrence were associated when analyzed in the sensitivity analysis (p=0.02). Cox regression showed that clinical T-stage was an independent predictor of local recurrence and adding ADC to the model did not increase its performance.

Conclusion

Pretreatment ADC values have no added value as a prognostic factor for local recurrence using a model including clinical T-stage.

Introduction

Patients with head and neck squamous cell carcinomas (HNSCC) can be primarily treated with (chemo)radiotherapy ((C)RT) or surgery. In advanced stages, a combination of these modalities is used. If an organ-sparing treatment such as (C)RT is considered, surgery is kept in reserve for salvage treatment should the primary (C)RT fail. However, effects of the (C)RT on the tumor and healthy tissue makes salvage surgery challenging. Reliable pretreatment identification of radio-insensitive tumors could enable patient selection for primary surgery, avoiding the surgical challenges and morbidity induced by previous irradiation [1].

Quantitative analysis of diffusion-weighted MRI (DW-MRI) is one of the methods to identify patients who will not respond to (C)RT [2]. Several studies investigated the correlation of pretreatment apparent diffusion coefficient (ADC) and local tumor recurrence after (C)RT and found contradicting results [3–9]. These, mostly single center, studies reported on a very limited number of patients and used a wide variety of methods to delineate the tumor and calculate pretreatment ADC. Most importantly, many of these studies did not account for clinical factors such as T-stage, which is known to be an important prognostic factor for local recurrence [10, 11]. This resulted in mixed conclusions, although many studies concluded that tumors with a relatively high pretreatment ADC have a higher chance of local recurrence [3–5]. In our center, MRI with diffusion-weighted imaging has been incorporated in the diagnostic work-up of head and neck patients for many years, resulting in a wide availability of DW-MRI for retrospective analyses. To determine the clinical relevance of ADC as a prognostic variable, it needs to be evaluated in relation to established important prognostic factors such as clinical T-stage. Therefore, the question this research tries to answer is: What is the added value of ADC to clinical T-stage as a prognostic factor for local recurrence of HNSCC?

Methods and materials

This retrospective study was approved by the institutional review board and the need for informed consent was waived. Reporting was done in accordance to the STROBE statement [12].

Study population
This retrospective cohort study included consecutive patients from a single tertiary care hospital with biopsy proven primary HNSCC of the oropharynx, hypopharynx, larynx or oral cavity, who were treated with (chemo)radiotherapy with curative intent between

April 2009 and December 2015. Patients were included if an MRI with diffusion-weighted sequences acquired within 2.5 months prior to the start of treatment was available.

After the diagnosis of HNSCC, all patients were discussed in a tumor board meeting to determine the stage of the tumor and the suggested treatment. The clinical T-stage was extracted from the report of this meeting for use in our study. Staging of patients was done according to the seventh edition of the American Joint Commission on Cancer [13]. Patients with T1 tumors were excluded due to reduced visibility of these tumors on MRI. Clinical charts were retrospectively reviewed for the baseline characteristics of the study population (table 1).

Table 1. Baseline patient characteristics

Variable	N	%
Age in years	63 (40-87)*	
Sex		
Female	55	25
Male	162	75
Tumor site		
Larynx	69	32
Hypopharynx	35	16
Oropharynx	102	47
Oral cavity	11	5
AJCC tumor stage[a]		
T2	81	37
T3	71	33
T4a	53	24
T4b	12	6
Nodal stage		
N0	95	44
N1	26	12
N2a	2	1
N2b	53	24
N2c	41	19
HPV status oropharyngeal tumors		
Positive	18	18
Negative	62	61
Unknown	22	21
Days between MRI and start treatment*	18 (1-63)	
Treatment		
Radiotherapy only	120	55
Radiotherapy + Cisplatin	66	31
Radiotherapy + Cetuximab	31	14
Months of followup*	34 (2 -102)	

* Median (range), [a] AJCC: American Joint Committee on Cancer 7th edition.

Treatment protocol

Treatment consisted of 5-7 weeks of radiotherapy with an effective dose of 70 Gy to the primary tumor. Patients with an indication for concomitant chemotherapy were administered 100 mg cisplatin/m^2 at day 1, 22 and 43. If chemotherapy with cisplatin was not possible due to severe comorbidities, patients received cetuximab instead of cisplatin. After completion of the treatment, the follow-up consisted of visits to the multidisciplinary outpatient clinic according to national guidelines; every 2 months in the first year, every 3 months in the second year after treatment, every 4 months in the third year and every 6 months in the fourth and fifth year after treatment. These consultations contained physical examination and fiberoptic endoscopy. Additional imaging was performed on indication, for example when there were complaints or clinical findings suspect for local or regional recurrence.

Local recurrence

Local recurrence was defined as a biopsy proven squamous cell carcinoma of the primary tumor region within three years after completion of the (chemo)radiotherapy treatment. Additionally, patients clinically suspected of local recurrence within three years after treatment who did not have a (conclusive) biopsy were considered to have a local recurrence if the suspected recurrence progressed between follow up visits or if the patient died while the diagnostic work-up of the suspected recurrence was still ongoing. The time to recurrence was defined as the time between the last day of therapy and the first day of suspicion of recurrence as recorded in the subject's chart.

Subjects who did not meet above criteria and who completed three years of follow up were considered to have had no local recurrence. Patients who were lost to follow up without a local recurrence as previously defined were considered to have local control for the period between end of treatment and their last follow up visit.

MRI and diffusion-weighted MRI

All patients underwent pretreatment MRI with diffusion-weighted imaging on a 1.5 Tesla Philips Intera or 3.0 Tesla Philips Ingenia MRI scanner. The obtained images were used for target and organs-at-risk delineation for radiotherapy treatment planning. The MRIs were acquired routinely over a period in which new protocols were developed (Table S1, in supplemental material). Available b-values differed across scans with all containing at least a high (b800 or b1000 s/mm^2) and a low (b0 s/mm^2) b-value. ADC maps were calculated from the DW images using all available b-values and a mono-exponential model.

Tumor delineation

The tumor delineation procedure was semi-automatic. The endoscopy report and available pretreatment imaging was used to determine the location of the tumor. First, a seed point was placed in the tumor on the axial DW-MRI with the highest available b-value (b800 or b1000 s/mm^2). Using the seed point and a threshold of 50% of the maximum signal intensity the tumor was segmented. Secondly, this segmentation was transferred to the corresponding ADC map. The contrast of the ADC map was set to window: 1500 x 10^{-6} mm^2/s and level: 1000 x 10^{-6} mm^2/s. Areas with relatively high ADC values at the outer-contour of the segmentation were considered not to be part of the tumor and were manually removed from the volume of interest by a single observer blinded to the local recurrence status of subjects. High or very low ADC values inside the tumor were not excluded to maintain tumor heterogeneity. This final segmentation was used to determine tumor volume and ADC values. The ADC value of an individual tumor was defined as the median ADC value of all voxels included in the segmentation.

ADC and HPV

A possible relationship exists between human papilloma virus (HPV), ADC and local recurrence in oropharyngeal carcinoma [14–17]. In order to test if this relationship had any influence on the final results ADC values of the patients with HPV positive oropharyngeal tumors were compared to those with HPV negative oropharyngeal tumors.

Sensitivity analysis

In order to test the robustness of the results and their susceptibility to variation in delineation a sensitivity analysis was performed. In this analysis, we reduced the segmentations by 1 mm in x, y and z directions. Theoretically, this reduced volume should contain a higher percentage of tumor and less of the edges of the tumor and surrounding tissues, reducing partial volume effects. This reduced segmentation was used to recalculate tumor ADC and to repeat the rest of the applied analyses.

Statistical analysis

Differences in ADC values between patients with local control and local recurrence were determined using an independent samples t-test. The most discriminating ADC cutoff value (i.e. the value with the highest combined sensitivity and specificity) was extracted from the receiver operating characteristic (ROC) curve of ADC and local recurrence. Local disease-free survival was visualized by Kaplan-Meier curves using the Log-rank test to determine significant differences between the high and low ADC groups. Finally, two models were created; a Cox regression model containing T-stage as a categorical predictor and a Cox regression model containing both ADC (continuous) and T-stage. The area under the ROC curve (AUC) of both models, using probabilities for local control

at 3 years, was calculated to determine the added value of ADC. P-values smaller than 0.05 were considered statistically significant. Statistical analyses were performed with SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp).

Results

Patients
A total of 295 patients met the inclusion criteria. Final analysis included 217 patients. Reasons for exclusion were: failed scans due to technical errors, artifacts interfering with the depiction of the tumor and the inability to distinguish the tumor from the surrounding tissue. A diagram showing the flow of patients is provided in the supplemental material (Fig S1).

Median time between the MRI and the start of treatment was 18 days (range 1 – 63 days). Median total follow-up time after end of treatment was 34 months (range 2 – 102 months). Fourteen patients (6.5%) were lost to follow-up within three years. Fifty-eight patients (27%) developed a local recurrence within three years. See table 1 for baseline characteristics. See Fig 1 for an example of the DW-MRI of one of the patients.

Predictive value of ADC and T-stage for local control
No significant difference in tumor ADC between the groups with local recurrence and those with local control three years after (C)RT was found: The mean of median ADC values ± SD was $(1.01 ± 0.18)$ 10^{-3} mm^2/s in the group with local recurrence and $(1.00 ± 0.24)$ 10^{-3} mm^2/s (p=0.72) in the group without local recurrence (Fig 2).
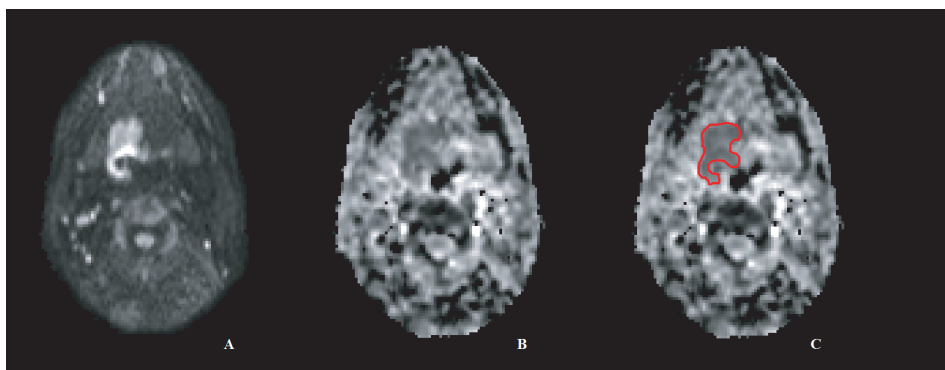

Figure 1. Images of a patient with a T4aN2c oropharyngeal carcinoma. DW-MRI acquired in 2011. A: DW-MRI b800 s/mm^2 image, B: corresponding ADC map, C: corresponding ADC map with the contour used to determine the median ADC value of the tumor.
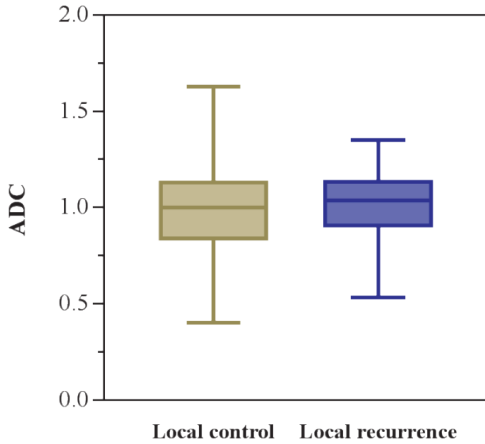
Figure 2. ADC values in 10^{-3} mm^2/s of patients with local control and local recurrence. The box depicts the percentile range of p25-p75 and whiskers depict the range of ADC values.

The most discriminating cutoff value of ADC was 0.90 10^{-3} mm^2/s. The group of patients with an ADC value higher than this value had no significant difference in the rate of local recurrence compared to the group with ADC values lower than this value (p=0.09) (Fig 3).

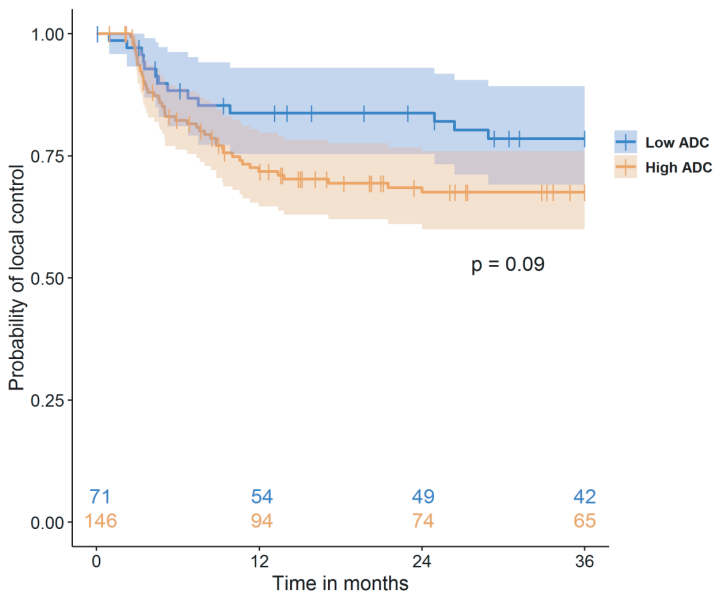
Figure 3. Survival curve with CI95% showing local disease free survival of groups with ADC values higher or lower than 0.90 · 10^{-3} mm^2/s. The number at risk for each group are displayed in the lower portion of the graph.

The rate of local recurrence was different for each T-stage; patients with T2 or T3 tumors had a significant lower rate of tumor recurrence than the patients with T4a or T4b tumors (p<0.01) (Fig 4).
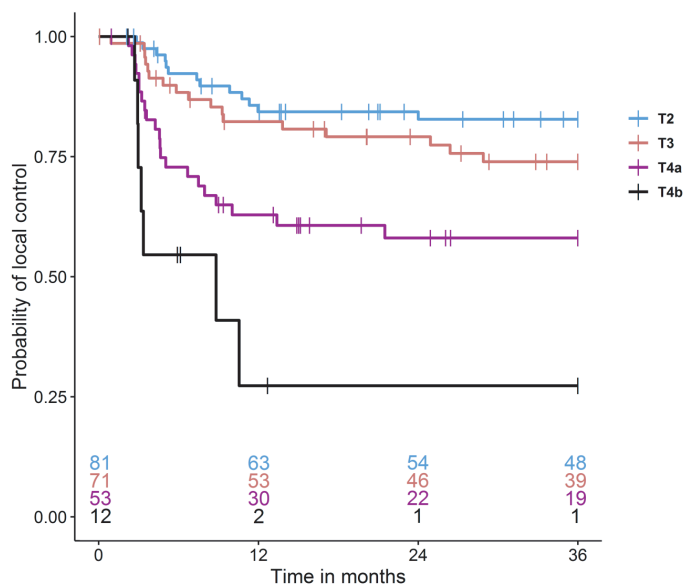

Figure 4. Survival curve showing local disease free survival by T-stage. The number at risk for each group are displayed in the lower portion of the graph.

Cox regression showed that T-stage was associated with local recurrence (table S2). The accuracy of the model expressed as the AUC was 0.66 (CI95% 0.57-0.74). In the multivariable Cox regression model containing both T-stage and ADC values, ADC was not independently associated with local recurrence (table S3). The AUC of this model was 0.66 (CI95% 0.58-0.74) while the AUC based on solely ADC was 0.53 (CI95% 0.45-0.62) (Fig 5).

HPV and ADC

Of the 102 patients with oropharyngeal carcinoma, 18 had HPV positive tumors and 62 were negative for HPV. The HPV status of the 22 remaining patients was unknown. HPV positive tumors had a significantly lower ADC value compared to HPV negative tumors, $0.81 \cdot 10^{-3}$ mm^2/s compared to $0.97 \cdot 10^{-3}$ mm^2/s (p<0.01), respectively. No local recurrences were detected in the HPV positive group and 18 recurrences occurred in the HPV negative group. Removing the 18 HPV positive patients from the total sample of 217 patients and retesting the predictive value of ADC and volume did not change the final conclusions.
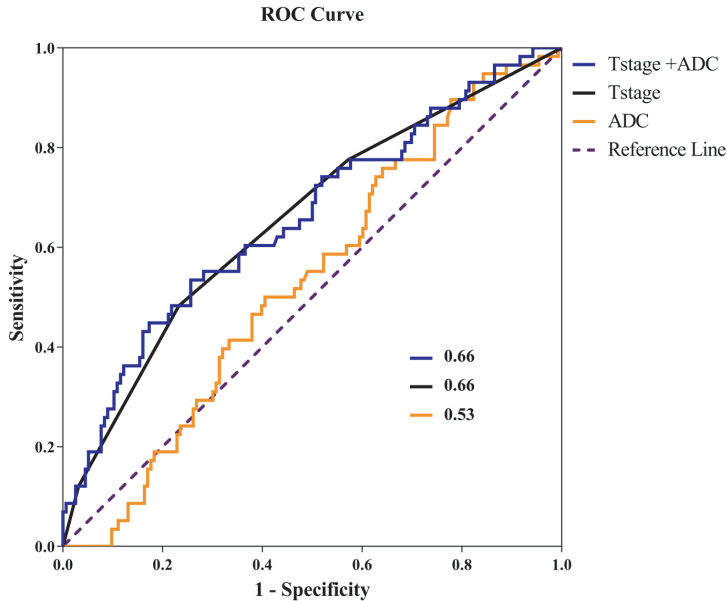
ROC Curve

Figure 5. ROC curve displaying the performance of the models containing T-stage and ADC as individual parameters (black and yellow) and the performance of the model containing both variables (blue).

Sensitivity analysis

After a reduction of all segmentations by 1 mm in all directions, new ADC values were determined. In 29 subjects with initially small tumors, the reduced segmentations were too small to determine ADC. The mean of the ADC values ±SD for the 188 remaining patients was (0.97 ± 0.20) 10^{-3} mm^2/s. Prior to the reduction the mean ADC ±SD was (1.01 ± 0.21) 10^{-3} mm^2/s for these 188 patients. Correlation between the ADC values of both analyses was high (r=0.94). The most discriminating cutoff value of ADC changed to $0.85 \cdot 10^{-3}$ mm^2/s from $0.90 \cdot 10^{-3}$ mm^2/s. Using this new cutoff value in this subgroup of patients resulted in a significant difference in the rate of local recurrence between both groups, with patients in the high ADC group having the highest rate of recurrence (p=0.02) (Fig 6). Cox regression analysis results, using these new ADC values, remained the same as before with T-stage being the only independent factor associated with local recurrence. For these 188 patients, the model with only T-stage showed an AUC of 0.66 (CI95% 0.57 – 0.75) while the model with T-stage and ADC had an AUC of 0.67 (CI95% 0.59 – 0.76) (table S4).
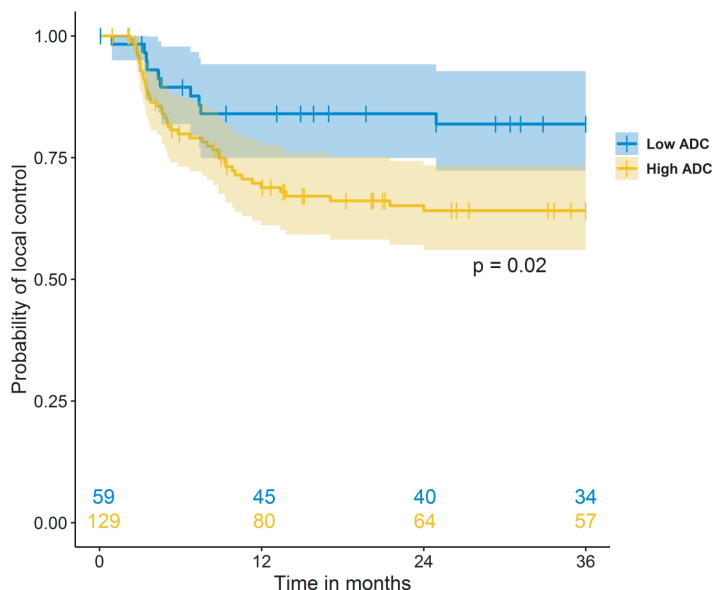
Figure 6. Survival curve with CI95% showing local disease free survival using ADC values obtained in the sensitivity analysis. The ADC cutoff value separating both groups is $0.85 \cdot 10^{-3}$ mm^2/s.

Discussion

In this retrospective cohort study, we found that pretreatment ADC has no added value to clinical T-stage as a prognostic factor for local recurrence of HNSCC after (chemo) radiotherapy. Furthermore, in our primary analysis median ADC is not correlated with local recurrence within 3 years. Clinical T-stage however, is correlated with local recurrence, with higher stage tumors showing a higher rate of local recurrence.

The absence of an association between ADC and local recurrence did not confirm the results of previous studies, which concluded that ADC was able to predict local recurrence after (C)RT [3–5]. However, it is in line with other research which reported that local recurrence could not be predicted by pretreatment ADC [7–9]. Reports with mixed conclusion on the value of pretreatment ADC measurements exist as well [6, 18].

The conflicting results reported in previous work might be due to the different tumor delineation methods. Most studies calculated ADC on a single axial slice of the MRI [6,7]. ADC values change depending on the delineation method used, this might be due to the heterogeneity of HNSCC [19]. We showed the effect of different delineation methods by removing 1 mm isotropically of the original delineation. This had two effects; first, some tumors were too small to analyze after shrinkage and had to be excluded from

analysis; secondly, ADC values of all tumors changed due to the new delineations. It suggests that a hard cutoff value for ADC, reported by previous studies, might not be generalizable to other study populations unless delineation protocols are reproducible and study populations are interchangeable. For this study, we used a semi-automatic method for segmentation of the primary tumor in order to increase reproducibility, but some subjective decisions concerning the delineation are unavoidable.

Another point of difference between our study and previous work is that predominantly, these earlier studies do not take known prognostic factors such as tumor stage or tumor volume into account. This step is important as it provides clinical relevance of the added prognostic value of ADC. Clinical T-stage is arguably more easily obtained than tumor ADC and is already widely used in daily practice to determine prognosis and treatment strategies for individual patients. ADC should significantly add to this existing model in order to justify its use as a prognostic factor for local recurrence.

One study that also found pretreatment ADC to be predictive of local recurrence is the study of Lambrecht et al. [18]. This study included 161 patients and performed multivariable multivariate analysis including, amongst others, tumor ADC and tumor volume. Additionally, similar to our methodology, they created two models, one with ADC included and one without ADC and compared the performance of the models. They report an AUC, used to determine the discriminatory capacity of the first model, of 0.62 (CI 0.56 – 0.70), while for the model without ADC the AUC is 0.60 (CI 0.55 – 0.67). These result are very similar to our findings and it supports our conclusion that ADC has no added value as a prognostic factor for local recurrence compared to, more easily obtainable, clinical parameters.

Finally, all of these studies, except for Lambrecht et al. [18], suffered from small sample sizes. The second largest sample size was obtained by King et al. [7] who investigated 50 patients. One of the strengths of our study is the large sample size of 217 subjects and a total of 58 local recurrence.

Although the biological origin of ADC remains partly unknown, it is related to cell density, stromal components of the tumor and infiltration of inflammatory cells and also correlates to HPV status [15, 20–22]. High stromal proliferation is associated with poor outcome, and also with high ADC [23, 24]. HPV positive tumors, which are known for their prognostic favorable outcome compared to their HPV negative counterparts, have significantly lower ADC compared to HPV negative tumors [15–17]. This was confirmed in our study and suggests that ADC could be able to serve as a predictor of outcome but only as a surrogate marker for HPV status. This might have biased studies which showed a predictive value for ADC but failed to correct for HPV status.

An interesting alternative to pretreatment ADC as a predictor of local recurrence might be the change in ADC (ΔADC) between pretreatment ADC and ADC values obtained during treatment. Some studies report that this ΔADC has prognostic power, theorizing that regardless of pretreatment ADC values, different tumors respond differently to (C)RT and that this effect can only be determined after treatment has started [7, 25].

We used clinical T-stage as a variable to compare ADC to. This study confirms that this seemingly straightforward model, based on clinical observations, is able to differentiate patients at risk for local recurrence from patients with local control reasonably well.

Limitations

Our study had some limitations. We excluded many patients due to insufficient visibility of the tumor on DW-MRI. Furthermore, the use of different DW-MRI protocols might influence our measurement of ADC values [26]. In our retrospective study, the DW-MRI protocols changed over the years. This might result in different ADC values for similar tumors, possibly obscuring the prognostic value of ADC. Nonetheless, the use of different DW-MRI protocols results in a model applicable to clinical practice. Finally, the obtained results were not validated in an external cohort, yet the similarity in results between our study and that of the only other large sample size study supports our findings [18].

Conclusion

In this retrospective cohort study of 217 patients, we found pretreatment ADC to have no added value to clinical T-stage as a prognostic factor for local recurrences within three years after (chemo)radiotherapy for HNSCC.

Acknowledgements

Supplemental material
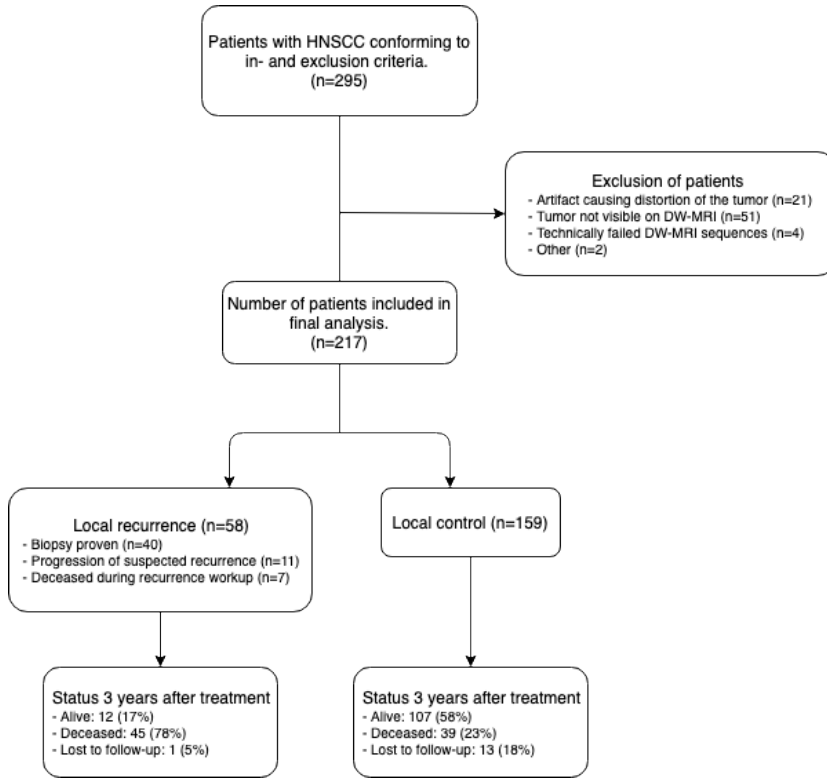
Figure S1 Flowchart

Table S1. MRI protocol detail

Sequence	Echo Time (ms)	Repetition Time (ms)	Resolution (mm)		
			x	y	z
T1	10 - 120	600 - 800	1	1	2
T2	100 - 220	2300 - 4200	0.5 - 1	1	2 - 5
DW-MRI	68 - 70	3700 - 5900	1.4 - 2.2	1.4 - 2.5	3 - 4

Table S2. Cox regression. Effect of T-stage local recurrence.
T stage according to American Joint Committee on Cancer 7th edition.

Variable	Beta coefficient	Standard error	p-value	Hazard ratio	Hazard ratio CI95%
T2 (reference)	..	..	..	..	..
T3	0.44	0.37	0.24	1.55	0.75 – 3.18
T4a	1.20	0.35	<0.01	3.31	1.67 – 6.57
T4b	2.05	0.47	<0.01	7.78	3.08 – 19.7

Table S3. Cox regression. Effect of T-stage and ADC on local recurrence.
ADC in 10^{-3} s/mm^2. T stage according to American Joint Committee on Cancer 7th edition.

Variable	Beta coefficient	Standard error	p-value	Hazard ratio	Hazard ratio CI95%
ADC	0.26	0.63	0.69	1.29	0.37 – 4.47
T2 (reference)	..	..	..	..	..
T3	0.43	0.37	0.24	1.54	0.75 – 3.18
T4a	1.13	0.35	<0.01	3.11	1.56 – 6.22
T4b	2.06	0.47	<0.01	7.88	3.11 – 19.9

Table S4. Cox regression using reduced margins (n=188). Effect of T-stage and ADC on local recurrence.
ADC in 10^{-3} s/mm^2. T stage according to American Joint Committee on Cancer 7th edition.

Variable	Beta coefficient	Standard error	p-value	Hazard ratio	Hazard ratio CI95%
ADC	0.63	0.70	0.37	1.88	0.47 – 7.50
T2 (reference)	..	..	..	..	..
T3	0.51	0.41	0.21	1.67	0.75 – 3.73
T4a	1.14	0.39	<0.01	3.12	1.47 – 6.63
T4b	2.05	0.50	<0.01	7.80	2.94 – 20.7

References

1. Lee SC, Shores CG, Weissler MC (2008) Salvage surgery after failed primary concomitant chemoradiation. Curr Opin Otolaryngol Head Neck Surg 16:135–40 . doi: 10.1097/MOO.0b013e3282f495b6

2. Driessen JP, van Kempen PMW, van der Heijden GJ, et al (2015) Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. Head Neck 37:440–448 . doi: 10.1002/hed.23575

3. Hatakenaka M, Nakamura K, Yabuuchi H, et al (2011) Pretreatment apparent diffusion coefficient of the primary lesion correlates with local failure in head-and-neck cancer treated with chemoradiotherapy or radiotherapy. Int J Radiat Oncol Biol Phys 81:339–345 . doi: 10.1016/j.ijrobp.2010.05.051

4. Hatakenaka M, Shioyama Y, Nakamura K, et al (2011) Apparent diffusion coefficient calculated with relatively high b-values correlates with local failure of head and neck squamous cell carcinoma treated with radiotherapy. AJNR Am J Neuroradiol 32:1904–1910 . doi: 10.3174/ajnr.A2610

5. Ohnishi K, Shioyama Y, Hatakenaka M, et al (2011) Prediction of local failures with a combination of pretreatment tumor volume and apparent diffusion coefficient in patients treated with definitive radiotherapy for hypopharyngeal or oropharyngeal squamous cell carcinoma. J Radiat Res 52:522–530

6. Lombardi M, Cascone T, Guenzi E, et al (2017) Predictive value of pre-treatment apparent diffusion coefficient (ADC) in radio-chemiotherapy treated head and neck squamous cell carcinoma. Radiol Med 122:345–352 . doi: 10.1007/s11547-017-0733-y

7. King AD, Mo FKF, Yu K-H, et al (2010) Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol 20:2213–2220 . doi: 10.1007/s00330-010-1769-8

8. Chawla S, Kim S, Dougherty L, et al (2013) Pretreatment diffusion-weighted and dynamic contrast-enhanced MRI for prediction of local treatment response in squamous cell carcinomas of the head and neck. AJR Am J Roentgenol 200:35–43 . doi: 10.2214/AJR.12.9432

9. King AD, Chow K-K, Yu K-H, et al (2013) Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology 266:531–538 . doi: 10.1148/radiol.12120167

10. Brockstein B, Haraf DJ, Rademaker AW, et al (2004) Patterns of failure, prognostic factors and survival in locoregionally advanced head and neck cancer treated with concomitant chemoradiotherapy: A 9-year, 337-patient, multi-institutional experience. Ann Oncol 15:1179–1186

11. Brandwein-Gensler M, Teixeira MS, Lewis CM, et al (2005) Oral squamous cell carcinoma: histologic risk assessment, but not margin status, is strongly predictive of local disease-free and overall survival. Am J Surg Pathol 29:167–78

12. Vandenbroucke JP, von Elm E, Altman DG, et al (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 4:e297 . doi: 10.1371/journal.pmed.0040297

13. Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17:1471–4 . doi: 10.1245/s10434-010-0985-4

14. Ravanelli M, Grammatica A, Tononcelli E, et al (2018) Correlation between Human Papillomavirus Status and Quantitative MR Imaging Parameters including Diffusion-Weighted Imaging and Texture Features in Oropharyngeal Carcinoma. AJNR Am J Neuroradiol 39:1878–1883 . doi: 10.3174/ajnr.A5792

15. Driessen JP, van Bemmel AJM, van Kempen PMW, et al (2015) Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. Head Neck 38 Suppl 1:E613-8 . doi: 10.1002/hed.24051

16. Kimple RJ, Smith MA, Blitzer GC, et al (2013) Enhanced radiation sensitivity in HPV-positive head and neck cancer. Cancer Res 73:4791–800 . doi: 10.1158/0008-5472.CAN-13-0587

17. Isayeva T, Li Y, Maswahu D, Brandwein-Gensler M (2012) Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review. Head Neck Pathol 6 Suppl 1:S104-20 . doi: 10.1007/s12105-012-0368-1

18. Lambrecht M, Van Calster B, Vandecaveye V, et al (2014) Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiother Oncol 110:429–434 . doi: 10.1016/j.radonc.2014.01.004

19. Perrot XT De, Lenoir X V, Ayllo XMD, et al (2017) Apparent Diffusion Coefficient Histograms of Human Papillomavirus – Positive and Human Papillomavirus – Negative Head and Neck Squamous Cell Carcinoma : Assessment of Tumor Heterogeneity and Comparison with Histopathology. 2153–2160

20. Driessen JP, van Bemmel AJM, van Kempen PMW, et al (2016) Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. Head Neck 38:613–8 . doi: 10.1002/hed.24051

21. Sun Y, Tong T, Cai S, et al (2014) Apparent Diffusion Coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer. PLoS One 9:e109371 . doi: 10.1371/journal.pone.0109371

22. Choi SY, Chang Y-W, Park HJ, et al (2012) Correlation of the apparent diffusion coefficiency values on diffusion-weighted imaging with prognostic factors for breast cancer. Br J Radiol 85:e474-9 . doi: 10.1259/bjr/79381464

23. de Kruijf EM, van Nes JGH, van de Velde CJH, et al (2011) Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. Breast Cancer Res Treat 125:687–96 . doi: 10.1007/s10549-010-0855-6

24. Wang K, Ma W, Wang J, et al (2012) Tumor-stroma ratio is an independent predictor for survival in esophageal squamous cell carcinoma. J Thorac Oncol 7:1457–61 . doi: 10.1097/JTO.0b013e318260dfe8

25. Vandecaveye V, Dirix P, De Keyzer F, et al (2010) Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol 20:1703–1714 . doi: 10.1007/s00330-010-1734-6

26. Kolff-Gart AS, Pouwels PJW, Noij DP, et al (2015) Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. AJNR Am J Neuroradiol 36:384–90 . doi: 10.3174/ajnr.A4114

Part 2

Imaging during therapy

Chapter 5

Predictive value of diffusion weighted MRI in head and neck cancer patients during (chemo) radiotherapy (PREDICT): a single center observational study

Boris Peltenburg, Tim Schakel, Patricia Doornaert, Frank A. Pameijer, Niels Raaijmakers, Marielle E.P. Philippens, Chris H.J. Terhaard, Remco de Bree

Abstract

Background

Head and neck squamous cell carcinomas (HNSCC) could benefit from a predictive imaging biomarker such as the apparent diffusion coefficient obtained through diffusion weighted MRI. If the outcome of the radiation treatment can be predicted at an early stage, adaptive treatment strategies or a switch of treatment modality may be organized.

Methods/design

The PREDICT study is a prospective observational study. Diffusion weighted MR images of 100 HNSCC patients will be acquired in the second, third, fourth and fifth week of the radiation treatment. ADC measurements will be obtained from the different scans. Follow up will consist of two years of clinical follow up after the treatment has ended. Any locoregional recurrences will be recorded and compared on a group level to the change in ADC during treatment.

Discussion

Some evidence exists suggesting that ADC has predictive value for HNSCC patients during radiotherapy. However, sample sizes are small and focus on subgroups of patients. The PREDICT study once completed will provide a large patient cohort with long term follow up. This will allow us to determine if ADC can be used as a stratifying tool after radiotherapy has already begun.

Background

Head and neck squamous cell carcinomas (HNSCC) are malignant tumors of the upper aerodigestive tract including the oral cavity, pharynx, larynx, nasal cavity and paranasal sinuses. Surgical resection has been a valid method for the curative treatment of HSNCC for many years but often comes at the cost of functional and cosmetic morbidity, especially in the case of locally advanced tumors. Nowadays, in an attempt to decrease the morbidity, non-surgical treatments are increasingly applied with comparable complete remission rates. Treatment strategies include intensified radiotherapy schemes (accelerated or hyperfractionated) and combinations of chemotherapy and radiotherapy. However, treatment with (chemo)radiotherapy ((C)RT) is associated with acute and long-term side effects, leading to compromised quality of life. Common complaints are xerostomia (mouth dryness), impairment in speech and difficulties in eating and swallowing. Fibrosis of the neck with reduced mobility and pain is also not unusual.

Although (C)RT is often an effective treatment for HNSCC, loco–regional recurrence rates can be as high as 55% after (C)RT [1, 2], depending on several factors such as tumor stage, tumor location and HPV status. For some patients with functionally irresectable HNSCC (resectable but high morbidity of surgical treatment expected), a non-surgical treatment with salvage surgery for eventual residual or recurrent disease in reserve is preferred. However, the complication rate of salvage surgery after (C)RT is high, with wound healing problems as a well-known complication in irradiated patients [3, 4].

Because salvage treatment after (C)RT has a questionable prognosis and a high incidence of complications, (C)RT may not be the choice of treatment in all patients with advanced HNSCC. A reliable predictor for outcome after (C)RT is needed to select patients with resectable tumors who are unlikely to benefit from primary (C)RT treatment [5].

Early identification of non-responders to (C)RT would allow for treatment intensification or would justify a switch to surgical treatment with consequently a lower risk of complications caused by post-radiotherapy effects.

Imaging biomarkers can be determined prior to treatment, thus functioning in essence as a prognostic marker, much the same as T-stage [6]. Although both pretreatment DWI- and ^{18}F-FDG-PET/CT-parameters appeared to have predictive value for treatment failure [7–9], room for improvement remains. In order to predict individual treatment outcomes of (C)RT, quantifying imaging biomarkers during treatment might be a more promising strategy [10, 11]. A recent systematic review showed that early tumoral changes from (C)RT can be captured by functional MRI and ^{18}F-FDG-PET/CT and could allow for personalized treatment adaptation [12].

5

By obtaining the biomarkers after the start of treatment, subclinical alterations in tumor physiology and biochemistry are already ongoing. This might happen long before changes in tumor volume can be visualized [13]. It is possible that effective (C)RT exhibits many more of these changes expressed by the biomarkers than ineffective treatment. The sooner these changes can be measured the better, as the early treatment phase corresponds to a pivotal moment in the treatment of HNSCC patients. At this point, a switch from (C)RT to surgery is still viable and side-effects associated with irradiation are at a minimum.

PET-CT with the use of FDG or other tracers shows some promise as a modality capable of predicting treatment response, but it has some limitations [14, 15]. For example, inflammation induced by the (C)RT surrounding the tumor hinders accurate measurement of tracer uptake and tumor volume [16].

Another imaging modality which is reported to have predictive value, is diffusion weighted magnetic resonance imaging (DW-MRI). DW-MRI characterizes tissue based on differences in tissue water mobility. Diffusion weighted imaging is an attractive technique because it can be performed within a few minutes and can be easily incorporated into routine head and neck MR imaging protocols. Moreover, the costs and burden are lower than the costs of PET-CT.

Differences in tissue water mobility, as characterised by DW-MRI, can be quantified using an Apparent Diffusion Coefficient (ADC). Hypercellular tissue is characterized by a low ADC, while hypocellular tissue like necrosis or apoptosis is characterized by a high ADC. This characteristic of DWI makes differentiating tumour recurrence from inflammation or necrosis possible [17].

Because a treatment-induced loss of tumor cells increases water mobility at the microscopic level, treatment response corresponds to an increase in ADCs. Additionally, a decrease in ADC might indicate remaining tumor cells. Hypothetically, DW-MRI could thus be used as a non-invasive tool to select patients who have a high risk of not achieving or maintaining locoregional control and therefore benefit most from adaptation of treatment [18].

This supports the assumption that DW-MRI can be used for accurate response prediction. In several small studies, it was found that treatment-induced changes in ADC over time (ΔADC) are highly predictive of treatment [12, 19–24]. Complete responders typically show significantly higher ADC-increase than partial responders within the first few week of CRT. There is no consensus on the optimal time point to determine the ADC changes,

although it is suggested that the optimal timing to perform functional imaging to predict locoregional control and overall survival is 2-3 weeks after treatment initiation [12].

By evaluating the performance of MR imaging parameters, obtained before and during treatment, and combining them with clinical factors such as TNM-stage, HPV status and smoking habit, a multimodal prediction model can be developed. This model could predict a patients' individual probability of locoregional control after treatment. This could lead to a treatment that is made-to-measure for each patient, potentially improving clinical outcomes and raising quality of life. Several other studies are currently in the process of collecting MRI data and patient characteristics of HNSCC populations [25, 26]. Combining the results of these studies with the current study might enhance such a multimodal predictive model.

In conclusion, there are several studies indicating that diffusion-weighted MRI early during non-surgical treatment has prognostic value. For HNSCC, only small studies have been performed and none of these determined an intratreatment cut-off level or optimal scanning moment. A study seems warranted to explore the value of diffusion-weighted MRI for the early response evaluation during therapy in which optimal timing and the predictive value of ADC changes are assessed.

Methods

Objectives
The primary objective of the study is to determine the value of ADC during treatment in predicting local recurrence after (C)RT.

Secondary objectives are:
- To evaluate the changes in the appearance of HNSCC tumors on MRI during radiotherapy.
- To determine the optimal time point for MRI scanning during treatment in order to predict treatment outcome.
- To develop a multimodal prediction model that predicts treatment outcome by combining clinical and imaging parameters.

Study design
The PREDICT is a prospective observational single institute study of 100 consecutive HNSCC patients scheduled for (C)RT in which the predictive value of DW-MRI for locoregional control will be examined. The project is a collaboration between the

department of Radiotherapy and the department of Head and Neck Surgical Oncology of the University Medical Center Utrecht. The study was approved by the Medical Ethics Committee of the University Medical Center Utrecht (16-331, NL57164.041.16) and was registered in the Dutch Trial registry (NL5955, www.trialregister.nl). Written informed consent will be obtained from all participants.

Study population
Patients eligible for participation in the study are scheduled for primary radiotherapy with or without concomitant chemotherapy for a technically resectable HNSCC of the oral cavity, oropharynx, hypopharynx or larynx. Only patients with a curative intent of treatment are included.

Radiotherapy treatment consists of 70 Gy to the primary tumor in a conventional schedule (35x2Gy over 7 weeks) or 69.5 Gy in a slightly accelerated schedule (10x2Gy + 15x1.5Gy + 15x1.8Gy over 5 weeks). Chemotherapy, if indicated, is administered according to local protocols and consists of Cisplatin 100mg/m^2 intravenously administered at three time points: the start of week 1, week 4 and week 7 of the conventional radiotherapy program. Patients with a T1 oral cavity, oropharyngeal or hypopharyngeal are excluded from study participation, as well as patients with a T1 or T2 laryngeal carcinoma due to limited visibility of these tumors on MRI. Further exclusion of patients is based on: an age younger than 18 years, pregnancy and the existence of contraindications for MRI examination.

Patients who complete one or none of the intratreatment MRI are withdrawn from the study.

Study protocol
Patients meeting the inclusion criteria will be recruited from the Department of Head and Neck Surgical Oncology. Prior to treatment, the first MRI shall be acquired as part of standard care. Subsequently, all patients will receive radiotherapy with or without concomitant chemotherapy following standard procedure. During the radiotherapy course four additional MRI examinations will be performed. The scans will be planned for week 2, 3, 4 and 5 of the radiotherapy course.
Study follow up will be 2 years after finishing treatment. After this patients will receive 3 more years of follow up outside of the study participation. All procedures during follow up will be as per standard posttreatment HNSCC protocol. This includes a sixth MRI scan at 3 months after end of treatment for response evaluation and as a baseline MRI for follow-up.

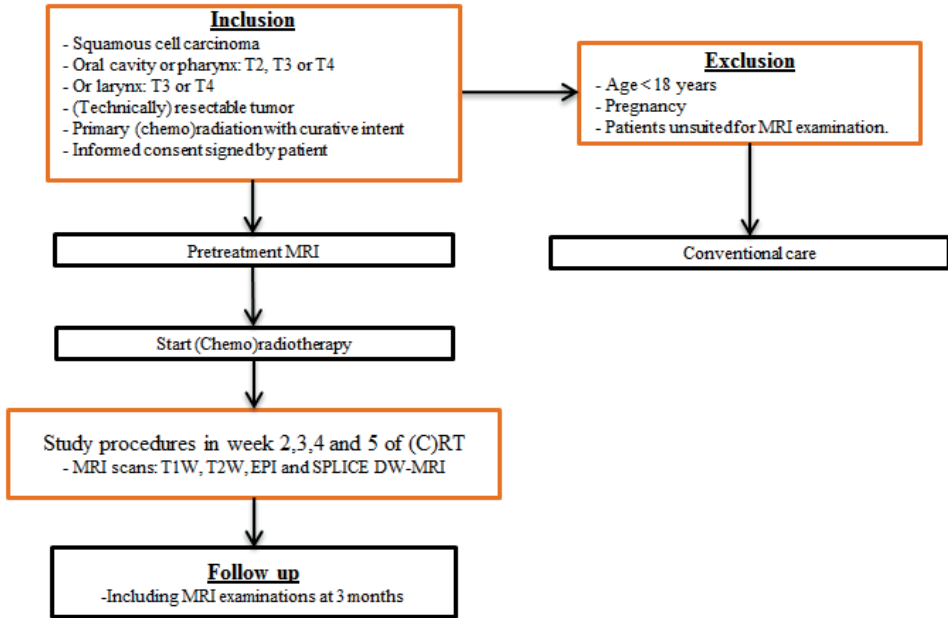The flow of patients is visualized in figure 1.

Figure 1. Flow of patients. Steps with a black outline are part of the conventional care. Steps with an orange outline are part of the PREDICT study.

MRI

In order to optimize comparability between time points, all imaging is performed on the same scanner (3.0T Philips Ingenia, Philips Medical Systems, Best) using the same scanning protocol for each acquisition. After initial calibration, two diffusion weighted sequences are acquired. A regular DW-MRI using a conventional single shot echo planar imaging (EPI) technique is obtained with b-values of 0, 50, 200 and 800 s/mm^2. The acquisition of this sequence is fast, has a high signal-to-noise ratio but suffers from geometrical distortions. The second DW-MRI is acquired with a TSE technique and b-values of 0, 200 and 800 s/mm^2 [27]. Finally, anatomical T1 and T2 weighted images both using mDixon fat separation are acquired. See Table 1 for complete imaging parameters.

Table 1. Imaging parameters of the study MRI examinations.

		TE (ms)	TR (ms)
Transversal T2 TSE mDIXON	0.94 x 0.94 x 3.00	100	9023
Transversal DWI SPLICE	1.46 x 1.46 x 5.00	74	5233
Sagittal T1 FFE cine	1.42 x 1.42 x 10.0	1.45	3.2
Transversal T1 TSE mDIXON	0.73 x 0.73 x 2.00	6.2	678
Transversal DWI EPI	2.02 x 2.02x 4.00	68	2588

Outcome assessment

Locoregional recurrence is defined as a biopsy proven squamous cell carcinoma of the primary tumor region within two years after completion of the (C)RT treatment. Additionally, in the absence of a biopsy, patients are considered to have recurrent disease if a suspected recurrence progresses between follow up visits or if the patient dies while the diagnostic workup of the suspected recurrence was still ongoing. The time to recurrence is defined as the time between the last day of therapy and the first day the clinical suspicion of recurrence is recorded in the subject's chart.

Subjects who do not meet above criteria and who complete two years of follow up are considered to have locoregional control. Patient with a surgical resection of the suspected recurrence are also scored local control if histopathological examination confirms absence of disease.

Patients who are lost to follow up without a local recurrence as previously defined are considered to have locoregional control for the period between end of treatment and their last follow up visit at which point they are censored from the study [28]. The same is true for patients who die during follow up of causes not related to the primary HNSCC. The period of locoregional control in these cases is defined as the time between of treatment and date of death.

Follow up

Follow up will consist of the standard clinical follow up procedure for HNSCC patients. The general follow up guideline in The Netherlands indicates routine follow up every 2-3 months in the first and second year and every 4-6 months in the third, fourth and fifth year. Locoregional recurrence status is assessed based on clinical examination including flexible endoscopy [29].

Final study analyses will be performed two years after treatment, since chance of recurrence is very low after two years. Local recurrence or local control status will be obtained from standard follow up procedures.

Sample size calculation

To determine the primary objective of the study, the difference of ADC values for the group *without* recurrences will be compared to the values for the group *with* recurrences. Previous studies have shown that expected ΔADC differences between both groups range from 40-60% with reported standard deviations of 31% [24, 30].

With a power of 90% and a significance of 5%, the study should at least include 15 patients who would eventually be diagnosed with a local recurrence. From previous

studies and personal experience with the study population, we expect a recurrence rate of 15% to 30%. To account for potentially high cure rates and low recurrence rates, one hundred patients will be included in order to attain a minimum of 15 included patients with a recurrence [31].

Statistical analysis

Primary analysis
Assessment of tumor and lymph node ADC will be performed after all data is acquired. Assessors will be blinded to the treatment outcome of the patients. Different ADC parameters (i.e. maximum ADC, minimum ADC, median ADC) are obtained from DW-MRI. The most valuable parameter will be determined by the area under the receiver operating curve (AUC_{ROC}). The AUC_{ROC} of the change in ADC values between time points (ΔADC_{x-y}) will be analyzed to detect the added value of using a ΔADC to predict treatment outcome as opposed to a single ADC measurement.

Secondary analysis
Functional diffusion maps will be created to identify the pattern of ADC changes in individual tumors [32].

Analyses of all ΔADCs by the AUC_{ROC} will yield information of the most optimal time point for MR imaging in order to predict treatment outcome.

The multimodal prediction model will have clinical and imaging parameters entered based on knowledge from previous studies on the subject and expert opinion. An AUC measurement of the ROC curve of the model will determine the discriminating effect of the model.

Discussion

Radiotherapy with or without chemotherapy is an effective treatment for many HNSCC, but for some patients (C)RT is not able to prevent recurrence of disease. The aim of the PREDICT study is to assess if DW-MRI acquired early during treatment can facilitate the decision to switch from ineffective (C)RT to modified treatment (e.g. early treatment (de) escalation or switch to another treatment modality) to increase effectivity and reduce unnecessary toxicity in patients with good prognosis. A timely switch is important as it prevents surgical complications due to post radiotherapy tissue changes and avoids

5

unnecessary toxicity for the patient. Moreover, this information can be used for patient counseling.

In this study, MRI examinations are performed during treatment. The effect of the (C)RT on the tumor, its cellular structure and the surrounding tissues is therefore directly determined. By using DW-MRI, evaluation of the treated area is less affected by treatment effects such as inflammation [33]. If the current study provides a reliable imaging parameter, a new standard of care for head and neck cancer patients could be developed. The responsiveness of an individual patient to (C)RT treatment might become the focal point of treatment strategy.

With the advent of hybrid MRI/radiotherapy devices, weekly MR imaging is becoming clinical practice for many radiotherapy departments [34, 35]. The additional MR exams are expected to be used for position verification, target verification and possibly for adaptive radiotherapy planning. Moreover, the additional imaging could also be used for early response assessment. It is therefore of great importance that a base of knowledge is created in order to help clinicians decide on the continuation or cessation of the treatment of individual patients. The PREDICT study hopes to contribute to this base of knowledge by systematically investigating the predictive value of DW-MRI.

Notable studies with a similar methodology and a comparable patient population are the MeRInO (NCT02497573, Scotland), PREDICTION (NL3946, Netherlands) and the PREDICT-HN (NCT03491176, USA) trials [25, 26]. The MeRInO trial is set to include 80 patients with HNSCC of the oropharynx and tries to find a threshold change in ADC that can predict treatment outcome. The PREDICTION trial uses FDG-PET-CT and DW-MRI (EPI and non-EPI technique) to predict the locoregional response in the primary tumor and cervical node metastases. Their study is set to include 20 patients with T2, T3 or T4 HNSCC. Finally, the aim of the PREDICT-HN trial is to assess the prognostic value of weekly MR signal changes and circulating tumor cells during radiation therapy. PREDICT-HN aims at including up to 100 patients.

Including the PREDICT study, the four studies will include 300 patients with HNSCC and subject them to serial MR imaging. This large sample size is expected to create a more accurate prediction model containing multiple clinical and imaging parameters than these studies could produce separately. Although ADC values differ significantly between MR imaging systems and sequences [36], the change in ADC (ΔADC) on serial MRI scanning is expected to be comparable. Therefore, the pooling of data might be possible despite the difference in MR imaging parameters and device specifications. Ideally, the results of these four trials form the basis of a new guideline for HNSCC

were early response assessment to (C)RT allows for the optimal treatment of individual patients. In the future, this might result in increased retention of functionality, less recurrent disease (with probably better overall survival) and overall a higher quality of life for HNSCC patients.

5

References

1. Forastiere AA, Goepfert H, Maor M, et al (2003) Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. N Engl J Med 349:2091–8. https://doi.org/10.1056/NEJMoa031317

2. Spector GJ, Sessions DG, Lenox J, et al (2004) Management of stage IV glottic carcinoma: therapeutic outcomes. Laryngoscope 114:1438–46. https://doi.org/10.1097/00005537-200408000-00024

3. Ganly I, Patel S, Matsuo J, et al (2005) Postoperative complications of salvage total laryngectomy. Cancer 103:2073–2081. https://doi.org/10.1002/cncr.20974

4. Eerenstein SEJ, van der Putten L, de Bree R, et al (2007) Salvage surgery following chemo-radiation for head and neck cancer: Does it yield satisfactory results? Radiother Oncol 82:S17. https://doi.org/10.1016/S0167-8140(07)80056-9

5. Alkureishi LWT, de Bree R, Ross GL (2006) RADPLAT: an alternative to surgery? Oncologist 11:469–480. https://doi.org/10.1634/theoncologist.11-5-469

6. Brockstein B, Haraf DJ, Rademaker AW, et al (2004) Patterns of failure, prognostic factors and survival in locoregionally advanced head and neck cancer treated with concomitant chemoradiotherapy: a 9-year, 337-patient, multi-institutional experience. Ann Oncol Off J Eur Soc Med Oncol 15:1179–86. https://doi.org/10.1093/annonc/mdh308

7. Martens RM, Noij DP, Koopman T, et al (2019) Predictive value of quantitative diffusion-weighted imaging and 18-F-FDG-PET in head and neck squamous cell carcinoma treated by (chemo) radiotherapy. Eur J Radiol 113:39–50. https://doi.org/10.1016/j.ejrad.2019.01.031

8. Lambrecht M, Van Calster B, Vandecaveye V, et al (2014) Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiother Oncol 110:429–434. https://doi.org/10.1016/j.radonc.2014.01.004

9. Differding S, Hanin F-X, Grégoire V (2015) PET imaging biomarkers in head and neck cancer. Eur J Nucl Med Mol Imaging 42:613–22. https://doi.org/10.1007/s00259-014-2972-7

10. Kyle SD, Law WP, Miles KA (2013) Predicting tumour response. Cancer Imaging 13:381–390. https://doi.org/10.1102/1470-7330.2013.9039

11. Phillip Law W, Miles KA (2013) Incorporating prognostic imaging biomarkers into clinical practice. Cancer Imaging 13:332–341. https://doi.org/10.1102/1470-7330.2013.9003

12. Martens RM, Noij DP, Ali M, et al (2019) Functional imaging early during (chemo)radiotherapy for response prediction in head and neck squamous cell carcinoma; a systematic review. Oral Oncol 88:75–83. https://doi.org/10.1016/j.oraloncology.2018.11.005

13. Yen T-C, Lin C-Y, Wang H-M, et al (2006) 18F-FDG-PET for evaluation of the response to concurrent chemoradiation therapy with intensity-modulated radiation technique for Stage T4 nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 65:1307–1314. https://doi.org/10.1016/j.ijrobp.2006.02.031

14. Hentschel M, Appold S, Schreiber A, et al (2011) Early FDG PET at 10 or 20 Gy under chemoradiotherapy is prognostic for locoregional control and overall survival in patients with head and neck cancer. Eur J Nucl Med Mol Imaging 38:1203–1211. https://doi.org/10.1007/s00259-011-1759-3

15. Garibaldi C, Ronchi S, Cremonesi M, et al (2017) Interim (18)F-FDG PET/CT During Chemoradiation Therapy in the Management of Head and Neck Cancer Patients: A Systematic Review. Int J Radiat Oncol Biol Phys 98:555–573. https://doi.org/10.1016/j.ijrobp.2017.02.217

16. Hentschel M, Appold S, Schreiber A, et al (2009) Serial FDG-PET on patients with head and neck cancer: implications for radiation therapy. Int J Radiat Biol 85:796–804. https://doi.org/10.1080/09553000903039180

17. Vandecaveye V, De Keyzer F, Nuyts S, et al (2007) Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: correlation between radiologic and histopathologic findings. Int J Radiat Oncol Biol Phys 67:960–971. https://doi.org/10.1016/j.ijrobp.2006.09.020

18. Dirix P, Vandecaveye V, De Keyzer F, et al (2009) Dose painting in radiotherapy for head and neck squamous cell carcinoma: value of repeated functional imaging with (18)F-FDG PET, (18)F-fluoromisonidazole PET, diffusion-weighted MRI, and dynamic contrast-enhanced MRI. J Nucl Med 50:1020–1027. https://doi.org/10.2967/jnumed.109.062638

19. Kim S, Loevner L, Quon H, et al (2009) Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res 15:986–994. https://doi.org/10.1158/1078-0432.CCR-08-1287

20. King AD, Mo FKF, Yu K-H, et al (2010) Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol 20:2213–2220. https://doi.org/10.1007/s00330-010-1769-8

21. Chen Y, Liu X, Zheng D, et al (2014) Diffusion-weighted magnetic resonance imaging for early response assessment of chemoradiotherapy in patients with nasopharyngeal carcinoma. Magn Reson Imaging 32:630–637. https://doi.org/10.1016/j.mri.2014.02.009

22. Galban S, Brisset J-C, Rehemtulla A, et al (2010) Diffusion-weighted MRI for assessment of early cancer treatment response. Curr Pharm Biotechnol 11:701–708

23. Vandecaveye V, Dirix P, De Keyzer F, et al (2010) Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol 20:1703–1714. https://doi.org/10.1007/s00330-010-1734-6

24. Matoba M, Tuji H, Shimode Y, et al (2014) Fractional change in apparent diffusion coefficient as an imaging biomarker for predicting treatment response in head and neck cancer treated with chemoradiotherapy. AJNR Am J Neuroradiol 35:379–385. https://doi.org/10.3174/ajnr.A3706

25. Ng SP, Bahig H, Wang J, et al (2018) Predicting treatment Response based on Dual assessment of magnetic resonance Imaging kinetics and Circulating Tumor cells in patients with Head and Neck cancer (PREDICT-HN): matching "liquid biopsy" and quantitative tumor modeling. BMC Cancer 18:903. https://doi.org/10.1186/s12885-018-4808-5

26. Paterson C, Allwood-Spiers S, McCrea I, et al (2017) Study of diffusion weighted MRI as a predictive biomarker of response during radiotherapy for high and intermediate risk squamous cell cancer of the oropharynx: The MeRInO study. Clin Transl Radiat Oncol 2:13–18. https://doi.org/10.1016/j.ctro.2016.12.003

27. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP (2017) Technical Note: Diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method. Med Phys 44:4188–4193. https://doi.org/10.1002/mp.12363

28. Rich JT, Neely JG, Paniello RC, et al (2010) A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg 143:331–6. https://doi.org/10.1016/j.otohns.2010.05.007

29. (2014) National guideline Head and Neck Cancer

30. Vandecaveye V, Dirix P, De Keyzer F, et al (2010) Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol 20:1703–1714. https://doi.org/10.1007/s00330-010-1734-6

31. Rodriguez Del Aguila M, Gonzalez-Ramirez A (2014) Sample size calculation. Allergol Immunopathol (Madr) 42:485–492. https://doi.org/10.1016/j.aller.2013.03.008

32. Yrondi A, Peran P, Sauvaget A, et al (2018) Structural-functional brain changes in depressed patients during and after electroconvulsive therapy. Acta Neuropsychiatr 30:17–28. https://doi.org/10.1017/neu.2016.62

33. Driessen JP, van Kempen PMW, van der Heijden GJ, et al (2015) Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. Head Neck 37:440–448. https://doi.org/10.1002/hed.23575

34. Raaymakers BW, Jurgenliemk-Schulz IM, Bol GH, et al (2017) First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol 62:L41–L50. https://doi.org/10.1088/1361-6560/aa9517

35. Mutic S, Dempsey JF (2014) The ViewRay system: magnetic resonance-guided and controlled radiotherapy. Semin Radiat Oncol 24:196–199. https://doi.org/10.1016/j.semradonc.2014.02.008

36. Kolff-Gart AS, Pouwels PJW, Noij DP, et al (2015) Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. AJNR Am J Neuroradiol 36:384–90. https://doi.org/10.3174/ajnr.A4114

Chapter 6

Interobserver agreement of the delineation of
head and neck squamous cell carcinoma on
MRI decreases as (chemo)radiotherapy progresses

Boris Peltenburg, Patricia Doornaert, Frank A. Pameijer, Alexis Kotte,
Marielle E.P Philippens, Remco de Bree, Chris H.J. Terhaard

Abstract

Introduction
Patients with head and neck squamous cell carcinoma (HNSCC) treated with radiotherapy can potentially benefit from online adaptive radiotherapy planning using MRI. However, changes in tumor visibility on MRI during therapy may hamper the contouring process and decrease the reliability of adaptive planning. As the ground truth is missing, interobserver agreement is used as a measure for precision. This study aimed at determining the change in interobserver agreement of the delineation of head and neck tumors on T2 weighted MRI during (chemo)radiotherapy.

Methods
Twenty patients with HNSCC treated with (chemo)radiotherapy received four MRI exams; one prior to treatment and three during the first weeks of treatment. Three observers delineated the tumor on the intratreatment scans. Interobserver agreement was determined using conformity index and by measuring the distance between delineations.

Results
The conformity index decreased with each progressing week: 0.58 week 2, 0.52 week 3, 0.42 week 4. Average tumor volume also decreased from 15.1 ml prior to treatment to 5.4 ml in week 4. The maximum distance between delineations remained constant throughout treatment at around 15 mm.

Conclusion
Interobserver agreement generally decreases as radiotherapy progresses. However, the degree of reduction varies from patient to patient. For clinical introduction of adaptive radiotherapy based on MRI, guidelines and additional training are needed to improve the consistency of target definition.

Introduction

Patients with head and neck squamous cell carcinoma (HNSCC) often have primary tumors in close proximity to vulnerable organs at risk (OARs). Irradiation of these OARs could cause irreversible damage. Limiting the radiation dose to OARs such as the salivary glands and swallowing structures, reduces the chance of radiation-induced loss of function and improves quality of life [1,2].

The standard duration of radiotherapy treatment of a patient with head and neck cancer is five to seven weeks. During this period changes in tumor volume and position of OARs can be expected [3,4]. With the introduction of high field hybrid MRI-radiotherapy systems, such as the MR-Linac, online adaptation of the treatment plan has become available [5,6]. Weekly or even daily plan adaptation, to account for tumor volume changes, could result in a lower radiation dose to the OARs while maintaining the optimal dose to the tumor [7].

The objective of online plan adaptation is to quickly reoptimize the dose distribution of the treatment plan. Ideally, only one sequence would be sufficient to provide the information needed for a complete plan adaptation. A T2 weighted sequence might be suitable for this task as it can be acquired in a reasonable time and provide anatomical information on the tumor as well as on the OARs.

Adapting the treatment plan requires accurate identification of the tumor border. A lack of contrast between tumor and surrounding tissue causes uncertainty about the location of that border [8]. This uncertainty can be expressed by the interobserver agreement of the tumor delineation. Online plan adaptation requires radiation oncologists to delineate tumors that have already received part of the treatment. Treatment-induced changes such as edema and fibrosis may increase uncertainty during the delineation process and therefore a lower interobserver agreement. Determining the degree of interobserver agreement is essential to test if plan adaptation can be introduced safely and effectively. In order to reduce the systematic error caused by interobserver agreement, margins can be added to the Gross Tumor Volume (GTV) [9]. For daily or weekly plan adaptation the interobserver agreement could vary during therapy and as a result the margins to the GTV would change correspondingly.

Therefore, this study aimed at determining the change in interobserver agreement of the delineation of head and neck tumors on T2 weighted MRI during (chemo)radiotherapy.

Methods

Study population

The study population consisted of the first 20 consecutive patients included in the PREDICT study (NL5955, www.trialregister.nl). The inclusion criteria of the PREDICT study are T2, T3 or T4 oral cavity, oropharyngeal, hypopharyngeal or laryngeal squamous cell carcinoma treated with (chemo)radiotherapy with curative intent. As a part of this study, patients received four MRIs during their treatment. MRIs were acquired with one week intervals starting in the second week of treatment. A baseline MRI, obtained as part of the clinical workup, was available for all patients. This study was approved by the Medical Ethics Committee of the UMC Utrecht and all patients signed informed consent for the acquisition of the MRIs and the use of the images for scientific purposes.

Treatment

All patients were treated with radiotherapy over the course of 5-7 weeks, with an effective dose of 66 to 70 Gy to the primary tumor. Depending on tumor stage, patient preference and age, some of the patients received concomitant chemotherapy consisting of 100 mg/m^2 cisplatin administered at day 1, 22 and 43 of the radiotherapy treatment.

MRI

MR imaging was done on a 3.0T MRI machine (Philips Ingenia, Philips Medical Systems, Best). The protocol for the MR scans performed during treatment was similar to that of the scan that was acquired prior to treatment. The details of the T2 weighted sequence are shown in table 2. Imaging was performed in the treatment position using their radiotherapy treatment mask with the head resting on a personalized base. Two parallel surface coils on the lateral side of the patients together with an anterior body array and a posterior coil in the table were used to acquire the images.

Table 2. MRI parameters of the T2 weighted image.

Sequence	Voxel size (mm^3)	TE$_{eff}$ (ms)	TR (ms)	Matrix	Acquisition time
Transverse T2 TSE mDIXON	0.94 x 0.94 x 3.00	100/79	9023	348 x 235	343 seconds

TE$_{eff}$ = Effective Echo Time, TR = Repetition Time.

Observers

Three observers delineated the tumor on the images acquired during treatment. Two of the observers were radiation oncologists with 30 (A) and 20 (C) years of experience

treating HNSCC. The third observer was a radiologist with 30 years of experience with imaging of HNSCC (B). All observers had experience with delineating HNSCC tumors on MR images.

MR-Linac simulated workflow

In order to create a scenario that would simulate the online plan adaptation on a MRI-Linac, some restrictions were imposed. First, the observers were only provided T2 weighted images. Secondly, the observers had to delineate the three scans acquired at weeks 2, 3 and 4 of the therapy in chronological order. Thirdly, observers were provided only the most recent delineation of the tumor as a reference. Lastly, the available patient information was limited to the TNM stage, location and treatment of the primary tumor. No additional pretreatment imaging or posttreatment response information was available to the observers. Additionally, information about the clinical status of the patients during treatment was not provided to the observers.

Tumor delineation

Prior to the start of treatment, the GTV delineation was retrospectively extracted from the treatment plan. This GTV was delineated following standard clinical procedures and guidelines. All available imaging was used to generate the GTV delineation. This included CT, MRI with contrast enhanced and diffusion weighted sequences, and for most patients ^{18}F-FDG PET-CT.

This GTV was provided to each of the observers for the delineation of the first intratreatment scan. At each subsequent MR scan, the observers were only provided their own most recent delineation. This meant that for the week-2 scan they were adapting the pretreatment GTV and for the week-3 scan the final week-2 delineation. The delineation of week 3 was in turn provided for the last adaptation at the week 4 time point.

The observers were blinded to the delineation of the other observers. Furthermore, delineations could not be adjusted retrospectively.

Outcome assessment

The delineations of the three observers were compared for each separate week. An average was taken of the delineated volume of all three observers to obtain the tumor volume on the intratreatment scans. Interobserver agreement was measured by the generalized conformity index (CI) of the delineations [10].

The CI takes the overlap between the delineations of all three observers into account. CI is defined as follows:

$$CIgen = \frac{\sum_{pairsji} A_i A_j \vee}{\sum_{pairsji} A_i A_j \vee}$$

where A_i and A_j represent the different delineations. For a perfect overlap the CI score is 1.00, in case of no overlap the CI is 0.00.

Eventually, the difference between the delineation of each observer pair was determined by measuring the distance between points on delineation A to their closest neighbor on delineation B. This resulted in hundreds of distances for each patient, of which 95th percentile (p95) was chosen to represent the clinically most relevant separation between two delineations. Of the three pairings of observers (1 vs 2, 1 vs 3 and 2 vs 3) the largest p95 distance was chosen to represent the maximum distance between delineations.

Statistical analysis

Differences in generalized conformity index, volume and distances between delineations were determined by one-way ANOVA. P-values smaller than 0.05 were considered statistically significant. Statistical analyses were performed with SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp).

Results

Study population

Fourteen patients completed all three MRI exams during treatment. Due to logistical reasons or patient preference, some of the study participants did not complete all study MRI scans. Five patients did not complete the MRI planned for the second week of treatment and one patient did not complete the MRI exam in the third week. See table 1 for patient baseline characteristics.

Conformity index

On average the CI decreased with each consecutive week: 0.58 week 2, 0.52 week 3, 0.42 week 4. This decrease was only statistically significant between week 2 and week 4 (Mean difference: 0.16, CI95% 0.02 – 0.29, p = 0.018). The degree of reduction in agreement varied between patients (figure 1). For two patients (patient 3 and 8) the interobserver agreement was higher in week 4 than in week 2, but these differences were small: 0.52 vs 0.55 and 0.35 vs 0.38. Four patients had a decrease in conformity index of more than 0.2. The largest decrease in conformity index was seen in patient 16; interobserver agreement initially was high (0.76) for this patient, but in week 4 almost no overlap remained between the delineations of the observers (0.03). The images of

this patient and images of the patient with the highest agreement are shown in figure 2 and 3 respectively. The interobserver agreement of all patients for each week is shown in figure 4.

Table 1. Baseline characteristics. RT = radiotherapy, CRT = Chemoradiotherapy.

| Study No.	Sex	T-stage	N-stage	Location	HPV	Treatment	GTV (ml)
1	Female	T2	N1	Oral cavity	-	RT	4.7
2	Male	T4a	N2c	Oropharynx	Positive	CRT	17.7
3	Male	T3	N0	Larynx	-	RT	5.3
4	Male	T4a	N2b	Oropharynx	Negative	CRT	32.2
5	Male	T2	N0	Oropharynx	Negative	RT	3.7
6	Male	T3	N0	Larynx	-	RT	3.0
7	Male	T3	N2c	Oropharynx	Positive	CRT	32.7
8	Male	T3	N0	Oropharynx	Negative	RT	13.6
9	Male	T2	N3	Oropharynx	Positive	CRT	11.9
10	Female	T2	N2b	Oropharynx	Negative	RT	7.8
11	Male	T2	N0	Oropharynx	Negative	RT	11.4
12	Male	T3	N2c	Larynx	-	CRT	15.8
13	Female	T3	N2c	Larynx	-	CRT	12.1
14	Male	T4a	N2b	Oropharynx	Negative	RT	19.2
15	Male	T2	N3	Oropharynx	Negative	CRT	15.5
16	Male	T3	N1	Oropharynx	Positive	CRT	34.8
17	Male	T4a	N3	Hypopharynx	-	CRT	14.2
18	Female	T2	N3	Oropharynx	Negative	CRT	10.3
19	Male	T2	N3	Hypopharynx	-	RT	18.3
20	Female	T4a	N3	Hypopharynx	-	RT	16.9
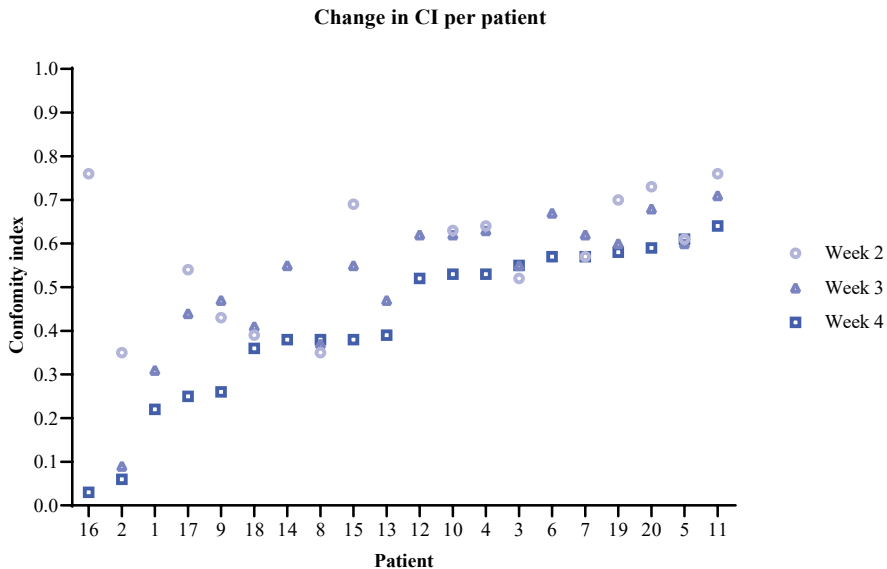
Figure 1. The CI of each individual patient for each week. Patient are sorted from lowest to highest CI at week 4.
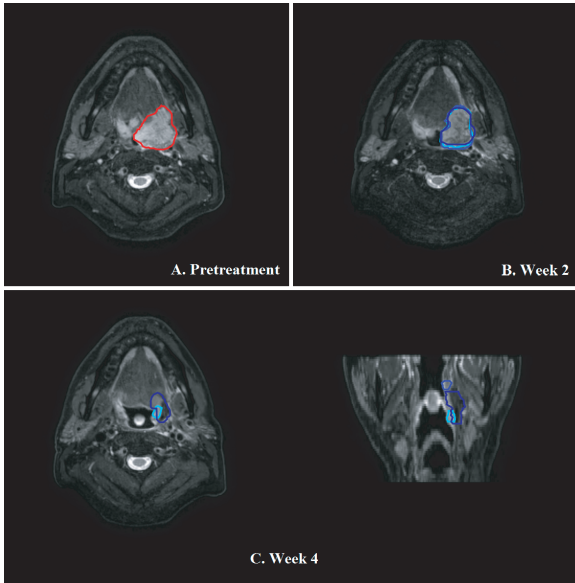
Figure 2. T2 weighted MR images of the patient with the lowest interobserver agreement at week 4 (patient 16). A: MRI prior to the start of treatment with delineation of clinical GTV in red. B: MR images obtained in week 2 of treatment, the three observers are represented by different shades of blue. C: Transverse and coronal images obtained in week 4 of treatment.
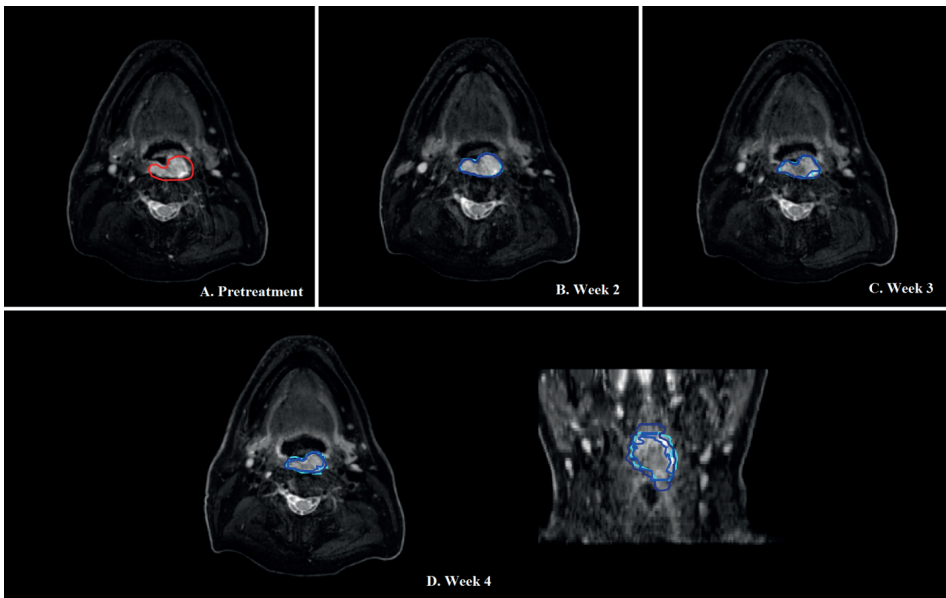

Figure 3. T2 weighted MR images of the patient with the highest interobserver agreement at week 4 (patient 11). A: MRI prior to the start of treatment with delineation of clinical GTV in red. B: MR images obtained in week 2 of treatment, the three observers are represented by different shades of blue C: MRI of week 3. C: Transverse and coronal images obtained in week 4 of treatment.
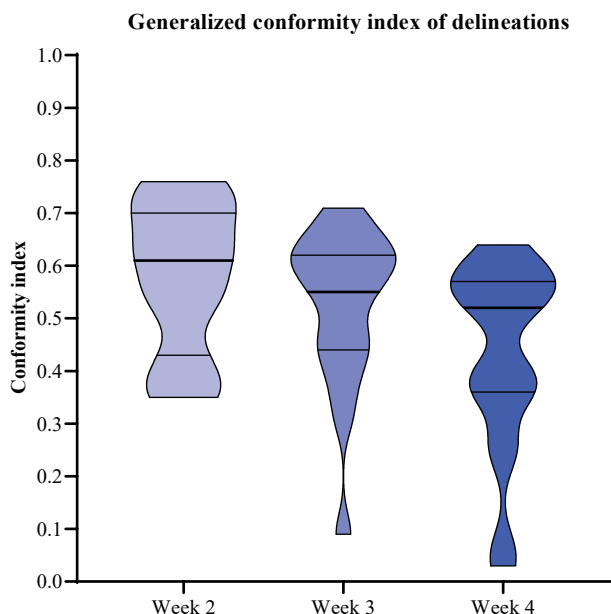
Generalized conformity index of delineations

Figure 4. CI for each week during treatment. The middle line represents the median of the group for the corresponding week. The thin lines represent the quartiles.

Volume

GTV volumes at the start of treatment were on average 15.1 ml (Table 1). During treatment, the average volume decreased from 11.8 ml in week 2 to 7.8 ml in week 3 to 5.4 ml in week 4 (Figure 5). The difference in volume between week 2 and week 4 was statistically significant (CI95% 1.35 – 11.3, p = 0.008).

Patient 16 showed the greatest absolute reduction in size. The tumor of this patient initially had a size of 34.8 ml. At the week 4 time point, the average volume delineated by the three observers for this patient was 3.0 ml. The smallest relative reduction was seen in patient 6 where 68% of the original tumor volume remained, going from 3.0 ml to 2.1 ml.

Correlation of volume and conformity index

No clear correlation can be seen between volume and CI. Small tumors (<12 ml) show variable CI. However, it seems that larger tumors (>12 ml) have a higher CI (figure 6).
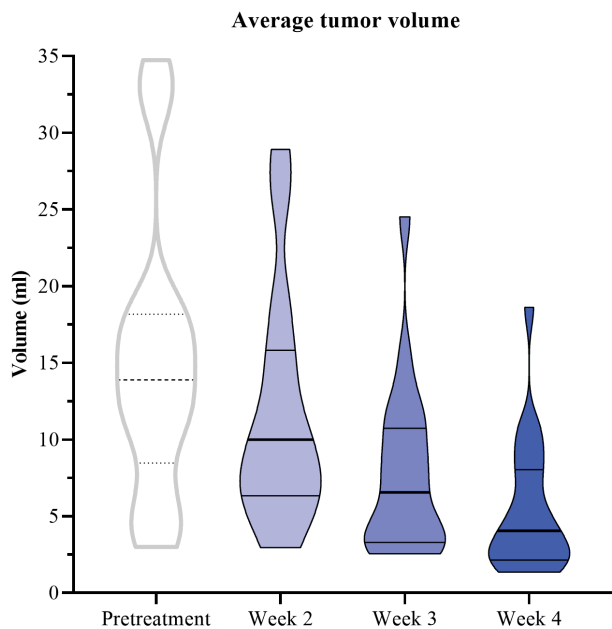
Figure 5. Average tumor volumes before and during treatment. The middle line represents the median of the group for the corresponding week. The thin lines represent the quartiles. The pretreatment volumes were not delineated by the observers but were obtained from the clinical treatment plan.
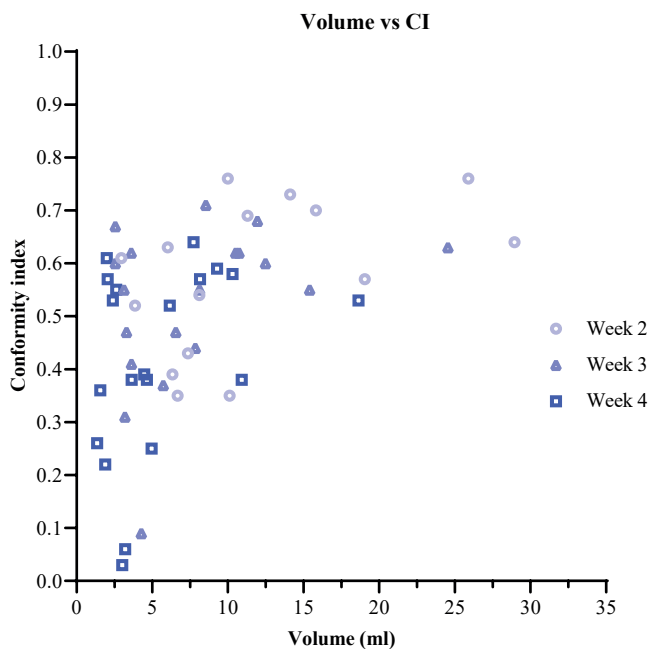

Figure 6. Conformity indexes of all patients at each of the three weeks with the corresponding average tumor volume at that time point.

Distance between observers

The most extreme difference between observers was again seen in patient 18. The maximum distance between observers for this patient was 32 mm. With the exclusion of the outlier, the average maximum distance between delineations remained constant throughout the weeks; 6 mm for week 2, 6 mm for week 3 and 7 mm for week 3 and ranged from 2 mm to 15 mm. The same was true for the largest maximum distances; at week 2 all delineations were within 13 mm from each other, for week 3 this was 15 mm. With the exclusion of patient 18 at week 4 this was 13 mm (figure 7).
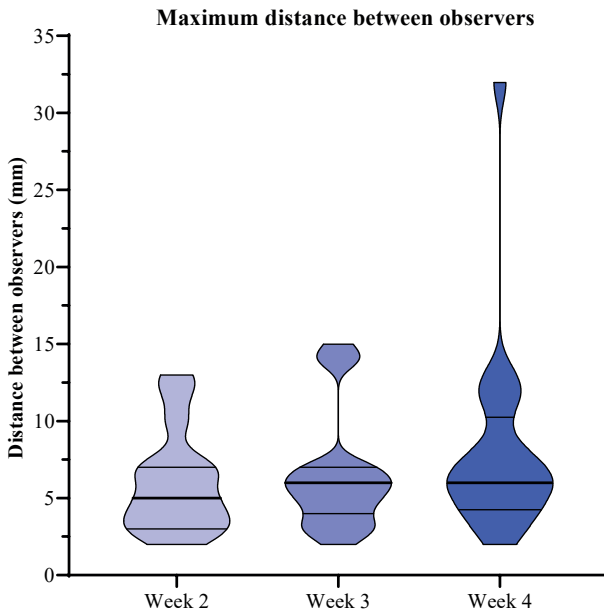

Figure 7. Maximum distances between observers for each week. Each middle line represents the median of the group for the corresponding week. The thin lines represent the quartiles.

Correlation of distance and conformity index

As the conformity index decreases, the distance between the individual delineations of observers increases (figure 8).
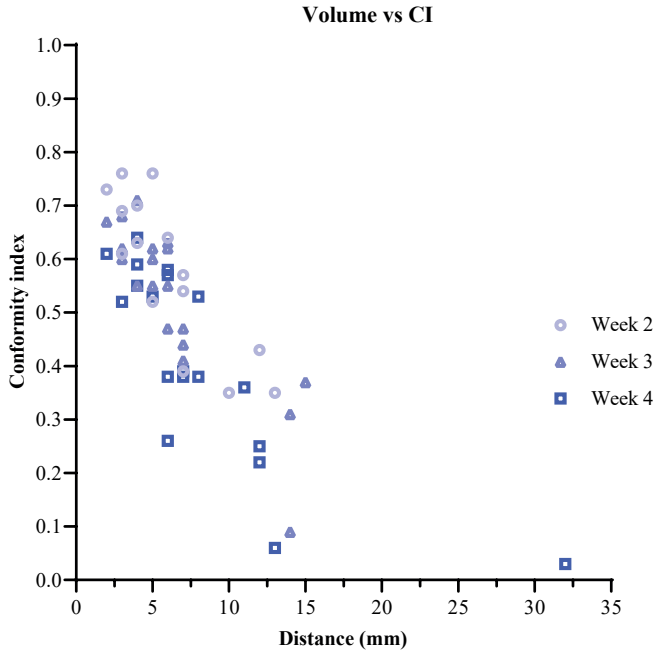
Figure 8. Conformity indexes of all patients at each of the three weeks with the corresponding maximum distance between the delineations at that time point.

Discussion

This study shows that interobserver agreement of delineations of HNSCC decreases as the radiotherapy treatment progresses, although the extent of this decrease is different for each patient. In other words, physicians tend to have different opinions about tumor position and location after the treatment has started to influence tumor visibility. In spite of this decreased agreement, the absolute difference in separation between the delineations remains stable as the weeks progress.

Prior to the start of radiotherapy, agreement between observers is not perfect [11–13], regardless of the modality used for the delineations. Even when provided with pathology validated guidelines, the interobserver agreement, expressed by the CI, for delineating tumors on MRI prior to treatment was only 0.55 [14]. Our study found a similar CI for the delineation of the week 2 MRI. This can be expected, as the tumor has had the shortest time to be affected by the therapy at the week 2 time point. At the week 4 time point, where the most effect of the treatment in our study was be expected, the CI was only 0.42. This decrease in interobserver agreement supports the findings of Apolle et al. [15]

who report on the delineation of five patients with HNSCC by five radiation oncologists using CT and FDG-PET. They report a decrease in interobserver agreement from 0.51 prior to treatment, to 0.38 around the 4th week of treatment.

In our study, tumor volume on average decreased from 15.1 ml prior to treatment to 5.4 ml in week 4 of the treatment. This indicates that around 64% of the original tumor volume was no longer visible. This corresponds with earlier evidence that on CT, HNSCC tumors seem to have a median total loss of 69% at the end of treatment [3]. The effect of chemoradiotherapy on the volume of HSNCC tumors was also investigated by Paudyal et al [4]. In their study of 34 HNSCC patients they found volume reductions of 44-58% at the third week of treatment.

The reduction in tumor volume forms an argument for the use of adaptive radiotherapy. If a part of the original volume no longer contains visible tumor, then this area could potentially be treated with a lower radiation dose thereby making OARs sparing easier. However, to prevent undertreatment of the tumor, the decrease in interobserver agreement that goes alongside the volume reduction has to be addressed. Without proper correction for the variation in delineations, adaptive planning cannot be performed in daily clinical practice. In our study, the maximum distance between the delineations of the observers was 15 mm. This distance remained constant throughout the treatment. It is furthermore comparable to the distance between delineation and tumor extent, as shown by Jager et al [14]. In their study, three observers delineated 27 laryngeal or hypopharyngeal tumors on MRI prior to treatment. Using histopathology as a reference standard they found that the maximum underestimation of the tumor extension was 13 mm. Using this figure as a margin the treated volume would increase to counteract the effects of the reduced interobserver agreement. Progressive changes in interobserver agreement have to be considered if adaptive planning for HNSCC tumors is to become standard of care. Besides using margins to correct for the variability, other methods could be considered. The formation of delineation guidelines or training using specific cases could help increase interobserver agreement [16]. The use of autocontouring or the use of different imaging sequences such as diffusion weighted MRI would also help observers in creating a uniform definition of the tumor borders. It should be noted however that increased interobserver agreement only improves the precision of adaptive treatment, not the accuracy [17]. Careful analysis of the pattern of recurrence after radiotherapy is an important step in order to gain insight in the accuracy of delineations and the effectiveness of adaptive treatment [18].

6

Limitations

This study is limited by incomplete data. Six out of the 20 patients completed only 2 intratreatment MRI exams. Therefore the power of the analyses is somewhat decreased. Furthermore, the study only uses T2 weighted imaging to simulate current MR-Linac online capabilities. However, adaptive radiotherapy can conceivably also be performed using off-line planning. In these cases additional MRI sequences might be available which could potentially help to increase interobserver agreement. In addition, observers were not provided intratreatment information about the patients. This information might have helped during delineation and could have influenced agreement. Lastly, the study included 4 HPV positive patients. Evidence suggests HPV positive patients respond differently to CRT than HPV negative patients. This can be seen in our study as the patient with the most extreme results was a patient with a HPV positive tumor. The group of HPV positive patients was too small to facilitate a subanalysis. Future studies should aim to include only HPV positive or HPV negative patients in their results.

In summary, adaptive radiotherapy planning might in future utilize the decrease in tumor volume in order to reduce radiation dose to OARs. However, this reduction in target volume may, in some patients, be closely related with a substantial decrease in interobserver agreement. For online adaptive radiotherapy planning to be successful, this decrease has to be limited, for example by providing guidelines or additional training for physicians.

References

1.	Chambers MS, Garden AS, Kies MS, Martin JW. Radiation-induced xerostomia in patients with head and neck cancer: pathogenesis, impact on quality of life, and management. Head Neck. 2004 Sep;26(9):796–807.

2.	Scrimger R. Salivary gland sparing in the treatment of head and neck cancer. Expert Rev Anticancer Ther. 2011 Sep;11(9):1437–48.

3.	Barker JL, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int J Radiat Oncol Biol Phys. 2004 Jul 15;59(4):960–70.

4.	Paudyal R, Oh JH, Riaz N, Venigalla P, Li J, Hatzoglou V, et al. Intravoxel incoherent motion diffusion-weighted MRI during chemoradiation therapy to characterize and monitor treatment response in human papillomavirus head and neck squamous cell carcinoma. J Magn Reson Imaging. 2017;45(4):1013–23.

5.	Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. Radiother Oncol. 2008 Jan;86(1):25–9.

6.	Raaymakers BW, Jurgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, van Asselen B, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol. 2017 Nov;62(23):L41–50.

7.	Castelli J, Simon A, Lafond C, Perichon N, Rigaud B, Chajon E, et al. Adaptive radiotherapy for head and neck cancer. Acta Oncol. 2018 Oct;57(10):1284–92.

8.	Eisbruch A, Gregoire V. Balancing risk and reward in target delineation for highly conformal radiotherapy in head and neck cancer. Semin Radiat Oncol. 2009 Jan;19(1):43–52.

9.	van Herk M, Remeijer P, Rasch C, Lebesque J V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys. 2000 Jul 1;47(4):1121–35.

10.	Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. Phys Med Biol. 2009 May 7;54(9):2863–73.

11.	Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. Int J Radiat Oncol Biol Phys. 2007 Jul 1;68(3):763–70.

12.	Cardoso M, Min M, Jameson M, Tang S, Rumley C, Fowler A, et al. Evaluating diffusion-weighted magnetic resonance imaging for target volume delineation in head and neck radiotherapy. J Med Imaging Radiat Oncol. 2019 Jun;63(3):399–407.

13.	Pavic M, Bogowicz M, Würms X, Glatz S, Finazzi T, Riesterer O, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. Acta Oncol. 2018 Aug;57(8):1070–4.

14.	Jager EA, Ligtenberg H, Caldas-Magalhaes J, Schakel T, Philippens ME, Pameijer FA, et al. Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer. Acta Oncol. 2016 Nov;55(11):1305–12.

15.	Apolle R, Appold S, Bijl HP, Blanchard P, Bussink J, Faivre-Finn C, et al. Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. Acta Oncol. 2019 Jul 4;1–8.

16.	Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol. 2016 Jun;60(3):393–406.

6

17. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accredit Qual Assur. 2007 Jan 9;12(1):45–7.

18. Kataria T, Gupta D, Goyal S, Bisht SS, Basu T, Abhishek A, et al. Clinical outcomes of adaptive radiotherapy in head and neck cancers. Br J Radiol. 2016 Jun;89(1062):20160085.

Part 3

Imaging after therapy

Chapter 7

Prospective comparative study of MRI including diffusion-weighted images versus FDG PET-CT for the detection of recurrent head and neck squamous cell carcinomas after (chemo)radiotherapy

Juliette P. Driessen, Boris Peltenburg, Marielle E.P. Philippens, Julia E. Huijbregts, Frank A. Pameijer, Remco de Bree, Luuk M. Janssen, Chris H.J. Terhaard

Abstract

Objective

This prospective study aims to test if MRI including diffusion weighted images can replace FDG PET-CT in the diagnosis of patients with suspicion of local recurrent head and neck squamous cell carcinomas after (chemo)radiation.

Methods

Seventy-five patients suspected of local recurrence underwent an MRI and an FDG PET-CT. Qualitative assessment of the images was performed. Reference standard was the results of biopsy or the absence of a recurrence during follow up.

Results

Seventy patients were included. Fifty percent had local recurrence. FDG PET-CT had accuracy of 71% compared to 73% for MRI. The sensitivity and specificity were 97% compared to 69% and 46% compared to 77% for FDG PET-CT and MRI, respectively.

Conclusions

MRI showed similar diagnostic accuracy, superior specificity but inferior sensitivity compared to FDG PET-CT. Based on current results, we consider MRI including diffusion weighted sequences unable to replace FDG PET-CT as a single imaging modality when local recurrent disease of HNSCC after (C)RT is suspected.

Introduction

Patients with head and neck squamous cell carcinomas (HNSCC) are often treated with radiotherapy with or without chemotherapy ((C)RT). Depending on subsite and tumor stage, loco-regional recurrence rates vary from less than 5% up to 55% after (C)RT [1,2]. Early detection of local recurrences is one of the main objectives during follow-up as delayed detection reduces the chance of successful salvage surgery and may decrease survival rates. However, discrimination between local recurrence and post-radiation effects is known to be a difficult clinical problem [3]. Post-radiation effects such as fibrosis, edema and inflammation may mimic tumor recurrence.

In clinical practice, patients with suspicion of local recurrent pharyngeal or laryngeal carcinoma undergo examination under general anaesthesia with taking of biopsies. Actually, a negative biopsy does not exclude a local recurrence due to sampling error. Unnecessary biopsies in previously radiated areas are undesirable as they can lead to wound healing problems [4]. An accurate selection strategy that reduces the number of patients requiring a biopsy without compromising early detection of residual disease is therefore of great interest. Several studies have shown the value of fluorine ^{18}F-Fluorodeoxyglucose (FDG) PET-CT in the detection of local recurrence after (C)RT [5,6]. This technique is reported to have high negative predictive value, but is limited by false positive results due to FDG avidity in inflammation and tissue changes after radiation therapy [7,8].

A different imaging technique in head and neck cancer is diffusion weighted magnetic resonance imaging (DW-MRI). DW-MRI is described to accurately discriminate malignant lesions from benign [9,10], and DW-MRI might be superior to PET-CT in the detection of local recurrences [11–13]. The present prospective study aimed to investigate if MRI including DW-MRI could replace PET-CT for the detection of local recurrent HNSCC after (C)RT in patients with clinical suspicion of local residual or recurrence disease by comparing the diagnostic accuracy of (DW-)MRI and FDG PET-CT.

Methods

The institutional ethical committee approved this study and written informed consent was obtained from all participants. Prior to the inclusion of the first patient the prospective trial was registered in the Netherlands Trial Registry (http://www.trialregister.nl, NTR3172).

Patients

Seventy-five patients were consecutively and prospectively included in this study between April 2011 and November 2014. Patients clinically suspected of local recurrence after (C)RT for HNSCC underwent as standard procedure an FDG PET-CT and an additional investigational MRI with diffusion-weighted MRI. Inclusion criteria were patients with laryngeal, hypopharyngeal or oropharyngeal cancer with clinical suspicion of local recurrence between 3 months and 3 years after the end of primary (chemo)radiation with curative intent. Oropharyngeal tumors received 69 Gy in 33 fractions in 6 weeks. Small glottic tumors received 60 Gy in 25 fractions in 5 weeks. In all other cases a total dose of 70 Gy in 35 fractions in 7 weeks was given. In case of concurrent chemotherapy, 100 mg/m^2 intravenous cisplatinum was added at day 1, 22 and 43. In case of contraindication for chemotherapy the radiation was combined with cetuximab. Clinical suspicion of local recurrence was defined by presentation with new, persistent or progressive symptoms, or suspicious findings during physical examination. Patients in whom, based on physical examination, local recurrence was so obvious that there was no reasonable doubt were not included in this study. Reporting was done in accordance to the STROBE statement [14].

Imaging

MRI was performed on a 3 Tesla unit (Intera NT, Philips Medical Systems, Best, The Netherlands) using a dedicated head and neck coil. The conventional MRI included a transverse T1-weighted turbospin-echo (TSE) before and after gadolinium, a transverse and coronal T1-weighted spectral presaturation with inversion recovery (SPIR) after administration of gadolinium and a transverse and coronal proton density SPIR. Echo-planar DW-MRI was performed in the transverse plane. Four diffusion gradient b-values (0, 100, 500 and 1000 s/mm^2) were applied in three orthogonal directions, minimizing the effects of diffusion anisotropy. See table 1 for the imaging parameters.

The FDG PET-CTs were performed on a whole body PET-CT scanner (Biograph mCT, Siemens Medical Systems, Erlangen, Germany) approximately one hour after injection of 2.0 MBq/kg of ^{18}F-FDG after a 6-hour fasting period. First, dedicated head and neck imaging was performed with the arms placed beside the body to minimize artifacts in the head and neck area. Subsequently, a whole body scan was performed ranging from the shoulders to the upper thigh with the patient's arms placed above the head. PET acquisition was preceded by a low dose CT scan (40 mAs, slice thickness 3 mm). CTs were non-contrast enhanced. PET images were acquired in 3D mode with Time of Flight for 4 minutes per bed position for the head and neck scan and 3 minutes per bed position for the whole body scan. High resolution PET reconstructions were made with FWHM of 5 mm (whole body scan) or FWHM 4 mm (head/neck scan), 4 iterations and 21 subsets.

Table 1. Imaging parameters. ST: slice thickness, S: number of slices, IG: intersection gap, TR: repetition time, TE; echo time, FS: fat suppression, M: matrix, FOV: field of view, TSE: turbo spin echo, Gd: gadolinium, PD: proton density, SPIR: spectral presaturation with inversion recovery, DW: diffusion weighted, EPI: echo planar imaging.

	ST (mm)	S	IG (mm)	TR (ms)	TE (ms)	FS	M	FOV (mm^2)
Transverse T1w TSE	4	33	0	653	16	-	268x211	240x218
Transverse PD TSE SPIR	4	33	0	2818	25	SPIR	240x183	240x200
Coronal PD TSE SPIR	3	30	0.3	2486	25	SPIR	220x171	220x196
Transverse T1w TSE Gd	5	43	0.8	653	16	-	240x209	240x220
Transverse T1w TSE SPIR Gd	4	33	0	703	16	SPIR	240x192	240x222
Coronal T1w TSE SPIR Gd	3	30	0.3	703	16	SPIR	220x175	220x196
Transverse DWI EPI SPIR	4	27	1	2588	68	SPIR	116x116	230266

Image assessment

The MRI, including DW-MRI, and the FDG PET-CT scan were independently reviewed. The FDG PET-CT scan was assessed by a nuclear medicine physician with 5 years of experience in head and neck PET-CT scanning (J.E.H.). The MRI was assessed in consensus by a radiologist, with more than 15 years of experience in head and neck imaging, and an ENT resident, with 4 years of experience in head and neck (DW-)MRI (F.A.P. and J.P.D.). MRI was interpreted by a combination of all available MRI images (conventional images, DW-MRI images and ADC map). Both DW-MRI and FDG PET-CT were assessed based on qualitative/visual analysis: measurement of apparent diffusion coefficient (ADC) or standard uptake value (SUV) was not performed. For the DW-MRI, hyperintense signal on the heavily diffusion-weighted image with a b-value of 1000 s/mm^2 with corresponding low signal in the ADC map was considered as local recurrence. A low signal was defined as visibly low diffusion coefficients compared to surrounding tissue. Absence of this low signal on the ADC map was considered as local control. Only the primary tumor site was evaluated, lymph nodes were not a subject of this study. Clinical information and all previous imaging, including pre-treatment imaging, were available of the patients. Reviewers of the MRI were blinded to the FDG PET-CT scan and vice versa, all reviewers were blinded to the clinical outcome. Cases could be scored as local recurrence, local control or inconclusive. A conservative approach was adopted with 'inconclusive' scorings treated as 'local recurrence' for all following steps.

Added value

Adding the results of both modalities was done in two ways: 1 - An approach which will increase the sensitivity, where patients were considered to have a local recurrence on imaging if one or both of the modalities were positive; and 2 - an approach which will increase the specificity, where only patients who had a positive FDG PET-CT combined with a positive MRI were considered to have local recurrence.

Reference standard

The local outcome of the patients was determined as follows: If at least one of the imaging modalities was suspicious for a local recurrence, investigation under general anesthesia was performed with biopsy of the primary tumor region. If this biopsy was negative, or if both imaging modalities were suggestive of local control, disease-free follow-up of 6 months was considered reference standard for local control. During the 6 months follow-up, patients had routine bimonthly examinations by their otolaryngologist including direct flexible laryngoscopy. If a renewed suspicion of recurrence arose, additional imaging was performed outside of study protocol. Recurrent disease was scored if patients had biopsy-proven local recurrence within this follow-up of 6 months after initial imaging for suspicion of recurrent disease.

Sample size and statistical analysis

For the use of MRI to be clinically relevant, the positive and negative predictive value of the MRI should be comparable to the current results of FDG PET-CT. We expected 70% of the DW-MRI to be positive for local recurrence based on previous research [12]. To prove a positive predictive value of ≥ 70%, comparable to the current results of FDG PET-CT, with a precision of ~10% we needed to include 75 patients [6,15]. A conservative approach was adopted by treating 'inconclusive' scorings the same as a 'local recurrence' scoring.

Diagnostic accuracy, sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) were calculated for MRI with diffusion-weighted imaging and FDG PET-CT. Sensitivity and specificity were compared with the McNemar test. Data was processed using SPSS (SPSS Inc. v20.0, Chicago, IL, USA) and MedCalc (MedCalc Software v 12.5, Ostend, Belgium).

Results

Patients

Seventy-five patients were included in this study. Five patients were excluded after inclusion due to incomplete imaging or death of the patient without a biopsy or autopsy

after imaging. The seventy remaining patients had clinical suspicion of local recurrence based on new or persistent symptoms (including pain, stridor, dysphonia or dysphagia) or findings at physical examination (e.g. edema, irregular mucosal surface or vocal cord impairment). Fifty-one patients were male (73%), the median age was 61 years (range 42-81). Thirty-four patients (49%) had laryngeal, 12 (17%) hypopharyngeal and 24 (34%) oropharyngeal cancer. Primary T stage ranged from 1 to 4b, N stage 0 to N2c. All patients were treated with primary fractionated radiotherapy: radiotherapy alone (n=50, 71%), combined with cisplatin (n=11, 16%) or combined with cetuximab (n=9, 13%). Patients and tumor characteristics are presented in table 2.

Table 2. Baseline patient characteristics * Median (range), † Percentage within subsite, AJCC: American Joint Committee on Cancer 7th edition.

Variable	N (range)	%
Age (y) *	63 (40-87)	
Sex		
Female	55	25
Male	162	75
Tumor site†		
Larynx	69	32
Hypopharynx	35	16
Oropharynx	102	47
Oral cavity	11	5
AJCC tumor stage		
T2	81	37
T3	71	33
T4a	53	24
T4b	12	6
Nodal stage		
N0	95	44
N1	26	12
N2a	2	1
N2b	53	24
N2c	41	19
HPV status		
Positive	20	9
Negative	81	37
Unknown	116	54
Treatment		
Radiotherapy	120	55
Chemoradiotherapy	66	31
Radiotherapy + Cetuximab	31	14

7

Histopathology

Thirty-five patients (50%) had local recurrence, of which 31 were histologically proven by the first biopsy after imaging. One patient had initially negative biopsy, but due to persistent clinical suspicion for recurrence within the follow-up of the study protocol, a second endoscopy with biopsy was performed after 4 weeks which confirmed recurrent disease. One patient with local control on FDG PET-CT and (DW-)MRI received a biopsy outside of the study protocol. Finally, one patient with a positive MRI and a positive FDG PET-CT refrained from biopsy. Within the follow-up of this study, this patient had clinical progressive local disease combined with progressive alterations seen on repeated imaging (CT and conventional MRI). Therefore, this patient was considered as having recurrent disease. The remaining thirty-five patients all completed the follow-up period and had no biopsy proven recurrence within 6 months after inclusion for this study. Twenty-one had negative biopsy, and fourteen of these patients had no biopsy, but all these patients had at least 6 months' disease-free follow-up (figure 1).
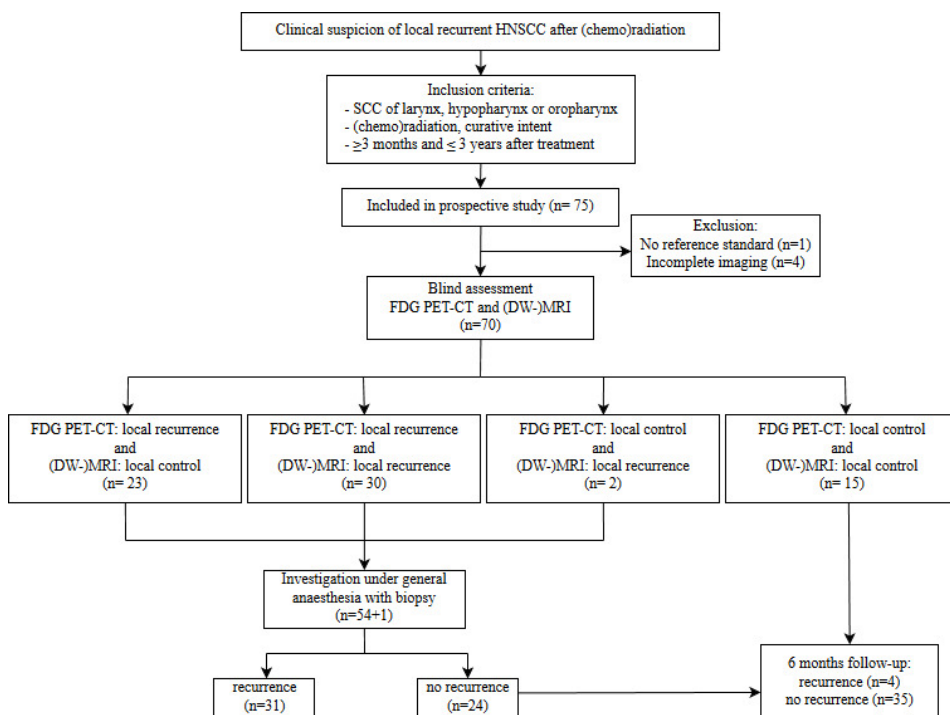

Figure 1. Flow-chart of study design. One patient with local control on FDG PET-CT and (DW-)MRI received a biopsy outside of the study protocol, one patient with positive imaging refrained from biopsy and 54 patients received the biopsy as per protocol. SCC: squamous cell carcinoma, DW-MRI: diffusion-weighted MRI

Imaging

Median time between MRI and FDG PET-CT was one day (0-31 days). All images were of sufficient quality to be evaluated. For the FDG PET-CT the plasma glucose of each patient was <10 mmol/l, except for one who had a plasma glucose of 12.5 mmol/l. Seventy-three percent (51/70) of the FDG PET-CTs were positive for local recurrence compared to only 46% (32/70) of the MRI. Two FDG PET-CTs were inconclusive. For the calculation of the diagnostic accuracy, inconclusive imaging results were regarded as positive. FDG PET-CT had a similar diagnostic accuracy compared to MRI (71% for FDG PET CT versus 73% for MRI (p=0.85)). The sensitivity of FDG PET-CT was significantly superior compared to MRI (97% versus 69%; p<0.01). The specificity of FDG PET-CT was significantly inferior compared to MRI (46% versus 77%; p<0.01). MRI had a false negative rate of 31% (11/35) compared to 3% for FDG PET-CT (1/35). See table 3 and 4 for the complete results. See figures 2 and 3 for examples of the images obtained from two of the included patients.

Table 3. Results of DW-MRI and FDG PET-CT. For the combination of modalities the inconclusive PET-CT scores are considered to be PET-CT positive.

Imaging	Local recurrence	Local control
MRI positive	24	8
MRI negative	11	27
PET-CT positive	33	18
PET-CT negative	1	16
PET-CT inconclusive	1	1
Both modalities positive	23	7
MRI positive PET-CT negative	1	1
MRI negative PET-CT positive	11	12
Both modalities negative	0	15

Table 4. Diagnostic accuracy of DW-MRI and FDG PET-CT.

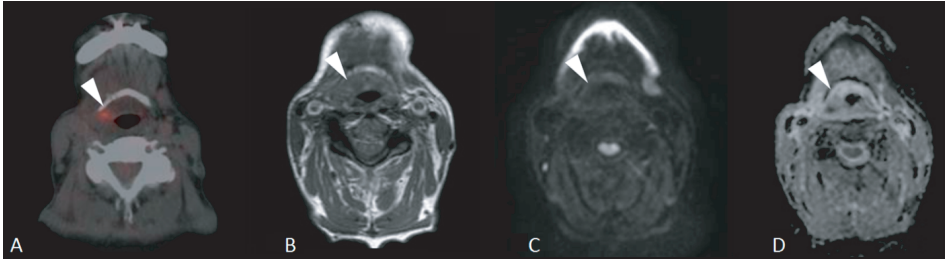|	FDG PET-CT (95% CI)	MRI (95% CI)	Both modalities combined	
			Sensitive (95% CI)	Specific (95% CI)
Sensitivity	97% (85-100)	69% (51-83)	100% (90-100)	66% (48-81)
Specificity	46% (29-63)	77% (60-90)	43% (26-61)	80% (63-92)
Accuracy	72%	73%	71%	73%
Positive predictive value	64%	75%	64%	77%
Negative predictive value	94%	71%	100%	70%

Figure 2. (A-D) Example of an included patient. 42-year-old-man with T2N1M0 oropharyngeal carcinoma in the right vallecula. Seven months after radiotherapy he presented with asymmetry in the vallecula and an ulcer in the right vallecula. (**A**) FDG PET-CT shows metabolic activity at the primary site (arrow). This scan was considered as a recurrent carcinoma. (**B**) T1w MRI shows swelling of the vallecula at the primary tumor site (arrow). (**C**) DW-MRI B1000 shows no increased signal intensity and no diffusion restriction on ADC therefore it was considered as a local control based on the DW-MRI (**D**, arrow). Histology confirmed recurrence of a squamous cell carcinoma.
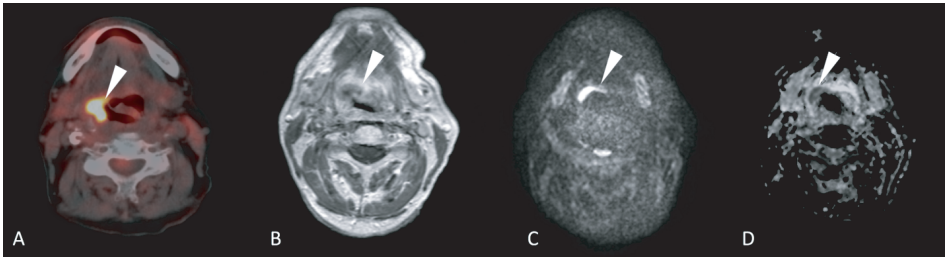

Figure 3. (A-D) Example of an included patient. 81-year-old-man with T4aN2cM0 oropharyngeal carcinoma in the right base of tongue. Thirteen months after chemoradiotherapy with Cetuximab he presented with progressive otalgia. (**A**) FDG PET-CT shows metabolic activity at the primary site, and therefore it was considered a recurrent carcinoma (arrow). (**B**) T1w contrast-enhanced MRI shows an isointense region at the primary tumor site (arrow). (**C**) DW-MRI B1000 shows increased signal intensity with corresponding diffusion restriction on ADC, and was therefore also considered as a local recurrence based on DW-MRI (**D**, arrow). Histology confirmed recurrence of a squamous cell carcinoma.

Added value

Combining both modalities and using the sensitive method resulted in a sensitivity of 100% and a specificity of 43%. Using the specific method resulted in a sensitivity of 66% and specificity of 80%. The accuracy of the combined tests was almost similar to the individual modalities. The complete results of this analysis are presented in table 3 and 4.

Discussion

The present study compares FDG PET-CT and MRI including diffusion weighted images regarding the detection of recurrence of oropharyngeal, hypopharyngeal or laryngeal cancer. It shows that MRI has a similar diagnostic accuracy as FDG PET-CT (73% compared to 71%; p=0.85). MRI has superior specificity but inferior sensitivity compared to FDG PET-CT (specificity 77% versus 46%, sensitivity 69% versus 97%; p<0.01). During follow-up after (chemo)radiation, early detection of residual or recurrent disease is one of the main objectives as salvage surgery might still be a curative option. In this scenario, a high sensitivity is the most important feature of an imaging modality. False negative results of a selection strategy can have tremendous effects; they will cause delay in detection of recurrences, which will potentially influence the chance of successful salvage surgery. Therefore, we consider FDG PET-CT to be superior to MRI including diffusion weighted sequences in the early diagnosis of recurrence of HNSCC after (C)RT.

The complementary effect of both modalities is limited. We found only a small difference in test characteristics between the combined test and the individual modalities. Combining all positive findings of FDG PET-CT and MRI improves the sensitivity for recurrences to 100% where FDG PET-CT alone missed one recurrence. Less desirable was that the same method scored 20 patients erroneously as having a local recurrence: one patient more than FDG PET-CT alone. It is debatable whether the difference in early detection of recurring disease in one patient at the cost of one extra patient with an unnecessary biopsy, justifies a combined use of FDG-PET and MRI.

Few studies report on the diagnostic accuracy of DW-MRI in HNSCC after (C)RT using a visual assessment method for the DW-MRI images. Those that do, show excellent results with accuracies up to 94% [12,16]. Compared to literature our results concerning DW-MRI are somewhat disappointing. This might be because we included patients based on 'clinical suspicion of local recurrence'. This is prone to subjectivity, which might be physician, institution and study dependent. Also the design of this study facilitates the inclusion of very early recurrences and therefore very small lesions which may be under or at the border of the detection limit. DW-MRI suffers from relative low spatial resolution, making it less suitable for detection of very small tumor residues. Partial volume effects will prevent diffusion restriction to stand out at the ADC map. Furthermore, forty-nine percent of the patients in our study had had a laryngeal localization of their primary tumor. The larynx, compared to other regions of the head and neck is especially known for movement and susceptibility artefacts. FDG PET-CT, due to its faster acquisition of individual slices, will be more forgiving of movement artefacts such as swallowing and breathing. In addition, differences in methodology between our research and previous

7

studies, such as the use of repeated imaging to detect recurrent or residual tumors by Tshering Vogel et al. [12] compared to our single acquisition approach, might explain the comparatively lower observed accuracy of (DW-)MRI of our study.

Several studies reported on the effectiveness of DW-MRI in detecting recurrences using quantitative measurements instead of visual review [13,17–19]. All report an ADC cutoff value for differentiating treatment induced tissue alterations from recurring tumors. Quantitative assessment yields accuracies of 90% and higher. This is similar to the accuracies reported by the studies where only visual review was used.

In our study, we chose visual i.e. qualitative assessment of DW-MRI instead of quantitative ADC measurements because it most closely resembles clinical practice. Radiologists usually visually score MRIs and base their final decision on the existence of recurrence on a combination of information given by all sequences, rather than based on one solitary measure of for example ADC. Also, quantitative ADC measurements is dependent on the placement of a region of interest within an ADC map, and therefore highly variable with low repeatability. Furthermore, using ADC as a discriminating tool for recurrence or benign lesions also comes with challenges and questions such as which ADC threshold to use and its reproducibility on other type of scanners and protocols [20]. Even though quantitative ADC measures are described to have significantly different mean values in benign compared to malignant lesions, there is extensive overlap between ADC values of malignant and benign lesions [11,21]. This limits its use on individual patients. Qualitative e.g. visual assessment has limitations as well; as it is subjective and might have a learning curve; however, our data of the diagnostic accuracy divided in three time-frames showed no time trends.

Although the high sensitivity of FDG PET-CT is arguably the most important test characteristics when trying to detect a tumor recurrence, the high specificity of DW-MRI could be important in reducing the amount of unnecessary examinations under general anesthesia in future patients. A possible use of DW-MRI is in direct combination with PET using PET/MRI scanners. This is shown by Becker et al. who found an excellent overall diagnostic accuracy using FDG PET/MRI in patients suspected of a local tumor recurrence [22].

One strength of this study is the design, in which biopsy and follow-up of six months served as the reference standard. The extra time frame of six months' disease-free follow-up was incorporated to prevent sampling errors of negative biopsies. Indeed, there was one subject in which biopsy was negative at first but turned out positive after 4 weeks during a second endoscopy because of persistent clinical suspicion of local disease. One

could argue that 6 months might be too short, however it cannot be expected that these imaging modalities can detect subclinical recurrence which become manifest more than 6 months after imaging.

Our study has limitations. First, the diffusion-weighted MRI was evaluated in combination with the anatomical images of the conventional MRI. This resembles clinical practice, as the diffusion-weighted images suffer from low resolution and lack anatomical landmarks, these images should always be viewed in context of other MRI sequences. Therefore, our results reflect MRI including DW-MRI, and does not study the added diagnostic value of diffusion weighted images to conventional MRI sequences. Secondly, pretreatment imaging of the primary tumor consisted of conventional CT, MRI and sometimes FDG PET-CT. None of the patients had diffusion-weighted MRI as pre-therapeutic imaging, neither was there any post-therapeutic baseline imaging available. If these additional examinations had been available they possibly could have made the interpretation of the MRI or FDG PET-CT, at the time of a suspected recurrence, more accurate. Furthermore, the DW-MRI sequence used was echo-planar (EPI) DW-MRI. Image distortion, especially in the head and neck region, might occur with EPI DW-MRI. Other techniques with less image distortion might increase DW-MRI accuracy [23].

The FDG PET-CT protocol in the hospital where this study was performed did not contain intravenous contrast. Contrast-enhanced CT images might increase the sensitivity or specificity of FDG PET-CT. Suenaga et al. [24] showed that the addition of a contrast-enhanced CT to a FDG PET-CT protocol has minimal added value in patients with recurring HNSCC. However, in their study only 32% of patients received some form of radiotherapy as the initial treatment. It is possible that contrast-enhanced FDG PET-CT might be more accurate in this patient group than non-contrast-enhanced FDG PET-CT. Finally, not all patients had biopsy as a reference standard, since 14 patients were considered as local control based on negative DW-MRI and negative FDG PET-CT combined with 6 months disease-free follow-up. Because of the high negative predictive value of FDG PET-CT we felt that it was ethically incorrect to expose these patients to unnecessary general anesthesia with biopsy.

Conclusion

In conclusion, non-contrast-enhanced FDG PET-CT is superior to MRI including DW-MRI in the detection of local recurrent oropharyngeal, hypopharyngeal or laryngeal cancer in patients with suspicion of recurrence after (chemo)radiation. Though having similar accuracies, MRI suffers from more false negative results than FDG PET-CT. When early

7

detection of recurrences is the main goal, false negatives lead to delayed detection and might lead to irresectability and decreased survival rates. The effect of combining FDG PET-CT and MRI is limited but might reduce false negative findings. Based on current results, we consider MRI including diffusion weighted sequences unable to replace FDG PET-CT as a single imaging modality when local recurrent disease of HNSCC after (C)RT is suspected.

References

1. Forastiere AA, Goepfert H, Maor M, Pajak TF, Weber R, Morrison W, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. N Engl J Med. 2003 Nov 27;349(22):2091–8.

2. Spector GJ, Sessions DG, Lenox J, Newland D, Simpson J, Haughey BH. Management of stage IV glottic carcinoma: therapeutic outcomes. Laryngoscope. 2004 Aug;114(8):1438–46.

3. Zbaren P, Weidner S, Thoeny HC. Laryngeal and hypopharyngeal carcinomas after (chemo)radiotherapy: a diagnostic dilemma. Curr Opin Otolaryngol Head Neck Surg. 2008 Apr;16(2):147–53.

4. Valentino J, Spring PM, Shane M, Arnold SM, Regine WF. Interval pathologic assessments in patients treated with concurrent hyperfractionated radiation and intraarterial cisplatin (HYPERRADPLAT). Head Neck. 2002 Jun;24(6):539–44.

5. Krabbe CA, Pruim J, Dijkstra PU, Balink H, van der Laan BF, de Visscher JG, et al. 18F-FDG PET as a routine posttreatment surveillance tool in oral and oropharyngeal squamous cell carcinoma: a prospective study. J Nucl Med. 2009 Dec;50(12):1940–7.

6. Terhaard CH, Bongers V, van Rijk PP, Hordijk GJ. F-18-fluoro-deoxy-glucose positron-emission tomography scanning in detection of local recurrence after radiotherapy for laryngeal/ pharyngeal cancer. Head Neck. 2001 Nov;23(11):933–41.

7. de Bree R, Putten L v d, Brouwer J, Castelijns JA, Hoekstra OS, René Leemans C. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Vol. 45, Oral Oncology. 2009. p. 386–93.

8. Purohit BS, Ailianou A, Dulguerov N, Becker CD, Ratib O, Becker M. FDG-PET/CT pitfalls in oncological head and neck imaging. Insights Imaging. 2014 Oct;5(5):585–602.

9. Koh D-M, Padhani AR. Diffusion-weighted MRI: a new functional clinical technique for tumour imaging. Br J Radiol. 2006 Aug;79(944):633–5.

10. Padhani AR. Diffusion magnetic resonance imaging in cancer patient management. Semin Radiat Oncol. 2011 Apr;21(2):119–40.

11. Abdel Razek AAK, Kandeel AY, Soliman N, El-shenshawy HM, Kamel Y, Nada N, et al. Role of diffusion-weighted echo-planar MR imaging in differentiation of residual or recurrent head and neck tumors and posttreatment changes. AJNR Am J Neuroradiol. 28(6):1146–52.

12. Tshering Vogel DW, Zbaeren P, Geretschlaeger A, Vermathen P, De Keyzer F, Thoeny HC. Diffusion-weighted MR imaging including bi-exponential fitting for the detection of recurrent or residual tumour after (chemo)radiotherapy for laryngeal and hypopharyngeal cancers. Eur Radiol. 2013 Feb;23(2):562–9.

13. Vandecaveye V, De Keyzer F, Nuyts S, Deraedt K, Dirix P, Hamaekers P, et al. Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: correlation between radiologic and histopathologic findings. Int J Radiat Oncol Biol Phys. 2007 Mar;67(4):960–71.

14. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007 Oct 16;4(10):e297.

15. Brouwer J, de Bree R, Comans EFI, Akarriou M, Langendijk JA, Castelijns JA, et al. Improved detection of recurrent laryngeal tumor after radiotherapy using (18)FDG-PET as initial method. Radiother Oncol. 2008 May [cited 2016 Jan 20];87(2):217–20.

7

16. Noij DP, Jagesar VA, de Graaf P, de Jong MC, Hoekstra OS, de Bree R, et al. Detection of residual head and neck squamous cell carcinoma after (chemo)radiotherapy: a pilot study assessing the value of diffusion-weighted magnetic resonance imaging as an adjunct to PET-CT using 18 F-FDG. Oral Surg Oral Med Oral Pathol Oral Radiol. 2017 Sep;124(3):296–305.e2.

17. Vaid S, Chandorkar A, Atre A, Shah D, Vaid N. Differentiating recurrent tumours from post-treatment changes in head and neck cancers: does diffusion-weighted MRI solve the eternal dilemma? Clin Radiol. 2017 Jan;72(1):74–83.

18. Abdel Razek AAK, Gaballa G, Ashamalla G, Alashry MS, Nada N. Dynamic Susceptibility Contrast Perfusion-Weighted Magnetic Resonance Imaging and Diffusion-Weighted Magnetic Resonance Imaging in Differentiating Recurrent Head and Neck Cancer From Postradiation Changes. J Comput Assist Tomogr. 2015;39(6):849–54.

19. Desouky S, AboSeif S, Shama S, Gaafar A, Gamaleldin O. Role of dynamic contrast enhanced and diffusion weighted MRI in the differentiation between post treatment changes and recurrent laryngeal cancers. Egypt J Radiol Nucl Med. 2015 Jun;46(2):379–89.

20. Kolff-Gart AS, Pouwels PJW, Noij DP, Ljumanovic R, Vandecaveye V, de Keyzer F, et al. Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. AJNR Am J Neuroradiol. 2015 Feb;36(2):384–90.

21. Gouhar GK, El-Hariri MA. Feasibility of diffusion weighted MR imaging in differentiating recurrent laryngeal carcinoma from radionecrosis. Egypt J Radiol Nucl Med. 2011 Jun;42(2):169–75.

22. Becker M, Varoquaux AD, Combescure C, Rager O, Pusztaszeri M, Burkhardt K, et al. Local recurrence of squamous cell carcinoma of the head and neck after radio(chemo)therapy: Diagnostic performance of FDG-PET/MRI with diffusion-weighted sequences. Eur Radiol. 2018 Feb;28(2):651–63.

23. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother Oncol. 2013 Dec;109(3):394–7.

24. Suenaga Y, Kitajima K, Ishihara T, Sasaki R, Otsuki N, Nibu K-I, et al. FDG-PET/contrast-enhanced CT as a post-treatment tool in head and neck squamous cell carcinoma: comparison with FDG-PET/non-contrast-enhanced CT and contrast-enhanced CT. Eur Radiol. 2016 Apr;26(4):1018–30.

Chapter 8

The interobserver agreement in the detection of recurrent HNSCC using MRI including diffusion weighted MRI

Boris Peltenburg*, Juliette P. Driessen*, Marielle E.P. Philippens, Frank A. Pameijer, Jonas Castelijns, Jan Willem Dankbaar, Bart Dorgelo, Charlotte A.H. Lange, Frederick J.A. Meijer, Walter. M. Palm, Stefan D. Roosendaal, Stefan C.A. Steens, Berit M. Verbist, Inge Stegeman, Luuk M. Janssen, Remco de Bree, Chris H.J. Terhaard

authors contributed equally

Abstract

Introduction
For the detection of local recurrences of head and neck squamous cell carcinomas (HNSCC) after (chemo)radiation, diagnostic imaging is generally performed. Diffusion weighted magnetic resonance imaging (DW-MRI) has been proven to be able to adequately diagnose the presence of cancer. However, evaluation of DW-MR images for recurrences is difficult and could be subject to individual interpretation.

Aim
To determine the interobserver agreement, intraobserver agreement and influence of experience of radiologists in the assessment of DW-MRI in patients clinically suspected of local recurrent HNSCC after (chemo)radiation.

Methods
Ten experienced head and neck radiologists assessed follow-up MRI including DW-MRI series of 10 patients for the existence of local recurrence on a two point decision scale (local recurrence or local control). Patients were clinically suspected for a recurrence of laryngeal (n=3), hypopharyngeal (n=3) or oropharyngeal (n=4) cancer after (chemo) radiation with curative intent. Fleiss' and Cohen's Kappa were used to determine interobserver agreement and intraobserver agreement, respectively.

Results
Interobserver agreement was κ=0.55. Intraobserver agreement was κ=0.80. Prior experience within the field of radiology and with DW-MRI had no significant influence on the scoring.

Conclusion
For the assessment of HNSCC recurrence after (chemo)radiation by DW-MRI, moderate interobserver agreement and substantial intraobserver agreement was found.

Introduction

Improved radiotherapy techniques have led to increased use of radiotherapy, with or without concurrent chemotherapy, as the primary treatment for HNSCC [1]. (C)RT has obvious advantages of organ preservation compared to surgical intervention [2]. However even with (C)RT, recurrence or residual disease occurs in up to 40% of patients with stage III and IV disease [3, 4]. Early detection of local HNSCC recurrences greatly influences the chance of successful surgical salvage treatment. Delayed detection might lead to irresectability of residual disease or local regional metastasis and decreased survival rates [5–7].

The diagnosis of recurrent head and neck squamous cell carcinomas (HNSCC) is usually made based on a combination of diagnostic imaging and physical examination (under general anesthesia) including biopsies. ^{18}F-fluorodeoxyglucose-PET-CT (FDG PET-CT) is often cited as a useful imaging modality due to its high sensitivity and high negative predictive value but may suffer from variable tracer uptake in irradiated areas [8–12].

MRI techniques are of increasing interest in the detection of HNSCC due to their advantage of superior soft tissue contrast compared to CT and possibility of tissue characterization. Diffusion weighted MRI (DW-MRI) has proven to be useful in the diagnosis of recurring HNSCC [13]. However, the head and neck region has many transitions between air, tissue and bone, which reduce image quality and can lead to susceptibility artifacts [14, 15]. Furthermore, discrimination between tumor recurrence and post-radiation effects, although possible on DW-MRI, has been proven to be difficult, as post-radiation effects may mimic tumor recurrence after (C)RT [16]. These inherent difficulties of DW-MRI combined with differences in experience among head and neck radiologists could result in different interpretations of the same diffusion images. The degree of uniformity in classifying recurrence by different observers might influence the applicability of DW-MRI in this patient group in daily clinical practice. The interobserver agreement diagnosing HNSCC using DW-MRI has been evaluated previously but no data is available for the detection of recurrent HNSCC [17–19].

Therefore, the aim of this study is to determine the interobserver agreement, intraobserver agreement and influence of experience of radiologists in the assessment of DW-MRI in patients clinically suspected of local recurrent HNSCC after (chemo)radiation.

8

Materials and Methods

Patients

This study included ten patients with HNSCC treated with (C)RT. All patients were clinically suspected of having local residual or recurrent disease. This clinical suspicion was based on findings at physical examination or having persistent suspicious symptoms such as referred otalgia or hoarseness. The cases were selected from a prospective comparative study conducted in our institution investigating the diagnostic accuracy of DW-MRI versus PET-CT in detection of recurrence in patients with suspicious symptoms after (C)RT (RETURNED study, conducted in the University Medical Center Utrecht Trial registry number: NTR3172. Full article in submission) [20]. This study consisted of 75 patients included from April 2011 till November 2014. Inclusion criteria for this study were primary (chemo)radiotherapy with curative intent, at least 3 months but no more than 3 years prior to inclusion. Exclusion criteria were contraindications for MRI and age under 18 years. Chemotherapy consisted of either cisplatin or carboplatin or a combination of both drugs. After completion of this study, five cases with a proven recurrence and five cases with proven local control were randomly selected out of all patients using the random sample cases function in SPSS (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp, USA). This distribution resembled the prevalence of recurrences as in the complete study, which was 50%. Of the patients with a proven local recurrence the smallest lesion was 15x24x32mm, measured on MRI. This study was approved by the local ethics board of our institution and all patients gave written informed consent.

See Table 1 for further baseline characteristics of the included patients. The results of this study are reported using the GRRAS and STROBE guidelines [21].

Table 1. Patient characteristics

Case	Age	Sex	Initial tumor location	Initial tumor stage	Initial therapy	6 months follow up
1	55	Female	Supraglottic larynx	T3N0M0	Radiotherapy	Local control
2	49	Female	Supraglottic larynx	T3N0M0	Radiotherapy	Local recurrence
3	61	Male	Hypopharynx	T3N2bM0	Chemo + Radiotherapy	Local control
4	65	Female	Oropharynx	T4aN2bM0	Chemo + Radiotherapy	Local recurrence
5	71	Female	Oropharynx	T3N2bM0	Cetuximab + RT	Local control
6	62	Male	Hypopharynx	T4bN2cM0	Chemo + Radiotherapy	Local recurrence
7	75	Male	Hypopharynx	T3N2bM0	Radiotherapy	Local recurrence
8	60	Male	Oropharynx	T4aN0M0	Radiotherapy	Local control
9	76	Male	Supraglottic larynx	T3N0M0	Radiotherapy	Local control
10	53	Male	Oropharynx	T3N2bM0	Cetuximab + RT	Local recurrence

Diffusion weighted MRI

MR images obtained as part of the RETURNED study were collected. Imaging was performed on a 3 Tesla unit (Achieva 3TX, Philips Medical Systems, Best, The Netherlands) using a dedicated head and neck receive coil [22]. The conventional MRI included a transverse T1-weighted turbospin echo (TSE) before and after gadolinium, a transverse and coronal T1-weighted after admission of gadolinium using spectral presaturation with inversion recovery (SPIR) as fat suppression and a transverse and coronal proton density short tau inversion recovery (STIR). Echo-planar DWI was performed in the transverse plane. Four levels of diffusion weighting (b-values: 0, 100, 500 and 1000 s/mm^2) were applied in three orthogonal directions, minimizing the effects of diffusion anisotropy. Apparent diffusion coefficient (ADC) maps were calculated from the DW images using all four b-values and a mono-exponential model. See Table 2 for the imaging parameters.

Table 2. Imaging parameters. ST: slice thickness, S: number of slices, IG: intersection gap, TR: repetition time, TE; echo time, FS: fat suppression, M: matrix, FOV: field of view, TSE: turbo spin echo, Gd: gadolinium, PD: proton density, SPIR: spectral presaturation with inversion recovery, DW: diffusion weighted, EPI: echo planar imaging.

	ST (mm)	S	IG (mm)	TR (ms)	TE (ms)	FS	M	FOV (mm^2)
Transverse T1w TSE	4	33	0	653	16	-	268x211	240x218
Transverse PD TSE SPIR	4	33	0	2818	25	SPIR	240x183	240x200
Coronal PD TSE SPIR	3	30	0.3	2486	25	SPIR	220x171	220x196
Transverse T1w TSE Gd	5	43	0.8	653	16	-	240x209	240x220
Transverse T1w TSE SPIR Gd	4	33	0	703	16	SPIR	240x192	240x222
Coronal T1w TSE SPIR Gd	3	30	0.3	703	16	SPIR	220x175	220x196
Transverse DWI EPI SPIR	4	27	1	2588	68	SPIR	116x116	230266

Reference standard specification

In the RETURNED study, every patient suspected of local recurrence on either DW-MRI or PET-CT underwent endoscopy with biopsy under general anesthesia. This biopsy was only obtained after all imaging studies had been completed. The reference standard was histologically confirmed squamous cell carcinoma. In order to exclude biopsy sample

error, patients with a negative biopsy had to remain recurrence free for 6 months of clinical follow up after biopsy. If a biopsy during the follow up showed a recurrence of HNSCC, the patient was considered to initially have local recurrence. Similarly, patients not suspected of local recurrence on DW-MRI or PET-CT were followed for 6 months to reduce the possibility of a recurrence undetected by imaging.

Observers

Ten head and neck radiologists, representing eight different medical institutions, were recruited from a special interest group focused on DW-MRI of the head and neck. These reviewers were asked to specify their years of experience as a radiologist and if their specific clinic included diffusion weighted sequences in routine head and neck scanning protocols.

The observers were informed about patient characteristics, clinical information such as TNM stage, primary tumor location, current symptoms, findings at physical examination and previous treatment. No specific instructions on the interpretation of the DW-MRI sequences was provided.

Observers had access to T1, proton density and diffusion weighted images combined with T1 weighted gadolinium contrast images on a standardized DICOM viewer (RadiAnt DICOM Viewer 1.9.16.7446, Medixant, Poznan, Poland. Available from http://www. radiantviewer.com). ADC maps and diffusion weighting 1000 sec/mm^2 (b1000) were the provided diffusion sequences. Images of the primary tumor before treatment were not supplied.

The reviewers were blinded to the pathology outcome and were unaware of ratio of local control/local recurrence of the ten cases. All patient data was anonymized and reviewed independently by the radiologists. Furthermore, the observers were prohibited from discussing the cases with other physicians or each other. Cases were scored having either local recurrence or local control. If local HNSCC recurrence was scored, the location of the recurrence had to be specified in order to ensure correct interpretation and prevent accidentally correct scoring.

Intraobserver variability

One of the radiologists (F.A.P.) had previously reviewed the ten cases as a principal reviewer in the RETURNED study. This reviewer was blinded to his earlier scores. Time between the first and second scored set varied with a range of 9 to 45 months. The first set of scores was compared to the second set of scores of this reviewer.

Statistical analysis

Interobserver agreement was determined by Fleiss' kappa (κ) using Minitab® Statistical Software v17.2.1 (Minitab Inc, State College, Pennsylvania). Missing scores were omitted from the interobserver analysis. Cohen's kappa (κ) was used to determine intraobserver agreement. Interpretation of κ was done in accordance with the suggestions made by Landis and Koch with values of κ indicating the degree of agreement as: <0.00 poor; 0.00–0.20 slight; 0.21–0.40 fair; 0.41-0.60 moderate; 0.61–0.80 substantial; 0.81–1.00 almost perfect [23]. Difference between groups with distinct experience levels was tested by the Mann-Whitney *U* test using SPSS.

Sample size

In order to estimate the required sample size the R-package 'kappaSize' was used (kappaSize version 1.1, M.A. Rotondi. Available from https://CRAN.R-project.org/package =kappaSize). With the power-based method, a H0 hypothesis of 0.01 and a H1 hypothesis of 0.6 aiming for a power of 0.8 and alpha of 0.05, the required number of cases scored by three observers was eight. Increasing the number of observers reduced the number of required cases with all other variables being equal.

Results

All observers independently scored the ten cases with exception of one radiologist who was unable to score a single case because of technical difficulties. This resulted in a total of 99 distinct scores. 'Local recurrence' was scored in 52 cases of which nine were false positive when compared to the reference standard. The observers scored 47 cases as 'local control' with six of these being false negative.

Table 3 reports the median for accuracy, positive predictive value, negative predictive value, sensitivity and specificity of the observers.

In order to check if the observers had detected a recurrence at the same anatomical location as proven by the reference standard (histopathology), they were instructed to specify the location of the recurrence. All of the locations provided by the radiologists corresponded with the pathology-proven location.

Table 3. Test characteristics for the 10 observers. PPV= positive predictive value. NPV= negative predictive value. The ranges in brackets indicate the worst and the best result among the radiologists for the corresponding category.

	Sensitivity	Specificity	PPV	NPV	Accuracy
Median (range)	100% (60-100)	80% (60-100)	78% (71-100)	100% (67-100)	80% (70-100)

Interobserver agreement

Scores given per case are shown in Figure 1. Overall interobserver agreement was κ=0.55 with a 95% confidence interval of 0.46 – 0.64. Images from cases with the highest (case 10) and lowest (case 5) interobserver agreement are shown in Figure 2 and 3, respectively.
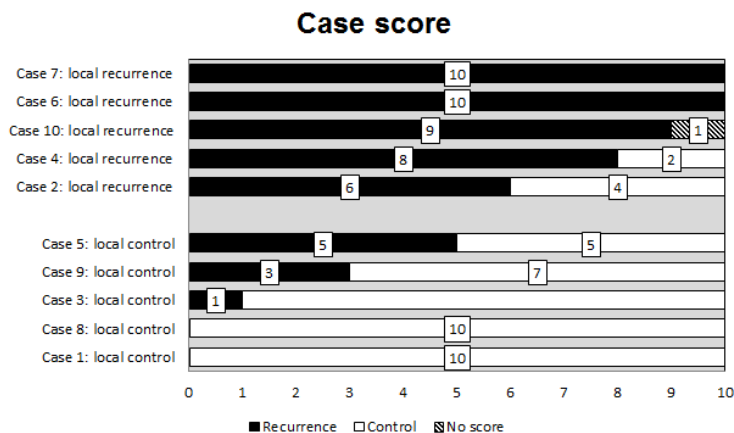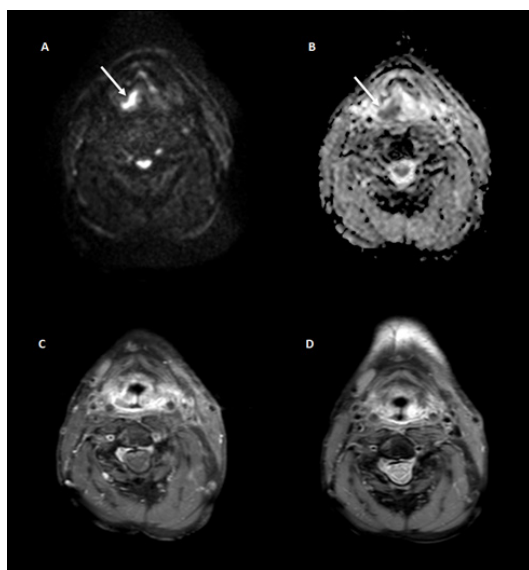

Figure 1. Observer scores per case.

Figure 2. Axial MR images with different contrasts of one of the cases with high interobserver agreement (case 10). Clinical information available to the observers included the following details: *53 year old male with primary T1N0M0 oropharyngeal carcinoma. Sixteen months after treatment with CRT, fiberoptic endoscopy showed an abnormality of the right oropharynx extending into the hypopharynx.* The superimposed white arrows indicate regions with low apparent diffusion indicative for diffusion restriction which is suspected for malignancy. The patient had a biopsy proven local recurrence. A: Diffusion weighted b = 1000 s/mm2, B: ADC, C: T1w SPIR + Gd contrast, D: PD SPIR.
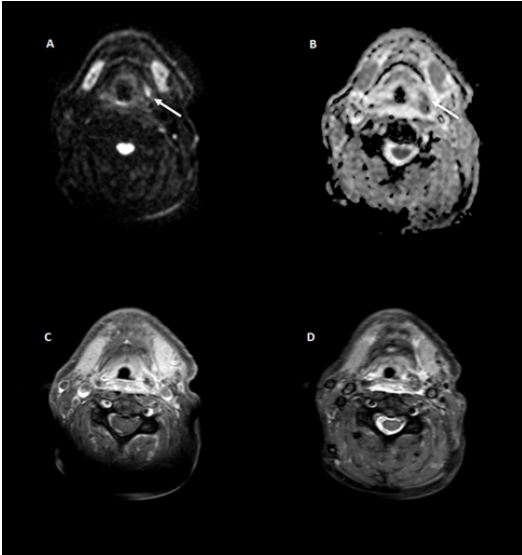
Figure 3. Axial MR images with different contrasts of the case with the lowest interobserver agreement (case 5). Clinical information available to the observers included the following details: *74 year old female with T4aN2bM0 carcinoma of the left tonsillar region extending to the lateral pharyngeal wall treated with radiotherapy. Now suspected of recurrence.* The white arrow indicates a region with low apparent diffusion indicative for diffusion restriction which is suspected for malignancy. However, for this patient biopsy did not show carcinoma and no recurrence occurred during follow up. The diffusion restriction could be based on postradiation effects. It must be noted that the location of the diffusion restriction does not match the localization of the primary tumor. This is a possible explanation why five observers scored this case as a local control despite the restricted diffusion. A: Diffusion weighted b = 1000 s/mm^2, B: ADC, C: T1w SPIR + Gd contrast, D: PD SPIR.

Intraobserver variability

One reviewer scored every case twice. Scoring was similar in 9 of 10 cases. Intraobserver variability was κ=0.80 indicating substantial agreement between both sets.

The first set was completed by the reviewer without errors; the second set contained one false positive diagnosis of a recurrence of HNSCC.

Experience analysis

Six radiologists represented five institutions which included DW-MRI in the standard oncologic follow-up head and neck protocol. The four remaining radiologists from three institutions used HNSCC protocols where DW-MRI was not generally included. There was no significant difference in the number of errors made between both groups. Out of ten cases the first and second group respectively scored on average of 1.67 and 1.25 cases incorrectly.

The average number of cases scored incorrectly by radiologists with less than 10 years of experience as a radiologist was 1.33 (n=6), and 1.75 (n=4) for those with more than 10 years of experience. This difference was not statistically significant.

Discussion

This study investigates the interobserver agreement, intraobserver agreement and influence of experience of radiologists assessing DW-MR images of patients suspected of local recurrent HNSCC after (C)RT. A moderate (κ=0.55) interobserver agreement was found between the ten observers if standard clinical and DW-MRI sequences are provided. The process of reviewing head and neck MR images is generally described to be challenging [24]. Modern MRI sequences such as DW-MRI are increasingly subject of research and might help to discriminate between recurrent disease and radiation effects. The diagnostic performance of MRI including DW images in the detection of recurrent HNSCC after CRT is described to be superior to other imaging modalities with remarkable high positive and negative predictive values [13, 25]. The applicability of these results in daily clinical practice is highly dependent on the ability of radiologists to score images correctly.

The radiologists in our study had access to conventional MRI sequences as well as DW-MRI sequences. This resembles standard clinical practice.

Since the observers in this study represented almost all academic and oncological centers treating HNSCC in the Netherlands, our results reflect the actual diagnostic skill reading DW- MR images in the Netherlands and adds to the applicability of the results to a clinical setting.

Although the interobserver agreement seems limited, the findings of this study provide a cautionary note when relying on MRI including DW-MRI for the diagnosis of recurrent HNSCC. In at least four of the ten cases (five cases if one missing score is excluded), the ten observers were in complete agreement. This suggests that in some cases uniform scores on MRI with DW-MRI are possible.

The cases with mixed scores might be particularly difficult cases to score. It is conceivable that a radiologist would refrain from giving a definite conclusion if confronted with these cases in daily practice. In this study the observers found more false positives than false negatives (9 vs 6 out of 99 distinct scorings). This might be due to the study design requiring reviewers to score either local recurrence or local control. As a missed HNSCC recurrence is arguably worse for an individual patient than a false positive diagnosis of recurrence, reviewers were possibly more likely to score a 'local recurrence' when they felt imaging was inconclusive. In short, the interobserver agreement of MRI with DW-MRI is very good in many cases but overall can only be considered moderate.

Other studies on interobserver agreement of diagnostic imaging in patients with a recurrence of HNSCC showed similar results as compared to our study. Brouwer et al. [26] and Van der Putten et al. [27] examined the interobserver agreement of radiologists using FDG-PET to determine a recurrence of HNSCC. Both showed moderate agreement with a kappa of 0.54 and 0.55 respectively. Fakhry et al. [28] also investigated FDG-PET while comparing it to FDG-PET-CT. Both modalities showed high interobserver agreement; however, this is reported using intraclass correlation and not a kappa statistic, which makes a direct comparison difficult.

Schouten et al. [29] used DW-MRI and PET-CT to detect residual lymph node metastases after CRT. They used a dichotomous system and Likert scale in order to classify the scans. This resulted in several values for interobserver agreement with kappa having a range of 0.53-0.64 for the PET-CT and 0.35-0.58 when scoring DW-MR images.

It seems that interobserver agreement when scoring for a recurrence of HNSCC is moderate regardless whether FDG-PET, PET-CT or DW-MRI is used.

In our study, the intraobserver variability based on the results from one radiologist was substantial. When reviewing the first set, the observer made no errors, during the second assessment one error was made. However, when reviewing the first set the radiologist had access to diagnostic images of the primary tumor which were not provided in the present study. The case incorrectly scored by this observer was the case with the lowest interobserver agreement (case 5). The case was scored erroneously as a local recurrence by five observers in total. After reviewing the images the mistake is understandable as there appears to be reduced ADC indicating diffusion restriction on several images, suggesting recurrence of malignancy. The high intraobserver agreement has to be interpreted with caution as it is based on the assessment of just one radiologist.

Observer experience with DW-MRI was not significantly related with the number of errors of individual radiologists. For DW-MRI imaging of the brain, liver and prostate, this effect has been shown [30–32]. Furthermore, experienced reviewers tended to have a higher interobserver agreement when scoring for prostate cancer than less experienced reviewers [33]. In this study we could not confirm this effect for the diagnosis or recurrence of HNSCC.

Limitations

The radiologists were presented with two options for scoring each case. In daily practice, imaging is sometimes unclear and is deemed to be inconclusive. As patients with 'inconclusive' imaging often receive the same diagnostic procedures as patients with positive imaging, we decided to disallow 'inconclusive' scorings. However, this limited freedom of choice possibly influenced our results on interobserver agreement. As kappa is a statistic dependent on the chance agreement, studies with a high expected chance of agreement will produce a lower kappa than studies with the same results but a lower expected chance of agreements.

8

Radiologists did not have access to imaging of the primary tumor. Information concerning localization and appearance had to be obtained from the clinical information. It is possible that individual observers with access to pretreatment imaging of the primary tumor might be more accurate in scoring for recurrences.

Conclusion

This study reports on an overall moderate interobserver agreement for the detection of recurrence of HNSCC using DW-MRI. Despite the relatively new technique in HNSCC, this study shows that the MR images including DW-MRI are often similarly scored by different radiologists. However, in some select cases agreement is very low, suggesting that in these cases MRI with DW-MRI is unable to reliably diagnose recurring disease.

References

1. Grégoire V, Langendijk JA, Nuyts S (2015) Advances in Radiotherapy for Head and Neck Cancer. J Clin Oncol 33:3277–3284 . doi: 10.1200/JCO.2015.61.2994

2. Dandekar M, D'Cruz A (2014) Organ preservation strategies: Review of literature and their applicability in developing nations. South Asian J Cancer 3:147–150 . doi: 10.4103/2278-330X.136764

3. Bourhis J, Le Maitre A, Baujat B, et al (2007) Individual patients' data meta-analyses in head and neck cancer. Curr Opin Oncol 19:188–194 . doi: 10.1097/CCO.0b013e3280f01010

4. Doweck I, Denys D, Robbins KT (2002) Tumor volume predicts outcome for advanced head and neck cancer treated with targeted chemoradiotherapy. Laryngoscope 112:1742–1749 . doi: 10.1097/00005537-200210000-00006

5. Merkx MAW, van Gulick JJM, Marres HAM, et al (2006) Effectiveness of routine follow-up of patients treated for T1-2N0 oral squamous cell carcinomas of the floor of mouth and tongue. Head Neck 28:1–7 . doi: 10.1002/hed.20296

6. Wong LY, Wei WI, Lam LK, Yuen APW (2003) Salvage of recurrent head and neck squamous cell carcinoma after primary curative surgery. Head Neck 25:953–959 . doi: 10.1002/hed.10310

7. Digonnet A, Hamoir M, Andry G, et al (2013) Post-therapeutic surveillance strategies in head and neck squamous cell carcinoma. Eur Arch Otorhinolaryngol 270:1569–80 . doi: 10.1007/s00405-012-2172-7

8. Wang SJ (2015) Surveillance radiologic imaging after treatment of oropharyngeal cancer: a review. World J Surg Oncol 13:94 . doi: 10.1186/s12957-015-0481-1

9. Isles MG, McConkey C, Mehanna HM (2008) A systematic review and meta-analysis of the role of positron emission tomography in the follow up of head and neck squamous cell carcinoma following radiotherapy or chemoradiotherapy. Clin Otolaryngol 33:210–222 . doi: 10.1111/j.1749-4486.2008.01688.x

10. Kostakoglu L, Fardanesh R, Posner M, et al (2013) Early detection of recurrent disease by FDG-PET/CT leads to management changes in patients with squamous cell cancer of the head and neck. Oncologist 18:1108–1117 . doi: 10.1634/theoncologist.2013-0068

11. Purohit BS, Ailianou A, Dulguerov N, et al (2014) FDG-PET/CT pitfalls in oncological head and neck imaging. Insights Imaging 5:585–602 . doi: 10.1007/s13244-014-0349-x

12. Yao M, Smith RB, Graham MM, et al (2005) The role of FDG PET in management of neck metastasis from head-and-neck cancer after definitive radiation treatment. Int J Radiat Oncol Biol Phys 63:991–999 . doi: 10.1016/j.ijrobp.2005.03.066

13. Vandecaveye V, de Keyzer F, Vander Poorten V, et al (2006) Evaluation of the larynx for tumour recurrence by diffusion-weighted MRI after radiotherapy: initial experience in four cases. Br J Radiol 79:681–687 . doi: 10.1259/bjr/89661809

14. Chawla S, Kim S, Wang S PH, Chawla S, Kim S, et al (2009) Diffusion-weighted imaging in head and neck cancers. Futur Oncol 5:959–975 . doi: 10.2217/fon.09.77.Diffusion-weighted

15. Wang J, Takashima S, Takayama F, et al (2001) Head and neck lesions: characterization with diffusion-weighted echo-planar MR imaging. Radiology 220:621–630

16. Vandecaveye V, De Keyzer F, Nuyts S, et al (2007) Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: correlation between radiologic and histopathologic findings. Int J Radiat Oncol Biol Phys 67:960–971 . doi: 10.1016/j.ijrobp.2006.09.020

8

17. Noij DP, Boerhout EJ, Pieters-Van Den Bos IC, et al (2014) Whole-body-MR imaging including DWIBS in the work-up of patients with head and neck squamous cell carcinoma: A feasibility study. Eur J Radiol 83:1144–1151

18. Varoquaux A, Rager O, Lovblad K-O, et al (2013) Functional imaging of head and neck squamous cell carcinoma with diffusion-weighted MRI and FDG PET/CT: quantitative analysis of ADC and SUV. Eur J Nucl Med Mol Imaging 40:842–852 . doi: 10.1007/s00259-013-2351-9

19. Verhappen MH, Pouwels PJW, Ljumanovic R, et al (2012) Diffusion-weighted MR imaging in head and neck cancer: Comparison between half-Fourier acquired single-shot turbo spin-echo and EPI techniques. Am J Neuroradiol 33:1239–1246

20. Driessen J, Philippens M, Huijbregts J, et al (2016) Prospective Comparative Study of Diffusion-Weighted MRI Versus FDG Positron Emission Tomography/Computed Tomography for the Detection of Recurrent Head and Neck Squamous Cell Carcinomas After (Chemo)Radiation Therapy. Int J Radiat Oncol 96:E337–E338 . doi: 10.1016/j.ijrobp.2016.06.1475

21. Kottner J, Audige L, Brorson S, et al (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud 48:661–671 . doi: 10.1016/j.ijnurstu.2011.01.016

22. Verduijn GM, Bartels LW, Raaijmakers CPJ, et al (2009) Magnetic resonance imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors. Int J Radiat Oncol Biol Phys 74:630–636 . doi: 10.1016/j.ijrobp.2009.01.014

23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

24. de Bree R, Putten L v d, Brouwer J, et al (2009) Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 45:386–393

25. Tshering Vogel DW, Zbaeren P, Geretschlaeger A, et al (2013) Diffusion-weighted MR imaging including bi-exponential fitting for the detection of recurrent or residual tumour after (chemo) radiotherapy for laryngeal and hypopharyngeal cancers. Eur Radiol 23:562–569 . doi: 10.1007/s00330-012-2596-x

26. Brouwer J, de Bree R, Comans EFI, et al (2008) Improved detection of recurrent laryngeal tumor after radiotherapy using (18)FDG-PET as initial method. Radiother Oncol 87:217–20 . doi: 10.1016/j.radonc.2008.02.001

27. van der Putten L, Hoekstra OS, de Bree R, et al (2008) 2-Deoxy-2[F-18]FDG-PET for detection of recurrent laryngeal carcinoma after radiotherapy: interobserver variability in reporting. Mol Imaging Biol 10:294–303 . doi: 10.1007/s11307-008-0154-3

28. Fakhry N, Lussato D, Jacob T, et al (2007) Comparison between PET and PET/CT in recurrent head and neck cancer and clinical implications. Eur Arch Otorhinolaryngol 264:531–538 . doi: 10.1007/s00405-006-0225-5

29. Schouten CS, de Graaf P, Alberts FM, et al (2015) Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT. Oral Oncol 51:541–7 . doi: 10.1016/j.oraloncology.2015.01.017

30. Fukumoto W, Nakamura Y, Higaki T, et al (2015) Additional Value of Diffusion-weighted MRI to Gd-EOB-DTPA-enhanced Hepatic MRI for the Detection of Liver Metastasis: the Difference Depending on the Experience of the Radiologists. Hiroshima J Med Sci 64:15–21

31. Fiebach JB, Schellinger PD, Jansen O, et al (2002) CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. Stroke 33:2206–2210

32. Gaziev G, Wadhwa K, Barrett T, et al (2016) Defining the learning curve for multiparametric magnetic resonance imaging (MRI) of the prostate using MRI-transrectal ultrasonography (TRUS) fusion-guided transperineal prostate biopsies as a validation tool. BJU Int 117:80–86 . doi: 10.1111/bju.12892

33. Rosenkrantz AB, Lim RP, Haghighi M, et al (2013) Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI. AJR Am J Roentgenol 201:W612-8 . doi: 10.2214/AJR.12.10173

8

Chapter 9

Optimal use of MRI for modern radiotherapy in head and neck cancer patients: a guidebook

Boris Peltenburg, Tristan van Heijst, Marielle E.P. Philippens,
Remco de Bree, Chris H.J. Terhaard, Niels Raaijmakers

Abstract

Patients with head and neck squamous cell carcinoma (HNSCC) are conventionally treated with radiotherapy with or without concomitant chemotherapy. Despite the innovations in treatment procedures over the years, the recurrence rate of HNSCC remains high with around one third of patients experiencing a loco-regional recurrence within 3 years after treatment. In order to improve outcome, recent developments in radiotherapy have focused on increasing visibility of the tumor by introducing MRI to the treatment process. The high soft tissue contrast of MRI together with its versatility can prove to be important at each stage of treatment, from diagnosis to the post therapy phase. However, optimal utilization of MRI entails more than simple acquisition of MRI at the different stages. By carefully designing the process of patient logistics, patient setup, scanning protocol and data analysis, a synergistic effect of any additional imaging can be reached. Our institution was one of the centers where a hybrid MRI-radiotherapy device was first developed and as a result, broad experience concerning head and neck MRI was gained over the years. This report is an account of these experiences. Its main aim is to provide a framework that can be used to ensure the optimal use of MRI in the head and neck radiotherapy setting. The first part of this article explains how we use a PDCA cycle to implement changes to the MR imaging protocol of HNSCC patients. In the second part, we share our current practices in MR imaging. The last part of the report provides an example of how we used the PDCA cycle to improve the detection of recurrent disease with MRI in our center.

It should be possible for radiotherapy centers to use the information presented in this article as a template for implementation, incorporation and analysis of MRI into their own work-flow.

Introduction

Head and neck cancer is the seventh most occurring cancer in the world [1]. Squamous cell carcinomas are the most common form of head and neck cancer. Head and neck squamous cell carcinoma (HNSCC) is mostly treated with either surgery or radiotherapy and has a high recurrence rate. Depending on the initial tumor stage, around a third of patients will experience a recurrence within three years after initial therapy [2]. Surgical resection of the primary tumor without loss of function (e.g. swallowing and speech) is often not possible due to the sensitive nature of the area these tumors occur in. In these cases, radiation therapy is often preferred as this gives the patient a chance to better retain functionality. However, radiation therapy has other issues compared to surgery, most notably radiation damage to healthy tissue. In recent years, radiotherapeutic techniques such as IMRT and VMAT have helped reduce radiation induced damage to healthy tissue. Moreover, studies indicated that the management of HNSCC can be further improved by adequate imaging of the tumor before, during and after treatment [3, 4]. The use of imaging to increase the accuracy of the radiation treatment has been dubbed image guide radiotherapy (IGRT). Conventionally, CT has been used in the radiotherapy setting as the imaging modality of choice due to the wide availability of CT scanners and because electron densities required for dose calculations can be extracted from Hounsfield Units in a CT. However, MRI may replace CT as the primary imaging modality thanks to its increased soft-tissue contrast, its ability to image anatomical and functional processes and the lack of ionizing radiation [5–7].

A recent development in the field of IGRT is the introduction of hybrid devices for MRI and radiotherapy such as the MR-Linac (Elekta, Surrey, UK) and MRIdian (ViewRay, Cleveland, USA) [8, 9]. Adoption of these devices will result in MR imaging of HNSCC patients before, during and after radiotherapy treatment to become standard practice in many clinics around the world. As MRI in radiotherapy has not yet become mainstream for head and neck radiotherapy, it is very important to share experience and knowledge on what to do with longitudinal MRI data of patients. Our institution was one of the centers where a hybrid MRI-radiotherapy device was developed and as a result a significant amount of experience concerning head and neck MRI has been acquired over the years [5, 10]. One of the factors found to be of importance was the interdependency of MR images acquired at different time points. Due to the longitudinal nature of MR exams before, during and after radiotherapy, changing any of the exams will influence its comparability to the others. This is true for many factors of the MRI whether it is the imaging protocol, the timing of the imaging or the positioning of the patient. To prevent this, any time an MRI exam is added, changed or removed from the normal HNSSC monitoring protocol, the reason for the change should be clearly determined and the potential influence on imaging at other imaging time points should be examined.

9

Using a standardized way to plan and assess changes to the monitoring protocol may prove to be essential in modern day radiotherapy. Therefore, this report aims at providing a framework that can be used to ensure the optimal use of MRI in the head and neck radiotherapy setting. The first part of this article explains how we use a PDCA cycle to implement changes to the MR imaging protocol of HNSCC patients. In the second part, we share our current practices in MR imaging. The last part of the report provides an example of how we used the PDCA cycle to improve the detection of recurrent disease with MRI in our center.

It should be possible for radiotherapy centers to use the information presented in this article as a template for incorporation and analysis of MRI into their own work-flow.

Part 1. Planning and evaluating changes to the head and neck radiotherapy imaging protocol using a PDCA cycle.

The introduction of online replanning and intratreatment response assessment made possible by hybrid MRI-radiotherapy devices is expected to result in more frequent MR imaging of patients with head and neck cancer. The additional MRI exams are expected to be helpful during the treatment process. For example they might facilitate intratreatment plan adaptations or they will form the basis for early response assessment. Regardless of the reason, any changes made to the timing and frequency of MRI scans, the imaging protocols or patient set-up should be carefully planned and systematically evaluated. Without this it is very difficult to determine if changes to the MRI protocol help improve or maybe even worsen treatment outcomes.

We therefore present a PDCA-cycle that can be used to systematically implement changes to the MR imaging protocol of HNSCC patients in the radiotherapy setting.

A PDCA cycle is a four-step method used for continuous improvement [11]. The four steps we use to implement changes to the MRI monitoring protocol are described below and shown in figure 1. In part 3 of this article we present an example of how we used the PDCA cycle to improve the detection of recurrent disease.

It is important to note that there are two types of MRI monitoring. Firstly, there is the monitoring of individual patients at diagnosis, during treatment and after treatment. Secondly, we identify the monitoring of an entire patient population over time. Consequently, population monitoring is more focused on group outcomes: hard outcomes such as survival and toxicity but also 'soft' outcomes such as patient reported

outcome measures (PROMs). A successful implementation of MRI protocol changes has to take both types of monitoring into account.

1. Plan

The PDCA is used when changes to the MRI monitoring protocol are needed. This can be the result of additions to the radiotherapy treatment protocol such as weekly plan adaptations, or because changes are needed to acquire new information such as early response prediction, or because the type and availability of MR imaging changes for example when a hybrid MRI-radiotherapy is acquired.

During the planning stage the first and most important step is to determine the specific aim of the change. Without a proper aim there is no way to evaluate if a change was successfully implemented.

Also the type of change is determined, examples of this are changes to the scanning protocol, changes to the timing and frequency of the imaging or changes to the amount of monitoring appointments. Then the effect on the complete longitudinal chain of regularly acquired MRI examinations is assessed. Changes in the frequency of MRI exams may for instance make imaging at other time points obsolete, or changes to the scanning protocol such as the addition of a new sequence may also have to be implemented for exams at other time points in order to improve comparability between the scans. Finally, the methods and metrics for evaluating the change are specified and the time frame determined. An improvement cycle may end after a predetermined number of months have passed or after a certain amount of analyzed patients has been reached.

In our center the planning phase is the responsibility of a head and neck working group. This working group consists of radiation oncologists, head and neck surgeons, clinical physicists, radiologists, pathologists and head and neck researchers. Decision on monitoring changes are reached by consensus. Aside from their membership of the monitoring working group, most members of this group are also responsible for changes in the treatment protocol for HNSCC patients.

2. Do

The plan from the previous phase is executed in the Do phase. Evaluation metrics are recorded. This includes outcome measurements. In this step, individual patients receive imaging for diagnosis, treatment response and recurrence detection. It is here that the monitoring protocol should work to the benefit of individual patients.

3. Check

Once the time frame for the data collection has passed the acquired information is analyzed as specified during the planning phase. Results are presented to the head and neck working group.

4. Act

The working group draws a conclusion from the results obtained during the check phase. If it is decided that the change has reached its predetermined objective and if any negative effects remain within tolerance the change is implemented in the new monitoring protocol. If this is not the case a decision has to be made: either the change is discarded completely or a new planning phase is starting using the experience gained from the previous PDCA cycle.

Conclusion

This PDCA provides a basis for implementing changes to the MRI monitoring protocol of HNSCC patients treated with radiotherapy. This framework can be used and copied by other centers in order to optimizing the use of MRI in their own treatment process based on their specific patient population and availability of assets. In part 2 of this article we share our own current practices in MR imaging.
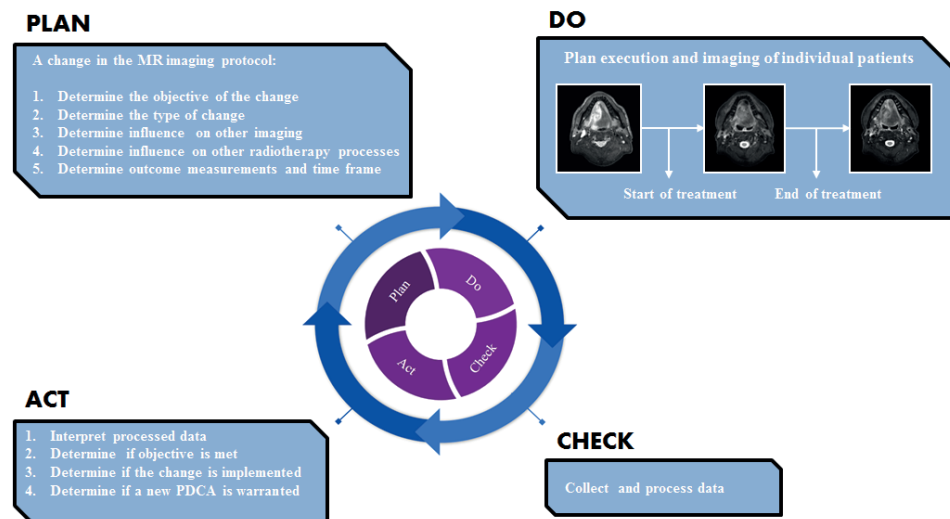
Figure 1. The PDCA cycle for implementing changes to the MR imaging protocol of HNSCC patients treated with radiotherapy.

Part 2. Current practices in patient monitoring

Our protocol for the optimal use of MRI for the monitoring of HNSCC patients treated with (chemo)radiotherapy is based on several key elements:
1. Uniform patient positioning for all imaging,
2. Identical imaging protocols for each MRI examination performed,
3. Minimization of patient discomfort.

Here, we will discuss how these key elements are implemented in our monitoring protocol.

Patient logistics prior to treatment
When patients are referred to our tertiary care facility, the suspicion of a primary HNSCC is often very high or in some cases already proven by histopathological examination of a biopsy. The first appointment is at the outpatient clinic of the department of Head and Neck Surgical Oncology. Here, patients are seen by a head and neck surgical oncologist and a radiation oncologist. If not yet performed, biopsies are taken at the outpatient clinic or at the operating room depending om location of the primary tumor. The lesion is clinically staged. An MRI and CT in radiotherapy positioning mask are planned for all patients likely to be treated with (chemo)radiotherapy. For some patients, the CT scan is replaced by a PET/CT based on the clinical tumor stage. Patients suspected of T1 laryngeal carcinoma do not receive an MRI as these tumors are often poorly visible. A weekly multidisciplinary meeting including radiation oncologists, head and neck surgeons, oncologists, pathologists and radiologists propose treatment strategies for individual patients. Based on their recommendation and patient's preference, the patient is referred to the appropriate department for treatment.

Layout of the radiotherapy department of the UMC Utrecht
Patients at the radiotherapy department of the UMC Utrecht can be treated on one of the nine clinical linear accelerators. Additionally, the department contains two clinical hybrid MR-Linacs (Elekta Unity). Two more MRI scanners are in use for treatment purposes, one is dedicated to MRI-guided brachytherapy and one facilitates the MRI high-intensity focused ultrasound treatment. Finally, diagnostic imaging and follow up imaging is done on one of two diagnostic MRI scanners (3.0T Ingenia and 1.5T Ingenia, Philips). Both scanners allow for patients to be scanned in radiotherapy position, i.e. with a rigid table surface and with the possibility to attach radiotherapy masks to limit motion. Routine imaging of head and neck cancer patients is performed on the 3.0T MRI.

9

Imaging protocol

If a patient is referred for radiotherapy, the initial MRI (already made in mask) is used for treatment planning. In the primary setting these MR images were used for the diagnosis of malignant disease, the general tumor size, the invasion of surrounding tissues and the presence of lymph node metastases. This poses a problem because diagnosis and radiotherapy planning essentially focus on different issues. The questions that have to be answered during diagnosis are more often dichotomous (i.e. is there a tumor?) whereas question answered during treatment planning are more dynamic (i.e. where is the tumor?). As it is the most efficient and patient friendly to combine both the diagnostic and treatment planning purpose in one MRI examination, the MRI protocol has to facilitate the different questions. For this, we held several multidisciplinary meetings which included head and neck radiologists, radiation oncologists and clinical physicists. This resulted in the following protocol which is sporadically evaluated and updated (Table 1).

Additionally, comparisons between tumor size and functional parameters over time can only be done reliably when imaging parameters are kept constant. This means that the imaging protocol should remain constant throughout the treatment episode of the patient meaning that pretreatment MRI, intratreatment imaging and follow up scans up to the point of a potential recurrence all have to be made with the same protocol. Therefore, the presented protocol also accommodates response assessment and recurrence diagnosis while still keeping the time a patient is positioned in the MRI to a minimum.

Table 1. Sequences of the MRI protocol used in our center. TE is echo time and TR is repetition time.

Sequence	Voxel size (mm)	TE (ms)	TR (ms)	Additional information
mDIXON				
Transverse T2 TSE	0.94 x 0.94 x 3.00	100	9023	
Transverse DWI SPLICE	1.46 x 1.46 x 5.00	74	5233	(b0, b200 and b800)
Sagittal T1 FFE cine	1.42 x 1.42 x 10.0	1.45	3.2	Scan duration: 60 seconds.
Transverse T1 3D FFE mDIXON	0.85 x 0.85 x 1.50	1.97	6.8	
Transverse T1 TSE mDIXON	0.73 x 0.73 x 2.00	6.2	678	
Dynamic SPIR	1.09 x 1.09 x 4.00	2.0	4.1	
Transverse T1 TSE mDIXON post contrast	0.73 x 0.73 x 2.00	13	743	
Coronal T1 TSE mDIXON post contrast	0.78 x 0.78 x 4.00	14	606	

Patient Setup

An important part of treatment planning is the correlation of different imaging modalities. For this, we used a rigid registration method. This method is aided by the uniform positioning of a patient during the different examinations. This uniform positioning is achieved by the use of a head and neck mask and individual base. The mask anchors the patients to the diagnostic device, the individual base provides a cushion for the back of the head to rest in. The combination of mask and base restricts free movement of the head, neck, shoulders and jaw. However, the shape of the mask does not allow for the use of dedicated head-and-neck coils, instead flexible surface coils are used. For the specifics of patient positioning, see our previous papers [5, 12].

During treatment

For the radiation treatment the patient is positioned in the mask as well. No standard MR imaging is performed during the treatment. However, imaging is available for patients participating in studies or if indicated. An example of an indication is a patient with HPV+ tumors with noticeable reduction of tumor mass detected either during physical examination or during cone beam positioning. Treatment of a patient with HNSCC tumor takes 5-7 weeks, as per international guidelines.

Post treatment

After the final fraction of the radiotherapy treatment has been delivered, the personalized mask, together with the individual base is stored at a central storage location in a cardboard box. A tracking code linked to the patient identification number is attached to the box. This tracking code contains information about the location of the box in the storage facility.

Three months after the end of treatment patients return to our radiotherapy department for regular follow up. As part of follow up an additional MRI examination is performed. The patient is positioned by using the stored mask and base. Some patients lose or gain weight in such a way during these three months that the original mask does not fit properly anymore. In these cases, the old base is still used but a new mask is made.

This MRI at 3 months after treatment is the last MRI that all patients receive as part of the standard treatment protocol. Naturally, if at some point later than three months after treatment a recurrence is suspected, the patient will receive an additional MRI. The original mask and base are again used to ensure consistent patient positioning.

The masks are stored for one and a half years before they are discarded. This is mainly due to storage size limitations. HNSCC recurrences can occur up to three years after treatment and the process of creating a larger storage facility has been started at our department.

9

MRI and patient comfort

The transition from CT to MRI as the primary imaging modality for HNSCC patients may have some negative effects on the comfort of patients. To obtain the required sequences, it is often necessary for the patient to remain stationary in the MRI scanner for more than 30 minutes. The acoustic noise generated inside the MRI scanner combined with the restricted mobility due to the mask and the narrow space of the MRI bore could create a feeling of anxiety in patients. In some cases claustrophobia can become such a problem that MRI acquisition becomes impossible. Additionally, the use of the head and neck mask limits measures for hearing protection to ear plugs as the headphones commonly used in MRI examinations to reduce noise are ineffective on a positioning mask. However, the headphones can still be used for communication and for music played to comfort the patient. Adequate precautions should be taken in order to avoid damage to hearing, especially when MRIs are more frequently acquired as noise induced hearing loss is cumulative. Individual synthetically molded earplugs might result in better shielding from the loud noises of the MRI. Furthermore, patient comfort should be increased wherever possible, distracting the patient with music or images shown on a monitor might help. More exotic solutions such as increasing the pleasantness of the MRI noises have also been suggested [13].

Part 3. Showcase of examples from our institute

Described here is an example of how we systematically planned and implemented improvements in the imaging protocol for HNSCC in order to obtain information about the patterns of failure both on an individual patient level as well as on a group level.

Pattern-of-failure

At the start of the project, patients suspected of a recurrence after radiation treatment were often examined with MRI in our center. This MRI exam was usually obtained on MRI scanners at the radiology department and its main purpose was to provide diagnostic evidence confirming or rejecting a recurrence. However, the head and neck working group agreed that the MRI exam could also provide information on the pattern-of-failure in patients with a recurrence. This pattern could possibly be determined if the location of the recurrence could be directly compared to the initial dose delivered to that area. The information provided would allow the distinction between in-field, marginal and out-of-field recurrences. If recurrences were found to be mainly in high dose areas (in-field), this would suggest errors during patient set-up or inadequate delivery of dose to the target area, or it could imply radioresistance of the tumor. Recurrences in low dose areas (out-of-field) could be caused by errors during delineation or by the presence of

malignant cells not visible on the available imaging. Marginal recurrences might suggest too small CTV margins, or motion during treatment. Obtaining pattern-of-failure would be to the benefit of individual patients as it would provide arguments for or against re-irradiation. Additionally, the entire group of HNSCC patients could benefit as pattern-of-failure analysis might bring to light areas for treatment improvement based on the in-field/out-of-field ratio. In order to obtain the pattern-of-failure analysis, it was decided an additional MRI had to be added to the HNSCC radiotherapy monitoring plan. For this, we used the PDCA cycle described in part 1.

1. Plan

The objective of the additional MRI was clear: it had to provide the answer if a recurrence was in-field or out-of-field. It was decided that the routine MRI exam for recurrence diagnosis should be acquired using the same patient set-up and image contrasts as performed as part of the treatment of the primary tumor. This would likely enhance comparability between examinations. Furthermore, a plan was made to store the personal radiotherapy immobilization mask used during treatment of the primary tumor of each patient. If at any point after the treatment the patient would be suspected of a recurrence, the new MRI exam could be performed with the patient in his or her original mask. This would ensuring minimal differences in patient positioning across the different exams. The mask could only be used if the images of the patient were obtained on the MRI scanners of the radiotherapy department, as the MRI machines at the radiology department were not equipped for scanning patients in the radiotherapy treatment position.

At first, the endpoint was set at proof of principle. If it was possible to determine the pattern-of-failure for just one patient the objective of the first PDCA would have been reached. Steps would follow to determine robustness and completeness data. The time point for evaluating the project was set at one year after the storage of the first mask.

2. Do

The project started with many logistical problems to solve. First, a storage area and system for the masks had to be developed. Second, the MRIs acquired using the sequence used for radiotherapy planning had to be reviewed by radiologists to ensure that it was suitable for diagnostic examination of HNSCC recurrences. One of the more peculiar problems that was not completely anticipated came in the form of the electronic ordering system. Before, head and neck surgical oncologists would do a large part of the follow up of HNSCC patients. When they suspected a recurrence they would order an MRI at the radiology department via the electronic ordering system. What seemed like a simple change to the order – changing the location from radiology to radiotherapy –

proved to be unusually difficult. After many changes to authorizations and scheduling protocols this problem was eventually solved. At the beginning of 2017, the first masks were stored for the purpose of reusing them should the patient be suspected of a local recurrence. Within 1 year around 150 masks were stored and in total ten patients had received an MRI examination in their original mask for a suspected recurrence of disease.

3. Check

For the ten patients, the MR images were registered to the CT used for the dose calculations and made for the treatment of the primary tumor. A radiologist delineated the tumors of patients that had a recurrence proven by histopathological examination. This delineation was then placed on the original dose map and a qualification of in-field, marginal or out-of-field was given.

As expected, the new protocol had not functioned perfectly. For example, some of the stored masks did not belong to HNSCC patients.

4. Act

During this step the head and neck working group was provided the images and delineations of the recurrences such as the one shown in figure 2 and 3.
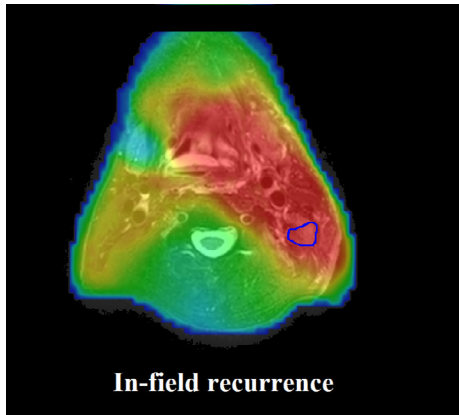
Figure 2. MR image of a patient with an in-field recurrence. The MR image made at the time of the recurrence is shown with an overlay of the dose map made during the treatment of the primary tumor. Dark blue areas represent a cumulative dose of 1 Gy, dark red areas represent a cumulative dose of 70 Gy. The dark blue circle indicates the location of the recurrence. The entire volume of recurrence was positioned inside the high dose area of the initial treatment. This suggests that the recurrence is an in-field recurrence.
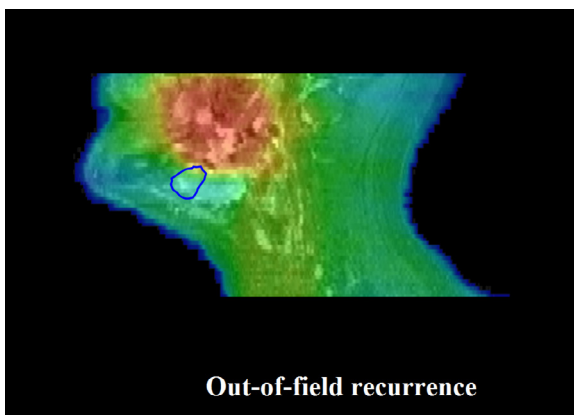
Figure 3. MR image of a patient with an out-of-field recurrence. The image is constructed similar to figure 2 but here the entire volume of recurrence was positioned outside the high dose area indicating an out-of-field recurrence.

The first PDCA cycle had proven that the created workflow for HNSCC patients with the addition of the new MRI scan could determine if recurrences were in-field, marginal or out-of-field. The changes were implemented and are now standard clinical practice in our center. In the next ongoing PDCA cycle, one of its objectives is to improve the quality and amount of recurrence MRI data collected. Further steps may change the protocol, for example by omitting the scanning in the original masks and applying a deformable registration of MRI to CT and by that to the radiation dose. In any case, such a change should be carefully planned and only implemented if it demonstrably improves the monitoring process.

Discussion

In the near future MR imaging of HNSCC patients treated with radiotherapy is expected to become more prevalent as a result of the introduction of hybrid MRI-radiotherapy devices. MR images acquired before, during and after treatment can provide additional information, thereby possibly improving; diagnosis, outcome prediction, delineation, response assessment and recurrence detection. However, efforts have to be made to ensure that additional MRI scans or changes to MR imaging protocols are effective. Care should be taken that patients are not exposed to unnecessary scans. This article provides a framework that can be used to ensure such optimal use of MRI in the head and neck radiotherapy setting. It shows that using a simple model, changes to the monitoring protocol can be planned and evaluated. Using the PDCA cycle improvements can be safeguarded changes without effect can be discarded.

9

References

1. Bray F, Ferlay J, Soerjomataram I, et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424. https://doi.org/10.3322/caac.21492

2. Ho AS, Kraus DH, Ganly I, et al (2014) Decision making in the management of recurrent head and neck cancer. Head Neck 36:144–151. https://doi.org/10.1002/hed.23227

3. Grégoire V, Langendijk JA, Nuyts S (2015) Advances in Radiotherapy for Head and Neck Cancer. J Clin Oncol 33:3277–3284. https://doi.org/10.1200/JCO.2015.61.2994

4. Ko C, Citrin D (2009) Radiotherapy for the management of locally advanced squamous cell carcinoma of the head and neck. Oral Dis 15:121–132. https://doi.org/10.1111/j.1601-0825.2008.01495.x

5. Verduijn GM, Bartels LW, Raaijmakers CPJ, et al (2009) Magnetic Resonance Imaging Protocol Optimization for Delineation of Gross Tumor Volume in Hypopharyngeal and Laryngeal Tumors. Int J Radiat Oncol Biol Phys 74:630–636. https://doi.org/10.1016/j.ijrobp.2009.01.014

6. Gage KL, Thomas K, Jeong D, et al (2017) Multimodal Imaging of Head and Neck Squamous Cell Carcinoma. Cancer Control 24:172–179. https://doi.org/10.1177/107327481702400209

7. Nooij RP, Hof JJ, van Laar PJ, van der Hoorn A (2018) Functional MRI for Treatment Evaluation in Patients with Head and Neck Squamous Cell Carcinoma: A Review of the Literature from a Radiologist Perspective. Curr Radiol Rep 6:2. https://doi.org/10.1007/s40134-018-0262-z

8. Raaymakers BW, Jurgenliemk-Schulz IM, Bol GH, et al (2017) First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol 62:L41–L50. https://doi.org/10.1088/1361-6560/aa9517

9. Mutic S, Dempsey JF (2014) The ViewRay system: magnetic resonance-guided and controlled radiotherapy. Semin Radiat Oncol 24:196–199. https://doi.org/10.1016/j.semradonc.2014.02.008

10. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP (2013) Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. Radiother Oncol 109:394–397. https://doi.org/10.1016/j.radonc.2013.10.004

11. Deming WE (2018) The new economics for industry, government, education. MIT press

12. Houweling AC, van der Meer S, van der Wal E, et al (2010) Improved immobilization using an individual head support in head and neck cancer patients. Radiother Oncol 96:100–3. https://doi.org/10.1016/j.radonc.2010.04.014

13. Ma D, Pierre EY, Jiang Y, et al (2016) Music-based magnetic resonance fingerprinting to improve patient comfort during MRI examinations. Magn Reson Med 75:2303–2314. https://doi.org/10.1002/mrm.25818

Chapter 10

Summary and Discussion

Summary

This thesis shows that before, during and after radiotherapy medical imaging can aid the treatment of HNSCC patients. This appears to be particularly true for MR imaging. It also shows that imaging has to be applied correctly in order to be effective.

In part one, the role of imaging prior to the start of treatment is discussed. The role of FDG PET-CT for the detection of metastatic lymph nodes is examined in chapter 2. Metastatic lymph nodes are routinely diagnosed using cytology acquired by USgFNAC. However, in patients that already received a PET-CT, for example to detect distant metastases, some information on the status of the lymph nodes is already available. Lymph nodes with a SUVmax value of lower than 2.2 never had detectable tumor cells in the cytology, while lymph nodes larger than 1 cm with a SUVmax higher than 4.9 were almost always positive. By using the information freely available from the PET-CT around half of all lymph nodes could be spared an unnecessary USgFNAC examination.

In chapter 3, the power of DW-MRI in the delineation process is examined. Although the conformity of the delineations was good, indicating that DW-MRI could be helpful during the pretreatment delineation phase, we also saw clear instances with a different interpretation of the images. This chapter also demonstrates the usefulness of a semi-automatic delineation process on PET.

Chapter 4 introduces ADC as a quantitative measurement. Although previous research suggested that ADC measured before treatment could be a prognostic factor for local recurrence, we could not confirm this. Moreover, a more easily obtainable disease characteristic, the TNM stage, is significantly more powerful as a prognostic factor than pretreatment ADC.

Part two starts by looking into ADC and in particular the change in ADC as it is measured during treatment. A prospective trial (the PREDICT trial) is described in chapter 5. PREDICT has the primary aim of predicting treatment outcome of HNSCC patients using the change in ADC induced by the radiation therapy. Ideally, this could lead to patient selection based on this change in ADC. Patients with a change in ADC indicating an inadequate response to the initiated radiotherapy would then receive intensified treatment or be switched to surgery. Conversely, patients with a good response would continue on the same or even less intensive radiotherapy course.

Chapter 6 illustrates a potential issue with MRI tumor identification and delineation during treatment. The change in the appearance of tumors on T2 weighted imaging hinders delineation and consequently reduces interobserver agreement. The most effective way to address this issue is probably to create clear and easy to follow guidelines on the delineation of tumors during treatment.

In the last part of this thesis, part 3, imaging in the posttreatment phase is presented. Chapter 7 discusses a prospective comparative study of MRI with diffusion weighted

sequences and PET-CT, the RETURNED study. Both are used to detect local recurrence. The diagnosis of recurrent HNSCC is significantly more challenging than the diagnosis of primary disease. This is due to the difficulty of differentiating between post treatment effects and recurrent tumor. In the study, MRI and PET-CT were both able to detect recurrences with an accuracy of around 70%. However, MRI showed a higher specificity but a lower sensitivity than PET-CT. Unfortunately, by adding the two modalities together the accuracy of detection did not improve much. Therefore we advocate that PET-CT, already routinely used for recurrence detection, should remain the modality of choice if there is clinical suspicion of recurrent tumor.

A reason for the lack of accuracy of the MRI with diffusion weighted images is sought in chapter 8. Here, ten radiologist tried to determine if the MRI images of ten RETURNED patients showed a local recurrence or local control. The interobserver agreement was moderate. The results furthermore showed that there were easy cases where all ten observers were in agreement, and that there were difficult cases where agreement was low. Other studies reported also a similar interobserver agreement for PET-CT and also the presence of easy and difficult cases. Perhaps that MRI examinations performed at regular intervals could improve the detection of recurrent MRI, however this was not examined in this thesis.

Finally, chapter 9 describes a framework for implementing MRI in the radiotherapy process. This is relevant as innovations in the field resulted in hybrid radiotherapy and MRI devices that are currently being introduced to the clinic. It provides answers to questions a radiation oncologist or clinical physicist might have when introducing MRI for HNSCC radiotherapy treatment in their institutions.

Discussion

Imaging of head and neck cancer patients

How can you treat something that you cannot see? A clearly visible tumor is essential for precise targeting. The use of MRI and CT and treatment techniques such as IMRT and VMAT have made precise targeting and accurate dose delivery possible in the modern day radiotherapy treatment of HNSCC patients [1–3]. As a result, patient positioning became more important. Any inter- or intrafraction changes to the positioning would potentially influence treatment effectiveness. By changing position the tumor may be, in essence, partially moving in and out of the high dose field. Margins added to the dose field, fixation techniques and imaging by cone beam CT are currently being used as solutions to positioning errors [4]. But what if the shape and size of the tumor is changing? Usually, a HNSCC patient will be treated with a fractionated radiotherapy scheme, spread out over several weeks [5,6]. The effects of the treatment on the tumor

10

are visible from early on [7]. Moreover, the treatment could not only affect the tumor but the internal anatomy of the area as well. In patients with HNSCC, healthy tissue and organs at risk are often in close proximity to the tumor. Organs at risk that are pushed to the side due to tumor growth or edema can start to return to their original position when the treatment starts to affect the tumor. In the case of a shrinking oropharyngeal tumor a salivary gland, normally meticulously shielded from too many radiation due to its damaging effect, may change position and start to receive high doses of radiation (Figure 1). Anatomy changes could even result in the tumor dropping out of the high dose area potentially receiving substandard treatment.
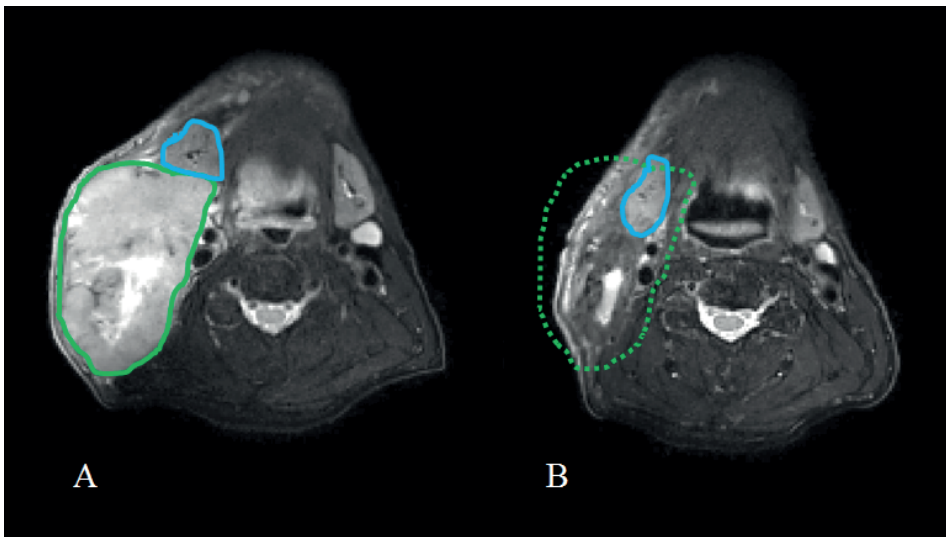
Figure 1. MR images of a patient with a T2N3M0 HPV positive oropharyngeal carcinoma.
A: T2 weighted image acquired prior to the start of treatment. The large metastatic lymph node on the right side of the patient (green outline) pushes the submandibular salivary gland (blue outline) anteriorly. B: T2 weighted images after 4 weeks of chemoradiotherapy treatment. The lymph node has recessed as a reaction to the treatment and the salivary gland now shifted back to its original position. If the metastatic lymph node was planned to receive a high dose the salivary gland would have moved into that high dose area.

For HNSCC, the current standard of care calls for a single treatment plan. This plan is based on the pretreatment imaging of the tumor and is only rarely adapted during treatment to account for changes in the internal anatomy. In any case, these changes are hard to detect as they are often not visible through physical examination or on the low contrast cone beam CT.

However, this all might soon change as weekly or even daily internal anatomy visualization and verification of the tumor position can be realized by the MR-Linac. Developed in Utrecht and in clinical use since 2017, the MR-Linac is a hybrid radiotherapy and MRI device that allows for simultaneous MR imaging and irradiation of the tumor [8,9]. Its ability to conveniently and accurately visualize the tumor during treatment can prove to be the key to even more precise targeting allowing higher dose to the tumor and reducing dose to the surrounding healthy tissues. Furthermore, by performing MRI examinations at a regular interval, a wealth of information regarding individual patients can be acquired. Not only changes in tumor size and shape can be tracked but also the way the therapy affects functional parameters such as diffusion can be examined [10–12]. Additionally, the results of individual HNSCC patients can be pooled and because potentially every HNSCC patient with an indication for radiotherapy can be treated on an MR-Linac, this can result in an extensive amount of MR imaging data becoming available for analysis [13]. The MR-Linac and its use in acquiring MR scans before, during and after radiotherapy touches on many themes discussed in this thesis. The delineation process and prognosis formation, discussed in part 1 and 2, may change due to the introduction of the MR-Linac and the additional scans regularly acquired during therapy may even change the way recurrences are diagnosed, as discussed in part 3. Ideally, some results presented in this thesis are soon eclipsed by the sheer amount of data and new insights obtained using the MR-Linac. However, for this to happen, several subjects of this thesis have to be taken into account. Most notably, the way tumor changes (both anatomically and functionally) and the interobserver agreement of observers while delineating tumors are measured.

Measurements of volume and ADC

As long as we are unable to detect all individual tumor cells on an MR image, measuring the volume of a tumor will always be an approximation [14]. In order to get this approximation as close to reality as possible we have to make subjective choices. First, a decision has to be made about the MRI sequence that is used for the approximation. Does a contrast-enhanced T1 image show the true tumor extent? Or is a T2 weighted image more suited for this purpose? This can further be influenced by the characteristics of the MRI device used such as the field strength, availability of imaging techniques in the software, hardware dedicated to radiotherapy and vendor. Secondly, we have to take into account that for head and neck tumors especially, geometric distortions can occur. The specific layout of the head and neck area with many air-filled cavities in close proximity to soft tissue and dense bone structures give rise to magnetic field inhomogeneities. The inhomogeneities in turn cause the geometric distortions [15–17]. Furthermore, imaging artifacts, caused for example by dental fillings, can occur [18]. Both geometric distortions and artifacts can interfere with determining the tumor

10

volume. Finally, someone has to determine the extent of the tumor while recognizing the limitations of the generated image and dealing with any distortions or artifacts. Each choice made previously will affect the interpretation of the images. This interpretation will differ from person to person even if they are similarly trained and can even differ for the same observer on different days [19]. This inter- and intraobserver variation leads to different approximations of the tumor volume. A computer algorithm might alleviate some of the day-to-day and person-to-person variations, but who will train the computer in learning what is tumor and what is not? Without a true measurement of tumor volume the computer will be no better at determining tumor volumes than the humans training him. Clearly a ground truth is needed. Work has been done on obtaining such a ground truth by examining surgical specimens on a microscopic level and comparing this to MR images [14,20]. Although new insights have been obtained through this research, many more comparisons have to be made before inter- and intraobserver variability can largely be eliminated.

Accurately measuring ADC is even more difficult than measuring volume. ADC is determined by diffusion weighted imaging using multiple sequences with different b-values. Changing the number or characteristics of the b-values may result in different ADC values. Even if the b-values are kept the same, different implementation of the diffusion weighted imaging and different models for calculating ADC exist, each with different end results [21]. Furthermore, diffusion weighted imaging can be highly susceptible to geometric distortions, depending on the technique used [22]. All these factors influence the measurement by a substantial amount, but there is an attribute of DW-MRI that makes determining ADC even more difficult than measuring volume. A person responsible for determining the volume of a tumor on MRI will do so by trying to find all the boundaries between tumor and healthy tissue. These boundaries are subject to his or her interpretation but once they are established, anything within the boundaries is considered tumor and the volume can be calculated. This does not work for determining ADC. ADC is different to volume in the sense that it is a numerical property of each voxel of the MRI. When measuring ADC, the goal is to find this numerical value for all the voxels in the tumor. For this the boundary between the tumor and the surrounding tissue again has to be determined. Therein lies the problem: If the ADC map is used to determine boundary than ADC forms both the selection variable as well as the outcome variable! It is similar to trying to determine the maximum age of all the children in a room full of people of all ages by first establishing that everyone above 12 years old is an adult and then concluding that the maximum age children can be is 12 years old. See figure 2 for an example of this using DW-MRI.

There are several ways to alleviate this problem, some of them are used in this thesis (see chapter 3 and 4) but they too only provide approximations of the tumor boundary. In any case, it is more important to acknowledge that for measuring ADC another choice,

the choice of how to establish the tumor boundary, has to be made. As stated earlier, each choice affects the interpretation of a scan and as a result contributes to inter- and intraobserver variation.
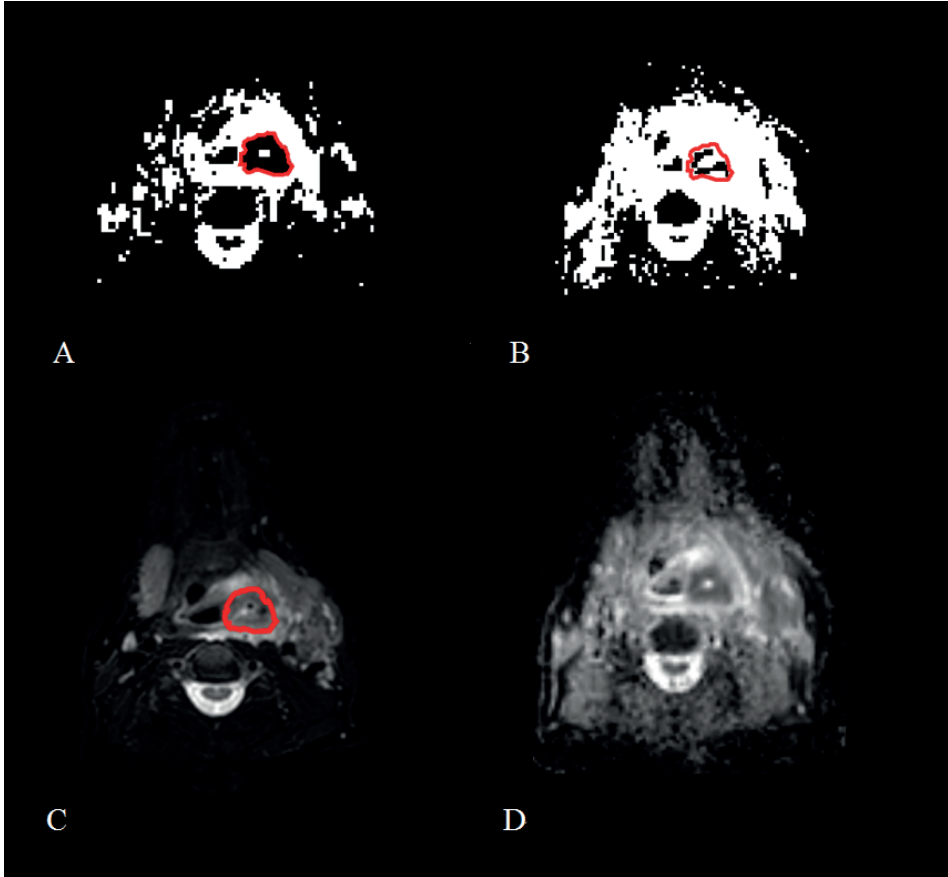

Figure 2. MR images of a patient with a T4aN3M0 hypopharyngeal carcinoma. A and B both represent ADC values but they differ in cutoff values. Every voxel in A with an ADC value of 1.6×10^{-3} mm^2/s or higher is white, voxels with a lower value are black. For B this cutoff value is 1.2×10^{-3} mm^2/s. The red outline is the interpretation of the tumor border using ADC image A. By comparing both maps it is clearly visible how two different interpretations of the tumor extent affect the outline. An observer would probably draw a much smaller red outline on image B than on image A. This would result in a different mean ADC value of the tumor. For comparison, image C and image D are the T2 weighted image and an ADC map with conventional window/level settings.

10

Volume and ADC as a prognostic markers: pretreatment vs. intratreatment

The inter- and intraobserver variation in determining ADC is perfectly illustrated by the research into the prognostic value of ADC in HNSCC patients, measured before treatment commences [11,23–28]. This is discussed in chapter 4 of this thesis. Not only are there conflicting results with studies reporting that pretreatment ADC can predict the outcome of radiotherapy and studies reporting that pretreatment ADC can *not* predict outcome, but the measured values of ADC vary wildly as well. Different choices in measuring ADC will undoubtedly have influenced the differences in reported ADC values.

There is another, seemingly unrelated, reason why ADC measured before treatment is not suited as a prognostic marker. Pretreatment ADC tells you nothing about the reaction of the tumor cells of individual patients to radiotherapy. This is probably true for almost all functional imaging parameters and to some extent even for the volume of tumors. For one patient a large tumor with a high cell density measured by ADC may disappear rapidly when treated with radiotherapy while another patient with seemingly similar tumor characteristics may prove to have a tumor that is unresponsive to the treatment. It is conceivable that patients like the first one will have a better outcome of treatment than patients with tumor behaving as the second one. It sounds almost too simple, but seeing how a tumor responds to radiotherapy will probably tell you more about the effect of the treatment than trying to predict it in advance.

The relation between variability in volume and ADC measurements and the inability of pretreatment parameters to predict a treatment effect is that they both may greatly benefit from imaging during the course of treatment. First, by measuring the ADC of an individual patient prior to treatment and then again during treatment a lot of the variability is reduced: the same sequences can be acquired, distortions and artifacts will be quite similar and for ADC the same method to calculate ADC can be used. Secondly, by comparing tumor characteristics like ADC and volume before treatment with the same characteristics during treatment, a fractional change can be determined (ΔADC) [29]. Again, it is probably more informative to know that the tumor has shrunk 20% during treatment and ADC has risen by 50% than it is to know that the tumor was 20 cm^3 and had a mean ADC of 0.97 x 10^{-3} mm^2/s prior to treatment.

Of course, the MR-Linac could play an important role here. The many MR images obtained during the treatment of HNSCC patients can possibly serve as adequate predictors of treatment outcome. For example, patients with shrinking tumors might be responding well to the therapy and would not require intervention whereas patients with tumors with an unchanged or even higher volume would require reassessment. This could entail an intensification of the treatment by delivering more dose to the tumor or even a change of treatment modality, switching to surgery without completing the ineffective radiotherapy course [30].

Consequently, if we have measured the volume and functional parameters such as ADC prior to treatment we only need to determine the volume and ADC during treatment. But for this to work, both the values should be equally reliable approximations of the real truth. It wouldn't make much sense to compare both values if one of them could not be reliably measured. Are we sure that we can measure volume and ADC with the same certainty during treatment as we did before treatment?

Tumor identification and interobserver agreement during radiotherapy

As shown in chapter 6 of this thesis, delineation of HNSCC becomes more difficult as treatment progresses. During the radiotherapy course, the contrast between tumor and normal tissue generally starts to disappear as tumors start to shrink and surrounding normal tissue changes in appearance most likely due to inflammation and formation of edema [31]. As can be expected, the more a tumor responds to the radiotherapy, the more difficult it becomes to differentiate between remaining malignant areas and non-malignant tissue. The effect is most pronounced in HPV positive oropharyngeal tumors as they tend to respond well to radiotherapy [32]. Already around the third week of treatment, HPV positive tumors can be nearly indistinguishable from surrounding tissue. This poses a problem, if tumors become less visible different observers start to disagree about the boundaries. If interobserver agreement is reduced the tumor characteristics such as volume and ADC cannot reliably be determined. Without reliable determination of the size or functionality of the tumor during treatment, we can not compare it to the situation prior to treatment and we lose the ability to reliably gauge the response to radiotherapy.

Accurate tumor delineation is not only important for volume and ADC measurements, it is also paramount for plan adaptations. Without a clearly visible tumor during treatment, there can be no accurate targeting of the high dose area. As plan adaptation is one of the future promises of the MR-Linac, this problem has to be overcome before it becomes a standard part of the radiotherapy treatment of HNSCC. One of the ways to do this is trough additional training of radiation oncologists. MR imaging during the treatment course is still rare and not a lot of experience has been gained in viewing and interpreting these images yet. The reduced interobserver agreement may be the result of a lack of guidelines for delineating tumors during treatment [33]. Simple rules such as: *do not delineate outside the pretreatment GTV* or *when in doubt stick to the original GTV* may be able to greatly reduce interobserver agreement. It is important to note that these guidelines can increase the precision of delineations but not the accuracy. In order to increase the accuracy of intratreatment delineations they should be compared to a reference standard, preferably histopathological examination. In radiotherapy for HNSCC it is arguably better for radiation oncologists to be a bit wrong but in high agreement than to have high disagreement with some being right and others being

10

wrong. In the first case the lack of accuracy can be analyzed by studying recurrence patterns and adjusted for accordingly, in the second case there is hardly any way to tell who is wrong and who is right and therefore no adjustment is possible.

Eventually, it is here that automation may prove to be most valuable [34]. An algorithm can be trained to always adhere to the same rules when delineating tumors. Furthermore, the large amount of MR imaging data soon to be acquired by many MR-Linac machines will allow for huge training sets for these algorithms to study. The problem remains that there will be no reference standard to correct the algorithm. But maybe the persistence of this problem can be assuaged by the quantitative nature of DW-MRI, using so called functional diffusion maps.

Functional diffusion maps and dose painting

A functional diffusion map takes the best of regular MR imaging and does away with one of the disadvantages. It keeps the reduction in ADC variability by using DW-MR images obtained prior to the start of treatment and DW-MR images obtained during treatment [35,36]. To reiterate: this provides a more stable and patient specific ΔADC instead of a single ADC value. It does not however require an observer or algorithm to determine the boundary between normal tissue and tumor. This is because a functional diffusion map determines the ΔADC of each voxel within a volume of interest, groups them and assigns a color accordingly. For example, the volume of interest can be the original GTV and the voxels can be grouped by highly negative ΔADC (blue), a ΔADC of around zero indicating no change (green) and by highly positive ΔADC (red). This can be done for several different time points as is shown in figure 3. The generated image shows the local change in ADC of the tumor as a reaction to the radiotherapy. As ADC can be a surrogate marker for cellular density, areas of increasing and decreased cell density can be visualized. This can prove to be invaluable to individual patient outcome prediction. For example, the tumor of a patient may show a large rise in median ADC suggesting a reduction in cellular density and consequently an adequate treatment of the tumor. However, the functional diffusion map may show a small but significant area within the GTV that shows no increase in ADC or even a decrease in ADC. This could potentially be a small part of the original tumor that is resistant to the radiation treatment. If research proves that patients with this area with no increase of ADC are more likely to have local recurrences after treatment, it would mean that a functional diffusion map can be a powerful new predictive tool.

From this follows a logical next step. If the small part with no increase of ADC is somehow more radioresistant, would it respond better if a higher dose was applied to that area? And conversely, would areas that show an early sharp decrease in ADC be receiving too much radiation? Can a lower dose in this area still destroy the tumor cells while sparing
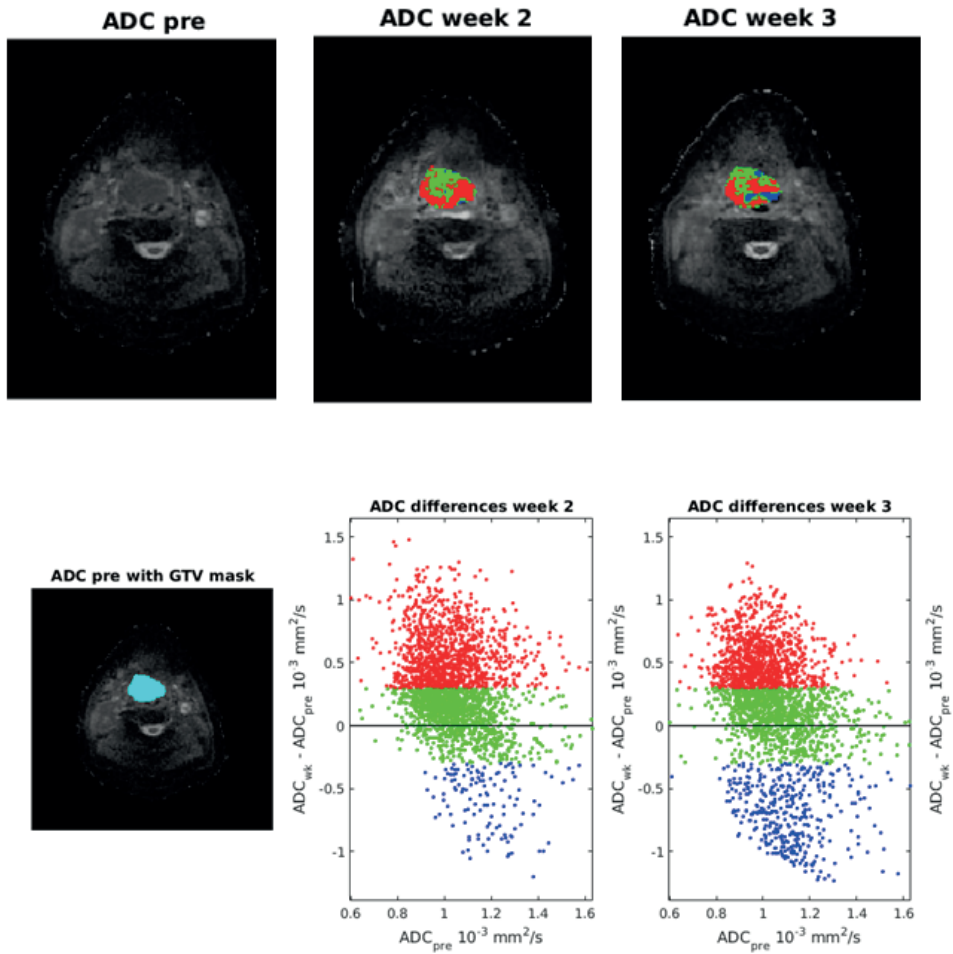
Figure 3. Functional diffusion map of a patient with a T3N2M0 oropharyngeal carcinoma. For all the voxels contained in the GTV mask (lower left image) the pretreatment ADC value is determined. For the scans acquired in week 2 and 3 of the radiotherapy treatment the ADC of the voxels is determined again. Voxels showing an increase in ADC indicating a decrease in cell density are colored red. Conversely, voxels showing a decrease in ADC indicating an increase in cell density are colored blue. Voxels without a significant ADC change are colored green.

10

more of the normal tissue? This process is called dose painting [37]. Clinical evidence for the effectiveness of dose painting for HNSCC tumors does not yet exist. However, with adequate research using the MR-Linac data it may in the future be possible to assign probabilities to a voxel that tell us how likely it is to be a voxel containing tumor cells and its likelihood to be radioresistant. In these cases we can start to change treatment schemes: increasing dose on radioresistant tumor areas and decreasing dose on areas

that are less likely to contain tumor cells. This so called dose-painting by numbers may result in less recurring disease and less toxicity from the radiotherapy treatment. This reduction of dose to certain parts of the tumor may be considered dangerous as it can potentially result in undertreatment of tumors but as long as there is sufficient equipoise, medical trials on the subject can be justified. For without thorough research of HNSCC treatments we will never improve the outcome of individual patients and concerning recurrent disease and treatment toxicity, HNSCC patients deserve less of both.

Patient comfort

A comment has to be made with regard to imaging and patient comfort. Undergoing imaging and especially MRI scans can be quite an uncomfortable experience for some patients. Not only do MRI examinations take a substantial amount of time, with average examinations taking around 30 minutes, but the unnatural space and restricted movement can induce feelings of anxiety in patients. One study even found that around 14% of adult patients in a large tertiary center in the United States required some form of oral sedation, intravenous sedation or general anesthesia in order to endure their MRI scan [38]. Anxiety is not only bad for the patient but bad for the imaging as well. Anxious patients often start to move in the MRI, greatly affecting the quality of the resulting images. For HNSCC patients treated with radiotherapy, the percentage of anxious patients could be even higher than reported. In order for the MR images to correspond to the treatment plan, HNSCC patients are imaged in a radiotherapy immobilization mask. This mask restricts any movement of the head, neck and shoulders. The patient is then positioned with their head in the center of the MRI bore for optimal acquisition. However, this position restricts any view of the outside world. Take into consideration that some patient have a tumor that is obstructing the normal flow of air and therefore already have difficulty with breathing, it is easy to imagine that some patients consider MRI examinations to be particularly unpleasant. Additionally, the radiotherapy mask interferes with the use of a headphone which can normally be used to play music through in order to provide some form of distraction from this patients. Without this headphone only in-ear plugs protect the HSNCC patient from the loud noises produced by the MRI machine due to the changing magnetic fields. This might not be a problem in the current practice, where most patient receive an MRI prior to treatment and maybe two or three more in the following weeks. But with the introduction of the MR-Linac patients could be exposed to the loud noises and experience anxiety for the MRI on a daily basis. It is therefore clear that in order for daily MRI acquisition to become a reality, something has to be done about patient comfort. Adequate protection of hearing can possibly be achieved by changing the in-ear protection. Or perhaps the design of the radiotherapy mask can be changed in such a way that it allows over ear headphones, restoring the ability for patients to listen to music as well. Furthermore, steps should be taken to

reduce anxiety. For example mirrors may allow patients to look out of the MRI device, ideally a television can be positioned in such a way that patients can catch up to their favorite television shows while being treated and scanned on the MR-Linac.

Conclusion and future perspective

In conclusion, medical imaging is of paramount importance for the current treatment of HNSCC with radiotherapy. It plays a role during diagnosis, staging, prognosis, tumor delineation, response prediction and assessment, and in recurrence detection. In the near future, imaging and especially MRI will become even more important thanks to the introduction of hybrid radiotherapy and MRI devices such as the MR-Linac. The stream of new information obtained by regular intratreatment MRI acquisition will undoubtedly result in improvements to the radiotherapy treatment of HNSCC patients. However, thought should be given to the idiosyncrasies of obtaining imaging of tumors at different stages of treatment. Actively treated tumors may be more difficult to detect and interobserver agreement for the delineation of these tumors may be low. Additionally, regular MR imaging may compromise patient comfort to an unacceptable level, with in the worst case patients abandoning treatment due to fear for the MRI machine. However, if these issues are addressed imaging could very well considerably improve patient outcomes.

10

References

1. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol. 2011 Feb;12(2):127–36.

2. Scorsetti M, Fogliata A, Castiglioni S, Bressi C, Bignardi M, Navarria P, et al. Early clinical experience with volumetric modulated arc therapy in head and neck cancer patients. Radiat Oncol. 2010 Oct 15;5(3):93.

3. Bhide SA, Nutting CM. Advances in radiotherapy for head and neck cancer. Oral Oncol. 2010 Jun;46(6):439–41.

4. Parvathaneni U, Laramore GE, Liao JJ. Technical advances and pitfalls in head and neck radiotherapy. J Oncol. 2012;2012:597467.

5. Mazeron R, Tao Y, Lusinchi A, Bourhis J. Current concepts of management in radiotherapy for head and neck squamous-cell cancer. Oral Oncol. 45(4–5):402–8.

6. Grégoire V, Langendijk JA, Nuyts S. Advances in Radiotherapy for Head and Neck Cancer. J Clin Oncol. 2015 Oct 10;33(29):3277–84.

7. Castadot P, Geets X, Lee JA, Christian N, Grégoire V. Assessment by a deformable registration method of the volumetric and positional changes of target volumes and organs at risk in pharyngo-laryngeal tumors treated with concomitant chemo-radiation. Radiother Oncol. 2010 May;95(2):209–17.

8. Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. Radiother Oncol. 2008 Jan;86(1):25–9.

9. Raaymakers BW, Jurgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, van Asselen B, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol. 2017 Nov;62(23):L41–50.

10. Vandecaveye V, Dirix P, De Keyzer F, de Beeck KO, Vander Poorten V, Roebben I, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010 Jul;20(7):1703–14.

11. King AD, Chow K-K, Yu K-H, Mo FKF, Yeung DKW, Yuan J, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology. 2013 Feb;266(2):531–8.

12. Schouten CS, de Bree R, van der Putten L, Noij DP, Hoekstra OS, Comans EFI, et al. Diffusion-weighted EPI- and HASTE-MRI and 18F-FDG-PET-CT early during chemoradiotherapy in advanced head and neck cancer. Quant Imaging Med Surg. 2014 Aug;4(4):239–50.

13. Kerkmeijer LGW, Fuller CD, Verkooijen HM, Verheij M, Choudhury A, Harrington KJ, et al. The MRI-Linear Accelerator Consortium: Evidence-Based Clinical Introduction of an Innovation in Radiation Oncology Connecting Researchers, Methodology, Data Collection, Quality Assurance, and Technical Development. Front Oncol. 2016;6:215.

14. Jager EA, Ligtenberg H, Caldas-Magalhaes J, Schakel T, Philippens ME, Pameijer FA, et al. Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer. Acta Oncol. 2016 Nov;55(11):1305–12.

15. Fransson A, Andreo P, Potter R. Aspects of MR image distortions in radiotherapy treatment planning. Strahlentherapie und Onkol Organ der Dtsch Rontgengesellschaft . [et al]. 2001 Feb;177(2):59–73.

16. Walker A, Liney G, Metcalfe P, Holloway L. MRI distortion: considerations for MRI based radiotherapy treatment planning. Australas Phys Eng Sci Med. 2014 Mar;37(1):103–13.

17. Adjeiwaah M, Bylund M, Lundman JA, Söderström K, Zackrisson B, Jonsson JH, et al. Dosimetric Impact of MRI Distortions: A Study on Head and Neck Cancers. Int J Radiat Oncol Biol Phys. 2019;103(4):994–1003.

18. Hargreaves BA, Worters PW, Pauly KB, Pauly JM, Koch KM, Gold GE. Metal-induced artifacts in MRI. AJR Am J Roentgenol. 2011 Sep;197(3):547–55.

19. 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;10:26.

20. Caldas-Magalhaes J, Kooij N, Ligtenberg H, Jager EA, Schakel T, Kasperts N, et al. The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncol. 2015;54(8):1181–7.

21. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988 Aug;168(2):497–505.

22. Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Technical Note: Diffusion-weighted MRI with minimal distortion in head-and-neck radiotherapy using a turbo spin echo acquisition method. Med Phys. 2017 Aug;44(8):4188–93.

23. Hatakenaka M, Nakamura K, Yabuuchi H, Shioyama Y, Matsuo Y, Ohnishi K, et al. Pretreatment apparent diffusion coefficient of the primary lesion correlates with local failure in head-and-neck cancer treated with chemoradiotherapy or radiotherapy. Int J Radiat Oncol Biol Phys. 2011 Oct;81(2):339–45.

24. Hatakenaka M, Shioyama Y, Nakamura K, Yabuuchi H, Matsuo Y, Sunami S, et al. Apparent diffusion coefficient calculated with relatively high b-values correlates with local failure of head and neck squamous cell carcinoma treated with radiotherapy. AJNR Am J Neuroradiol. 2011;32(10):1904–10.

25. Ohnishi K, Shioyama Y, Hatakenaka M, Nakamura K, Abe K, Yoshiura T, et al. Prediction of local failures with a combination of pretreatment tumor volume and apparent diffusion coefficient in patients treated with definitive radiotherapy for hypopharyngeal or oropharyngeal squamous cell carcinoma. J Radiat Res. 2011;52(4):522–30.

26. Lombardi M, Cascone T, Guenzi E, Stecco A, Buemi F, Krengli M, et al. Predictive value of pre-treatment apparent diffusion coefficient (ADC) in radio-chemotherapy treated head and neck squamous cell carcinoma. Radiol Med. 2017 May;122(5):345–52.

27. King AD, Mo FKF, Yu K-H, Yeung DKW, Zhou H, Bhatia KS, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010 Sep;20(9):2213–20.

28. Chawla S, Kim S, Dougherty L, Wang S, Loevner LA, Quon H, et al. Pretreatment diffusion-weighted and dynamic contrast-enhanced MRI for prediction of local treatment response in squamous cell carcinomas of the head and neck. AJR Am J Roentgenol. 2013 Jan;200(1):35–43.

29. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012 Apr;263(1):19–32.

30. Yuan J, Lo G, King AD. Functional magnetic resonance imaging techniques and their development for radiation therapy planning and monitoring in the head and neck cancers. Quant Imaging Med Surg. 2016 Aug;6(4):430–48.

31. Zbaren P, Weidner S, Thoeny HC. Laryngeal and hypopharyngeal carcinomas after (chemo) radiotherapy: a diagnostic dilemma. Curr Opin Otolaryngol Head Neck Surg. 2008 Apr;16(2):147–53.

32. Shaw R, Robinson M. The increasing clinical relevance of human papillomavirus type 16 (HPV-16) infection in oropharyngeal cancer. Br J Oral Maxillofac Surg. 2011 Sep;49(6):423–9.

10

33. Heukelom J, Fuller CD. Head and Neck Cancer Adaptive Radiation Therapy (ART): Conceptual Considerations for the Informed Clinician. Semin Radiat Oncol. 2019;29(3):258–73.

34. Yang J, Beadle BM, Garden AS, Schwartz DL, Aristophanous M. A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy. Med Phys. 2015 Sep;42(9):5310–20.

35. Hamstra D a, Chenevert TL, Moffat B a, Johnson TD, Meyer CR, Mukherji SK, et al. Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. Proc Natl Acad Sci U S A. 2005;102(46):16759–64.

36. Moffat BA, Chenevert TL, Meyer CR, McKeever PE, Hall DE, Hoff BA, et al. The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. Neoplasia. 2006 Apr;8(4):259–67.

37. Bentzen SM, Gregoire V. Molecular imaging-based dose painting: a novel paradigm for radiation therapy prescription. Semin Radiat Oncol. 2011 Apr;21(2):101–10.

38. Murphy KJ, Brunberg JA. Adult claustrophobia, anxiety and sedation in MRI. Magn Reson Imaging. 1997;15(1):51–4.

Nederlands wetenschappelijke samenvatting

Dit proefschrift beschrijft de rol van medische beeldvorming voor, tijdens en na radiotherapie bij de behandeling van HNSCC patiënten. Vooral voor MR-beeldvorming lijkt een belangrijke rol weggelegd. Dit proefschrift laat verder zien dat beeldvorming op een correcte wijze moet worden toegepast om effectief te zijn.

In deel één wordt de rol van beeldvorming voorafgaand aan de behandeling besproken. De rol van FDG PET-CT bij de detectie van metastatische lymfeklieren wordt onderzocht in hoofdstuk 2. Metastatische lymfeklieren worden routinematig gediagnosticeerd met behulp van cytologie verkregen door echogeleide punctie. Bij patiënten die al een PET-CT hebben ontvangen, bijvoorbeeld om metastasen op afstand te detecteren, is er dankzij de PET-CT ook al enige informatie beschikbaar over de status van de lymfeklieren. Lymfeklieren met een SUVmax-waarde lager dan 2,2 bleken geen detecteerbare tumorcellen in de cytologie te hebben, terwijl lymfeklieren groter dan 1 cm met een SUVmax hoger dan 4,9 bijna altijd tumorcellen bevatten. Door gebruik te maken reeds vervaardigde PET-CT kon ongeveer de helft van alle lymfeklieren een onnodig echogeleid punctieonderzoek worden bespaard.

In hoofdstuk 3 wordt de kracht van DW-MRI tijdens het intekenproces onderzocht. Hoewel de conformiteit van de intekeningen goed was, wat aangeeft dat DW-MRI nuttig zou kunnen zijn tijdens het intekenproces voorafgaande aan de behandeling, zagen we ook duidelijke voorbeelden waarbij de beelden door de verschillende intekenaars anders werden geïnterpreteerd. Dit hoofdstuk toont ook het nut van een semiautomatisch intekenproces op PET.

Hoofdstuk 4 introduceert ADC als een kwantitatieve variabele. Hoewel eerder onderzoek suggereerde dat ADC gemeten voor de behandeling een prognostische factor kon zijn voor lokaal recidief, konden we dit niet bevestigen. Bovendien is een gemakkelijker te verkrijgen klinische variabele, het TNM-stadium, als voorspellende factor aanzienlijk krachtiger dan ADC gemeten voorafgaande aan de behandeling.

Deel twee begint met opnieuw onderzoek naar ADC, maar dan in het bijzonder de verandering in ADC gemeten tijdens behandeling. Een prospectieve studie (de PREDICT-studie) wordt beschreven in hoofdstuk 5. PREDICT heeft als primair doel het voorspellen van de behandelingsresultaten van HNSCC-patiënten aan de hand van ADC gemeten tijdens de therapie. Idealiter zou dit kunnen leiden tot selectie van patiënten op basis van deze verandering in ADC. Patiënten met een verandering in ADC die een onvoldoende respons op de geïnitieerde radiotherapie aangeeft, zouden dan een

intensievere behandeling kunnen krijgen of de behandeling zou over kunnen gaan naar een operatie. Omgekeerd zouden patiënten met een goede respons hetzelfde of zelfs een minder intensief radiotherapieprogramma kunnen blijven volgen. Hoofdstuk 6 illustreert een potentieel probleem met identificatie van een tumor op MRI en het maken van intekeningen tijdens de behandeling. De uiterlijke veranderingen van tumoren op T2-gewogen beeldvorming belemmert de intekening en vermindert bijgevolg de overeenstemming tussen de personen die moeten intekenen. De meest effectieve manier om dit probleem aan te pakken, is waarschijnlijk duidelijke en gemakkelijk te volgen richtlijnen voor de afbakening van tumoren tijdens de behandeling.

In het laatste deel van dit proefschrift, deel 3, wordt beeldvorming in de fase na de behandeling gepresenteerd. Hoofdstuk 7 bespreekt een prospectieve vergelijkende studie van MRI met diffusiegewogen sequenties en PET-CT. Deze studie heet de RETURNED-studie. Zowel diffussie gewogen MRI en PET-CT worden gebruikt om lokale recidieven te detecteren. Hierbij komt kijken dat de diagnose van recidiverende HNSCC aanzienlijk lastiger te bepalen is dan de diagnose van primaire ziekte. Dit komt omdat het moeilijk is om onderscheid te maken tussen effecten op weefsel als gevolg van de behandeling en terugkerende tumor. In het onderzoek konden MRI en PET-CT beide recidieven detecteren met een nauwkeurigheid van ongeveer 70%. MRI vertoonde echter een hogere specificiteit maar een lagere sensitiviteit dan PET-CT. De twee modaliteiten leken elkaar helaas ook niet te versterken, de nauwkeurigheid van de detectie verbeterde nauwelijks wanneer de uitkomst van de MRI en PET-CT samen werden genomen. Daarom pleiten wij ervoor dat PET-CT, dat al routinematig wordt gebruikt voor recidiefdetectie, de modaliteit van keuze moet blijven als er een klinisch vermoeden bestaat van een terugkerende tumor.

Een reden voor het gebrek aan nauwkeurigheid van de MRI met diffusiegewogen beelden wordt gezocht in hoofdstuk 8. Hier probeerden tien radiologen te bepalen of de MRI-beelden van tien RETURNED-patiënten een lokaal recidief of lokale controle vertoonden. De overeenkomst tussen de radiologen was matig. De resultaten toonden verder dat er eenvoudige gevallen waren waarbij alle tien waarnemers het met elkaar eens waren en dat er moeilijke gevallen waren waarin de overeenstemming laag was. Andere studies rapporteerden een vergelijkbare interobserver overeenstemming voor PET-CT. Tevens werd de aanwezigheid van gemakkelijke en moeilijke gevallen door deze studies beschreven. Misschien kunnen MRI-onderzoeken die met regelmatige tussenpozen worden uitgevoerd de detectie van terugkerende MRI verbeteren, maar dit is in dit proefschrift niet onderzocht.

Hoofdstuk 9 ten slotte, beschrijft een raamwerk voor het implementeren van MRI in het radiotherapieproces. Dit is relevant omdat innovaties in het veld resulteerden in hybride radiotherapie en MRI-apparaten die momenteel in de kliniek worden geïntroduceerd. Het geeft antwoord op vragen die een bestralingsoncoloog of klinisch fysicus zou kunnen hebben bij de introductie van MRI voor HNSCC radiotherapie in hun instellingen.
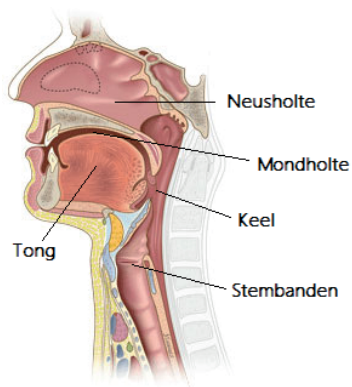
Samenvatting voor geïnteresseerden

Dit proefschrift is ingedeeld naar de drie fasen die een persoon met hoofd-halskanker moet doorlopen mits hij of zij behandeld wil en kan worden. Het gaat om de periode voor de start van behandeling, de behandeling zelf en de periode na het beëindigen van de behandeling. Sommige patiënten zullen deze drie fases ook als afzonderlijk ervaren, voor anderen zal het onderscheid minder duidelijk zijn. Tijdens elke fase zijn er verschillende belangrijke vragen waar artsen en patiënten samen het antwoord op proberen te vinden. Heb ik kanker? Waar zit het? En wat kan ik verwachten? Op deze vragen is helaas niet altijd een sluitend antwoord te vinden, maar in veel gevallen kan medische beeldvorming helpen bij het vinden van antwoorden. Dit hoofdstuk gebruikt enkele van de verschillende vragen die tijdens de drie fases door artsen en patiënten gesteld kunnen worden om een samenvatting te geven van het proefschrift dat voor u ligt.

Voor de behandeling

Heb ik kanker?

De neusholte, mondholte, tong, keel en de stembanden zijn bekleed met schubvormige cellen genaamd plaveiselcellen (figuur 1). Deze cellen worden regelmatig door het lichaam vernieuwd, echter bij de aanmaak van nieuwe plaveiselcellen maakt het lichaam soms een fout en creëert een gemuteerde cel. Vaak wordt de gemuteerde cel door het lichaam weer opgeruimd, maar in sommige gevallen lukt dat niet en begint de cel zich ongecontroleerd te vermenigvuldigen. Als deze vermenigvuldigende cellen andere weefsels in het hoofd-halsgebied binnendringen spreken we over hoofd-halskanker.

De ongecontroleerde groei van cellen kan zo groot zijn dat er een zichtbare zwelling ontstaat. Dankzij de Latijnse invloeden in de geneeskunde wordt deze zwelling ook wel tumor genoemd. Een tumor kan in het hoofd-halsgebied voor specifieke klachten zorgen afhankelijk van de locatie van de tumor. Keelpijn, oorpijn, een hese stem of slikklachten kunnen allen wijzen op de aanwezigheid van een hoofd-hals tumor. Deze klachten kunnen echter ook passen bij andere,

Figuur 1. Zijaanzicht van het hoofd-halsgebied. De aangegeven plekken tonen de locaties waar hoofd-halskanker vaak ontstaat.

onschuldige aandoeningen. De eerste stap in de behandeling van hoofd-halskanker is dus een goede diagnose.

Een hoofd-halstumor kan, in tegenstelling tot ander tumoren, goed zichtbaar zijn. Een zwelling in de mond of tong kan door een arts relatief eenvoudig geïnspecteerd worden. Als een tumor verder in het hoofd-halsgebied zit is gespecialiseerde apparatuur vaak nodig om een beeld te vormen. De arts heeft hiervoor een flexibele camera welke door de neus ingebracht wordt en de neus, keel en het gebied rond de stembanden in beeld kan brengen.

Als aanvulling op de visuele inspectie kan een arts gebruik maken van medische beeldvorming. Door middel van medische beeldvorming is het mogelijk om dieper in het lichaam te kijken. De beelden die gemaakt worden kunnen artsen helpen bij het diagnosticeren van hoofd-halskanker. De diagnose van hoofd-halskanker kan voor het gemak in drie onderdelen ingedeeld worden:

1. De diagnose van kanker en de plek waar de tumor voor het eerst ontstond. Dit wordt ook wel de diagnose van de primaire tumor genoemd.
2. De diagnose van lymfkliermetastasen.
3. De diagnose van metastasen in andere organen.

Hoofdstuk 2 van dit proefschrift gaat in op het tweede onderdeel, de diagnose van lymfekliermetastasen. De ongeremd delende cellen van de hoofd-hals tumor kunnen zich in sommige gevallen verspreiden over het lichaam, we spreken dan over uitzaaiing of metastase. Een van de plekken waar de kankercellen als eerste terecht kunnen komen zijn de lymfeklieren. Deze klieren zijn normaal gezien belangrijk bij het afvoeren van afvalstoffen, maar kunnen ook een plek vormen waar kankercellen terecht komen en verder groeien. Het is mogelijk om de behandeling van hoofd-halskanker aan te passen om ook deze lymfkliermetastasen te behandelen, hiervoor moet echter wel goed in kaart gebracht worden welke lymfeklieren kankercellen bevatten. De methode die gebruikt wordt om dit te doen is door cellen op te zuigen uit de lymfeklieren met een naald en deze cellen onder de microscoop te bekijken, een zogenaamd naaldbiopt.

Cellen kunnen zich ook verder door het lichaam verspreiden. Bij sommige patiënten is het dan ook nodig om op zoek te gaan naar metastasen in andere organen. Bij dit derde onderdeel van de diagnose van hoofd-halskanker krijgen patiënten een PET-CT (zie kader 1). De PET-CT geeft niet alleen informatie over het derde onderdeel, maar ook over het tweede onderdeel (de lymfkliermetastasen). In hoofdstuk 2 is te lezen dat deze informatie in sommige gevallen goed overeenkomt met de informatie die met het naaldbiopt verkregen wordt. In deze gevallen kan dus zowel het tweede als het derde

onderdeel van de diagnose gedaan worden met één scan (figuur 2). Bij sommige patiënten kan dan het naaldbiopt achterwege gelaten worden omdat dit geen nieuwe informatie oplevert. Dat bespaart kosten, maar is bovenal fijn voor de patiënt.

PET-CT is een beeldvormende techniek die gebruik maakt van een radioactieve stof. Om kankercellen op te sporen wordt vaak gebruik gemaakt van een stof die zich gebieden met kanker ophoopt. De radioactiviteit is vervolgens te meten en wordt vaak zichtbaar gemaakt als oranje-rode kleur. Als je op deze scans dus een helder oranje gebied ziet, dan zouden daar mogelijk kankercellen kunnen zitten.

Kader 1

Waar zit de tumor?

Na het vaststellen van kanker zijn er verschillende behandelopties. De drie meest voorkomende opties zijn chirurgie, radiotherapie en radiotherapie met chemotherapie.
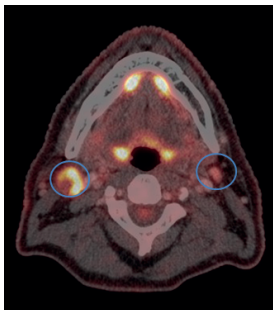
Figuur 2. Een PET-CT beeld van een patiënt. Beide cirkels bevatten een lymfeklier. De klier aan de linkerzijde van het plaatje bevat kankercellen de cirkel rechts niet. Dit komt overeen met de heldere delen die dankzij de PET-CT te zien zijn.

De arts en patiënt besluiten samen wat de meest geschikte optie voor behandeling is. Dit hangt onder andere af van de omvang van de tumor, aanwezigheid van uitzaaiingen en de voorkeur van de patiënt. Dit proefschrift richt zich hoofdzakelijk op de behandeling met radiotherapie.

Tijdens de radiotherapiebehandeling wordt de tumor van buitenaf bestraald met ioniserende straling. Deze straling zorgt ervoor dat de tumorcellen kapot gaan. De straling maakt echter niet alleen de tumorcellen kapot, ook gezonde cellen worden door de straling aangetast. Schade aan gezonde weefsels is een belangrijke bijwerking van de radiotherapie. Om deze schade te minimaliseren is het belangrijk om de tumor zo nauwkeurig mogelijk te bestralen. Om dit voor elkaar te krijgen moet de arts met zekerheid de vraag 'waar zit de tumor?' kunnen beantwoorden. Medische beeldvorming speelt ook hier een belangrijke rol.

Om het belang van medische beeldvorming in te zien moeten we het eerst hebben over intekeningen. De apparaten die de straling maken noemen we versnellers. Moderne versnellers maken gebruik van een intekening om tumoren te bestralen. In een proces wat bedrieglijk eenvoudig lijkt gebruiken artsen beelden verkregen via medische beeldvorming om een lijn rondom de tumor te tekenen. Deze kan vervolgens gebruikt worden om een bestralingsplan te maken dat door de versneller uitgevoerd kan worden. Het intekenproces is zo bedrieglijk eenvoudig omdat de opzet simpel is: maak een plaatje van de tumor en teken hem in. De uitvoer hiervan is echter vaak bijzonder lastig.

Dit heeft er mee te maken dat de tumor vaak niet met 100% zekerheid te onderscheiden is van het omliggende gezonde weefsel. We zeggen ook wel dat het contrast tussen de tumor en het gezonde weefsel laag is. Vergelijk het met een wolk in de lucht op een zonnige ochtend. De rand van de witte wolk steekt duidelijk af tegen de blauwe lucht. Het contrast tussen wolk en omgeving is hoog. Verandert het weer echter en verschijnen er meer wolken dan is het opeens veel lastiger om een individuele wolk van de andere te onderscheiden. Het contrast tussen een wolk en de omgeving is dan laag (figuur 3). Een vergelijkend laag contrast tussen tumor en omliggend weefsel kan grote gevolgen hebben. Het lage contrast kan er voor zorgen dat de arts stukjes tumor mist met zijn intekening. Hierdoor zal de versneller dit stukje niet goed bestralen waardoor de tumor mogelijk niet helemaal kapot gaat. Echter, maakt de arts zijn intekening te groot dan
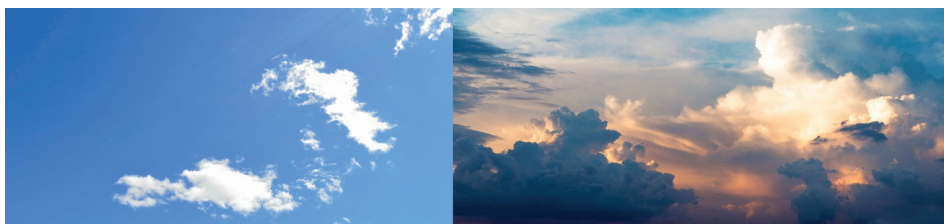
Figuur 3. Het contrastprobleem. Bij een hoog contrast (links) is de wolk duidelijk te onderscheiden van de omgeving. Bij een laag contrast (rechts) is het een stuk lastiger om een individuele wolk te onderscheiden.

zal relatief veel gezond weefsel bestraald worden. Het weefsel gaat hierdoor kapot waardoor de patiënt tijdens en na de behandeling ernstige bijwerkingen kan krijgen, die in sommige gevallen zelfs permanent kunnen zijn. De arts moet dus altijd balanceren tussen te veel intekenen en te weinig intekenen.

Het contrastprobleem is in meer of mindere mate altijd aanwezig (figuur 4). Dit betekent dat een arts die een intekening maakt soms op basis van zijn opleiding en ervaring moet beslissen of een stukje weefsel tumor is of niet. Hierdoor is een intekening dus nooit geheel objectief. Een andere arts kan bij hetzelfde plaatje er een andere mening op nahouden wat de grenzen van de tumor zijn. Dit fenomeen kan samengevat worden in het begrip 'interobserver agreement', de mate waarin artsen (observers) het met elkaar eens zijn (agreement).
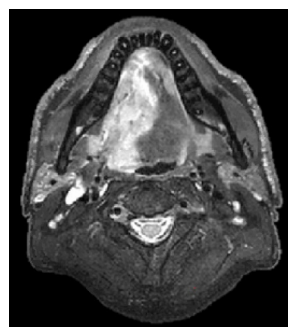
Figuur 4. Het contrastprobleem bij een patiënt. De tumor is zichtbaar als witte vlek in de tong (bovenaan het beeld). Maar waar de tumor precies eindigt en normaal weefsel begint is op sommige plekken lastig te zien.

Om dit alles nog lastiger te maken kan het contrast tussen tumor en omgeving van patiënt tot patiënt sterk verschillen. Daarbij maakt het bovendien ook nog uit met wat voor methode het plaatje van de tumor wordt gemaakt. In hoofdstuk 3 van dit proefschrift komen we naast de eerder besproken PET-CT voor het eerst in aanraking met diffusiegewogen MRI (DW-MRI). Beide beeldvormende technieken worden hier gebruikt om de tumor in te tekenen. Een geautomatiseerd programma tekende de tumor in op de PET-CT beelden en drie mensen tekenden de DW-MRI plaatjes in. Het blijkt dat de interobserver agreement tussen de mensen onderling en tussen de mensen en het geautomatiseerde programma goed is. We weten hierdoor dat artsen ongeveer op dezelfde manier de tumor intekenen op de DW-MRI beelden en dat dit ook nog goed overeenkomt met de PET-CT beelden. Hierdoor hoeven we bij vervolgonderzoeken en bij de behandeling minder rekening te houden met de techniek die gebruikt is om de beelden te verkrijgen.

Wat kan ik verwachten?

Er is nog een laatste vraag die voor het starten van de behandeling gesteld kan worden. Namelijk, wat kan ik als patiënt van de uitkomst van de behandeling verwachten? In hoofdstuk 4 kijken we of het mogelijk is om met behulp van DW-MRI te voorspellen of de radiotherapiebehandeling succesvol gaat zijn. Een succesvolle behandeling definiëren we hier als een behandeling waarna de kanker minimaal 3 jaar weg blijft. Het voorspellen van de uitkomst van een behandeling is niet heel anders dan het voorspellen van de uitslag van een voetbalwedstrijd. Je kan nooit helemaal zeker zijn, maar bepaalde variabelen kunnen je helpen om je voorspelling meer zekerheid te geven. Als je weet dat voetbalclub A miljoenen kan besteden aan spelers en trainers terwijl club B het moet doen met vrijwilligersvergoedingen dan kan je de uitslag van een wedstrijd tussen beide clubs wel met enige zekerheid voorspellen. Toch weet je de uitkomst niet zeker, het kan zijn dat de spelers van club B een geweldige wedstrijd spelen en club A weten te verslaan. Bij tumoren weten we dat de grootte van de tumor een goede voorspeller van de uitkomst van de behandeling is. Bij grotere tumoren is er een grotere kans dat de tumor terugkeert na de behandeling. De grootte van een tumor kan grofweg uitgedrukt worden met het T-stadium van de tumor waarbij een hoger T-stadium een grotere tumor betekent. In hoofdstuk 4 vergelijken we dit T-stadium met het gedrag van water in de tumor verkregen met behulp van DW-MRI. Specifiek kijken we hier naar de ADC (zie kader 2). Het blijkt dat de ADC voorafgaande aan de behandeling helemaal geen goede variabele is om

> Diffusie gewogen MRI of DW-MRI maakt het mogelijk om de vrije beweging van watermoleculen tussen de cellen te meten. De mate van vrije beweging wordt weergegeven als de 'schijnbare diffusie coëfficiënt' of volgens de Engelse afkorting 'ADC'. In een tumor is de vrije beweging van water vaak beperkt en als gevolg is de ADC ook laag.

Kader 2

de uitkomst van de behandeling te voorspellen. Als ADC al een voorspellende waarde zou hebben dan is de kracht van de tumorgrootte om de uitkomst te voorspellen vele malen groter. Om op de vergelijking met voetbal terug te komen, het is alsof je de uitslag van een wedstrijd zou proberen te voorspellen door te bepalen welk team de meest gelukkig getrouwde spelers heeft. Theoretisch kan dit een effect hebben, maar de spelers van club B mogen allemaal de meest gelukkig getrouwde spelers ter wereld zijn, het weegt niet op tegen het enorme verschil in budget met club A. Zo ook met ADC gemeten voor de behandeling en tumorgrootte. Wil je van tevoren de uitslag van een radiotherapiebehandeling voorspellen, kijk dan vooral naar hoe groot een tumor is en niet naar wat de ADC-waarde is.

Tijdens behandeling

Werkt de behandeling?

ADC werkt dus niet voor het starten van de behandeling als voorspeller. Maar misschien werkt het als voorspeller tijdens de behandeling. Dit is het onderwerp van de studie die in hoofdstuk 5 beschreven wordt. De theorie hierachter is dat ADC informatie geeft over de cellen in een stukje weefsel. Omdat elke patiënt anders is en elke tumor anders is geeft ADC weinig zinnige voorspellende informatie voorafgaande aan de radiotherapiebehandeling. Als echter de behandeling gestart is gaan er cellen kapot. Als het weefsel met de nu kapotte cellen opnieuw onderzocht zou worden met DW-MRI dan meten we een andere ADC-waarde dan voor de behandeling. Theoretisch is het mogelijk dat een groot verschil tussen de ADC-waarde voor behandeling en ADC-waarde tijdens behandeling betekent dat er veel cellen kapot zijn gegaan en dat de bestraling dus effectief is. Een klein verschil zou dan betekenen dat er weinig cellen kapot zijn met als conclusie dat de bestraling niet goed werkt. Als dat zo is dan kunnen arts en patiënt mogelijk aan de hand van het verschil in ADC bepalen of de radiotherapie behandeling werkt. Dat zou een geweldige uitkomst zijn. Bij patiënten waar de therapie effectief is zou in toekomst de behandeling ingekort kunnen worden waardoor gezonde weefsels minder schade oplopen. Tegenovergesteld zouden patiënten waarbij de behandeling niet effectief is langer of met een hogere dosis straling behandeld kunnen worden. Dit betekent meer schade voor omliggende weefsels, maar kan wel nodig zijn om de tumorcellen wel kapot te maken. In andere gevallen kunnen deze patiënten er voor kiezen om te stoppen met de radiotherapie behandeling en de tumor operatief te laten verwijderen.

De PREDICT studie is nog niet afgerond, maar een aantal patiënten heeft al aan de studie deel genomen. Deze mensen hebben zonder eigenbelang tijd en energie vrijgemaakt om toekomstige patiënten te helpen. De MRI-beelden van een aantal van de PREDICT deelnemers die hier aanvullende toestemming voor hadden gegeven zijn gebruikt om de onderzoeksvraag van hoofdstuk 6 te beantwoorden. Dit hoofdstuk staat wederom in het thema van het contrastprobleem: het probleem waarbij de tumor moeilijk af te grenzen is van het omliggende gezonde weefsel. Dit probleem is nog groter tijdens de behandeling, waar dankzij het effect van de bestraling tumor en gezond weefsel nog meer op elkaar gaan lijken.

Om te begrijpen waarom het contrastprobleem nog relevant is tijdens de behandeling kunnen we het volgende voorbeeld nemen: Een denkbeeldige patiënt met hoofd-halskanker heeft een tumor ter grootte van een golfbal in de keel. De tumor is zo groot dat hij er voor zorgt dat de patiënt zijn mond nog maar een klein stukje open kan doen. De patiënt en arts besluiten dat de tumor behandeld gaat worden met radiotherapie. In voorbereiding hierop heeft de arts de tumor ingetekend op MRI beelden die voorafgaand aan de behandeling gemaakt zijn. De tumor was erg goed in te tekenen met een hoog contrast tussen tumor en omliggend weefsel (zoals een witte wolk tegen een blauwe lucht). De behandeling zal in totaal zeven weken duren, maar al tijdens de tweede week voelt de patiënt dat hij zijn mond veel verder kan openen. De arts vermoedt dat de tumor kleiner aan het worden is als gevolg van de bestraling en besluit een MRI te maken (figuur 5). Op deze MRI ziet de arts weliswaar dat de tumor kleiner is geworden, eerder een grote knikker dan een golfbal, maar de precieze grens tussen tumor er gezond weefsel is een stuk vager geworden (het contrast is meer zoals een wolk op een half-bewolkte dag). De arts wil graag de tumor opnieuw intekenen, hij denkt dat de intekening kleiner gemaakt kan worden. Maar het contrastprobleem veroorzaakt een lastige keuze: tekent de arts de tumor opnieuw in dan wordt er minder gezond weefsel bestraald en houdt de patiënt mogelijk minder klachten over aan de radiotherapie, maar de arts is dankzij het contrastprobleem minder zeker van zijn intekening en patiënt en arts lopen het risico dat de kleinere intekening ervoor zorgt dat een slecht zichtbaar stukje tumor niet voldoende bestraald wordt. In het ergste geval betekent dit dat de patiënt niet geneest. De beelden van de PREDICT-patiënten hebben geholpen om dit probleem in kaart te brengen. Het blijkt inderdaad dat naarmate de radiotherapie vordert artsen het moeilijker vinden om de tumor in te tekenen, maar gelukkig blijven de verschillen tussen de intekeningen van de artsen klein. Dit betekent dat radiotherapie in de toekomst ook tijdens de behandeling aangepast kan worden. Hopelijk zorgt dit ervoor dat patiënten genezen van de kanker met minder bijwerkingen van de behandeling.
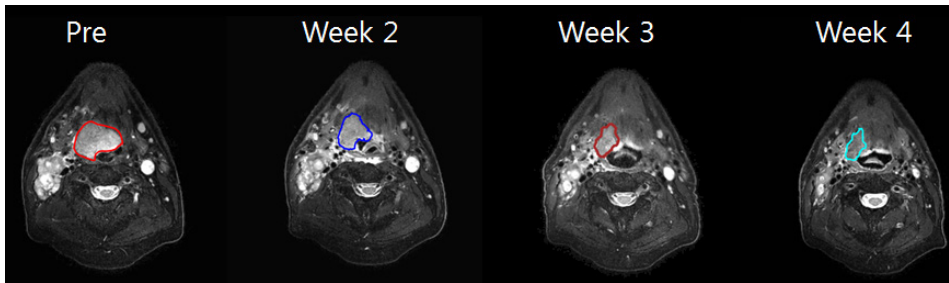
Figuur 5. De tumor van de patiënt wordt per week van de behandeling kleiner, maar het wordt telkens lastiger om de tumor van de omgeving te onderscheiden. De gekleurde kringen zijn intekeningen, deze geven aan waar de arts denkt dat de tumor zich bevindt.

Na de behandeling

Is de kanker weg?

Het afronden van de radiotherapiebehandeling luidt voor patiënten de derde fase in. Als alles goed is zijn alle tumorcellen kapot gemaakt. Het gezonde weefsel dat tijdens de behandeling beschadigd is geraakt kan krijgt nu de gelegenheid om te herstellen. In dagen tot weken die volgen op de behandeling is er vaak sprake van ontstekingsreacties in het gezonde weefsel. Na genezing zijn bij veel patiënten nog steeds de effecten van de radiotherapie te zien (en te voelen). Dit zorgt voor problemen als de patiënt met klachten komt die een mogelijke terugkeer van de kanker signaleren. Als de patiënt klachten heeft die passen bij een recidief van de tumor dan kan er nogmaals medische beeldvorming worden toegepast. Door de radiotherapie-effecten is het op deze beelden echter vaak nóg moeilijker om weefsel zonder kankercellen van een tumor te onderscheiden. De vraag die arts en patiënt dan stellen is: Is de kanker teruggekomen?

Deze vraag wordt behandeld in hoofdstuk 7 en 8 van dit proefschrift. In het onderzoek dat centraal staat in deze hoofdstukken werd uitgezocht of er een verschil was tussen PET-CT of DW-MRI bij het detecteren van terugkerende kanker. Het resultaat van dit onderzoek is dat het ontdekken van terugkerende kanker lastig is, ongeacht de gebruikte methode.

Om deze mogelijk wat onbevredigende boodschap wat te verzachten deel ik graag mijn persoonlijke verwachtingen voor de behandeling van patiënten met hoofd-halskanker voor de komende tien jaar. Ik hoop en verwacht dat medisch beeldvormende technieken zich sterk zullen blijven ontwikkelen. De nauwkeurigheid van artsen in het voorspellen van het beloop van de ziekte en de behandeling zal toenemen dankzij optimaal gebruik

van de informatie die PET-CT en MRI voorafgaand en tijdens de behandeling leveren. Verder zal de radiotherapiebehandeling zelf effectiever worden. Door MRI-beelden direct te combineren met de bestraling kan de tumor nog nauwkeuriger bestraald worden met daarbij minder schade aan omliggend gezond weefsel. Patiënten waarbij de radiotherapiebehandeling onvoldoende blijkt te werken zullen dit dankzij beeldvorming eerder te weten komen en kunnen voor een intensiever radiotherapieschema kiezen. Patiënten waarbij de tumor voor het einde van de behandeling al weg is kunnen hopelijk eerder stoppen met de therapie. Kwalitatief hoogwaardige MRI en PET-CT beelden en een nauwe samenwerking tussen de hoofd-hals chirurgie en radiotherapie zal er voor zorgen dat toekomstige patiënten sneller de meest effectieve therapie krijgen. Mijn verwachting is dat over tien jaar relatief gezien meer mensen zullen genezen van hoofd-halskanker, met daarbij minder restschade. Ik hoop dat dit proefschrift daar een klein beetje aan heeft kunnen bijdragen.

List of publications

Publications of this thesis

Chapter 2
Prediction of ultrasound guided fine needle aspiration cytology results by FDG PET-CT for lymph node metastases in head and neck squamous cell carcinoma patients
Boris Peltenburg, Bart de Keizer, Jan Willem Dankbaar, Mirthe de Boer, Stefan M. Willems, Marielle E. P. Philippens, Chris H. J. Terhaard & Remco de Bree
Acta Oncologica (2018), Vol. 57, p. 1687-1692

Chapter 3
Evaluation of diffusion weighted imaging for tumor delineation in head-and-neck radiotherapy by comparison with automatically segmented 18F-fluorodeoxyglucose positron emission tomography
Tim Schakel, Boris Peltenburg, Jan Willem Dankbaar, Carlos E. Cardenas, Michalis Aristophanous, Chris H.J. Terhaard, Johannes M. Hoogduin, Marielle E.P. Philippens
Physics and Imaging in Radiation Oncology (2018), Vol. 5, p. 13-18

Chapter 4
Pretreatment ADC is not a prognostic factor for local recurrences in head and neck squamous cell carcinoma when clinical T-stage is known
Boris Peltenburg, Juliette P. Driessen, Jeanine E. Vasmel, Frank A. Pameijer, Luuk M. Janssen, Chris H.J. Terhaard, Remco de Bree, Marielle E.P. Philippens
European Radiology (2019), doi: 10.1007/s00330-019-06426-y

Chapter 5
Predictive value of diffusion weighted MRI in head and neck cancer patients during (chemo) radiotherapy (PREDICT): a single center observational study
Boris Peltenburg, Tim Schakel, Patricia Doornaert, Frank A. Pameijer, Niels Raaijmakers, Marielle E.P. Philippens, Chris H.J. Terhaard, Remco de Bree
In preparation for submission

Chapter 6
Interobserver agreement of the delineation of head and neck squamous cell carcinoma on MRI decreases as (chemo)radiotherapy progresses
Boris Peltenburg, Patricia Doornaert, Frank A. Pameijer, Alexis Kotte, Marielle E.P Philippens, Remco de Bree, Chris H.J. Terhaard
In preparation for submission

Chapter 7

Prospective comparative study of MRI including diffusion-weighted images versus FDG PET-CT for the detection of recurrent head and neck squamous cell carcinomas after (chemo)radiotherapy

Juliette P. Driessen, Boris Peltenburg, Marielle E.P. Philippens, Julia E. Huijbregts, Frank A. Pameijer, Remco de Bree, Luuk M. Janssen, Chris H.J. Terhaard

European Journal of Radiology (2019), Vol. 111, p. 62-67

Chapter 8

The interobserver agreement in the detection of recurrent HNSCC using MRI including diffusion weighted MRI

Boris Peltenburg*, Juliette P. Driessen*, Marielle E.P. Philippens, Frank A. Pameijer, Jonas Castelijns, Jan Willem Dankbaar, Bart Dorgelo, Charlotte A.H. Lange, Frederick J.A. Meijer, Walter. M. Palm, Stefan D. Roosendaal, Stefan C.A. Steens, Berit M. Verbist, Inge Stegeman, Luuk M. Janssen, Remco de Bree, Chris H.J. Terhaard

* authors contributed equally

European Journal of Radiology (2018), Vol. 105, p. 134-140

Other publications

Target Volume Delineation Using Diffusion-weighted Imaging for MR-guided Radiotherapy: A Case Series of Laryngeal Cancer Validated by Pathology.

Hans Ligtenberg, Tim Schakel, Jan Willem Dankbaar, Lilian N. Ruiter, Boris Peltenburg, Stefan M. Willems, Nicolien Kasperts, Chris. H.J. Terhaard, Niels Raaijmakers, Marielle E.P. Philippens.

Cureus. 2018 Apr 11;10(4):e2465. doi: 10.7759/cureus.2465.

Cerebrovascular reactivity and white matter integrity.

Kevin Sam, Boris Peltenburg, John Conklin, Olivia Sobczyk, Julien Poublanc, Adrian P. Crawley, Daniel M. Mandell, Lakshmikumar Venkatraghavan, James Duffin, Joseph A. Fisher, Sandra E. Black, David J. Mikulis.

Neurology. 2016 Nov 29; 87(22): 2333–2339.

Curriculum Vitae

Boris was born on the 13th of December 1988 in Utrecht. He moved from Utrecht to Driebergen-Rijsenburg, a small municipality near the Utrechtse Heuvelrug with his older sister and parents at the age of four. Here he attended the primary school CBS Coolsma before receiving his secondary education at KSG de Breul in Zeist.

In 2007 Boris started studying Chemical Engineering in Eindhoven. Together with fellow students, he organized a study tour to Hong Kong and Indonesia. After the opportunity presented itself to start medical school in Utrecht he moved back to the city of his birth and started the bachelor Medicine in 2009.

During his time as a medical student Boris involved himself with other academic activities. From 2011-2012 he was chairman of the faculty council after which he helped with the revision of the medicine curriculum. In 2014 he got his first experience with research into medical imaging when he joined the research group of prof. Mikulis at the UHN in Toronto. Here he worked on MRI measurements of cerebrovascular reactivity under the supervision of Kevin Sam.

After finishing his medical education at the end of 2015 Boris started his PhD research at the department of Head and Neck Surgical Oncology and the department of Radiation Oncology at the UMC Utrecht.

Boris believes that the Netherlands has one of the best healthcare systems in the world, though one where there is still room for an immeasurable amount of improvement.

Since October 2019 Boris is a member of the team at SiRM – Strategies in Regulated Markets. Here he focuses his energy on improving the patientcare and increasing the job satisfaction of healthcare professionals in the Netherlands.

Acknowledgements

Dit proefschrift is tot stand gekomen dankzij de inzet en hulp van velen. Hiervoor is een bedankje zeker op de plaats.

Allereerst wil ik alle mensen bedanken die meegedaan hebben aan de wetenschappelijke studies PREDICT, CORRECT I en II en RETURNED. Terwijl zij zelf met de behandeling van hoofd-halskanker bezig waren hebben zij zich onbaatzuchtig ingezet voor toekomstige patiënten door scans te ondergaan of bloed af te geven. Dit vergt moed en een groot hart, hier heb ik dan ook veel bewondering voor. Zeer veel dank.

Beste prof. dr. R. de Bree en prof. dr. C.H.J. Terhaard, beste promotoren, beste Remco en Chris, graag wil ik jullie bedanken voor de kans om promotieonderzoek te doen onder jullie begeleiding. Ondanks jullie verschil in individuele begeleidingsstijl kon ik altijd op jullie steun en expertise rekenen. Ik bewonder jullie gedeelde passie voor patiënten en gedegen onderzoek.

Beste dr. M.E.P. Philippens, beste copromotor, beste Mariëlle, ik denk dat ik gedurende mijn promotie geen vraag heb gehad waar jij het antwoord niet op wist. Van de meest obscure details van de MRI tot aan de meest praktische patiëntgebonden vragen, jij kon me altijd weer op weg helpen als ik vastliep. Ik kijk ook met veel plezier terug op de keren dat we samen beelden bekeken of ADC-waarden probeerden te verzamelen, deze momenten waren altijd nuttig en vooral zeer gezellig.

Beste leden van de beoordelingscommissie, beste prof. dr. Lam, prof. dr. de Jong, prof. dr. Kaanders, dr. Takes, prof. dr. ir. Lagendijk bedankt dat jullie de tijd hebben genomen om dit proefschrift te lezen en te beoordelen.

Beste Hoofd-halswerkgroep, beste Hans, Emma, Homan, Inne, Joost, Jorine, Lilian, Sandra, Anna, Lianne, Nicolien, Tristan, Myrthe, Najiba, África, Rutger, Niels, Sanam en alle anderen, dank voor de vele uren discussie over het hoofd-halsonderzoek. Ik denk dat we regelmatig met meer vragen uit de bijeenkomsten kwamen dan we erin gingen, maar altijd met hernieuwde energie. Ik kijk ook met plezier terug op de hoofd-halscongressen en jonge onderzoekersdagen die ik met jullie mocht bezoeken.

Beste Tim, jij verdient een bijzonder bedankje. Zonder jouw hulp zat ik nu nog steeds met een vragende blik naar Volumetool te kijken. Bovendien was je elke keer weer bereid om me het verschil tussen EPI en SPLICE uit te leggen. Dank voor je hulp en tips.

Beste tutorstudenten, beste Maureen, Pravish, Bart, Anne Marije, Max, Mozdalefa, Sarah Dewi, Anna, Bram, Diede, Bethany, Femke en Lola. In 2016 werden door de opleiding geneeskunde de leukste, slimste en meest kritische studenten in werkgroep 1 geplaatst. Althans dat is mijn conclusie nadat ik jullie drie jaar heb mogen begeleiden. Bedankt voor de onmetelijke energie die ik van elk gesprek met jullie kreeg. Ik wens jullie een geweldige tijd toe in verdere studies of andere avonturen en ik hoop dat jullie je altijd net iets meer laten leiden door het hart dan door het hoofd.

Beste Lenny, ik ben je veel dank verschuldigd, niet alleen deelde je je kennis bij onze Journal Club, maar dankzij jou heb ik ook mijn nieuwe collega's bij SiRM leren kennen. Beste Drian, Fons, Geraline, Jolien, Jan-Peter Julia, Maarten, Michelle, Pam, Reinout, René, Saskia en Thijs ik kijk er naar uit om de komende tijd met jullie samen te werken.

Lieve vrienden, ik heb het immense geluk dat ik van meer mensen vriendschap ervaar dan dat ik hier op papier kan zetten. Dank voor jullie bijdrage aan dit proefschrift.

Lieve Bianca, vanaf ons gezamenlijke begin als coassistenten tot aan onze promotietrajecten in het UMC heb ik altijd mijn twijfels en geluk met je kunnen delen. Waar ik soms het gevoel had struikelend door het leven te gaan leek elke golf van tegenslag op jouw optimisme en positivisme uiteen te spatten. Ik heb ongelooflijk veel bewondering voor jouw kracht. Ik vind het een eer dat jij mijn paranimf bent.

Lieve Hester, enkele jaren geleden mocht ik jou als paranimf ondersteunen bij je verdediging. Nu zijn de rollen omgekeerd. Jij hebt de bijzondere gave om altijd precies de zaken te benoemen zoals ze zijn. Op die nuchterheid ben ik altijd jaloers geweest, maar het maakt met ook enorm trots op mijn grote zus. Het is een veilig gevoel dat jij mijn paranimf bent.

Beste lezer, dank voor je interesse in dit proefschrift. Of je het nu van voor tot achter leest of enkel dit dankwoord tot je neemt, je belangstelling stel ik op prijs.

Lieve Hiltrud, Theo en Saskia, dank voor jullie steun door de jaren en de altijd gezellige dagen en avonden in Maastricht. Dankzij jullie ben ik kerst en gourmetten gaan waarderen.

Lieve Theo, Marjolijn, André en Kirsten, bedankt voor al de liefde die ik van jullie heb mogen ontvangen. Sterke wortels zorgen voor krachtige vleugels.

Lieve Laura, mijn gevoel voor jou is niet in woorden te bevatten. Bedankt dat jij mijn coach, geweten, proeflezer, adviseur, entertainer, vertrouwenspersoon, vriendin en nog veel meer bent. Je bent voor mij de helderste ster aan het firmament.