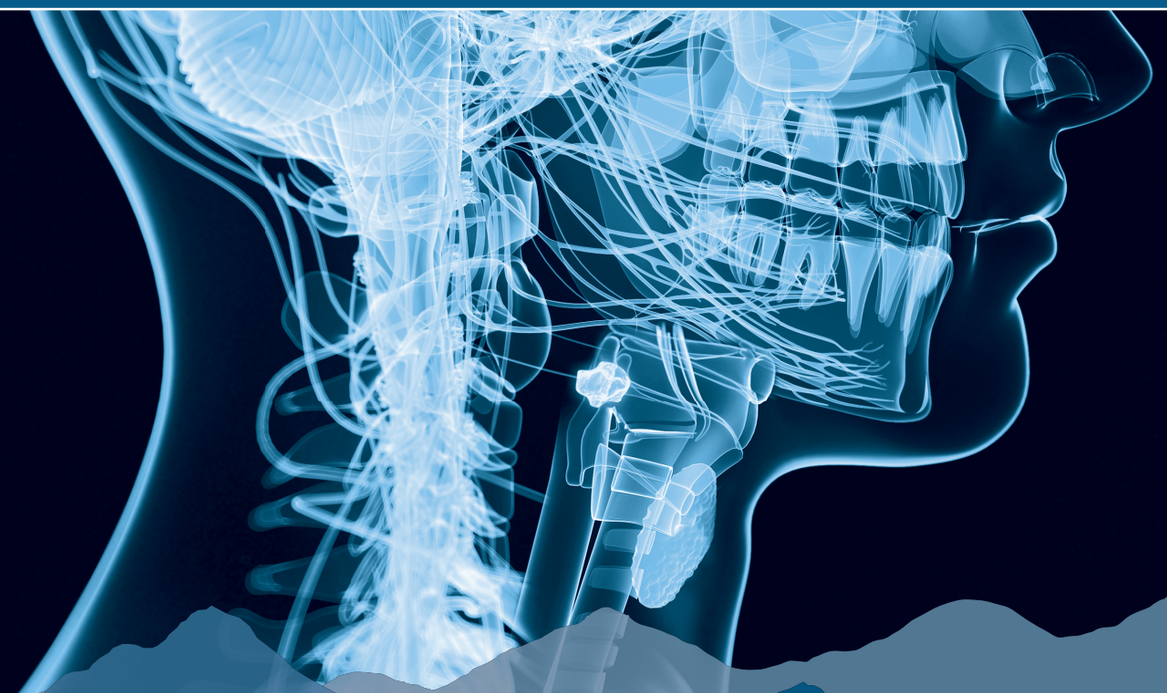
Functional MRI in head and neck cancer

Potential applications, reproducibility, diagnostic and prognostic capacity

Daniel Noij

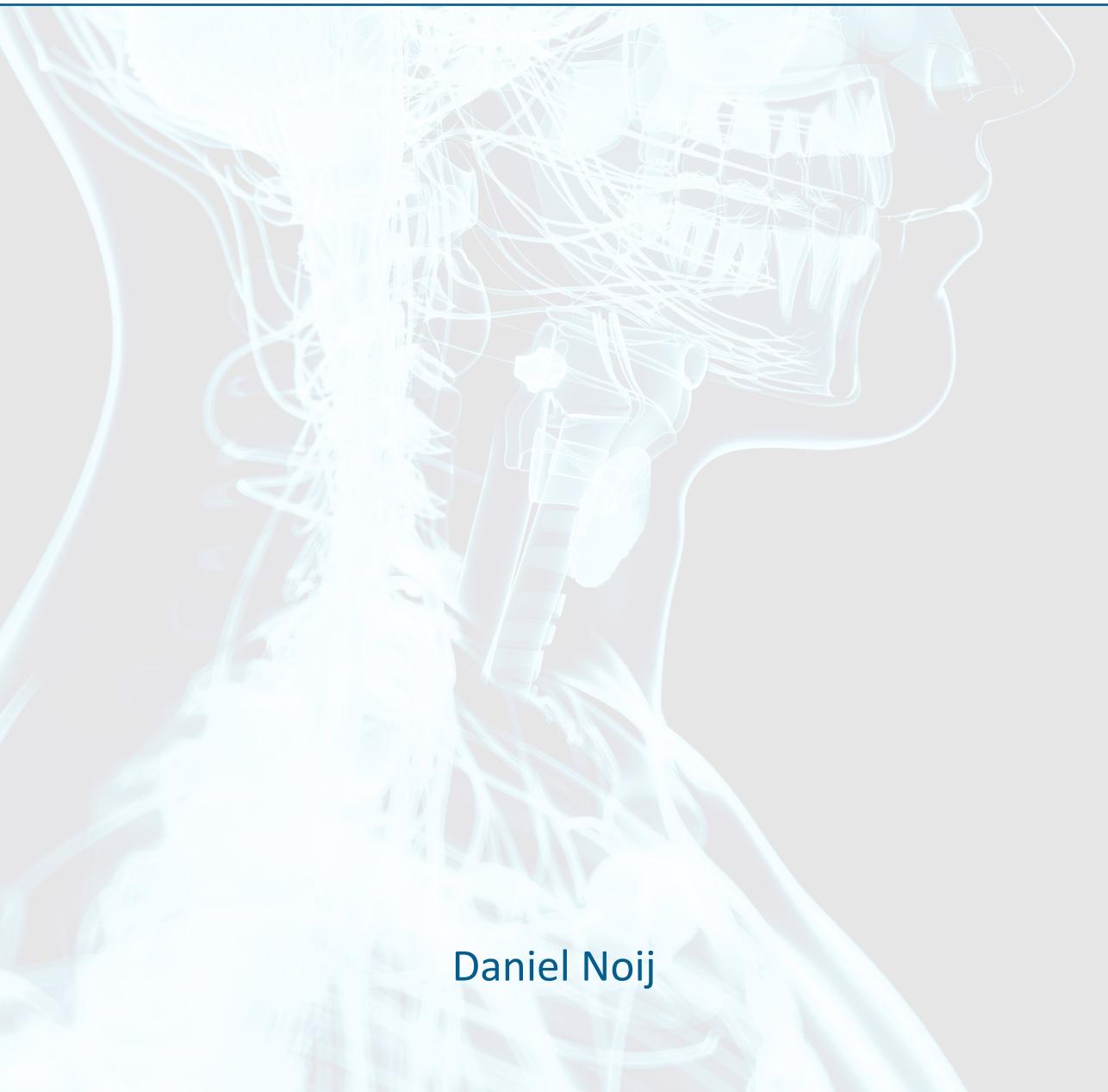
Functional MRI in head and neck cancer



Daniel Noij

Functional MRI in head and neck cancer

Potential applications, reproducibility, diagnostic and prognostic capacity



Daniel Noij

COLOFON

Cover and lay-out by:	Raimond Boessenkool
Printed by:	Ridderprint BV, www.ridderprint.nl
ISBN:	978-94-6375-148-3

Printing of this thesis was financially supported by: Guerbet Nederland B.V., Oldelft Benelux B.V., SECTRA Benelux, Chipsoft, Atos Medical and Sonova Nederland B.V.

The research described in this thesis was performed at the Department of Radiology and Nuclear Medicine at the VU University Medical Center, Amsterdam, the Netherlands.

VRIJE UNIVERSITEIT

Functional MRI in head and neck cancer

Potential applications, reproducibility, diagnostic and prognostic capacity

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op maandag 12 november 2018 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Daniel Peter Noij

geboren te Haarlemmermeer

promotoren:	prof.dr. J.A. Castelijns
	prof.dr. R. de Bree
copromotor:	dr. P. de Graaf

‘The finish is a the finish line’

John Degenkolb

CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 1.1

General introduction,
aims and outline of the thesis

GENERAL INTRODUCTION

Head and neck cancer

Head and neck cancer comprises approximately 3% of malignancies worldwide (1-3). Most malignancies in the head and neck area are squamous cell carcinomas originating at the mucosal surface of the upper aerodigestive tract, which includes oral cavity, oropharynx, nasopharynx, hypopharynx, and larynx. Traditionally the use of tobacco and alcohol consumption has been recognized as the most important risk factors which account for approximately 75% of the head and neck cancers (4, 5). The combined use of these substances appears to be synergistic in the development of head and neck squamous cell carcinoma (HNSCC). It appears that smoking mainly increases the risk of laryngeal cancer and alcohol consumption is more associated with pharyngeal an oral cavity cancer (6). Patients are generally over 50 years of age with a male to female ratio of 70:30 (1). Presenting symptoms include hoarseness for laryngeal cancers. For pharyngeal cancers patients often present when the tumor is larger and results in dysphagia, a sore throat or ear pain. Patients may also present with one or more painless cervical nodes (6).

At early stages treatment usually consist of a single modality, most often surgery or radiotherapy. Whereas advanced disease is treated with combinations of surgery, chemotherapy and/or radiotherapy with salvage surgery in reserve (7-9). Radiotherapy is associated with better organ preservation compared to surgery and results in less problems with swallowing and speech and comparable survival rates (10). For this reason (chemo)radiotherapy is increasingly used in head and neck cancer. Early complications of radiotherapy include: radiation dermatitis, both xerostomia as excessive mucus production and a painful mucositis which all can result in problems with adequate intake of food and liquids (10). Therefore patients of require nasogastric feeding or a percutaneous gastronomy. Adding chemotherapy results in more acute toxicity. An added survival benefit of 6.5% for concomitant chemotherapy has been found (11). Cisplatin-based regimens are most frequently used. More recently the anti-epidermidal growth factor receptor antibody cetuximab has shown to also increase overall survival when added to radiotherapy (12, 13). On the long-term radiotherapy results in less problems with body image and facial contour compared to extensive surgical procedures. Long-term side effects of radiotherapy may occur in up to 82% of patients and include xerostomia, soft tissue fibrosis, dysphagia and osteoradionecrosis of the mandible (10, 14). These long-term effects result in an impairment of the quality of life of long-term survivors of head and neck cancer (15, 16).

When residual disease after chemoradiotherapy is detected, salvage surgery may still be possible. The success rate of salvage surgery is the highest when it is performed as early as possible. In a study including patients with oropharyngeal carcinoma salvage surgery after 1-2 months was successful in 70% of early detected residues, compared to only 33% in later detected recurrences (8). The increasing use of non-invasive chemoradiotherapy warrants good surveillance protocols. Especially since symptoms caused by post-radiation changes and residual overlap. Both can present with hoarseness, pain, and swallowing complaints (17). Because taking repeated biopsies can exacerbate these symptoms and can cause infection, non-invasive methods of surveillance are preferred. Moreover, information obtained by imaging techniques may guide biopsies when indicated.

Five year overall survival (OS) in HNSCC is strongly dependent of tumor localization and disease stage varying from 59-78% for stage I, to 32-47% for stage IV M0 disease (18). Best survival is seen in laryngeal carcinoma (66% 5-year OS), and worst survival in hypopharyngeal carcinoma (32% 5-year OS) (18).

Recently the human papillomavirus (HPV), known for causing cervical cancer, has been recognized as a risk factor for developing oropharyngeal HNSCC (19-22). The HPV-16 genotype is considered the most important HPV subtype in oncogenesis (23). Patients with HPV-associated HNSCC are generally younger and often do not have a history of excessive alcohol and tobacco consumption (19, 20, 22). This may result in a delayed diagnosis. Generally these patients present with a small primary tumor and large lymph nodes (6). HPV-positive HNSCC is associated with a better prognosis than HPV-negative HNSCC (19, 20, 22).

Another important infectious causative of head and neck cancer, and more specifically nasopharyngeal cancer (NPC), is the Epstein-Barr virus (EBV) (24-26). Nasopharyngeal carcinoma differs from other HNSCC in the geographic distribution with a peak incidence in Southeast Asia (27). Another distinct feature of NPC is the excellent radiosensitivity compared to other forms of HNSCC (28).

Salivary gland tumors are a rare and very heterogeneous group of neoplasms. All salivary glands can be affected, however approximately 80% is located in the parotid gland. In the parotid gland approximately 25% of the lesions is malignant compared to up to 45% for submandibular gland tumors and up to 90% for sublingual gland tumors (29, 30). Pleomorphic adenomas are the most common benign salivary gland tumor (29, 30). These tumors are formed by epithelial and myoepithelial cells and harbor the ability to transform to malignant lesions (i.e. carcinoma ex pleomorphic adenoma). Warthin tumors are another relatively common benign salivary gland tumor (29, 30). This entity is a more homogeneous compact to the heterogeneous pleomorphic adenoma. Treatment of benign salivary gland tumors usually consists of surgical excision with wide surgical margins, often in the form of a superficial parotidectomy, to prevent disease recurrence (30). Benign salivary gland tumors need to be distinguished from malignant salivary gland tumors. This last group is characterized by local invasion, perineural spread and the ability to metastasize (31).

Imaging

In head and neck cancer imaging is used in: diagnosing disease, determining disease stage, monitoring treatment response and for follow-up after treatment. The most commonly used modalities are ultrasound, computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) (32).

Ultrasound is mainly used for nodal staging (33, 34). Advantages of ultrasound are the high spatial resolution, low costs, general availability and that the technique does not requires ionizing radiation. Moreover ultrasound can easily be combined with fine needle aspiration (FNA) to obtain material for cytological assessment of lymph nodes suspicious of containing malignant tissue (33, 34). Disadvantages of ultrasound are the operator dependency and risk of a sampling error with FNA. Another disadvantage is that deeper structures (e.g. retropharyngeal nodes) cannot be assessed with this modality.

Computed tomography is a fast imaging technique with superior bone detail. Advantages of CT are the high spatial resolution, low acquisition times resulting in the virtual absence of motion artefacts. With the current generation of multidetector row CT systems slice thicknesses of less than 1 mm are widely used in clinical practice (32). However CT lacks soft tissue contrast, the technique requires ionizing radiation, dental implants may result in severe artifacts especially when the oral cavity is assessed, and the iodinated contrast used in contrast-enhanced CT imaging is relatively contraindicated in patients with renal impairment. This makes CT the modality of choice in patients:

- with contra-indications to MRI (e.g. claustrophobia, pacemakers, vascular clips or foreign metal bodies);
- at risk of cortical bone invasion (e.g. in the mandibular or skull base);
- who are unable to lay flat for a long period of time, have difficulty breathing or have problems with of swallowing secretions.

Further, chest-CT is often used in patients who are at risk of pulmonary metastases.

Magnetic resonance imaging is the technique of choice for the visualization of soft tissue (e.g. muscle invasion, perineural spread, cartilage invasion and intracranial extension). Conventional MRI protocols for head and neck imaging usually include T1, T2, a sequence with fat-suppression (e.g. short-tau inversion recovery (STIR) or spectral attenuated inversion recovery (SPAIR)), and contrast-enhanced T1-weighted sequences (35). In MRI chelates of the paramagnetic element gadolinium are used as a contrast medium. Contrast accumulates in the extravascular space resulting in shortening of T1 and therefore signal enhancement of contrast-enhanced T1-weighted images (36). Malignant tissue induces increased vascularity resulting in increased contrast delivery and enhanced signal intensity. With the current 1.5T and 3T MRI systems slice thicknesses of 3-7 mm are clinically feasible to perform within an acceptable time (32). Due to the relatively long examination time, MRI is susceptible to motion artifacts caused by patient movement, swallowing and breathing. Besides providing anatomical information, MRI can also be used for acquiring functional imaging, which will be discussed below.

Functional imaging

The aforementioned techniques are considered anatomical since these techniques provide information on the location and size of a lesion, and the relation to nearby structures. Functional imaging techniques provide information on tissue functioning, with the anatomical detail being of secondary importance.

In MRI several functional imaging techniques are increasingly applied. In this thesis we focused on two types of functional imaging: Diffusion-weighted imaging (DWI) and perfusion weighted imaging, since these techniques are the main focus of imaging research in head and neck cancer.

Diffusion-weighted imaging is based upon measuring the random or Brownian motion of water molecules in tissue (37). By applying different diffusion gradients with different strengths tissues can be separated based on the amount of diffusion restriction. The amount of diffusion weighting depends on the timing and the strength of the gradient and can be quantified in a b-value. By comparing signal intensities at varying b-values, the signal decay can be quantified in an apparent diffusion coefficient. In order to calculate an ADC at least two different b-values are needed: typically a low (e.g. <150 s/mm^2) and a high b-value (e.g. >700 s/mm^2). The ADC can be calculated with the following formula:

$$\frac{S(b)}{S_0} = \exp\left(-b \cdot ADC\right)$$

Where S_b represents the signal intensity with diffusion gradient b, and S_0 represents the signal intensity without diffusion gradients.

Restricted diffusion (i.e. low ADC values) in oncologic imaging is often caused by hypercellular tissue with limited extracellular space, and is therefore associated with the presence of malignancy (38-42) (Table 1). On the contrary, areas with ample extracellular space and a relatively low cell density are characterized by a high ADC value. Necrosis and inflammation generally meet these criteria (41, 43, 44) (Table 1).

Table 1 *Qualitative interpretation criteria for DWI (41)*

Signal intensity on high b-value image	Signal intensity on ADC map	Interpretation
High	Low	Highly proliferative malignant tissue, abscess, viscous fluid or blood
High	High	T2 shine through, liquefactive necrosis
Low	High	Necrosis, fluids, adenocarcinoma with low cellularity
Low	Low	Fibrosis, fat, susceptibility artifact

In the head and neck area diffusion restriction may be physiological in some tissues (e.g. in the parotid and submandibular salivary glands, thyroid glands, palatine tonsils and benign lymph nodes). When assessing these tissues, it is important to also look for other signs of malignancy (e.g. asymmetry or ingrowth in surrounding structures) (45, 46). An increase in ADC during and after treatment is associated with a favorable prognosis (38, 47, 48).

As the word 'apparent' indicates, true diffusion is not measured with the ADC concept. Especially at lower b-values, pseudorandom or 'incoherent' diffusion also contributes to the imaging signal. This incoherent movement is mainly caused by perfusion at the capillary level. Le Bihan et al. introduced a bi-exponential intravoxel incoherent motion (IVIM) model to account for this (37):

$$\frac{S_b}{S_0} = (1 - f) \cdot e^{(-bD)} + f \cdot e^{(-bD*)}$$

Where S_b represents the signal intensity with diffusion gradient b, and S_0 represents the signal intensity without diffusion gradients. D is known as pure or slow diffusion coefficient, which is related to pure molecular diffusion. D* is the fast or pseudodiffusion coefficient that resembles the perfusion related incoherent microcirculation and is about a factor of 10 greater than D in biological tissue (37, 49-51). Finally, f is the perfusion or (micro) vascular volume fraction, which depends on capillary geometry and blood velocity (37, 49). With this model several quantitative parameters can be acquired in order to further characterize lesions without the admission of contrast material (Figure 1).

Most often DWI is performed with an echo-planar imaging (EPI) sequence (38-40, 42, 43, 47, 48, 52-67). Advantages of EPI are the high signal intensity and image contrast, and the relatively short acquisition time (68). The main disadvantage of EPI-based imaging is frequent presence of susceptibility artifacts resulting in image distortion. In the head and neck area air-tissue boundaries and the presence of metallic implants (e.g. dental fillings or implants) may result in magnetic inhomogeneities, which can cause severe susceptibility artifacts, especially when EPI-based DWI is used. To deal with this turbo spin-echo (TSE) based sequences can be used. This sequence does not suffer from image distortion; however this is at the cost of a reduced signal-to-noise ratio and prolonged acquisition time (40, 69-72).

Perfusion-weighted imaging describes a group of imaging techniques in which pharmacokinetic modelling after the administration of intravenous contrast is used. It is hypothesized that malignant tissue can be characterized based on differences in vascular properties. Malignant tissue induces vessel formation of vessels with poor functionality, high permeability, tortuosity and density (73). Changes in these vascular properties can be measured and may provide information on tumor response to treatment (74-79). High expression of hypoxia-associated markers as hypoxia inducible factor 1α and carbonic anhydrase have been associated with an adverse prognosis (80).

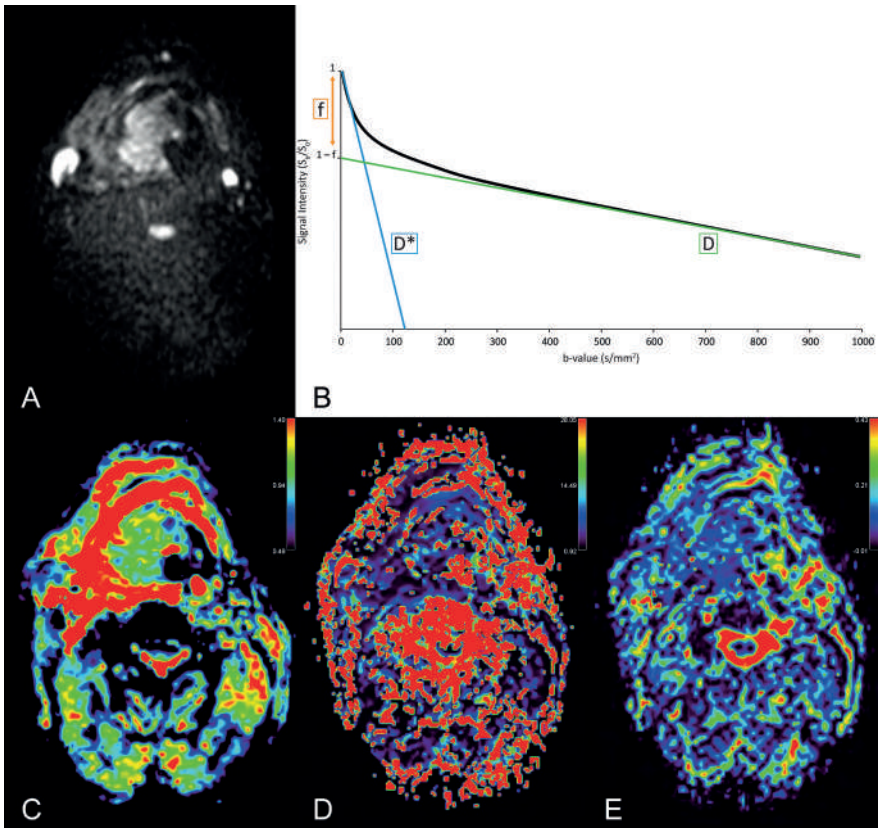


Figure 1 *Example of IVIM in a patient with a T4b squamous cell carcinoma originating from the hypopharynx. A) Show the original diffusion weighted image with a b-value of 1000 s/mm2. B) shows the bi-exponential decay of signal intensity over b-values. At low b-values the pseudodiffusion coefficient (D*) is the main contributor to the signal decay over b-values. With increasing b-values the pure diffusion coefficient (D) becomes the main contributor to the signal decay. When using the intercept of D and the signal intensity at b=0 s/mm2 f can be estimated. C), D) and E) are the D, D* and f map, respectively. The respective values can be acquired by drawing a region of interest of the tumor.*

The most frequently used technique is dynamic contrast-enhanced imaging (DCE) (74-79, 81-105). This technique is based on the serial acquisition of T1-weighted images before, during and after the injection of intravenous contrast. Then, extravasation of the contrast material can be measured. The rate of extravasation is mainly determined by the amount of blood flow, the thickness and viability of the capillary wall (36, 106). The resultant changes in T1-signal caused by changes in the amount of contrast material can be quantified using pharmacokinetic modelling. The Tofts model is the most commonly used pharmacokinetic model in head and neck cancer (36, 106). With this model four different quantitative parameters can be obtained (Figure 2):

1. K^{trans}, the volume transfer constant between plasma and interstitial space. This parameters provides information on capillary permeability;
2. k_{ep}, the rate constant between interstitial space and plasma. With this parameters the flow from the interstitial space back to the capillaries can be estimated;
3. v_e, the fractional volume of the extracellular, extravascular space;
4. V_p, the plasma volume fraction.

With the use of other models even more parameters may be obtained. For example use of the Brix model results in the parameters 'amplitude scaling constant'(AH) and 'time of arrival of contrast medium' (TA) (107).

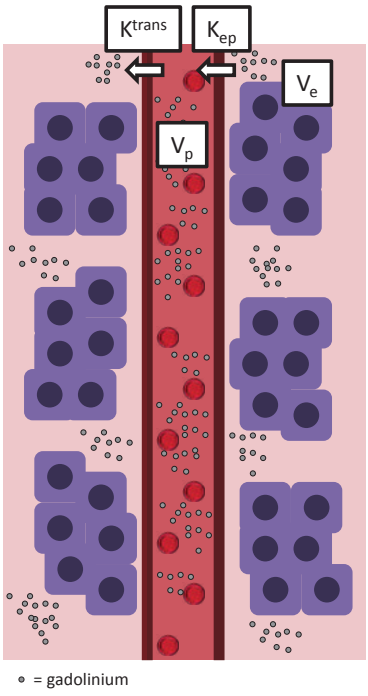


Figure 2 *Schematic picture of the Tofts model*

Another approach is to analyze the time-dependent change in signal intensity during and after the administration of intravenous contrast. This results in parameters such as maximum contrast-index, initial area under the curve, time to peak, wash in and wash out (84, 88, 91-93, 108). These parameters are considered to be not directly related to tissue pathology and are more difficult to compare between patients. Therefore these parameters are considered to be semi-quantitative parameters.

Perfusion-weighted imaging can also be performed with another, less frequently used, method based on T2*-based sequences. This is referred to as dynamic susceptibility contrast (DSC) MRI (109-113). The accumulation of contrast material causes a transient darkening of tissue. It is assumed that the degree of darkening is linearly related to the concentration of contrast material. This produces surrogate parameters indicative of tissue perfusion (e.g. blood volume and blood flow) (109, 114).

Positron emission tomography (PET) differs from the other imaging modalities because it is a pure functional imaging modality. It is often combined with CT for attenuation correction and to combine functional with anatomical imaging for topographical determination of high uptake lesions. Imaging is performed after the administration of a positron emitting tracer. These emissions can be measured and reconstructed in a three-dimension image. The most commonly used tracer is ^{18}F-Fluoro-deoxyglucose (^{18}F-FDG), which is a marker of tissue metabolism.

Malignant tissue has a relatively high metabolic rate compared to other tissues (115). This makes ^{18}F-FDG a relative specific marker for the detection of malignancy. However, severe inflammation (e.g. during and after radiotherapy) can also result in focal ^{18}F-FDG uptake which may compromise the sensitivity of ^{18}F-FDG-PET-CT during and after treatment with (chemo)radiotherapy (116-122). Another disadvantage of ^{18}F-FDG-PET-CT is the relatively low spatial resolution, which may result in false-negative findings in smaller lesions (123). Applications of ^{18}F-FDG-PET-CT are detecting nodal and distant metastases, unknown primary lesions and synchronous second primary lesions (e.g. lung cancer is relatively common in patients with HNSCC due to the overlapping risk factors) (124).

AIMS AND GENERAL OUTLINE OF THE THESIS

The general aim of this thesis was to assess the role of functional imaging in HNC: from making the diagnosis and differential diagnosis, to determining patient prognosis and the detecting residual disease.

The main shortcomings of currently accepted imaging strategies are:

1. Selecting patients who are likely to respond to (chemo)radiotherapy is not possible with current imaging modalities. Preferably this is done before treatment.
2. The relatively limited sensitivity and specificity of anatomic imaging in the detection of residual and recurrent disease after treatment with (chemo)radiotherapy.
3. That reliable response evaluation with ^{18}F-FDG-PET-CT can be performed from three months after treatment, when patients already received the complete treatment with the associated treatment toxicity.

We explored the potential of functional imaging techniques in head and neck cancer, as known from literature, in three reviews in **Chapter 2**. In **Chapter 2.1** we focused on the diagnostic and prognostic potential of DCE. In **Chapter 2.2** we used a comparable approach to assess the value of IVIM. In **Chapter 2.3** we focused on the use of various types of functional imaging in early follow-up in order to identify the most promising techniques.

Before any technique can be applied in clinical practice, reproducibility needs to be assessed. Therefore we determined DWI reproducibility in **Chapter 3**.

In **Chapter 4** we focused on the diagnostic potential of DWI. In **Chapter 4.1** we assessed the value of whole-body-MR imaging because the presence of distant metastases currently rules out the chance of curation. In these patients therapy should focus on patient comfort and toxicity needs to be limited. Another patient population at risk of over-treatment are patients presenting with a lymph node metastasis with an unknown primary tumor. If the primary tumor cannot be found during disease staging, then the whole pharyngeal axis is treated with radiotherapy with associated toxicity. In **Chapter 4.2** we therefore assessed if the use of DWI and ^{18}F-FDG-PET-CT alone or combined resulted in increased detection of occult primary tumors. Another clinical challenge is the detection of residual HNSCC after treatment with (chemo)radiotherapy. Our main aim in **Chapter 4.3** was to assess the added value of DWI to follow-up with ^{18}F-FDG-PET-CT and in **Chapter 4.4** to find the most optimal combination of imaging with the most relevant parameters.

Finally in **Chapter 5** we assessed the prognostic potential of DWI. In **Chapter 5.1** we focused on the combined use of DWI and contrast-enhanced imaging to determine if the use of intravenous contrast material can aid in determining diffusion restriction solely in the solid part of the primary tumor and lymph nodes. In **Chapter 5.2** we used histogram analysis to assess the prognostic value of pre-treatment DWI and ^{18}F-FDG-PET-CT in patients with HNSCC.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017;67(1):7-30.
2. Gatta G, Botta L, Sanchez MJ, et al. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. Eur J Cancer. 2015;51(15):2130-43.
3. European crude and age adjusted incidence by cancer, years of diagnosis 2000 and 2007: Analisys based on 83 population-based cancer registries. INFORMATION NETWORK ON RARE CANCERS IN EUROPE, 2015.
4. Hashibe M, Brennan P, Benhamou S, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst. 2007;99(10):777-89.
5. Conway DI, Hashibe M, Boffetta P, et al. Enhancing epidemiologic research on head and neck cancer: INHANCE - The international head and neck cancer epidemiology consortium. Oral Oncol. 2009;45(9):743-6.
6. Mehanna H, Paleri V, West CM, Nutting C. Head and neck cancer-part 1: epidemiology, presentation, and preservation. Clin Otolaryngol. 2011;36(1):65-8.
7. Alkureishi LW, de Bree R, Ross GL. RADPLAT: an alternative to surgery? Oncologist. 2006;11(5):469-80.
8. Yom SS, Machtay M, Biel MA, et al. Survival impact of planned restaging and early surgical salvage following definitive chemoradiation for locally advanced squamous cell carcinomas of the oropharynx and hypopharynx. Am J Clin Oncol. 2005;28(4):385-92.
9. Blanchard P, Baujat B, Holostenco V, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. Radiother Oncol. 2011;100(1):33-40.
10. Mehanna H, West CM, Nutting C, Paleri V. Head and neck cancer--Part 2: Treatment and prognostic factors. BMJ. 2010;341:c4690.
11. Pignon JP, le Maitre A, Maillard E, Bourhis J, Group M-NC. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol. 2009;92(1):4-14.
12. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med. 2006;354(6):567-78.
13. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. The Lancet Oncology. 2010;11(1):21-8.
14. Denis F, Garaud P, Bardet E, et al. Late toxicity results of the GORTEC 94-01 randomized trial comparing radiotherapy with concomitant radiochemotherapy for advanced-stage oropharynx carcinoma: comparison of LENT/SOMA, RTOG/EORTC, and NCI-CTC scoring systems. Int J Radiat Oncol Biol Phys. 2003;55(1):93-8.
15. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. J Clin Oncol. 2008;26(22):3770-6.
16. Mehanna HM, Morton RP. Deterioration in quality-of-life of late (10-year) survivors of head and neck cancer. Clin Otolaryngol. 2006;31(3):204-11.
17. de Bree R, Putten Lvd, Brouwer J, Castelijns JA, Hoekstra OS, René Leemans C. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 2009;45(4-5):386-93.
18. de Ridder M, Balm AJM, Baatenburg de Jong RJ, et al. Variation in head and neck cancer care in the Netherlands: A retrospective cohort evaluation of incidence, treatment and outcome. Eur J Surg Oncol. 2017;43(8):1494-502.

19. Huang SH, Xu W, Waldron J, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. J Clin Oncol. 2015;33(8):836-45.

20. Brana I, Siu LL. Locally advanced head and neck squamous cell cancer: treatment choice based on risk factors and optimizing drug prescription. Ann Oncol. 2012;23 Suppl 10:x178-85.

21. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. 2011;11(1):9-22.

22. O'Sullivan B, Huang SH, Su J, 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.

23. Fakhry C, Westra WH, Li S, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst. 2008;100(4):261-9.

24. Tsao SW, Yip YL, Tsang CM, et al. Etiological factors of nasopharyngeal carcinoma. Oral Oncol. 2014;50(5):330-8.

25. Lo KW, Chung GT, To KF. Deciphering the molecular genetic basis of NPC through molecular, cytogenetic, and epigenetic approaches. Semin Cancer Biol. 2012;22(2):79-86.

26. Pathmanathan R, Prasad U, Sadler R, Flynn K, Raab-Traub N. Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. N Engl J Med. 1995;333(11):693-8.

27. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev. 2006;15(10):1765-77.

28. Lee N, Harris J, Garden AS, et al. Intensity-modulated radiation therapy with or without chemotherapy for nasopharyngeal carcinoma: radiation therapy oncology group phase II trial 0225. J Clin Oncol. 2009;27(22):3684-90.

29. Spiro RH. Salivary neoplasms: overview of a 35-year experience with 2,807 patients. Head Neck Surg. 1986;8(3):177-84.

30. Guzzo M, Locati LD, Prott FJ, Gatta G, McGurk M, Licitra L. Major and minor salivary gland tumors. Crit Rev Oncol Hematol. 2010;74(2):134-48.

31. Xiao CC, Zhan KY, White-Gilbertson SJ, Day TA. Predictors of Nodal Metastasis in Parotid Malignancies: A National Cancer Data Base Study of 22,653 Patients. Otolaryngol Head Neck Surg. 2016;154(1):121-30.

32. de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol. 2009;45(11):930-5.

33. van den Brekel MW, Castelijns JA, Stel HV, et al. Occult metastatic neck disease: detection with US and US-guided fine-needle aspiration cytology. Radiology. 1991;180(2):457-61.

34. van den Brekel MW, Castelijns JA. What the clinician wants to know: surgical perspective and ultrasound for lymph node imaging of the neck. Cancer Imaging. 2005;5 Spec No A:S41-9.

35. Kuno H, Onaya H, Fujii S, Ojiri H, Otani K, Satake M. Primary staging of laryngeal and hypopharyngeal cancer: CT, MR imaging and dual-energy CT. Eur J Radiol. 2014;83(1):e23-e35.

36. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging. 1997;7(1):91-101.

37. 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;168(2):497-505.

38. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010;20(7):1703-14.

39. Srinivasan A, Dvorak R, Perni K, Rohrer S, Mukherji SK. Differentiation of benign and malignant pathology in the head and neck using 3T apparent diffusion coefficient values: early experience. AJNR Am J Neuroradiol. 2008;29(1):40-4.

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

41. Padhani AR, Koh DM, Collins DJ. Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology. 2011;261(3):700-18.

42. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46.

43. Vandecaveye V, De Keyzer F, Nuyts S, 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;67(4):960-71.

44. Abdel Razek AA, Kandeel AY, Soliman 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. 2007;28(6):1146-52.

45. Vandecaveye V, De Keyzer F, Dirix P, Lambrecht M, Nuyts S, Hermans R. Applications of diffusion-weighted magnetic resonance imaging in head and neck squamous cell carcinoma. Neuroradiology. 2010;52(9):773-84.

46. Choi YJ, Lee JH, Kim HO, et al. Histogram Analysis of Apparent Diffusion Coefficients for Occult Tonsil Cancer in Patients with Cervical Nodal Metastasis from an Unknown Primary Site at Presentation. Radiology. 2016;278(1):146-55.

47. King AD, Mo FK, Yu KH, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010;20(9):2213-20.

48. Chenevert TL, Meyer CR, Moffat BA, et al. Diffusion MRI: a new strategy for assessment of cancer therapeutic efficacy. Mol Imaging. 2002;1(4):336-43.

49. Le Bihan D. Intravoxel incoherent motion imaging using steady-state free precession. Magn Reson Med. 1988;7(3):346-51.

50. Burdette JH, Elster AD, Ricci PE. Acute cerebral infarction: quantification of spin-density and T2 shine-through phenomena on diffusion-weighted MR images. Radiology. 1999;212(2):333-9.

51. Lemke A, Laun FB, Simon D, Stieltjes B, Schad LR. An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. Magn Reson Med. 2010;64(6):1580-5.

52. Kim S, Loevner L, Quon H, et al. 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. 2009;15(3):986-94.

53. Ding Y, Hazle JD, Mohamed AS, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. NMR Biomed. 2015;28(12):1645-54.

54. Hauser T, Essig M, Jensen A, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. Neuroradiology. 2013;55(5):527-36.

55. Hauser T, Essig M, Jensen A, et al. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: Evaluation of lymph node metastasis. Eur J Radiol. 2014;83(5):783-7.

56. Lai V, Li X, Lee VH, Lam KO, Chan Q, Khong PL. Intravoxel incoherent motion MR imaging: comparison of diffusion and perfusion characteristics between nasopharyngeal carcinoma and post-chemoradiation fibrosis. Eur Radiol. 2013;23(10):2793-801.

57. Lai V, Li X, Lee VH, et al. Nasopharyngeal carcinoma: comparison of diffusion and perfusion characteristics between different tumour stages using intravoxel incoherent motion MR imaging. Eur Radiol. 2014;24(1):176-83.

58. Lu Y, Jansen JF, Stambuk HE, et al. Comparing primary tumors and metastatic nodes in head and neck cancer using intravoxel incoherent motion imaging: a preliminary experience. J Comput Assist Tomogr. 2013;37(3):346-52.

59. Marzi S, Piludu F, Vidiri A. Assessment of diffusion parameters by intravoxel incoherent motion MRI in head and neck squamous cell carcinoma. NMR Biomed. 2013;26(12):1806-14.

60. Sasaki M, Sumi M, Eida S, Katayama I, Hotokezaka Y, Nakamura T. Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors. PLoS One. 2014;9(11):e112866.

61. Sumi M, Nakamura T. Head and neck tumors: assessment of perfusion-related parameters and diffusion coefficients based on the intravoxel incoherent motion model. AJNR Am J Neuroradiol. 2013;34(2):410-6.

62. Sumi M, Van Cauteren M, Sumi T, Obara M, Ichikawa Y, Nakamura T. Salivary gland tumors: use of intravoxel incoherent motion MR imaging for assessment of diffusion and perfusion for the differentiation of benign from malignant tumors. Radiology. 2012;263(3):770-7.

63. Xiao Y, Pan J, Chen Y, Chen Y, He Z, Zheng X. Intravoxel Incoherent Motion-Magnetic Resonance Imaging as an Early Predictor of Treatment Response to Neoadjuvant Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma. Medicine (Baltimore). 2015;94(24):e973.

64. Xiao-ping Y, Jing H, Fei-ping L, et al. Intravoxel incoherent motion MRI for predicting early response to induction chemotherapy and chemoradiotherapy in patients with nasopharyngeal carcinoma. J Magn Reson Imaging. 2016;43(5):1179-90.

65. Zhang SX, Jia QJ, Zhang ZP, et al. Intravoxel incoherent motion MRI: emerging applications for nasopharyngeal carcinoma at the primary site. Eur Radiol. 2014;24(8):1998-2004.

66. Yu XP, Hou J, Li FP, et al. Intravoxel Incoherent Motion Diffusion Weighted Magnetic Resonance Imaging for Differentiation Between Nasopharyngeal Carcinoma and Lymphoma at the Primary Site. J Comput Assist Tomogr. 2016;40(3):413-8.

67. Guo W, Luo D, Lin M, et al. Pretreatment Intra-Voxel Incoherent Motion Diffusion-Weighted Imaging (IVIM-DWI) in Predicting Induction Chemotherapy Response in Locally Advanced Hypopharyngeal Carcinoma. Medicine (Baltimore). 2016;95(10):e3039.

68. Poustchi-Amin M, Mirowitz SA, Brown JJ, McKinstry RC, Li T. Principles and applications of echo-planar imaging: a review for the general radiologist. Radiographics. 2001;21(3):767-79.

69. Verhappen MH, Pouwels PJ, Ljumanovic R, et al. Diffusion-weighted MR imaging in head and neck cancer: comparison between half-fourier acquired single-shot turbo spin-echo and EPI techniques. AJNR Am J Neuroradiol. 2012;33(7):1239-46.

70. Schouten CS, de Bree R, van der Putten L, 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;4(4):239-50.

71. Sakamoto J, Imaizumi A, Sasaki Y, et al. Comparison of accuracy of intravoxel incoherent motion and apparent diffusion coefficient techniques for predicting malignancy of head and neck tumors using half-Fourier single-shot turbo spin-echo diffusion-weighted imaging. Magn Reson Imaging. 2014;32(7):860-6.

72. Reiser M, Faber SC. Recent and future advances in high-speed imaging. Eur Radiol. 1997;7 Suppl 5:166-73.

73. Azzi S, Hebda JK, Gavard J. Vascular permeability and drug delivery in cancers. Front Oncol. 2013;3:211.

74. Zahra MA, Hollingsworth KG, Sala E, Lomas DJ, Tan LT. Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy. The Lancet Oncology. 2007;8(1):63-74.

75. Cao Y. The promise of dynamic contrast-enhanced imaging in radiation therapy. Semin Radiat Oncol. 2011;21(2):147-56.

76. Chikui T, Kitamoto E, Kawano S, et al. Pharmacokinetic analysis based on dynamic contrast-enhanced MRI for evaluating tumor response to preoperative therapy for oral cancer. J Magn Reson Imaging. 2012;36(3):589-97.

77. Chikui T, Kawano S, Kawazu T, et al. Prediction and monitoring of the response to chemoradiotherapy in oral squamous cell carcinomas using a pharmacokinetic analysis based on the dynamic contrast-enhanced MR imaging findings. Eur Radiol. 2011;21(8):1699-708.

78. Hoskin PJ, Saunders MI, Goodchild K, Powell ME, Taylor NJ, Baddeley H. Dynamic contrast enhanced magnetic resonance scanning as a predictor of response to accelerated radiotherapy for advanced head and neck cancer. Br J Radiol. 1999;72(863):1093-8.

79. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. Med Phys. 2012;39(8):5277-85.

80. Silva P, Homer JJ, Slevin NJ, et al. Clinical and biological factors affecting response to radiotherapy in patients with head and neck cancer: a review. Clin Otolaryngol. 2007;32(5):337-45.

81. Agrawal S, Awasthi R, Singh A, Haris M, Gupta RK, Rathore RK. An exploratory study into the role of dynamic contrast-enhanced (DCE) MRI metrics as predictors of response in head and neck cancers. Clin Radiol. 2012;67(9):e1-5.

82. Van Cann EM, Rijpkema M, Heerschap A, van der Bilt A, Koole R, Stoelinga PJ. Quantitative dynamic contrast-enhanced MRI for the assessment of mandibular invasion by squamous cell carcinoma. Oral Oncol. 2008;44(12):1147-54.

83. Helbig M, Schlemmer HP, Lumer M, Knopp MV, Dietz A. [Dynamic magnetic resonance tomography (dMRT). It's value in advanced head-neck tumors treated with radiochemotherapy]. HNO. 2003;51(11):886-92.

84. Lee FK, King AD, Ma BB, Yeung DK. Dynamic contrast enhancement magnetic resonance imaging (DCE-MRI) for differential diagnosis in head and neck cancers. Eur J Radiol. 2012;81(4):784-8.

85. Bisdas S, Seitz O, Middendorp M, et al. An exploratory pilot study into the association between microcirculatory parameters derived by MRI-based pharmacokinetic analysis and glucose utilization estimated by PET-CT imaging in head and neck cancer. Eur Radiol. 2010;20(10):2358-66.

86. Wendl CM, Muller S, Meier J, et al. High resolution contrast-enhanced ultrasound and 3-tesla dynamic contrast-enhanced magnetic resonance imaging for the preoperative characterization of cervical lymph nodes: First results. Clin Hemorheol Microcirc. 2012;52(2-4):153-66.

87. Sumi M, Nakamura T. Extranodal spread in the neck: MRI detection on the basis of pixel-based time-signal intensity curve analysis. J Magn Reson Imaging. 2011;33(4):830-8.

88. Unetsubo T, Konouchi H, Yanagi Y, et al. Dynamic contrast-enhanced magnetic resonance imaging for estimating tumor proliferation and microvessel density of oral squamous cell carcinomas. Oral Oncol. 2009;45(7):621-6.

89. Ariyoshi Y, Shimahara M. Relationships between dynamic contrast-enhanced MRI findings and pattern of invasion for tongue carcinoma. Oncol Rep. 2006;15(5):1339-43.

90. Tomura N, Omachi K, Sakuma I, et al. Dynamic contrast-enhanced magnetic resonance imaging in radiotherapeutic efficacy in the head and neck tumors. Am J Otolaryngol. 2005;26(3):163-7.

91. Fischbein NJ, Noworolski SM, Henry RG, Kaplan MJ, Dillon WP, Nelson SJ. Assessment of metastatic cervical adenopathy using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2003;24(3):301-11.

92. Konouchi H, Asaumi J-i, Yanagi Y, et al. Evaluation of tumor proliferation using dynamic contrast enhanced-MRI of oral cavity and oropharyngeal squamous cell carcinoma. Oral Oncol. 2003;39(3):290-5.

93. Tomura N, Watanabe O, Kato K, et al. Irradiated carcinoma of the tongue: correlation of MR imaging findings with pathology. AJR Am J Roentgenol. 2002;178(3):705-10.

94. Baba Y, Furusawa M, Murakami R, et al. Role of dynamic MRI in the evaluation of head and neck cancers treated with radiation therapy. Int J Radiat Oncol Biol Phys. 1997;37(4):783-7.

95. Jansen JF, Carlson DL, Lu Y, et al. Correlation of a priori DCE-MRI and (1)H-MRS data with molecular markers in neck nodal metastases: Initial analysis. Oral Oncol. 2012;48(8):717-22.

96. Jansen JF, Schoder H, Lee NY, et al. Tumor metabolism and perfusion in head and neck squamous cell carcinoma: pretreatment multimodality imaging with 1H magnetic resonance spectroscopy, dynamic contrast-enhanced MRI, and [18F]FDG-PET. Int J Radiat Oncol Biol Phys. 2012;82(1):299-307.

97. Jansen JF, Schoder H, Lee NY, et al. Noninvasive assessment of tumor microenvironment using dynamic contrast-enhanced magnetic resonance imaging and 18F-fluoromisonidazole positron emission tomography imaging in neck nodal metastases. Int J Radiat Oncol Biol Phys. 2010;77(5):1403-10.

98. Shukla-Dave A, Lee NY, Jansen JF, et al. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases. Int J Radiat Oncol Biol Phys. 2012;82(5):1837-44.

99. Chawla S, Kim S, Dougherty L, 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;200(1):35-43.

100. Kim S, Loevner LA, Quon H, et al. Prediction of response to chemoradiation therapy in squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2010;31(2):262-8.

101. Hietschold V, Kittner T, Schreyer M, Appold S, Beuthien-Baumann B, Laniado M. MR perfusion measurements on pharyngeal tumors: comparison of quantification strategies. J Appl Clin Med Phys. 2004;5(4):96-111.

102. Chawla S, Kim S, Loevner LA, et al. Prediction of disease-free survival in patients with squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2011;32(4):778-84.

103. Ng SH, Lin CY, Chan SC, et al. Dynamic contrast-enhanced MR imaging predicts local control in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiotherapy. PLoS One. 2013;8(8):e72230.

104. Ai S, Zhu W, Liu Y, Wang P, Yu Q, Dai K. Combined DCE- and DW-MRI in diagnosis of benign and malignant tumors of the tongue. Front Biosci (Landmark Ed). 2013;18:1098-111.

105. Sumi M, Nakamura T. Head and neck tumours: combined MRI assessment based on IVIM and TIC analyses for the differentiation of tumors of different histological types. Eur Radiol. 2014;24(1):223-31.

106. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10(3):223-32.

107. Brix G, Semmler W, Port R, Schad LR, Layer G, Lorenz WJ. Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. J Comput Assist Tomogr. 1991;15(4):621-8.

108. Wendl CM, Müller S, Meier J, et al. High resolution contrast-enhanced ultrasound and 3-tesla dynamic contrast-enhanced magnetic resonance imaging for the preoperative characterization of cervical lymph nodes: First results. Clin Hemorheol Microcirc. 2012;52(2-4):153-66.

109. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. Radiographics. 2009;29(5):1433-49.

110. Wu RH, Bruening R, Ducreux D, Berchtenbreiter C, Jund R, Reiser M. Estimation of relative blood volume in head and neck squamous cell carcinomas. J Neuroradiol. 2004;31(3):190-7.

111. Bisdas S, Baghi M, Wagenblast J, et al. Tracer kinetics analysis of dynamic contrast-enhanced CT and MR data in patients with squamous cell carcinoma of the upper aerodigestive tract: comparison of the results. Clin Physiol Funct Imaging. 2009;29(5):339-46.

112. Razek AA, Elsorogy LG, Soliman NY, Nada N. Dynamic susceptibility contrast perfusion MR imaging in distinguishing malignant from benign head and neck tumors: a pilot study. Eur J Radiol. 2011;77(1):73-9.

113. Abdel Razek AA, Gaballa G. Role of perfusion magnetic resonance imaging in cervical lymphadenopathy. J Comput Assist Tomogr. 2011;35(1):21-5.

114. Cha S, Knopp EA, Johnson G, Wetzel SG, Litt AW, Zagzag D. Intracranial mass lesions: dynamic contrast-enhanced susceptibility-weighted echo-planar perfusion MR imaging. Radiology. 2002;223(1):11-29.

115. Warburg O. The Metabolism of Carcinoma Cells. The Journal of Cancer Research. 1925;9(1):148-63.

116. Nakamoto Y, Tatsumi M, Hammoud D, Cohade C, Osman MM, Wahl RL. Normal FDG distribution patterns in the head and neck: PET/CT evaluation. Radiology. 2005;234(3):879-85.

117. Gupta T, Master Z, Kannan S, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2011;38(11):2083-95.

118. Ryan WR, Fee WE, Jr., Le QT, Pinto HA. Positron-emission tomography for surveillance of head and neck cancer. Laryngoscope. 2005;115(4):645-50.

119. Malone JP, Gerberi MA, Vasireddy S, et al. Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography. Arch Otolaryngol Head Neck Surg. 2009;135(11):1119-25.

120. Nakajo M, Nakajo M, Kajiya Y, et al. FDG PET/CT and Diffusion-Weighted Imaging of Head and Neck Squamous Cell Carcinoma Comparison of Prognostic Significance Between Primary Tumor Standardized Uptake Value and Apparent Diffusion Coefficient. Clin Nucl Med. 2012;37(5):475-80.

121. Ho AS, Tsao GJ, Chen FW, et al. Impact of positron emission tomography/computed tomography surveillance at 12 and 24 months for detecting head and neck cancer recurrence. Cancer. 2013;119(7):1349-56.

122. van der Putten L, Hoekstra OS, de Bree R, et al. 2-Deoxy-2[F-18]FDG-PET for detection of recurrent laryngeal carcinoma after radiotherapy: interobserver variability in reporting. Mol Imaging Biol. 2008;10(5):294-303.

123. Fukui MB, Blodgett TM, Snyderman CH, et al. Combined PET-CT in the head and neck: part 2. Diagnostic uses and pitfalls of oncologic imaging. Radiographics. 2005;25(4):913-30.

124. Manca G, Vanzi E, Rubello D, et al. 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.

CHAPTER 2

POTENTIAL APPLICATIONS OF FUNCTIONAL IMAGING

CHAPTER 2.1

Contrast-enhanced perfusion magnetic resonance imaging for head and neck squamous cell carcinoma: a systematic review

Oral Oncology 2015;51:124-38

Daniel P Noij
Marcus C de Jong
Lieven G Mulders
J Tim Marcus
Remco de Bree
Christina Lavini
Pim de Graaf
Jonas A Castelijns

ABSTRACT

This systematic review gives an extensive overview of the current state of perfusion-weighed magnetic resonance imaging (MRI) for head and neck squamous cell carcinoma (HNSCC). Pubmed and Embase were searched for literature until July 2014 assessing the diagnostic and prognostic performance of perfusion-weighted MRI in HNSCC. Twenty-one diagnostic and 12 prognostic studies were included for qualitative analysis. Four studies used a T2* sequence for dynamic susceptibility (DSC)-MRI, 29 studies used T1-based sequences for dynamic contrast enhanced (DCE)-MRI. Included studies suffered from a great deal of heterogeneity in study methods showing a wide range of diagnostic and prognostic performance. Therefore we could not perform any useful meta-analysis. Perfusion-weighted MRI shows potential in some aspects of diagnosing HNSCC and predicting prognosis. Three studies reported significant correlations between hypoxia and tumor heterogeneity ($|\rho|>0.6$, $P<0.05$). Two studies reported synergy between perfusion-weighted MRI and positron emission tomography (PET) parameters. Four studies showed a promising role for response prediction early after the start of chemoradiotherapy. In two studies perfusion-weighted MRI was useful in the detection of residual disease. However more research with uniform study and analysis protocols with larger sample sizes is needed before perfusion-weighted MRI can be used in clinical practice.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the most common malignancy in the head and neck region with a world-wide incidence of approximately 550,000 cases (1). In the work-up of these patients multiple imaging modalities are used (e.g. ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) (2). This review focuses on the use of perfusion-weighted MRI in differentiating between HNSCC and other lesions and on the prognostic value of perfusion parameters.

Contrast-based MRI techniques targeting tissue perfusion are known as dynamic contrast enhanced (DCE) MRI and dynamic susceptibility contrast (DSC) MRI. We will refer to both as "perfusion-weighted MRI". In the head and neck area DCE-MRI is more commonly used, but we have also included DSC-MRI in this study to provide a more complete overview.

In short, DCE-MRI is based on the serial acquisition of multiple T1-weighted images before, during and after the injection of an intravenous contrast agent with a low molecular weight. After the injection, the contrast medium extravasates from the intravascular to the interstitial space, at a rate which is determined by the viability of the capillary wall to the contrast agent. The transfer of the contrast agent across the capillary wall can be quantified by applying a pharmacokinetic model to the acquired DCE-MRI data. The two-compartment model developed by Tofts et al. is commonly used for this purpose (3). The model provides a measure, called K^{trans} (volume transfer constant between plasma and interstitial space), which is an indirect measure of the capillary permeability. Other measures provided by pharmacokinetic modeling include the k_{ep} (rate constant between interstitial space and plasma), which is an indirect measure of flow from the interstitial place to the capillary and v_e (the fractional volume of the extracellular, extravascular space). We will refer to the above analysis method as "quantitative" analysis.

Other, more simple analysis methods do not make use of pharmacokinetic models, but analyze the time-dependent change in signal intensity of the DCE-MRI image, producing parameters such as maximum contrast index (of enhancement) (CI, or ME), initial area under the curve (iAUC) of the rate of enhancement. These parameters are not directly related to the tissue physiology, cannot be compared between patients, and are therefore semi-quantitative parameters.

Perfusion-weighted imaging performed with T2*-weighted imaging after contrast administration is referred to as DSC-MRI. When DSC-MRI is performed a transient darkening of tissue can be observed during passage of contrast media. By analyzing the signal time course of the DSC scan it is possible to provide relative values of perfusion parameters (e.g. blood volume (BV) and blood flow (BF)) (4).

With perfusion-weighted MRI it is possible to depict vascular properties of lesions. Malignant processes induce the formation of new vessels, which are characterized by poor functionality, with high permeability, tortuosity and density. Based on these characteristics, malignancies can be characterized on their vascular properties. Adequate

blood and oxygen supply to the tumor is essential for some therapies (e.g. radiotherapy and chemotherapy) to be effective. Perfusion-weighted MRI parameters can be associated with tumor hypoxia and thereby serve as predictors of treatment failure (5, 6).

The diagnostic and prognostic value of perfusion-weighted MRI in HNSCC has been assessed in several studies. However, study designs show great heterogeneity in the used perfusion parameters and outcome measures. In order to provide an overview of the value of the used parameters, a critical systematic review is warranted.

Our purpose was to determine and compare the diagnostic and prognostic performance of DCE-MRI and DSC-MRI in patients with HNSCC, with histopathology, other imaging modalities or clinical follow-up as reference standards.

METHODS

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews and meta-analyses as guidance (7, 8).

Search strategy

Pubmed and Embase were searched until July 2014 for published journal articles assessing the diagnostic and/or prognostic performance of perfusion-weighted MR imaging in HNSCC. We included studies in English, German or Dutch. When necessary, we contacted corresponding authors for additional data (e.g. to calculate sensitivity and specificity and for proposed cut-off values).

For our search we included keywords for the index test (MRI), the imaging technique (perfusion-weighted imaging) and the target condition (HNSCC). We did not include nasopharyngeal carcinoma because of its distinct treatment, epidemiology and prognosis. To increase the sensitivity of the search we did not include terms for the reference tests (histopathology, other imaging modalities or clinical follow-up). See Appendix A for our complete search strategy.

Study selection

Article titles and abstracts were independently reviewed for eligibility by two authors (DPN and MCJ) and discrepancies were resolved by consensus. We included studies if they met all of the following criteria: 1) the study population consisted of patients with HNSCC; 2) the study assessed diagnostic and/or prognostic performance of perfusion-weighted MRI in at least 10 patients with HNSCC; 3) histopathology, other imaging modalities or follow-up were used as the reference standard test. Studies were excluded if they met one of the following criteria: 1) the article was a review, meta-analysis or conference abstract; 2) if a study reported the same analysis performed on (potentially) overlapping study populations.

When a study assessed diagnostic accuracy (e.g. sensitivity and specificity) or correlations between perfusion-parameters and other diagnostic parameters it was considered a

diagnostic study. Studies were classified as prognostic if perfusion-parameters are related to prognostic parameters (e.g. survival rates and recurrence rates).

Data extraction
Two authors (DPN and MCJ) independently extracted data on study, patient, and imaging characteristics. If available, source data (true positive (TP), false positive (FP), true negative (TN), and false negative (FN)) were extracted from included studies. Discrepancies were resolved by consensus. If these data were unavailable, authors were contacted.

Quality assessment
We classified studies as diagnostic and/or prognostic based on the data we could extract. The quality assessment of studies of diagnostic accuracy included in systematic reviews (QUADAS-2) checklist was used to assess the quality of all included studies (9, 10). Two authors (DPN and MCJ) independently assessed the included articles for diagnostic study quality. For prognostic studies we also used the quality in prognosis studies (QUIPS) checklist (11, 12). Prognostic quality was assessed by two authors (DPN and LGM). Discrepancies were resolved by consensus.

Statistical analysis and data synthesis
Diagnostic and prognostic parameters were analyzed separately. We summarized the data of the studies where sensitivity and specificity were given or could be derived adequately in forest plots with 95% confidence intervals (95%CI) for diagnostic and prognostic parameters using RevMan (version 5.2; Copenhagen, Denmark). If per-patient data could be extracted, we used the cut-off in ROC analysis with the highest Youden Index (YI) using SPSS (version 20.0; Chicago, IL, USA). Forest plots were created with Photoshop CS6 (Adobe, San Jose, CA). *P*-values are reported as follows: NS (not significant), <0.05, <0.01 or <0.001. Correlations are reported as follows: -1.0 to -0.5= - - -; - 0.5 to -0.3= - -; -0.3 to -0.1= -; -0.1 to 0.1= +/-; 0.1 to 0.3= +; 0.3 to 0.5= + +; 0.5 to 1.0= + + +.

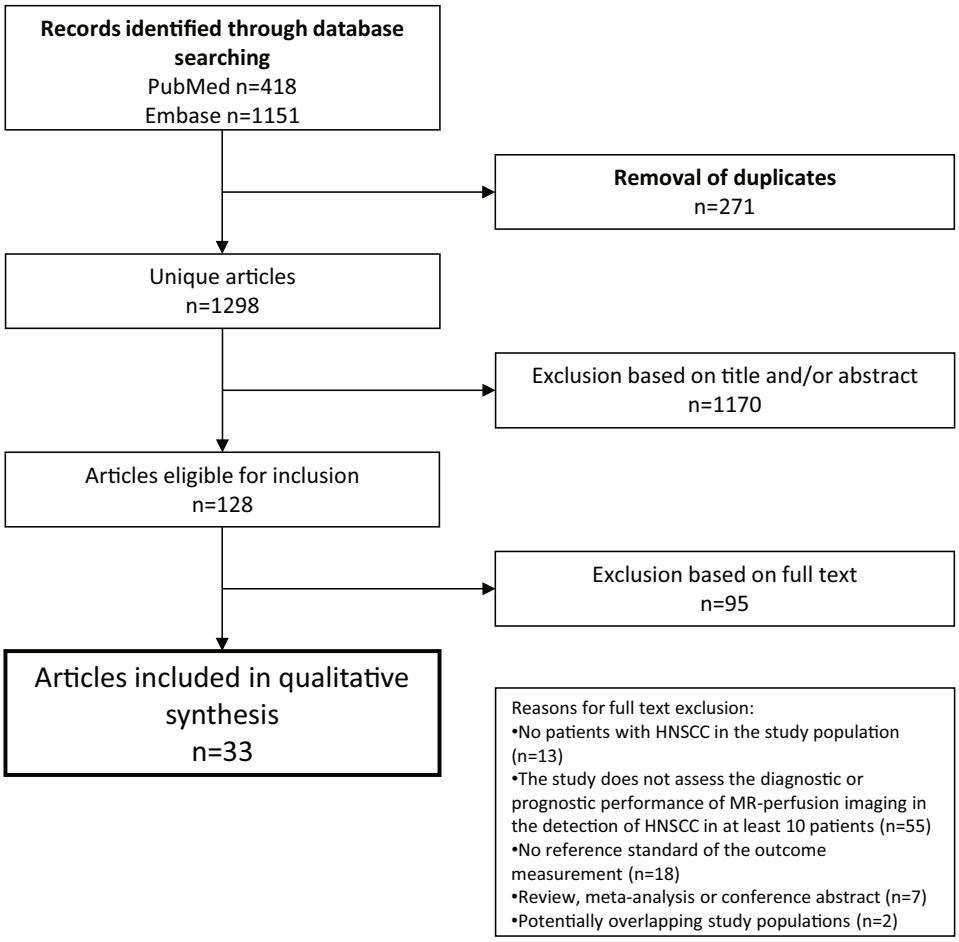


Figure 1 *Flow chart of study inclusion*

Abbreviations: HNSCC = Head and neck squamous cell carcinoma

RESULTS

Our search in Pubmed and Embase yielded 1,298 unique studies. Based on title and abstract 1,170 studies were excluded and another 95 studies were excluded based on the full-text (Figure 1). Thirty-three studies (22 diagnostic and 11 prognostic) were included for qualitative analysis. The total study population consisted of 738 patients from diagnostic studies and 348 patients from prognostic studies (Appendix B: patient characteristics). In 29 studies DCE-MRI was used (13-41), DSC-MRI was used in four studies (42-45).

In Appendix C abbreviations and relevant definitions are mentioned. In the DCE-MRI studies the output parameters of the studies were the quantitative parameters amplitude scaling constant (AH), elimination of contrast medium from the central compartment K_{el}, extraction ratio (E), intracellular water lifetime (w_i), k_{ep}, K^{trans}, permeability surface (PS), plasma volume fraction (v_p), time of arrival (TA) and v_e. Semi-quantitative parameters were AUC, maximal signal rise, maximum CI, maximum CI gain, maximum slope of increase ratio (MSIR), peak enhancement, peak time, relative enhancement (RE), relative slope, signal enhancement to noise ratio (SE/N), signal intensity (A), time to peak (TTP), time-signal intensity curve (TIC), wash out percentage, washout slope. In the DSC studies the output parameters were BF, BV, DSC% and relative BV (Appendix D and E).

Baseline study characteristics

Baseline study characteristics are reported in Table 1. Temporal resolution ranged from 1.3-30 seconds in T1-weighted DCE-MRI and from 1.2-3.8 seconds in T2*-weighted DSC-MRI (Table 1). In 18 studies the use of an arterial function (AIF) is mentioned (DCE-MRI: n=16; DSC-MRI: n=2) (13-17, 29-39, 42, 43). In five studies a population based AIF is used (29-33), in another eight a feeding artery is used (usually one of the carotid arteries or vertebral arteries) (14-16, 34, 35, 38, 42, 43), one study combines several methods (36), two studies use a model-based AIF (13, 17) and one study reports the use of an AIF, but this is not specified further (37). In 18 studies the use of a pharmacokinetic model is reported (DCE-MRI: n=17; DSC-MRI: n=1) (13-18, 29-39, 43). The Tofts model (n=6) (16, 29-32, 37) was most commonly used. In four studies two different scanners were used (28, 34, 35, 38). In none of these studies data is reported separately per scanner.

In diagnostic studies v_e (n=6) (14, 16, 17, 29-31), K^{trans} (n=6) (14, 16, 17, 29-31) and k_{ep} (n=5) (14, 17, 29-31) are the most frequently reported parameters. In prognostic studies K^{trans} (n=6) (32-35, 38, 39) and v_e (n=5) (32-35, 39) are most frequently reported.

Table 1 *Baseline characteristics*

Author, year	Study design	Study type	Image timing	Tesla	Voxel size (x*y*z) (mm)	Perfusion type	Perfusion sequences	Contrast administration	Pharmacokinetic model	Arterial input function	Temporal resolution	T1 map
Sumi (41), 2014	R	D	Pre	1.5	0.78*1.11*4-5	DCE	TSE T1	0.2 mmol/kg Gadopentate dimeglumine at 1.5 ml/s with a power injector; acquisition time 180s	...	...	10	...
Ai (40), 2013	R	D	Pre	1.5	1.05*0.94 *5	DCE	FSE T1	0.1 mmol/kg Gd-DTPA at 2 ml/s with a power injector; acquisition time 300s	...	...	30	yes
Chawla (34), 2013	R	P	Pre	1.5-3.0	1.02*1.02*5	DCE	Modified 3D-SPGR	0.1 mmol/kg Gd-DTPA at 1 ml/s with a power injector; acquisition time 10min	Generalized kinetic model	Semiautomatically via carotid arteries	2.5	yes
Ng (39), 2013)	P	P	Pre	3	2.12 *1.80 *5	DCE	3D GRE	0.1 mmol/kg Gd-DTPA at 3 ml/s with a power injector	Kety model	Adjacent carotid artery	3.3	yes
Agrawal (13), 2012	P	P	Pre and post	1.5	2.81*2.11*6	DCE	3D-SPGR	Gd-DTPA-BMA at 5 ml/s with a power injector; acquisition time 168s	Piecewise linear fitting model	Automatic AIF	5.25	yes
Chikui (33), 2012	...	P	Pre and post	1.5	1.72*1.82*5	DCE	3D-T1 FFE	0.1 mmol/kg Gadopentate dimeglumine at 2 ml/s with a mechanical injector; acquisition time 280s	Model free	Population based bi-exponential with slow bolus	3.5	yes
Jansen (29), 2012	P	D	Pre	1.5	...*...*5-6	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s; acquisition time 11.25-60s	Two compartments Tofts model	Population based	3.75-7.5	...
Jansen (30), 2012	R	D	Pre	1.5	0.70-0.78*1.41-1.56*5-7	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s; acquisition time 12-47.2s	Two compartments Tofts model	Population based	4.0-5.9	...
Lee (16), 2012	...	D	...	1.5	...	DCE	GRE	0.2ml/kg Gadoterate meglumine at 3 ml/s with a power injector; acquisition time 371s	Tofts-Kermode model	Common or external carotid artery	3.5	Calculated T1 map
Shukla-Dave (32), 2012	R	P	Pre	1.5	0.70-0.78*1.41-1.56*5-7	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s with a power injector; acquisition time 11.25-45	Tofts model	Population based bi-exponential	3.75-7.5	...
Wang (37), 2012	P	P	Pre and intra	3	2*2*2	DCE	3D GRE	Gd-DTPA	Modified two-compartment Tofts model and Mullani/Hermans model	yes (not further specified)	7.6	...
Wendl (19), 2012	R	D	Pre	3	1.2*0.9*3	DCE	3D VIBE	0.1 mmol/kg Gadoterate meglumine at 2 ml/s; acquisition time 90s	...	...	9	yes
Chawla (38), 2011	R	P	Pre	1.5-3.0	1.02*1.02*5	DCE	Fast 3D SGRE	0.1 mmol/kg Gd-DTPA at 1 ml/s with a power injector; acquisition time 10min	Shutter-speed model	Semi-automatically via carotid arteries	2.5	yes
Chikui (18), 2011	...	P	Pre (n=23)	1.5	1.72*1.82*5	DCE	DCE T1 FFE	0.1 mmol/kg Gadopentate dimeglumine at 2 ml/s	Brix model	No	3.5	yes

Table 1 *continued*

Author, year	Study design	Study type	Image timing	Tesla	Voxel size (x*y*z) (mm)	Perfusion type	Perfusion sequences	Contrast administration	Pharmacokinetic model	Arterial input function	Temporal resolution	T1 map
Sumi (41), 2014	R	D	Pre	1.5	0.78*1.11*4-5	DCE	TSE T1	0.2 mmol/kg Gadopentate dimeglumine at 1.5 ml/s with a power injector; acquisition time 180s	...	...	10	...
Ai (40), 2013	R	D	Pre	1.5	1.05*0.94 *5	DCE	FSE T1	0.1 mmol/kg Gd-DTPA at 2 ml/s with a power injector; acquisition time 300s	...	...	30	yes
Chawla (34), 2013	R	P	Pre	1.5-3.0	1.02*1.02*5	DCE	Modified 3D-SPGR	0.1 mmol/kg Gd-DTPA at 1 ml/s with a power injector; acquisition time 10min	Generalized kinetic model	Semiautomatically via carotid arteries	2.5	yes
Ng (39), 2013)	P	P	Pre	3	2.12 *1.80 *5	DCE	3D GRE	0.1 mmol/kg Gd-DTPA at 3 ml/s with a power injector	Kety model	Adjacent carotid artery	3.3	yes
Agrawal (13), 2012	P	P	Pre and post	1.5	2.81*2.11*6	DCE	3D-SPGR	Gd-DTPA-BMA at 5 ml/s with a power injector; acquisition time 168s	Piecewise linear fitting model	Automatic AIF	5.25	yes
Chikui (33), 2012	...	P	Pre and post	1.5	1.72*1.82*5	DCE	3D-T1 FFE	0.1 mmol/kg Gadopentate dimeglumine at 2 ml/s with a mechanical injector; acquisition time 280s	Model free	Population based bi-exponential with slow bolus	3.5	yes
Jansen (29), 2012	P	D	Pre	1.5	...*...*5-6	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s; acquisition time 11.25-60s	Two compartments Tofts model	Population based	3.75-7.5	...
Jansen (30), 2012	R	D	Pre	1.5	0.70-0.78*1.41-1.56*5-7	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s; acquisition time 12-47.2s	Two compartments Tofts model	Population based	4.0-5.9	...
Lee (16), 2012	...	D	...	1.5	...	DCE	GRE	0.2ml/kg Gadoterate meglumine at 3 ml/s with a power injector; acquisition time 371s	Tofts-Kermode model	Common or external carotid artery	3.5	Calculated T1 map
Shukla-Dave (32), 2012	R	P	Pre	1.5	0.70-0.78*1.41-1.56*5-7	DCE	Fast multiphase SGRE	0.1 mmol/kg Gd-DTPA at 2 ml/s with a power injector; acquisition time 11.25-45	Tofts model	Population based bi-exponential	3.75-7.5	...
Wang (37), 2012	P	P	Pre and intra	3	2*2*2	DCE	3D GRE	Gd-DTPA	Modified two-compartment Tofts model and Mullani/Hermans model	yes (not further specified)	7.6	...
Wendl (19), 2012	R	D	Pre	3	1.2*0.9*3	DCE	3D VIBE	0.1 mmol/kg Gadoterate meglumine at 2 ml/s; acquisition time 90s	...	...	9	yes
Chawla (38), 2011	R	P	Pre	1.5-3.0	1.02*1.02*5	DCE	Fast 3D SGRE	0.1 mmol/kg Gd-DTPA at 1 ml/s with a power injector; acquisition time 10min	Shutter-speed model	Semi-automatically via carotid arteries	2.5	yes
Chikui (18), 2011	...	P	Pre (n=23)	1.5	1.72*1.82*5	DCE	DCE T1 FFE	0.1 mmol/kg Gadopentate dimeglumine at 2 ml/s	Brix model	No	3.5	yes

Table 1 *continued*

Author, year	Study design	Study type	Image timing	Tesla	Voxel size (x*y*z) (mm)	Perfusion type	Perfusion sequences	Contrast administration	Pharmacokinetic model	Arterial input function	Temporal resolution	T1 map
Razek (44), 2011	P	D	Pre	1.5	...*...*5	DSC	T2*w-EPI	acquisition time 5min 0.2 mmol/kg Gadopentate dimeglumine at 5 ml/s with an automatic injector; acquisition time 110s	...	...	2	Not applicable
Abdel Razek (45), 2011	P	D	...	1.5	0.98-1.17*1.12-1.34*4	DSC	T2*w-EPI	0.1 mmol/kg Gadopentate dimeglumine at 4 ml/s with an automatic injector; acquisition time 110s	...	...	2	Not applicable
Bisdas (43), 2009	P	D	Pre	1.5	1.80*1.80*6	DSC	T2-EPI fs	0.2 mmol/kg Gadopentetate dimeglumine at 4 ml/s with a power injector; acquisition time 76s	Two-compartments distributed parameter model	Ipsilateral external or internal carotid artery	1.2	yes
Wu (42), 2004	P	D	Pre	1.5	3.29*1.80*5	DSC	T2* FLASH	0.2 mmol/kg Gadopentate dimeglumine at 5 ml/s; acquisition time 113s	...	Common or internal carotid artery or sternocleidomastoid muscle	3.8	Not applicable

Abbreviations: CRx = chemoradiotherapy; Cx = chemotherapy; DCE = dynamic contrast enhanced; DSC = dynamic susceptibility contrast; EPI = echo planar imaging; FFE = fast-field echo sequence; fGRE = fast gradient echo; FISP = fast imaging with steady state precession; FLASH = fast low angle shot; Fs= fat saturation; fSGRE = fast spoiled gradient echo; fSGRS = fast spoiled gradient recalled sequence; GRASS = gradient recalled acquisition in the steady state ; GRE = gradient echo; iCRx = induction chemoradiotherapy; Rx = radiotherapy; SE = spin echo; SGRE = spoiled gradient echo; SPGR = spoiled gradient recalled echo; SRTF= saturation-recovery-turbo-FLASH; Sx = surgery; TSE = turbo spin echo; VIBE = volumetric interpolated breath hold examination

Due to large heterogeneity in terms of used MRI acquisition pulse sequences, reference standards and postprocessing methods it was impossible to perform any quantitative analysis on the data pooled together. To avoid very large tables only significant relations (P<0.05) and/or strong correlations (|ρ|>0.5) are reported in Table 2 and 3 for diagnostic and prognostic studies respectively. A complete overview of study results is provided in Appendix D (diagnostic studies) and Appendix E (prognostic studies).

Sensitivity and specificity
Sensitivity and specificity of various DCE and DSC parameters for various outcome measurements could be extracted from 11 diagnostic and six prognostic studies (Table 2 and 3). Two forest plots were created (Figure 2 and 3).

Fifteen different DCE-parameters and two DSC-parameters were used to assess diagnostic accuracy (i.e. sensitivity and specificity) (Figure 2). Maximum contrast index (CI) (n=2) and maximum CI gain (n=2) are the most frequently investigated parameters for assessing diagnostic accuracy (21, 25). The CI is calculated with the following formula:

$$CI = \frac{signal\ intensity\ post\ contrast - signal\ intensity\ precontrast}{signal\ intensity\ precontrast}$$

For the parameter CI high sensitivity (i.e. 100% tumor proliferation microvessel density (MVD (21)) is reported in relation to cell proliferation. This is at the expense of specificity (i.e. 58% (21)) or vice versa (e.g. sensitivity=44% and specificity=95% (21)).

In the prognostic studies 10 different DCE-parameters are mentioned (Figure 3). The only parameter which is mentioned more than once in relation to clinical response is K^{trans} (34, 35). In these studies there also seems to be a trade-off between sensitivity and specificity (44/88% (34) and 89/63% (35)).

Bias assessment
The summary results of QUADAS-2 and QUIPS are reported in Figure 4 and 5, full results are reported in Appendix F and G.

The use of the QUADAS-2 tool yielded the following findings. In seven studies it was specified that consecutive patients were enrolled (14, 20, 26, 39, 40, 44, 45). Two studies had a case-control design (19, 42). In another two studies this was unclear (16, 45). In one study DCE-results were interpreted with knowledge of the reference test (14). In 21 studies this was unclear. In five studies the used reference standard raised concerns. Bisdas et al. used CT as reference standard (43). Chawla et al. (34) used clinical assessment for patient follow-up, but did not specify this further. In another study by Chawla et al. (38) and a study by Ng et al. (39) various clinical investigations were used in the follow-up of patients which may lead to verification bias. In the study performed by Jansen et al. (29)

MRI data was available for the pathologist when evaluating the resected nodes. In another study by Jansen et al. (30) both PET and PET-CT were used as reference standard. In the study by Van Cann et al. (14) histopathologic information was accessible for the radiologist assessing the DCE-images.

The use of the QUIPS tool for assessing the quality of prognostic studies yielded the following findings. The recruitment period is mentioned in only two of the prognostic studies (38, 39). In one study no inclusion and exclusion criteria are mentioned (37).

Table 2 *Summary of diagnostic study results*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Sumi (41), 2014					79
TIC pattern	Histopathological diagnosis	…	…	SCC >< benign: Sens: 76% Spec: 57%[a] SCC >< lymphoma: Sens: 94% Spec: 50% [b]	
Ai (40) 2013					46
TIC pattern	Histopathological diagnosis	…	…	SCC >< benign: Sens: 79% Spec: 91%[c]	
Jansen (29) 2012					12
Ktrans median	VEGF	Spearman rank correlation	NS	---	
Ktrans SD	VEGF	Spearman rank correlation	NS	+++	
Ktrans SD	Ki-67	Spearman rank correlation	<0.01	---	
Kep SD	VEGF[d]	Spearman rank correlation	<0.01	R = +++ Sens: 73% Spec: 100% AUC: 0.85 (CI95%: 0.64-1.00) YI: 0.73	
Ve median	VEGF	Spearman rank correlation	NS	---	
Ve SD	VEGF	Spearman rank correlation	NS	+++	
Ve SD	Ki-67	Spearman rank correlation	<0.01	---	
Jansen (30), 2012					16
Ktrans SD	Short term response	Logistic regression	NS	AUC: 0.50	
Lee (16), 2012					63
AUC90 mean	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 88% Spec: 68%	
AUC60 25%	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 100% Spec: 36%	
Wendl (19), 2012					10
TTP	Histopathological diagnosis	ROC analysis	NS	Sens: 100% Spec: 40% AUC: 0.68 (CI95%: 0.33-1.00) YI: 0.40	
Maximal signal rise	Histopathological diagnosis	ROC analysis	<0.01	Sens: 100% Spec: 100% AUC: 1.00 (CI95%: 1.00-1.00) YI: 1.00	
Sumi (20), 2011					43
Type 2 TIC	Presence of ENS	ROC analysis	…	Sens: 77% Spec: 100%	
Type 2 TIC	Presence of ENS	Spearman rank correlation	<0.001	+++	
Type 4 TIC	Presence of ENS	Spearman rank correlation	<0.001	---	
					27
Bisdas (17), 2010					
Ve	SUVmean	Spearman rank correlation	<0.05	++	
iAUC	SUVmean	Spearman rank correlation	<0.001	+++	
iAUC	SUVmax	Spearman rank correlation	<0.001	+++	
Jansen (31), 2010					13
Ktrans median	FMISO SUV	Mann Whitney U	<0.05	…	
Kep median	FMISO SUV	Spearman rank correlation	<0.05[e]	---	
Kep skewness	FMISO SUV	Mann Whitney U	<0.05	…	

Table 2 *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Unetsubo (21), 2009					28
Maximum CI	PCNA labelling index	ROC analysis	NS	Sens: 100% Spec: 31% AUC: 0.52 (CI95%: 0.28-0.75) YI: 0.31	
Maximum CI	MVD	ROC analysis	NS	Sens: 44% Spec: 95% AUC: 0.70 (CI95%: 0.49-0.90) YI: 0.39	
Maximum CI gain	PCNA labelling index	Spearman rank correlation	<0.05	++	
Maximum CI gain	PCNA labelling index	ROC analysis	NS	Sens: 100% Spec: 54% AUC: 0.63 (CI95%: 0.40-0.87) YI: 0.54	
Maximum CI gain	MVD	Spearman rank correlation	<0.01	++	
Maximum CI gain	MVD	ROC analysis	<0.01	Sens: 100% Spec: 58% AUC: 0.83 (CI95%: 0.67-0.99) YI: 0.58	
CI-gain/CI-max ratio	PCNA labelling index	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	PCNA labelling index	ROC analysis	<0.05	Sens: 100% Spec: 46% AUC: 0.75 (CI95%: 0.57-0.94) YI: 0.46	
CI-gain/CI-max ratio	MVD	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	MVD	ROC analysis	<0.01	Sens: 89% Spec: 79% AUC: 0.83 (CI95%: 0.68-0.98) YI: 0.68	
Van Cann (14), 2008					25
Ktrans	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 92% Spec: 85% AUC: 0.94 (CI95%: 0.86-1.00) YI: 0.76	
Kep	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 100% Spec: 92% AUC: 0.99 (CI95%: 0.97-1.00) YI: 0.92	
Ve	Medullary mandibular invasion	ROC analysis	NS	Sens: 67% Spec: 54% AUC: 0.49 (CI95%: 0.26-0.73) YI: 0.21	
Ariyoshi (22), 2006					20
AR SE/N	Tumor invasion peripheral area	Mann-Whitney U test	<0.05	…	
Tomura (23), 2005					27
MSIR[f]	Histologic grading	Mann-Whitney U test	<0.05	Sens: 77% Spec: 100%	
Hietschold (36), 2004					18
S1DE rel slope	SUV mean initial + 50Gy	Linear regression	≤0.05	0.19	
S1DE rel slope	SUV mean	Linear regression	≤0.05	0.21	
S2 flow	pO2 <5mmHg initial + 50 Gy	Linear regression	≤0.05	0.21	
S2DE flow	SUVmean	Linear regression	≤0.05	0.15	
S2DE flow	volume PET initial	Linear regression	≤0.05	0.29	
Fischbein (24), 2003					21
Peak time	Presence of tumor	T-test	<0.001	…	
Peak enhancement	Presence of tumor	T-test	<0.05	…	
Maximum slope	Presence of tumor	T-test	<0.01	…	
Washout slope	Presence of tumor	T-test	<0.05	…	
Konouchi (25), 2003					30
Maximum CI	PCNA[g]	Student's t-test → ROC analysis	<0.001	R= +++ Sens: 94% Spec: 92% AUC: 0.96 (CI95%: 0.90-1.00) YI: 0.86	

Table 2 *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Maximum CI gain	PCNA[h]	Student's t-test → ROC analysis	<0.01	R= +++ Sens: 56% Spec: 92% AUC: 0.79 (CI95%: 0.62-0.95) YI: 0.47	
Tomura (26), 2002					10
Peak time	Presence of tumor	…	…	Sens: 80%	
Abdel Razek (45), 2011					45
DSC%[i]	SCC >< NHL	ROC analysis	…	Sens: 95% Spec: 91% AUC: 0.97	
Bisdas (43), 2009					23
Blood flow (MR)	Blood flow (CT)	Parametric comparison test	<0.01	…	
T1 (MR)	T1 (CT)	Parametric comparison test	≤0.001	…	
T0 (MR)	T0 (CT)	Parametric comparison test	≤0.001	…	
PS (MR)	PS (CT)	Parametric comparison test	<0.01	…	
Vp (MR)	Vp (CT)	Parametric comparison test	≤0.001	…	
Wu (42), 2004					18

Correlations are reported as follows: -1.0 to -0.5 = - - -; - 0.5 to -0.3 = - -; -0.3 to -0.1 = -; -0.1 to 0.1 = +/-; 0.1 to 0.3 = +; 0.3 to 0.5 = + +; 0.5 to 1.0 = + + +

[a] Cut-off at TIC 2/3
[b] Cut-off at TIC 3/4
[c] Cut-off at TIC A/B
[d] Cut-off at VEGF -/+
[e] Mann Whitney U
[f] Cut-off at MSIR <2.5
[g] Cut-off at mean PCNA
[h] Cut-off at mean PCNA
[i] Cut-off at DSC% of 43.5%

Abbreviations: AR = ascending rate; AUC = area under the curve; CA-IX = carbonic anhydrase; CI = contrast index; CI95% = 95% confidence interval; DSC = dynamic susceptibility contrast; E = extraction ratio; ENS = extranodal spread; FMISO = 18F-fluoromisonidazol; HE = haematoxylin-eosin; HIF1α= Hypoxia-inducible factor 1-alpha; iAUC = initial area under the curve; Kep = rate constant between extracellular extravascular space and blood plasma; Ki-67 = marker of cellular proliferation; Ktrans = volume transfer constant between plasma and extracellular extravascular space; MSIR = maximum slope of increase ratio; MVD = microvessel density; NHL = non-Hodgkin lymphoma; NS = not significant; PCNA = proliferating cell nuclear antigen; PET = positron emission tomography; PS = permeability surface; R = regression coefficient; rBV= relative blood volume; ROC = receiver operating characteristic; S1DE = relative slope of the signal intensity corrected to TE = 0; S2DE = relative change of the spin-spin relaxation rate; SCC = squamous cell carcinoma; SD = standard deviation; SE/N = signal enhancement to noise ratio; Sens = sensitivity; Spec = specificity; SUV = standardized Uptake Value; T0 = lag time; T1 = intravascular blood volume; TIC = time-signal intensity curve; TLG= total lesion glycolysis; TTP = time to peak; Undiff = undifferentiated carcinoma; Ve = volume of extravascular extracellular space per unit volume of tissue; VEGF = vascular endothelial growth factor; YI = Youden Index

Subgroup analyses to identify possible confounders were carried out in only four prognostic studies. Shukla-Dave et al. (32) performed a subgroup analysis for patient who received chemoradiotherapy as primary treatment. Kim et al. performed separate analyses for patients who received conventional chemotherapy or cetuximab combined with radiotherapy (35). In one of the studies by Chawla et al. (34) a subgroup analysis was performed for the patients who received concurrent chemoradiotherapy without induction chemotherapy. In another study by Chawla et al. (38) the potential bias introduced by neck dissection was evaluated by performing a subgroup analysis for patients who underwent neck dissection within six months and those without. In only four studies the number of excluded patients is mentioned (18, 33, 38, 39).

Even though the included studies were heterogeneous, several studies assessed the same topics which are discussed separately below. We started by assessing diagnostic performance of perfusion-weighted MRI, followed by the correlation of DCE-MRI to other biomarkers, and ended with the prognostic capacity and detection of residual disease.

Differentiation between HNSCC and other malignant or benign lesions
Lee et al. (16) used histogram analysis of (semi-)quantitative DCE-parameters and found the semi-quantitative parameter AUC90 to be the most accurate parameter in separating HNSCC from undifferentiated carcinoma (i.e. more accurate than K^{trans}). Neither AUC90 nor K^{trans} could reliably differentiate HNSCC from lymphoma. Histogram analysis may be an appropriate method to differentiate between tumors, taking tumor heterogeneity into account.

Two studies used TIC to differentiate between various malignant en benign lesions. In the first study Ai et al. (40) assessed the TIC pattern of lesions of the tongue. HNSCC could be distinguished from benign lesions with a sensitivity and specificity of 79% and 91%, respectively. In the second study, by Sumi et al. (41), both benign and malignant lesions and malignant nodes were analyzed by comparing TIC patterns. In this study HNSCC could be distinguished from benign lesions with a sensitivity and specificity of 67% and 68%, respectively. The difference in diagnostic accuracy between the studies is mainly caused by the presence of Warthin tumors in the second study, these tumors have a TIC pattern similar to malignant lesions resulting in lower sensitivity and specificity. Unfortunately, both studies use different cut-offs to define TIC profiles, this limits the comparability.

In two studies using DSC-MRI by Abdel Razek et al. (44, 45) thresholds in DSC% were used to differentiate HNSCC from other malignant and benign lesions. DSC% is calculated as:

$$\left(\frac{S0 - S1}{S0}\right) \times 100\%$$

Where S0 represents the signal intensity of the lesion before descent, and S1 the signal intensity at the peak descent.

In both studies the DSC% could differentiate malignant from benign lesions with a sensitivity

and specificity up to 97% (95%CI, 84-100%) and 83% (95%CI, 52-98%), respectively. In one study HNSCC showed a (non-significant) low DSC% compared to other malignant lesions (44). However, vascular benign lesions were considered to be malignant based on DSC%. In the other study it was possible to reliably differentiate HNSCC nodes from non-Hodgkin lymphoma (45).

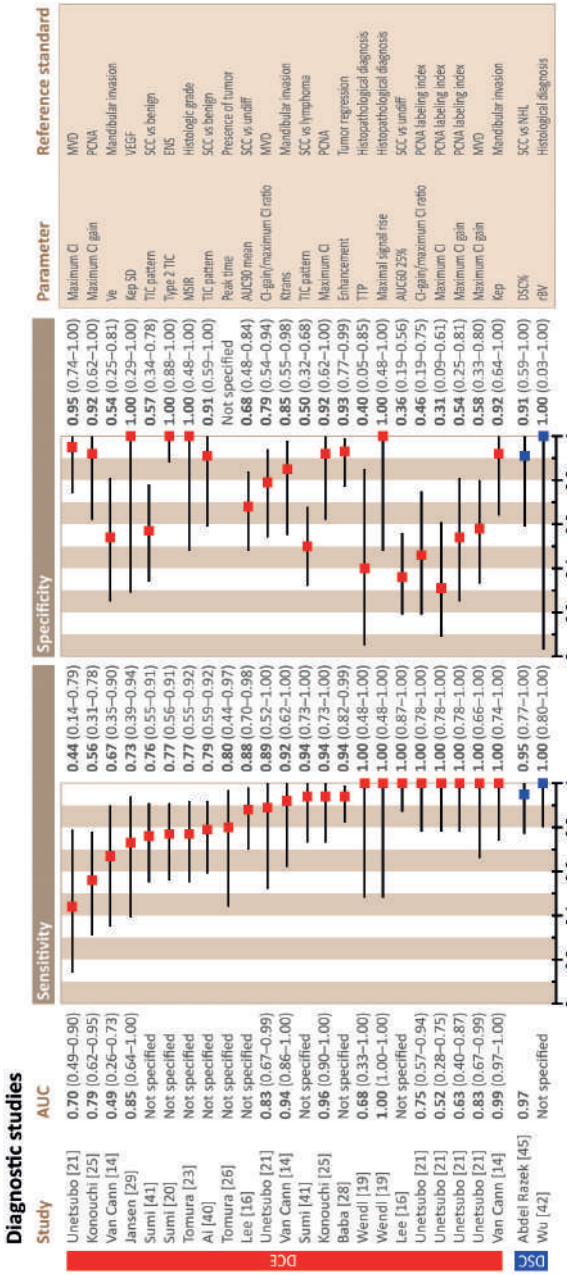


Figure 2 *Forest plot diagnostic studies sorted by sensitivity. Sensitivity and specificity are reported with their 95% CIs as horizontal lines.*

Abbreviations: AUC = area under the curve; CI = contrast index; DSC = dynamic susceptibility contrast; ENS = extranodal spread; K_{ep} = rate constant between extracellular extravascular space and blood plasma; K^{trans} = volume transfer constant between plasma and extracellular extravascular space; MSIR = maximum slope of increase ratio; MVD = microvessel density; NHL= non-Hodgkin lymphoma; PCNA = proliferating cell nuclear antigen; rBV = relative blood flow; SCC = squamous cell carcinoma; SD = standard deviation; TIC = time-signal intensity curve; TTP = time to peak; Undiff = undifferentiated carcinoma; V_e = volume of extravascular extracellular space per unit volume of tissue; VEGF = vascular endothelial growth factor

In conclusion various perfusion-weighted MRI parameters may be useful for pretreatment differentiation between benign and malignant tissue.

Correlation between DCE-MRI and (immuno)histological parameters

In a study on metastatic lymph nodes Jansen et al. (29) reported a significant positive correlation (ρ=0.81, *P*<0.001) between the standard deviation (SD) of k_{ep} and vascular endothelial growth factor (VEGF) receptor staining, a marker of angiogenesis. The authors stated that the width of SD was indicative of lymph node heterogeneity. Thus heterogeneous lymph nodes were likely to show angiogenesis.

In two other studies (21, 25) the CI was compared with the proliferating cell nuclear antigen (PCNA) labeling index, a marker of cell proliferation (46), in patients with oral squamous cell carcinoma. In both studies there was a positive correlation between maximum CI gain and PCNA index (r=0.544, *P*<0.01 (25) and r=0.378, *P*<0.01) (21). However, the results were not consistent regarding the correlation between maximum CI and PCNA index (r=0.866, *P*<0.001 (25) and r=0.129, *P*>0.05 (21)). Also, the reported values in the more recent study were considerably lower (10.1% for maximum CI, 13% for maximum CI gain and 25% for PCNA labeling index) (21). Unfortunately, disease stage was specified only in the more recent study and the authors did not give an explanation for the difference in results.

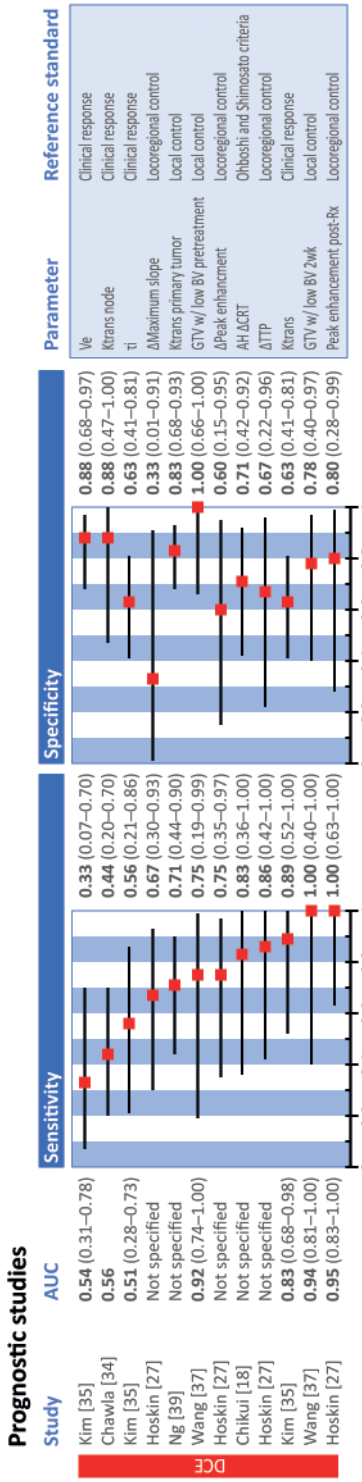


Figure 3 *Forest plot prognostic studies sorted by sensitivity. Sensitivity and specificity are reported with their 95% CIs as horizontal lines.*

Abbreviations: AH = amplitude scaling constant; BV = blood volume; CRT = chemoradiotherapy; GTV = gross tumor volume; K^{trans} = volume transfer constant between plasma and extracellular extravascular space; TTP = time to peak; V_e = volume of extravascular extracellular space per unit volume of tissue; Ti = intracellular water lifetime

Correlation between DCE-MRI and PET

Jansen et al. (31) found that ^{18}F-Fluoromisonidazole (^{18}F-MISO) standardized uptake value (SUV), a PET parameter of hypoxia, showed a strong negative correlation with median of k_{ep} (ρ=-0.58, *P*<0.05) in metastatic lymph nodes. Also, hypoxic nodes had a lower K^{trans}. This suggests that DCE-MRI may be used to identify hypoxic lymph nodes. The authors stated that hypoxia might play a role in heterogeneity of lymph nodes. However only the median of k_{ep} and skewness of k_{ep} were significantly different between hypoxic and non-hypoxic nodes and SD of k_{ep} was not. In an earlier mentioned study by Jansen et al. (29) DCE-MRI was compared with immunohistochemical markers of angiogenesis. These studies are difficult to compare because different reference standards were used: hypoxia as assessed by FMISO SUV and angiogenesis as determined by immunohistochemical staining of VEGF receptor.

Bisdas et al. (17) suggested that the lack of correlation between DCE-parameters (K^{trans} and k_{ep}) and fluorodeoxyglucose (FDG)-PET parameter standardized uptake value (SUV) indicates that both techniques provide complementary information. This was confirmed by Jansen et al. (30) in a study where DCE-MRI and FDG-PET were used in short-term response prediction. After the addition of SUV to SD of K^{trans} in a multivariable logistic regression model AUC rose from 0.50 to 0.96.

Correlation between DCE-MRI and tumor spread

Four studies assessed the relation between DCE-MRI and tumor spread (e.g. extranodal spread, tumor invasion and histological staging) (14, 20, 22, 26).

Tomura et al. (26) performed post-radiotherapy DCE-MRI in candidates for (hemi) glossectomy. Unfortunately, in only 10 of the 21 included patients DCE-MRI was performed. No relationship was observed between the DCE-parameter TTP and histologic grading. T2-weighted imaging outperformed DCE-MRI in the detection of residual viable tumor tissue. However, it was not specified why only 10 of 21 patients received DCE-MRI which makes it difficult to compare T2-weighted imaging with DCE-MRI. It should be noted that the main goal of this study was not to assess the value of DCE-MRI alone.

Van Cann et al. (14) used DCE-MRI for the assessment of tumor invasion of the cortex and medulla of the mandible. This is of clinical value because if medullary invasion is present segmental mandibular resection is warranted for curative treatment. With K^{trans} and k_{ep} the AUCs were 0.94 (95%CI, 0.86-1.00) and 0.99 (95%CI, 0.97-1.00) respectively. Unfortunately, the observers were not blinded to the histopathologic results, i.e. mandibular invasion.

Ariyoshi et al. (22) found the ascending rate of the signal enhancement to noise ratio (SE/N) of the peripheral area of SCC of the tongue to be significantly higher in patients with less invasive tumors (*P*<0.05). In the central area of the tumors this difference did not reach statistical significance.

Table 3 *Summary of prognostic study results*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
Chawla (34), 2013						24
Median 23.72 months (range 2.37-49.9m)	Ktrans node	Clinical response	Student t test	<0.01		
	Ktrans node	Clinical response	Univariate analysis	NS	Sens: 44% Spec: 88% AUC: 0.56	
Ng (39), 2013						58
Median 19.2 months (range 9-32.3 months)	Ktrans primary tumor	2-year control	Log rank test	<0.05	Hazard ratio: 0.34	
	Ktrans primary tumor	Local control	Univariable logistic regression	0.001	…	
	Ktrans primary tumor	Local control	Multivariable logistic regression → ROC analysis	<0.05	Sens: 71% Spec: 83%	
	Ktrans primary tumor	Local control	Student t test	0.01	Odds ratio: 0.06	
	Ve primary tumor	Local control	Univariable logistic regression	<0.05	…	
Agrawal (13), 2012						21
6 weeks	BF	Clinical response	Independent samples t-test	0.001	…	
	BF	T-stage	Independent samples t-test	<0.05	…	
	BV	clinical response	Independent samples t-test	0.001	…	
	BV	T-stage	Independent samples t-test	0.05	…	
	BV	control at primary	Independent samples t-test	0.05	…	
Chikui (33), 2012						29
…	Ktrans ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	Ve post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	Ve ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	AUGC post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
	AUGC ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
Shukla-Dave (32), 2012						74
Median 40 months (range 13-64 months)	Ktrans skewness	PFS	Cox regression analyses	<0.05	…	
	Ktrans skewness	OS	Cox regression analyses	<0.05	…	
	Stage IV only ktrans skewness	PFS	Cox regression analyses	<0.001	…	
	Stage IV only ktrans skewness	OS	Cox regression analyses	<0.001	…	
	Ve SD	PFS	Cox regression analyses	<0.05	…	
	Ve median	OS	Cox regression analyses	<0.01	…	
Wang (37), 2012						13
Median 19.6 months (range 14.1-36.4	Subvolume with low BF 2wk	Local control	Mann–Whitney U test	0.05	…	

Table 3 *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
months)						
	Subvolume with low BV pre-RT	Local control	Mann–Whitney U test	<0.05	…	
	Subvolume with low BV 2wk	Local control	Mann–Whitney U test	0.01	…	
	Mean ΔBV	Local control	Mann–Whitney U test	<0.05	…	
	GTV w/ low BV pretreatment	Local control	ROC analysis	…	Sens: 75% Spec: 100% AUC (SD): 0.92 (0.09) 95%CI: 0.74-1.00 YI: 0.75	
	GTV w/ low BV 2wk	Local control	ROC analysis	…	Sens: 100% Spec: 78% AUC (SD): 0.94 (0.07) 95%CI: 0.81-1.00) YI: 0.78	
	Change BV of GTV 2wk	Local control	ROC analysis	…	Sens: 85% Spec: 83% AUC (SD): 0.90 (0.08)	
	Change BV of pre GTV	Local control	ROC analysis	…	Sens: 85% Spec: 76% AUC (SD): Not specified	
Chawla (38), 2011						57
Median 30 months (range 13-48 months)	Ktrans	Disease-free survival	Hazard ratio	<0.05	3.8(95%CI: 1.0-13.9)	
Chikui (18(, 2011						23
54.4±5.9 days	AH pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH ΔCRT	Ohboshi and Shimosato criteria	Tukey's HSD	<0.05	Sens: 83% Spec: 71%	
Kim (35), 2010						33
Disease status at the end of CRT	Ktrans	Clinical response (CR/PR)	2-tailed t test with unequal variance	0.001	…	
	Ktrans	Clinical response (CR/PR)	ROC analysis	<0.01	Sens: 89% Spec: 63% AUC: 0.83 (95%CI: 0.68-0.98) YI: 0.51	
	Ve	Clinical response (CR/PR)	ROC analysis	NS	Sens: 33% Spec: 88% AUC: 0.54 (95%CI: 0.31-0.78) YI: 0.21	
	τi	Clinical response (CR/PR)	ROC analysis	NS	Sens: 56% Spec: 63%	

Table 3 *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
					AUC: 0.51 (95%CI: 0.28-0.73) YI: 0.18	
Hoskin (27), 1999						13
11.5 months (range 6-20 months)	ΔTTP	Locoregional control	…	…	Sens: 86% Spec: 67%	
	Peak enhancement post-Rx	Locoregional control	Student t test	<0.01	…	
	Peak enhancement post-Rx	Locoregional control	ROC analysis	<0.01	Sens: 100% Spec: 80% AUC: 0.95 (95%CI: 0.83-1.00) YI: 0.80	
	ΔPeak enhancment	Locoregional control	…	…	Sens: 75% Spec: 60%	
	ΔMaximum slope	Locoregional control	…	…	Sens: 67% Spec: 33%	

Abbreviations: A = signal intensity; AH = amplitude scaling constant; AUC = area under the curve; BF = blood flow; BV = blood volume; K21 = exchange rate constant; Ktrans = volume transfer constant between plasma and EES; NS = not significant; OS = overall survival; PFS = progression-free survival; Sens = sensitivity; Spec = specificity; TTP = time to peak; Ti = intracellular water lifetime; Ve = extravascular volume fraction; Vp = plasma volume fraction; YI = Youden Index

Extranodal spread (ENS) is an indicator of poor prognosis. Patients with ENS are at higher risk of developing distant metastases (47). Sumi et al. (20) used times-signal intensity curves (TIC) to assess the presence of ENS in 54 patients with HNSCC. Nodal areas displaying slow uptake of contrast (type 2 TIC) proved to be a significant predictor of ENS (P<0.001) which was independent of the short axis diameter of the node. The authors suggest this may be indicative of a epithelial to mesenchymal transition which is associated with invasive tumor growth (48). When using only TIC information sensitivity and specificity for the detection of ENS were 77% (95%CI, 56-91%) and 100% (95%CI, 88-100%), respectively. If a combination of short-axis diameter and TIC was used sensitivity increased to 96% (95%CI, 80-99%) and specificity remained the same.

Prognostic capacity of pretreatment DCE-MRI

Shukla-Dave et al. (32) investigated the distribution of quantitative parameters and showed that high skewness of K^{trans} is associated with lower progression-free and overall survival (P<0.05). The authors suggest that more heterogeneous tumors, as estimated by high skewness of K^{trans}, are associated with poor prognosis. However, the standard deviation of K^{trans}, another marker of heterogeneity, was neither a significant predictor of progression free nor overall survival. Unfortunately, the authors do not address this apparent discrepancy. Heterogeneity is associated with areas of hypoxia and necrosis in the tumor (49).

Chawla et al. (34) found significantly lower values of K^{trans} in solid parts of nodal masses of non-responders compared to responders to chemoradiotherapy. In a prospective study on primary oropharyngeal and hypopharyngeal tumors by Ng et al. (39) and a retrospective study on HNSCC lymph node metastases by Kim et al. (35) pretreatment DCE-MRI was used to predict the prognosis of patients treated with chemoradiotherapy. Low K^{trans} of the primary tumor (39) and malignant lymph nodes (35) were associated with poor prognosis. The findings of Chawla et al., Ng et al. and Kim et al. are in line with findings of Jansen et al. (31) who found lower values of K^{trans} in hypoxic nodes, since tumor hypoxia is associated with a poor prognosis.

In patients with mandibular invasion, a poor prognostic factor, higher values of K^{trans} were found by van Cann et al. (14). It can therefore not be concluded that low values of K^{trans} are always associated with poor prognosis.

Agrawal et al. (13) found BF and BV to be significantly higher in complete responders compared to partial responders to chemoradiotherapy (P=0.001). However higher tumor stage, which is associated with poor prognosis, is also associated with high BF and BV (i.e. T3-4 tumors had significantly higher BF and BV than T1-2 tumors (P=0.05)). It can therefore not be concluded that high BF and BV are associated with better prognosis. Pre- and post-treatment data were analyzed separately. Changes in BV and BF were not mentioned.

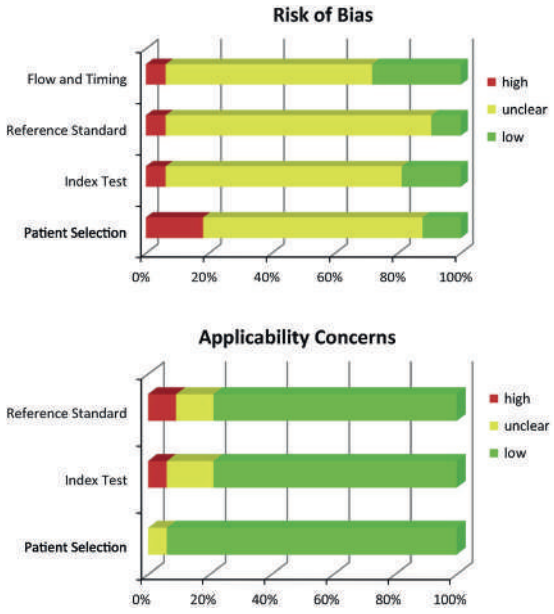


Figure 4 *QUADAS-2 results showing the risk of bias and applicability concerns as reviewed by the authors: presented as percentages across all the included studies.*

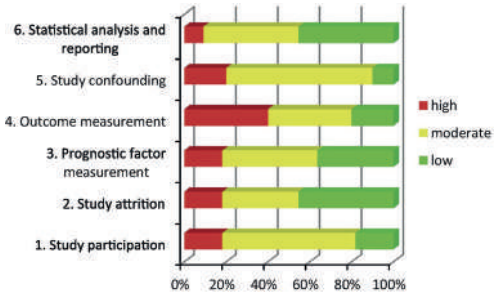


Figure 5 *QUIPS results showing the risk of bias on six domains as reviewed by the authors: presented as percentages across the included prognostic studies.*

Prognostic capacity of early follow-up DCE-MRI

Chikui et al. (33) used sequential DCE-MRI examinations (i.e. before and after chemoradiotherapy) to predict treatment response to chemoradiotherapy and found an increase in v_e (P=0.001) and K^{trans} (P<0.05) to be associated with good response. None of the pre- chemoradiotherapy parameters were significantly different between responders and non-responders. In another study Chikui et al. (18) used the Brix-model for pharmacological analysis. Only the AH and ΔAH were significantly different in non-responders compared to responders (i.e. low pretreatment AH and large increase in AH are associated with a good tumor response). AH is a parameter affected by several factors (e.g. the dose of contrast agent and the relative volume of the extracellular extravascular space).

Wang et al. (37) found patients with large subvolumes with poor perfusion (i.e. low BV) after chemoradiotherapy to have a poor prognosis. With the percentage of decrease of these subvolumes a better distinction between local control and local failure could be made than when analyzing total primary tumor volume. Patients with large areas of poor perfusion could therefore be candidates for local radiation dose intensification. Blood volume had a stronger association with local control than blood flow or the combination of blood volume and blood flow.

In a study from Hoskin et al. (27) DCE-MRI-findings before and immediately after radiotherapy were related to local tumor control with a median follow-up of 12 months. High post-radiotherapy relative signal intensity for individual pixels (E) (P<0.05) and an increase in TTP after therapy were significantly (P<0.05) associated with partial response. The authors recommend intensification of therapy in patients with high E after radiotherapy.

Detection of residual disease

Tomura et al. (23) used a ratio of the MSIR of the primary tumor versus that of muscle. Using an MSIR of 2.5 could differentiate between the presence and absence of viable tumor cells one week after radiotherapy with a sensitivity of 77% (95%CI, 55-92%) and a specificity of 100% (95%CI, 48-100%).

Baba et al. (28) found early tumor enhancement (30-90s) on dynamic MRI after radiation therapy to be indicative of residual tumor with a sensitivity of 94% (95%CI, 82-99%) and specificity of 93% (95%CI, 77-99%).

DISCUSSION

Quantitative compared to semi-quantitative DCE-parameters

The studies presented in this review utilize a variety of imaging endpoints, ranging from qualitative to pharmacokinetic-based quantitative parameters. Quantitative parameters offer the theoretical advantage of resembling tissue physiology more precisely than semi-quantitative parameters. However, an advantage of semi-quantitative parameters is that they are relatively simple to calculate compared to model based parameters like K^{trans}.

Of the included studies, two studies used both quantitative and semi-quantitative parameters. In the study by Lee et al. (16) only the semi-quantitative parameters $AUC90_{mean}$ and $AUC60_{25\%}$ could significantly differentiate squamous cell carcinoma from undifferentiated carcinoma. This may suggest that semi-quantitative parameters are more suitable for differentiating between various head and neck malignancies.

The second study, by Bisdas et al. (17), found stronger correlations with PET-parameter SUV for the semi-quantitative iAUC (ρ=0.66) than for the quantitative v_e (ρ=0.42). The authors suggested that the slightly lower correlation between v_e and PET-data was advantageous, as it implies that the information of both modalities is complementary. The mutually complementary information of DCE-MRI and PET data may therefore justify the use of hybrid PET-MRI systems.

More research focused on the direct comparison of the value of quantitative and semi-quantitative parameters is necessary to determine the optimal DCE-parameters.

Future applications of DCE-MRI

Even though K^{trans}, k_{ep} and v_e are the most commonly used parameters it remains challenging to perform a meta-analysis on these parameters: firstly, because different kinds of outcome measures were used and secondly due to large differences in calculation procedures and software platforms. As yet, there is no consensus on the calculation procedure of K^{trans}, resulting in a large heterogeneity of values, making it very difficult to compare the results, despite the claim of K^{trans} to be an absolute, scanner and scan-independent measure. Heye et al. demonstrated large varieties in the values of DCE-MRI parameters in two studies on patients with uterine fibroids (50, 51). The black-box nature of the used software may cause up to a 100-fold difference in parameters with the same name. This needs to be overcome to allow for the use of DCE-MRI in multi-centre trails (e.g. by standardization and a clear description of the used materials and mathematical analyses).

In head and neck cancer there usually are several arteries available to serve as input artery for the AIF. Based on biological properties we recommend the use of the external carotid artery (in lower tumors the common carotid artery or subclavian artery) instead

of the internal carotid artery because the external branches of the external carotid artery generally supply the primary HNSCC. These vessels are also used for selective intra-arterial chemotherapy perfusion, which indicates the role of these vessels in tumor perfusion (52)

Limitations

The included studies suffer from a number of limitations: firstly, the included studies were highly heterogeneous in terms of used sequences, postprocessing methods, malignancy criteria and reference standards. This made it neither possible to perform any meta-analysis nor to quantitatively assess the presence of bias. The used reference tests are not always optimal (e.g. when other imaging modalities are used). When (regular) clinical follow-up is used there is risk of verification bias (i.e. findings in clinical follow-up determine the use of additional tests). This can lead to an overestimation of test sensitivity and an underestimation of test specificity (53).

In some of the included studies multiple statistical tests were performed without correcting for these multiple comparisons. This may lead to an overestimation of significant results.

The QUADAS-2 and QUIPS tables indicate a moderate (and often unclear) risk of bias caused by patient selection. The number of eligible patients and in- and exclusion criteria are rarely mentioned. This may result in low bias for patient attrition in the QUIPS checklist, because all included patients often complete the entire study. In prognostic studies there is usually not accounted for confounders. This can be attributed to the relatively small patient populations which do not allow proper analysis of confounders.

Publication bias may lead to an overoptimistic appreciation of the value of perfusion-weighted MRI. Small studies with poor performance of perfusion-weighted MRI may be deemed not interesting enough to be published. We have not looked at the presence of publication bias with a test because the included studies are not homogeneous enough to perform such a test.

Even though most of the included studies are pilot studies with small patient populations, there still is room for improvement in describing patient inclusion, index test analysis (e.g. blinded analysis), reference standard analysis (e.g. blinded analysis, duration and type of follow-up) and describing flow and timing. For diagnostic studies we advise the use of the STARD checklist to allow for a clearer insight in used study methods and risk of bias (54). Further it would be easier to perform meta-analyses if studies reported per patient data (e.g. in an electronic supplement).

Conclusion

This systematic review gives an extensive overview of the current state of perfusion-weighted MRI for HNSCC. In conclusion perfusion-weighted MRI shows great potential in various aspects of diagnosing HNSCC and in the prediction of short-term prognosis; especially the correlation with hypoxia and tumor heterogeneity is promising (29, 30, 32). However, at this moment perfusion-weighted MRI is not considered to be reproducible enough to be used in clinical practice for HNSCC. More research with uniform study methods and with larger sample sizes is needed.

REFERENCE LIST

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62(1):10-29.
2. de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol. 2009;45(11):930-5.
3. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10(3):223-32.
4. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. Radiographics. 2009;29(5):1433-49.
5. Zahra MA, Hollingsworth KG, Sala E, Lomas DJ, Tan LT. Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy. The Lancet Oncology. 2007;8(1):63-74.
6. Cao Y. The promise of dynamic contrast-enhanced imaging in radiation therapy. Semin Radiat Oncol. 2011;21(2):147-56.
7. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.
8. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
9. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25.
10. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
11. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37.
12. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.
13. Agrawal S, Awasthi R, Singh A, Haris M, Gupta RK, Rathore RK. An exploratory study into the role of dynamic contrast-enhanced (DCE) MRI metrics as predictors of response in head and neck cancers. Clin Radiol. 2012;67(9):e1-5.
14. Van Cann EM, Rijpkema M, Heerschap A, van der Bilt A, Koole R, Stoelinga PJ. Quantitative dynamic contrast-enhanced MRI for the assessment of mandibular invasion by squamous cell carcinoma. Oral Oncol. 2008;44(12):1147-54.
15. Helbig M, Schlemmer HP, Lumer M, Knopp MV, Dietz A. [Dynamic magnetic resonance tomography (dMRT). It's value in advanced head-neck tumors treated with radiochemotherapy]. HNO. 2003;51(11):886-92.
16. Lee FK, King AD, Ma BB, Yeung DK. Dynamic contrast enhancement magnetic resonance imaging (DCE-MRI) for differential diagnosis in head and neck cancers. Eur J Radiol. 2012;81(4):784-8.
17. Bisdas S, Seitz O, Middendorp M, et al. An exploratory pilot study into the association between microcirculatory parameters derived by MRI-based pharmacokinetic analysis and glucose utilization estimated by PET-CT imaging in head and neck cancer. Eur Radiol. 2010;20(10):2358-66.
18. Chikui T, Kawano S, Kawazu T, et al. Prediction and monitoring of the response to chemoradiotherapy in oral squamous cell carcinomas using a pharmacokinetic analysis based on the dynamic contrast-enhanced MR imaging findings. Eur Radiol. 2011;21(8):1699-708.
19. Wendl CM, Muller S, Meier J, et al. High resolution contrast-enhanced ultrasound and 3-tesla dynamic contrast-enhanced magnetic resonance imaging for the preoperative characterization of cervical lymph nodes: First results. Clin Hemorheol Microcirc. 2012;52(2-4):153-66.

20. Sumi M, Nakamura T. Extranodal spread in the neck: MRI detection on the basis of pixel-based time-signal intensity curve analysis. J Magn Reson Imaging. 2011;33(4):830-8.

21. Unetsubo T, Konouchi H, Yanagi Y, et al. Dynamic contrast-enhanced magnetic resonance imaging for estimating tumor proliferation and microvessel density of oral squamous cell carcinomas. Oral Oncol. 2009;45(7):621-6.

22. Ariyoshi Y, Shimahara M. Relationships between dynamic contrast-enhanced MRI findings and pattern of invasion for tongue carcinoma. Oncol Rep. 2006;15(5):1339-43.

23. Tomura N, Omachi K, Sakuma I, et al. Dynamic contrast-enhanced magnetic resonance imaging in radiotherapeutic efficacy in the head and neck tumors. Am J Otolaryngol. 2005;26(3):163-7.

24. Fischbein NJ, Noworolski SM, Henry RG, Kaplan MJ, Dillon WP, Nelson SJ. Assessment of metastatic cervical adenopathy using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2003;24(3):301-11.

25. Konouchi H, Asaumi J-i, Yanagi Y, et al. Evaluation of tumor proliferation using dynamic contrast enhanced-MRI of oral cavity and oropharyngeal squamous cell carcinoma. Oral Oncol. 2003;39(3):290-5.

26. Tomura N, Watanabe O, Kato K, et al. Irradiated carcinoma of the tongue: correlation of MR imaging findings with pathology. AJR Am J Roentgenol. 2002;178(3):705-10.

27. Hoskin PJ, Saunders MI, Goodchild K, Powell ME, Taylor NJ, Baddeley H. Dynamic contrast enhanced magnetic resonance scanning as a predictor of response to accelerated radiotherapy for advanced head and neck cancer. Br J Radiol. 1999;72(863):1093-8.

28. Baba Y, Furusawa M, Murakami R, et al. Role of dynamic MRI in the evaluation of head and neck cancers treated with radiation therapy. Int J Radiat Oncol Biol Phys. 1997;37(4):783-7.

29. Jansen JF, Carlson DL, Lu Y, et al. Correlation of a priori DCE-MRI and (1)H-MRS data with molecular markers in neck nodal metastases: Initial analysis. Oral Oncol. 2012;48(8):717-22.

30. Jansen JF, Schoder H, Lee NY, et al. Tumor metabolism and perfusion in head and neck squamous cell carcinoma: pretreatment multimodality imaging with 1H magnetic resonance spectroscopy, dynamic contrast-enhanced MRI, and [18F]FDG-PET. Int J Radiat Oncol Biol Phys. 2012;82(1):299-307.

31. Jansen JF, Schoder H, Lee NY, et al. Noninvasive assessment of tumor microenvironment using dynamic contrast-enhanced magnetic resonance imaging and 18F-fluoromisonidazole positron emission tomography imaging in neck nodal metastases. Int J Radiat Oncol Biol Phys. 2010;77(5):1403-10.

32. Shukla-Dave A, Lee NY, Jansen JF, et al. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases. Int J Radiat Oncol Biol Phys. 2012;82(5):1837-44.

33. Chikui T, Kitamoto E, Kawano S, et al. Pharmacokinetic analysis based on dynamic contrast-enhanced MRI for evaluating tumor response to preoperative therapy for oral cancer. J Magn Reson Imaging. 2012;36(3):589-97.

34. Chawla S, Kim S, Dougherty L, 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;200(1):35-43.

35. Kim S, Loevner LA, Quon H, et al. Prediction of response to chemoradiation therapy in squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2010;31(2):262-8.

36. Hietschold V, Kittner T, Schreyer M, Appold S, Beuthien-Baumann B, Laniado M. MR perfusion measurements on pharyngeal tumors: comparison of quantification strategies. J Appl Clin Med Phys. 2004;5(4):96-111.

37. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. Med Phys. 2012;39(8):5277-85.

38. Chawla S, Kim S, Loevner LA, et al. Prediction of disease-free survival in patients with squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2011;32(4):778-84.

39. Ng SH, Lin CY, Chan SC, et al. Dynamic contrast-enhanced MR imaging predicts local control in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiotherapy. PLoS One. 2013;8(8):e72230.

40. Ai S, Zhu W, Liu Y, Wang P, Yu Q, Dai K. Combined DCE- and DW-MRI in diagnosis of benign and malignant tumors of the tongue. Front Biosci (Landmark Ed). 2013;18:1098-111.

41. Sumi M, Nakamura T. Head and neck tumours: combined MRI assessment based on IVIM and TIC analyses for the differentiation of tumors of different histological types. Eur Radiol. 2014;24(1):223-31.

42. Wu RH, Bruening R, Ducreux D, Berchtenbreiter C, Jund R, Reiser M. Estimation of relative blood volume in head and neck squamous cell carcinomas. J Neuroradiol. 2004;31(3):190-7.

43. Bisdas S, Baghi M, Wagenblast J, et al. Tracer kinetics analysis of dynamic contrast-enhanced CT and MR data in patients with squamous cell carcinoma of the upper aerodigestive tract: comparison of the results. Clin Physiol Funct Imaging. 2009;29(5):339-46.

44. Razek AA, Elsorogy LG, Soliman NY, Nada N. Dynamic susceptibility contrast perfusion MR imaging in distinguishing malignant from benign head and neck tumors: a pilot study. Eur J Radiol. 2011;77(1):73-9.

45. Abdel Razek AA, Gaballa G. Role of perfusion magnetic resonance imaging in cervical lymphadenopathy. J Comput Assist Tomogr. 2011;35(1):21-5.

46. Shin DM, Voravud N, Ro JY, Lee JS, Hong WK, Hittelman WN. Sequential increases in proliferating cell nuclear antigen expression in head and neck tumorigenesis: a potential biomarker. J Natl Cancer Inst. 1993;85(12):971-8.

47. Ljumanovic R, Langendijk JA, Hoekstra OS, Leemans CR, Castelijns JA. Distant metastases in head and neck carcinoma: identification of prognostic groups with MR imaging. Eur J Radiol. 2006;60(1):58-66.

48. Peinado H, Cano A. A hypoxic twist in metastasis. Nat Cell Biol. 2008;10(3):253-4.

49. Brizel DM, Sibley GS, Prosnitz LR, Scher RL, Dewhirst MW. Tumor hypoxia adversely affects the prognosis of carcinoma of the head and neck. Int J Radiat Oncol Biol Phys. 1997;38(2):285-9.

50. Heye T, Davenport MS, Horvath JJ, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part I. Perfusion characteristics in the female pelvis by using multiple computer-aided diagnosis perfusion analysis solutions. Radiology. 2013;266(3):801-11.

51. Heye T, Merkle EM, Reiner CS, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology. 2013;266(3):812-21.

52. Robbins KT, Kumar P, Harris J, et al. Supradose intra-arterial cisplatin and concurrent radiation therapy for the treatment of stage IV head and neck squamous cell carcinoma is feasible and efficacious in a multi-institutional setting: results of Radiation Therapy Oncology Group Trial 9615. J Clin Oncol. 2005;23(7):1447-54.

53. de Groot JA, Bossuyt PM, Reitsma JB, et al. Verification problems in diagnostic accuracy studies: consequences and solutions. BMJ. 2011;343:d4770.

54. Bossuyt. Towards complete and accurate reporting of studies of diagnostic accuracy the STARD initiative. BMJ. 2003.

2

CHAPTER 2.2

Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value

Oral Oncology 2017;68:81-91

Daniel P Noij
Roland M Martens
J Tim Marcus
Remco de Bree
C René Leemans
Jonas A Castelijns
Marcus C de Jong
Pim de Graaf

ABSTRACT

Intravoxel incoherent motion (IVIM) imaging is increasingly applied in the assessment of head and neck cancer (HNC). Our purpose was to determine the diagnostic and prognostic performance of IVIM in HNC by performing a critical review of the literature. Pubmed and EMBASE were searched until May 2016. Study and patient characteristics, imaging protocol and diagnostic or prognostic outcomes were extracted by two independent reviewers. The studied IVIM parameters were diffusion coefficient (D), pseudodiffusion coefficient (D*), and perfusion fraction (f). We included 10 diagnostic studies, five prognostic studies and two studies assessing both. Studies were very heterogeneous in terms of applied b-values, imaging protocols, outcome measurements and reference standards; therefore, we did not perform a meta-analysis. The most commonly used sequence was "spin-echo planar imaging". A median of 10.5 b-values (range, 3-17) were used. All but three studies included at least four b-values below b=200 s/mm^2. By combining IVIM-parameters squamous cell carcinomas, lymphomas, malignant salivary gland tumors, Warthin's tumors and pleomorphic adenomas could be differentiated with a sensitivity of 85-87% and specificity of 80-100%. Low pre-treatment D or f and an increase in D during treatment were associated with a favorable response to treatment. D* appeared to be the parameter with the lowest prognostic value. Future research should focus on finding the optimal IVIM protocol, using uniformly accepted study methods and larger patient populations.

INTRODUCTION

Head and neck cancer (HNC) accounts for approximately 4% of the cancer case worldwide, making HNC the sixth most common cancer by incidence rate (1, 2). HNC mainly consists of tumors arising in the oral cavity, nasopharynx, oropharynx, hypopharynx, larynx and salivary glands.

Squamous cell carcinomas (SCC) account for over 90% of HNC (3). Alcohol and tobacco use are the most important risk factors (4). While early stage disease is usually treated by surgery or radiotherapy, advanced stage disease is generally treated by surgery and adjuvant radiotherapy with or without chemotherapy or combined chemotherapy and radiotherapy. Salvage surgery is then held in reserve for residual or recurrent disease (2, 5-7). While chemotherapy is mainly used in a concomitant setting with radiotherapy, in selected cases it can also be applied as neoadjuvant treatment (2). There is increasing evidence that in some geographic regions up to 80% of the oropharyngeal SCC is associated with the human papillomavirus (HPV), especially in relatively young patients who do not drink or smoke (8). Oropharyngeal SCC associated with HPV has a different tumor biology and is associated with a better prognosis than HPV-negative SCC (5-7). Therefore, it is proposed to de-escalate treatment in HPV-associated oropharyngeal SCC in patients who do not smoke.

Nasopharyngeal carcinoma (NPC) takes a unique place in epithelial HNC because of the very distinct geographical distribution ranging from 1:100.000 in Western Europe to >20:100.000 in parts of Southeast Asia (2, 9). Further it harbors an association with the Epstein-Barr virus (EBV) which is not seen in other HNC (9). Nasopharyngeal carcinoma has a different tumor biology as compared to other HNC.

Imaging is increasingly used for diagnosing and staging of HNC, monitoring the effect of treatment and in the detection of distant metastases and recurrent disease (10-12). In this systematic review we focus on the use of intravoxel incoherent motion (IVIM) magnetic resonance imaging (MRI) for diagnosis in HNC.

In general, water diffusion is restricted in malignant tissue. With diffusion-weighted imaging (DWI) this restricted diffusion can be imaged and quantified. The main advantage of DWI compared to other functional imaging techniques (e.g. dynamic contrast-enhanced MRI and positron emission tomography) is that it requires neither the administration of contrast medium or radioactive tracer nor the use of ionizing radiation.

One of the proposed methods to quantify diffusion is by considering diffusion as a mono-exponential phenomenon. In this way diffusion can be quantified in an apparent diffusion coefficient (ADC) (13). The word “apparent” implicates that in this way true diffusion is not measured. Especially at low b-values other parameters as blood volume and blood flow also contribute to the ADC (14, 15). The ADC concept provides a quantifiable measure with promising results in HNC, e.g. in discriminating metastatic from benign lymph nodes with an accuracy of >85% and in the detection of recurrent disease with an accuracy of >78% (16).

The signal decay after the diffusion-encoding gradients is not only caused by diffusion, but also by pseudorandom, or "incoherent", perfusion at the capillary level. To account for this, Le Bihan et al. introduced the bi-exponential IVIM model (13):

$$\frac{S_b}{S_0} = (1 - f) \cdot e^{(-bD)} + f \cdot e^{(-bD*)}$$

Where S_b represents the signal intensity with diffusion gradient b, and S_0 represents the signal intensity without diffusion gradients. D is known as pure or slow diffusion coefficient which is related to pure molecular diffusion. D* is the fast or pseudodiffusion coefficient that resembles the perfusion related incoherent microcirculation and is about a factor of 10 greater than D in biological tissue (13). Finally, f is the perfusion or (micro)vascular volume fraction which depends on capillary geometry and blood velocity (13). In this way pure tissue diffusion may be quantified and also perfusion characteristics may be assessed without the admission of contrast-material. Commonly D is first estimated using a linear fit using only high b-values (i.e., above 200 s/mm^2 (17)) and then f and D* are calculated using a non-linear least-squares algorithm.

With IVIM being increasingly used in HNC, a critical systematic review of the diagnostic and prognostic value of this technique is warranted. The purpose of this study was therefore to determine the diagnostic and prognostic performance of IVIM in HNC. Histopathology, other imaging modalities or clinical follow-up were used as reference standards.

METHODS AND MATERIALS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews and meta-analyses was used as a guidance (18).

Search strategy

This systematic search was conducted in Pubmed and Embase until May 2016 for original articles on the diagnostic and/or prognostic capability of IVIM in HNC. We did not apply language restrictions. We approached corresponding authors for additional data if necessary (e.g. to compute sensitivity and specificity). The only included search terms were "(IVIM OR ((intra-voxel OR intravoxel) AND incoherent AND motion))" in order to be as sensitive as possible. In the Pubmed search we used text words [tw] in the absence of MeSH-terms on this subject.

Two authors (DPN and RMM) independently selected relevant articles based on title and abstracts and discrepancies were resolved by consensus.

The inclusion criteria were: 1) The study population consisted of at least 10 patients with malignant lesions in the head and neck area; 2) The study assessed diagnosing malignancy, response prediction to therapy, detection of residual/ recurrent disease. Or data of these subjects could be extracted from the article; 3) Histopathology, clinical follow-up or another imaging modality was used as reference standard test.

Exclusion criteria were: 1) The publication was a review, meta-analysis, only published as abstract or if it was another non-primary publication (e.g. editorial, technical note); 2) The study reported on (potentially) overlapping study populations.

Data extraction

Data on the study and patient characteristics, the imaging protocol and diagnostic outcomes were extracted by two independent reviewers (DPN and RMM) and discrepancies were resolved by consensus. If available, source data (i.e. true positive (TP), false positive (FP), true negative (TN), and false negative (FN)) were extracted or recalculated. If unavailable, the corresponding author of the article was contacted to provide additional data.

Quality assessment

Two authors (DPN and RMM) independently assessed all studies for study quality and discrepancies were resolved by consensus. All included studies were assessed for quality by using the QUality Assessment of studies of Diagnostic Accuracy included in Systematic reviews (QUADAS-2) checklist (19). The quality of prognostic studies was also assessed with the QUality In Prognostic Studies (QUIPS) checklist (20, 21).

Statistical analysis

Diagnostic accuracy data is presented with 95% confidence intervals (95%CI) if presented by the authors, or when we were able to reconstruct a 2x2 table. Receiver operating characteristic (ROC) analysis was performed if per-patient data could be extracted using SPSS Statistics (version 20.0; Chicago, IL, USA). The Youden Index (YI) was used to determine the optimal cut-off. *P*-values were reported as NS (not statistically significant, i.e., $P \geq 0.05$), ≤ 0.05, ≤ 0.01, and ≤ 0.001.

RESULTS

The search in Pubmed and Embase retrieved 429 unique studies. After excluding 383 studies on title or abstract we reviewed the full text of 46 studies. Finally, 17 studies were included (10 diagnostic, 5 prognostic and 2 both) for qualitative analysis (22-38) (Figure 1). Due to heterogeneity in applied b-values, imaging protocols, outcome measurements and reference standards we decided not to perform any quantitative meta-analysis.

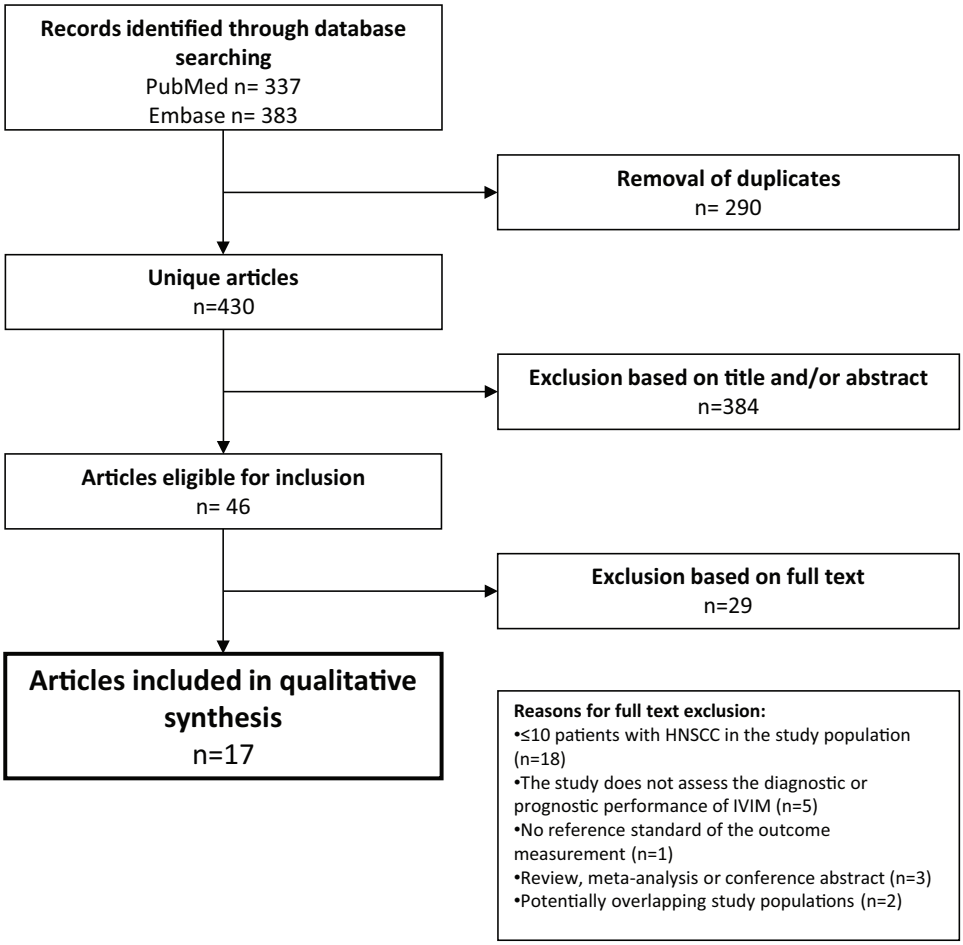


Figure 1 *Flow chart of study inclusion*

Study characteristics

In total the selected studies included 882 patients (22-38). The most common included head and neck malignancies were nasopharyngeal carcinoma (n=417) and squamous cell carcinoma at other sites (n=220). Patient characteristics are mentioned in Table 1. Definitions of included parameters are provided in Appendix A. In 414 patients with NPC, mean values of D ($0.732 \cdot 10^{-3}$ mm^2/s), D* ($84.34 \cdot 10^{-3}$ mm^2/s) and f (18.5%) could be extracted from the articles.

In SCC patients mean data were available for D (n=160, mean=$0.863 \cdot 10^{-3}$ mm^2/s), D* (n=82, mean=$32.51 \cdot 10^{-3}$ mm^2/s) and f (n=130, mean=22.3%). These differences where significant (P<0.001 for D and D* and P=0.006 for f).

Imaging was performed at 1.5 T in nine studies) (22, 29-34, 36, 37) or 3 T in eight (23, 25-28, 36, 38) (24). Diffusion-weighted imaging was commonly performed with single-shot spin-echo echo planar imaging (SS-SE-EPI) (n=13) (23-30, 32-38). In four of these studies the use of fat saturation is mentioned with CHEmical Shift Selective (CHESS) (31), chemical shift-based fat suppression (23) or Spectral Presaturation with Inversion Recovery (SPIR) (27, 28). One study used short tau inversion recovery EPI (STIR-EPI) (22) and another study used Half-fourier acquisition single-shot turbo spin-echo (HASTE) (31). Diffusion-weighted imaging was acquired with a median of 10.5 b-values (range, 3-17). The used b-values ranges were 0-800 s/mm^2 (n=7) (23, 25, 26, 30, 32, 34, 38), 0-850 s/mm^2 (n=1) (29) or 0-1000 s/mm^2 (n=9) (22, 24, 27, 28, 31, 33, 35-37). Inclusion of low b-values (i.e. b<200 s/mm^2 (17)) is important for bi-exponential fitting of the signal intensity over b. One study did only include b=0 s/mm^2 as b-value below b=200 s/mm^2 (33). Another study did include three b-values below b=200 s/mm^2 (22). The other studies included a median of seven b-values below b=200 s/mm^2 (range, 4-11) (23, 25-32, 34-36, 38). The Levenberg-Marquardt algorithm was most commonly used for signal fitting (n=11) (25-28, 31, 34-36, 38). In 12 studies it was mentioned that a two-step-fit was used with an estimation of D in the first step with D* and f being determined in the second step (23, 24, 26, 30-38). Median scan time was 4:51 minutes (range, 01:30-12:00 minutes). An overview of imaging characteristics is mentioned in Table 2.

Table 1 *Patient characteristics*

Author, year	Study design	Included patients	Mean age (SD, range)	% Male	Tumor type	AJCC stage	T stage	N stage	M stage	Treatment	Reference standard
Dikaios (22), 2014	...	40	58 (8, 43-79)	...	HNSCC	...	...	Unilateral	...	...	HP
Ding (23), 2015	PS	31	57[a] (44-78)	94	SCC: Tonsil (n=16), Base of tongue (n=15)	III (n=4), IV (n=27)	...	N0 (n=1), N1 (n=3), N2 (n=17)	...	CRT	RC
Hauser (25), 2013	RS	22	55 (9, 34-69)	73	SCC: Oropharynx (n=15), Hypopharynx (n=4), Larynx (n=2), NPC (n=1)	IV	...	N0 (n=1), N1 (n=1), N2 (n=18), N3 (n=2)	...	CRT	RC + CE
Hauser (26), 2014	RS	14	55 (8, 43-69)	60	SCC: Oropharynx (n=9), Hypopharynx (n=3), Larynx (n=3)	...	...	N+	...	CRT	RC + CE
Guo (24), 2016	RS	28	55 (18, 44-72)	100	Hypopharynx	III and IV	...	N2 and N3	M0	IC	RC + HP
Lai (27), 2013	PS	83	52 (12-90)	72	NPC, posttreatment fibrosis	...	T1 (n=17), T2 (n=7), T3 (n=22), T4 (n=7)	...	...	CRT	HP
Lai (28), 2014	PS	80	51 (14, 12-90)	73	NPC	I (n=10), II (n=11), III (n=44), IV (n=15)	T1 (n=33), T2 (n=7), T3 (n=31), T4 (n=9)	N0 (n=15), N1 (n=17), N3 (n=5)	M0 (n=76), M1 (n=4)	...	HP
Lu (29), 2013	RS	16	55 (38-64)	94	SCC: Oropharynx (n=11), Oral cavity (n=4), NPC (n=1)	III (n=1), IV (n=15)	...	N+	...	Surgery (n=2), CRT (n=14)	HP
Marzi (30), 2013	RS	37	57 (30-76)	84	SCC: Oropharynx (n=15), NPC (n=11), Hypopharynx/Larynx (n=8), Oral Cavity (n=2), Maxillary sinus (n=1),	...	T1 (n=3), T2 (n=13), T3 (n=8), T4 (n=13)	N0 (n=3), N1 (n=3), N2 (n=28), N3 (n=3)	...	...	HP
Sakamoto (31), 2014	PS	33	59 (14-89)	48	Basaloid SCC (n=1), Carcinoma ex pleomorphic adenoma (n=1), adenoid cystic carcinoma (n=1), SCC (n=18), Verrucous carcinoma (n=2)	...	...	...	...	...	HP
Sasaki (32), 2014	RS	94	62 (15, 3-91)	60	Various benign and malignant head and neck tumors	...	...	...	...	...	HP
Sumi (33), 2012	RS	113	60 (21-91)	55	Various benign and malignant head and neck tumors	...	...	...	...	...	HP
Sumi (34), 2012	PS	31	61 (21-82)	52	SG: Benign (n=20), Malignant (n=11)	...	...	...	...	Surgery	HP
Xiao (35), 2015	PS	48	42[a] (13-65)	69	NPC	III (n=19), IV (n=29)	T1 (n=5), T2 (n=7), T3 (n=14), T4 (n=22)	N0 (n=3), N1 (n=9), N2 (n=25), N3 (n=11)	M0 (n=45), M1 (n=3)	NAC	RC
Xiao-Ping (36), 2015	PS	50	49 (11)	78	NPC	III (n=12), IV (n=38)	T1 (n=1), T2 (n=10), T3 (n=19), T4 (n=20)	N0 (n=4), N1 (n=9), N2 (n=17), N3 (n=20)	M0 (n=31), M1 (n=19)	NAC + (C)RT	RC
Yu (37), 2016	RS	102	...	...	NPC (n=80), Lymphoma (n=22)	...	T1 (n=8), T2 (n=29), T3 (n=19), T4 (n=26)	...	...	...	HP
Zhang (38), 2014	PS	60	51 (16-69)	74	NPC	...	...	...	...	...	HP

[a] Median age

Abbreviations: AJCC = American Joint Committee on Cancer; CE = clinical evaluation; CRT = chemoradiotherapy; HNSCC = head and neck squamous cell carcinoma; HP = histopathology; NAC = neoadjuvant chemotherapy; NPC = nasopharyngeal carcinoma; RC = RECIST; SCC = squamous cell carcinoma; SG = salivary gland tumor

Quality assessment
Results of QUADAS-2 and QUIPS are mentioned in Table 3 and 4.

The QUADAS-2 yielded the following findings. There was a high risk of a biased patient selection in six studies (22, 24, 26, 27, 29, 32). In two studies this was due to a case-control design (22, 27). In six studies there were inappropriate exclusions (22, 24, 26, 27, 29, 32). For example by classifying lymph node as benign or malignant based only on imaging criteria, which may result in missing small lymph node metastases (26); by only including patients with lymph node metastases when the primary tumor is assessed separately, while patients with an N0 neck could have been included as well (24, 29); or by only including patients of whom the tumor was excised, while also including biopsy-proven malignancies would have led to a larger patient population (32). In none of the 10 studies where a threshold was used this was pre-specified, which may result in an overestimation of diagnostic value (23, 24, 28, 29, 31, 34-38). In two studies the observers were not blinded to the reference standard (29, 30) and in 11 studies this was unclear (22, 24-28, 32-34, 37, 38). In five studies imaging was used as reference standard instead of histopathology (22, 23, 25, 35, 36). The interval between index test and reference standard was not mentioned in 10 studies (22, 27-34, 36).

For the prognostic studies the assessment with QUIPS yielded the following results. In four studies reasons for loss to follow-up are mentioned without describing participants who were lost to follow-up (23, 25, 26, 36). Xiao et al. excluded patients with insufficient follow-up images (35). None of the included studies attempted to correct for possible confounders (23-26, 29, 35, 36). Lu et al. used both parametric and non-parametric statistical tests on the same data which raised concerns on the adequacy of the statistical analysis (29). In three studies we were unable to reconstruct a two-by-two table based on the given sensitivity and specificity (24, 35, 36).

Diagnostic study results
Study results are mentioned in Table 5 and Appendix B. In 9 studies the primary tumor was assessed separately (24, 27, 28, 30-32, 34, 37, 38), in two studies lymph nodes were assessed separately (22, 28) and in four studies a combination of primary tumor and lymph nodes was assessed (28, 29, 32, 33).

Sasaki et al. (32) provided an appendix with per-patient IVIM data. The authors used both the aforementioned IVIM formula and an IVIM analysis based on a geometric method using only 3 b-values:

$$GeoD = \frac{(\ln S_{200} - \ln S_{800})}{600}$$

$$Geof = \frac{1 - S_{inter}}{S_0}$$

$$GeoP = \frac{(\ln S_0 - \ln S_{inter})}{200}$$

S_{inter} is the interception of the logarithmic regression line obtained using b-values of 200 and 800 s/mm^2 with the y-axis (32). Because the GeoP is fundamentally different than D*

Table 2 *Imaging characteristics*

Author, year	Field Strength (T)	Diffusion sequence	TR/TE (ms)	FOV (mm)	Matrix	Section thickness (mm)	b-values	IVIM fit	IVIM parameters	Scan time
Dikaios (22), 2014	1.5	STIR-EPI	8700/88	200	128x128	4	0, 50, 100, 300, 600, 1000	MP, NR	D, D*, f	06:10
Ding (23), 2015	3	SS-SE-EPI	3600/100	256	128x128	3.5	0, 20, 40, 60,80, 100, 120, 150, 200, 400, 600, 800	BeF, SLF, NLLS	D, D*, f	…
Guo (24), 2016	3	SS-SE-EPI	2500/79	230	256x256	5	0, 10, 20, 30, 50, 70, 100, 150, 200, 400, 800, 1000	BeF using LMa	D, D*, f	05:08
Hauser (25), 2013	3	SS-SE-EPI	1300/50	240	80x80	3	0, 50, 100, 150, 200, 250, 700, 800	BeF using LMa	D, f	04:51
Hauser (26), 2014	3	SS-SE-EPI	1300/50	240	80x80	3	0, 50, 100, 150, 200, 250, 700, 800	BeF using LMa	D, f	04:51
Lai (27), 2013	3	SS-SE-EPI	7996/43	230	256x256	3	0, 10, 20, 30, 40, 60, 100, 120, 160, 200, 300, 500, 1000	BeF using LMa	D, D*, f	12:00
Lai (28), 2014	3	SS-SE-EPI	7996/43	230	256x256	3	0, 10, 20, 30, 40, 60, 100, 120, 160, 200, 300, 500, 1000	BeF using LMa	D, D*, f	12:00
Lu (29), 2013	1.5	SS-SE-EPI	4000/90	200-220	128x128	6-8	0, 13, 17, 23, 30, 40, 53, 70, 92, 122, 161, 212, 280, 369, 488, 644, 850	BeF using NLLS	D, D*, f	04:00
Marzi (30), 2013	1.5	SS-SE-EPI	4500/77	260-280	128x128	4	0, 25, 50, 75, 100, 150, 300, 500, 800	BeF	D, D*, f, f_{asym}	06:13
Sakamoto (31), 2014	1.5	HASTE-DWI	3000/101	230x173	192x144	4-5	0, 20, 40, 60, 80, 100, 150, 200, 500, 1000	BeF using LMa	D, D*, f	01:30
Sasaki (32), 2014	1.5	SS-SE-EPI	1625/81	200	112x90	4	0, 10, 20, 30, 50, 80, 100, 200, 300, 400, 800	Geo, LS	D, D*, f, P	01:53
Sumi (33), 2012	1.5	SS-SE-EPI	4283/87	200	112x90	4	0, 500, 1000	SIF	D, PP	02:08
Sumi (34), 2012	1.5	SS-SE-EPI	1625/81	200	112x90	4	0, 10, 20, 30, 50, 80, 100, 200, 300, 400, 800	BeF using LMa	D, D*, f	01:53
Xiao (35), 2015	3	SS-SE-EPI	4495/69	230	256x256	5	0, 10, 20, 30, 40, 50, 100, 150, 200, 350, 500, 650, 800, 1000	BeF using LMa	D, D*, f	06:00
Xiao-Ping (36), 2015	1.5	SS-SE-EPI	4225/106	220	128x130	5	0, 50, 80, 100, 150, 200, 400, 600, 800, 1000	BeF using LMa	D, D*, f	02:53
Yu (37), 2016	1.5	SS-SE-EPI	4225/106	220	128x130	5	0, 50, 80, 100, 150, 200, 400, 600, 800, 1000	BeF using LMa	D, D*, f	…
Zhang (38), 2014	3	SS-SE-EPI	3000/58	240	128x128	4	0, 10, 20, 30, 50, 80, 100, 150, 200, 300, 400, 600, 800	BeF using LMa	D, D*, f	03:45

Abbreviations: BeF = bi-exponential fit; D* = pseudodiffusion coefficient; D = diffusion coefficient; f = perfusion factor; f_{asym} = perfusion factor estimated using an asymptotic method; Geo = geometric method; HASTE-DWI = half-Fourier acquisition single-shot turbo spin-echo diffusion-weighted imaging; LMa = Levenberg Marquard algorithm; LS = least squares method; MD = median model; MP = maximum probability model; NLLS = non-linear least-squares; NR = nonlinear regression model; P = perfusion parameter that is heavily weighted towards extravascular space; PS = prospective; PP = perfusion-related parameter; RS = retrospective; SS-SE-EPI = single-shot spin-echo echo planar imaging; SIF = simplified IVIM formula; STIR-EPI = short tau inversion recovery echo planar imaging; SLF = simplified linear fit

Table 3 *Results of QUADAS-2 for bias assessment of all included studies: '✓' indicates a low risk of bias; '?' an unclear risk and 'x' indicates a high risk of bias*

	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Dikaios (22), 2014	x	?	?	?	?	✓	✓
Ding (23), 2015	?	x	?	✓	✓	✓	✓
Guo (24), 2016	x	x	✓	✓	✓	✓	✓
Hauser (25), 2013	✓	?	?	x	✓	✓	✓
Hauser (26), 2014	x	?	✓	x	✓	✓	✓
Lai (27), 2013	x	?	✓	✓	x	✓	✓
Lai (28), 2014	✓	x	✓	?	✓	✓	✓
Lu (29), 2013	x	x	✓	✓	✓	✓	✓
Marzi (30), 2013	?	x	✓	x	✓	✓	✓
Sakamoto (31), 2014	?	x	✓	?	✓	✓	✓
Sasaki (32), 2014	x	?	✓	✓	✓	✓	✓
Sumi (33), 2012	?	?	✓	✓	✓	✓	✓
Sumi (34), 2012	✓	x	✓	✓	✓	✓	✓
Xiao (35), 2015	✓	x	?	✓	✓	✓	✓
Xiao-Ping (36), 2015	✓	x	?	✓	✓	✓	✓
Yu (37), 2016	?	x	✓	✓	✓	✓	✓
Zhang (38), 2014	✓	x	✓	✓	✓	✓	✓

the authors used another symbol; GeoP is more weighted to the extravascular space compared to D* which is more weighted towards the vascular space. Based on the data provided in the appendix we used ROC analysis to determine the diagnostic accuracy of IVIM (with both methods) to discriminate SCC from lymphoma as these were the most prevalent malignancies in this study (32). Further we compared IVIM values between primary SCC and SCC lymph node metastases from Sasaki et al. (32) to validate the results of Lu et al. (29) of which the results will be discussed below.

The abilities of IVIM to differentiate between various benign and malignant head and neck lesions were determined by five studies (31-34, 37). Sakamoto et al. (31) found D to be the single most valuable parameter with a sensitivity of 87% (95%CI=66-97%) and specificity of 80% (95%CI=44-97%). When combining D and D* sensitivity and specificity increased to 91% (95%CI=72-99%) and 90% (95%CI=56-100%) respectively. Included benign lesions were mainly pleomorphic adenomas (4/10) and vascular malformations (4/10) and malignant lesions were predominantly SCCs (18/23). Yu et al. focused on differentiating between NPC and lymphoma (37). All IVIM values (D, D* and f) were significantly lower in lymphoma. The highest diagnostic accuracy was achieved when D* and f were combined with sensitivity and specificity being 85% (95%CI=76-92%) and 100% (95%CI, 83-100%), respectively. These results are in line with Sumi et al. (33) who reported that IVIM values (D and the perfusion-related parameter (PP)) differed significantly ($P \leq 0.001$) between six types of head and neck lesions (lymphoma, SCC, malignant salivary gland (SG) tumor, Warthin tumor, pleomorphic adenoma and schwannoma). Lymphomas had significantly lower D and PP values than SCC ($P \leq 0.001$). Malignant SG tumor D was significantly lower than D of pleomorphic adenoma and significantly higher than D of

Table 4 *QUIPS results for bias assessment of prognostic studies: '✓' indicates a low risk of bias; '?' a moderate risk and 'x' indicates a high risk of bias*

	1. Study participation	2. Study attrition	3. Prognostic factor measurement	4. Outcome measurement	5. Study confounding	6. Statistical analysis and reporting
Ding (23), 2015	✓	?	?	?	x	✓
Guo (24), 2016	x	✓	?	?	x	x
Hauser (25), 2013	✓	x	?	x	x	?
Hauser (26), 2014	x	?	x	?	x	?
Lu (29), 2013	✓	x	?	x	x	x
Xiao (35), 2015	✓	?	?	?	x	x
Xiao-Ping (36), 2015	✓	?	?	?	x	x

Table 5 *Diagnostic accuracy*

	n	Parameter	Outcome measurement	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)
Primary tumor							
Lai (27), 2013	83	D	NPC PT vs posttreatment fibrosis	$1.06 \cdot 10^{-3}$ mm^2/s	100 (93-100)	100 (88-100)	1.00 (1.00-1.00)
Lai (27), 2013	83	D*	NPC PT vs posttreatment fibrosis	$85.3 \cdot 10^{-3}$ mm^2/s	100 (93-100)	91 (76-98)	0.99 (0.97-1.00)
Lai (27), 2013	83	f	NPC PT vs posttreatment fibrosis	0.13	66 (52-79)	100 (82-100)	0.89 (0.82-0.96)
Lai (28), 2014	80	D	NPC T-stage: low vs high	$0.76 \cdot 10^{-3}$ mm^2/s	83 (68-93)	53 (36-69)	0.65
Lai (28), 2014	80	f	NPC T-stage: low vs high	0.13	81 (65-91)	93 (80-98)	0.90 (0.82-0.96)
Lai (28), 2014	80	D*	NPC T-stage: low vs high	$101.0 \cdot 10^{-3}$ mm^2/s	73 (57-86)	83 (67-93)	0.83
Sakamoto (31), 2014	33	D+D*	Malignant vs benign	…	91 (72-99)	90 (56-100)	0.96 (0.90-1.00)
Sakamoto (31), 2014	33	D	Malignant vs benign	$0.98 \cdot 10^{-3}$ mm/s^2	87 (66-97)	80 (44-97)	0.91 (0.81-1.00)
Sakamoto (31), 2014	33	f	Malignant vs benign	0.19	74 (52-90)	50 (19-81)	0.52 (0.29-0.74)
Sakamoto (31), 2014	33	D*	Malignant vs benign	$8.42 \cdot 10^{-3}$ mm^2/s	61 (39-80)	90 (56-100)	0.75 (0.58-0.97)
Sasaki (32), 2014	35	Fit D	SCC vs lymphoma	$0.84 \cdot 10^{-3}$ mm^2/s	74 (56-87)	100 (74-100)	0.87 (0.77-0.97)
Sasaki (32), 2014	35	Geo D	SCC vs lymphoma	$0.87 \cdot 10^{-3}$ mm^2/s	74 (56-87)	100 (74-100)	0.87 (0.76-0.97)
Sasaki (32), 2014	35	Fit f	SCC vs lymphoma	0.11	50 (32-68)	83 (52-98)	0.61 (0.45-0.78)
Sasaki (32), 2014	35	Geo P	SCC vs lymphoma	$0.57 \cdot 10^{-3}$ mm^2/s	50 (32-68)	83 (52-98)	0.61 (0.44-0.78)
Sasaki (32), 2014	35	Geo f	SCC vs lymphoma	0.11	50 (32-68)	83 (52-98)	0.61 (0.44-0.78)
Sasaki (32), 2014	23	Geo D + Geo P	SG malignant vs benign	…	100 (54-100)	94 (71-100)	…
Sasaki (32), 2014	23	Geo D + Geo f	SG malignant vs benign	…	100 (54-100)	94 (71-100)	…
Sasaki (32), 2014	23	Fit D + Fit D*	SG malignant vs benign	…	100 (54-100)	94 (71-100)	…
Sasaki (32), 2014	23	Fit D + Fit f	SG malignant vs benign	…	50 (12-88)	100 (80-100)	…
Sasaki (32), 2014	23	Geo D_{4b} + Geo P_{4b}	SG malignant vs benign	…	83 (36-100)	88 (64-99)	…
Sasaki (32), 2014	23	Geo D_{4b} + Geo f_{4b}	SG malignant vs benign	…	83 (36-100)	88 (64-99)	…
Sumi (34), 2012	31	D*	SG malignant vs benign	$10<D^*<23 \cdot 10^{-3}$ mm^2/s	73 (39-94)	65 (41-85)	…
Sumi (34), 2012	31	D	SG malignant vs benign	$0.8<D<1.1 \cdot 10^{-3}$ mm^2/s	64 (31-89)	100 (83-100)	…
Sumi (34), 2012	31	D + D*	SG malignant vs benign	…	100 (72-100)	100 (83-100)	…

Table 5 *continued*

	n	Parameter	Outcome measurement	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)
Sumi (34), 2012	31	D + f	SG malignant vs benign	…	82 (48-98)	100 (83-100)	…
Sumi (34), 2012	31	D* + f	SG malignant vs benign	…	82 (48-98)	65 (41-85)	…
Yu (37), 2016	102	D	NPC vs lymphoma	$0.66 \cdot 10^{-3}$ mm^2/s	55 (44-66)	100 (83-100)	0.80 (0.71-0.87)
Yu (37), 2016	102	D*	NPC vs lymphoma	$7.89 \cdot 10^{-3}$ mm^2/s	83 (73-90)	85 (62-97)	0.90
Yu (37), 2016	102	f	NPC vs lymphoma	0.29	41 (31-53)	95 (75-100)	0.66 (0.54-0.74)
Yu (37), 2016	102	fD*	NPC vs lymphoma	$1.99 \cdot 10^{-3}$ mm^2/s	85 (76-92)	100 (83-100)	0.96 (0.90-0.99)
Zhang (38), 2014	60	D	PT vs adenoid hypertrophy	$0.75 \cdot 10^{-3}$ mm^2/s	83	65	0.85
Lymph node							
Dikaios (22), 2014	40	mp-DWI	SCC NR benign vs metastatic	…	80	94	0.95 (0.80-1.00)
Dikaios (22), 2014	40	mp-DWI	SCC NR benign vs metastatic	…	80	88	0.92 (0.85-1.00)
Lai (28), 2014	80	D	NPC N stage: low vs high	$0.76 \cdot 10^{-3}$ mm/s^2	88 (75-95)	66 (48-82)	0.86 (0.76-0.93)
Lai (28), 2014	80	f	NPC N stage: low vs high	0.15	85 (72-94)	61 (42-77)	0.69
Lai (28), 2014	80	D*	NPC N stage: low vs high	$103.9 \cdot 10^{-3}$ mm^2/s	71 (56-83)	55 (36-72)	0.63
Primary tumor + lymph node							
Lai (28), 2014	80	D	NPC low vs high AJCC stage	$0.78 \cdot 10^{-3}$ mm/s^2	93 (84-98)	76 (53-92)	0.91 (0.83-0.97)
Lai (28), 2014	80	f	NPC low vs high AJCC stage	0.15	88 (77-95)	86 (64-97)	0.87
Lai (28), 2014	80	D*	NPC low vs high AJCC stage	$100.4 \cdot 10^{-3}$ mm^2/s	48 (35-62)	90 (70-99)	0.72
Lu (29), 2013	16	f+D	SCC PT vs LN	…	63 (35-85)	81 (54-96)	0.76
Lu (29), 2013	16	f	SCC PT vs LN	0.22	63 (35-85)	75 (48-93)	0.71
Lu (29), 2013	16	D*	SCC PT vs LN	$43.2 \cdot 10^{-3}$ mm^2/s	63 (35-85)	63 (35-85)	0.53
Lu (29), 2013	16	D	SCC PT vs LN	$0.80 \cdot 10^{-3}$ mm^2/s	56 (30-80)	94 (70-100)	0.74
Sasaki (32), 2014	34	Fit f	SCC PT vs LN	0.06	100 (85-100)	27 (6-61)	0.62 (0.41-0.84)
Sasaki (32), 2014	34	Geo f	SCC PT vs LN	0.05	96 (78-100)	27 (6-61)	0.53 (0.31-0.75)
Sasaki (32), 2014	34	Geo P	SCC PT vs LN	$0.26 \cdot 10^{-3}$ mm^2/s	96 (78-100)	27 (6-61)	0.53 (0.31-0.75)

Abbreviations: AJCC = American Joint Committee on Cancer; AUC = area under the curve; D* = pseudodiffusion coefficient; D = diffusion coefficient; D_{4b} = diffusion coefficient acquired with four b-values; f = perfusion factor; f_{4b}= perfusion factor acquired with four b-values; Fit = least-squares method; Geo = geometric method; LN = lymph node; mp-DWI = multiparametric analysis of diffusion-weighted imaging parameters; NPC = nasopharyngeal carcinoma; NR = nonlinear regression; P = perfusion parameter that is heavily weighted towards extravascular space; P_{4b} = perfusion parameter that is heavily weighted towards extravascular space acquired with four b-values; PT = primary tumor; SCC = squamous cell carcinoma; SG = salivary gland tumors; 95%CI = 95% confidence interval

Warthin tumors ($P \leq 0.01$). These results were verified in two other studies (32, 34). In all three studies malignant SG tumors had intermediate D and D* values compared to Warthin tumors (lower values) and pleomorphic adenoma (higher values). The combined use of D and D* could separate malignant from benign SG tumors with a sensitivity of 100% (95%CI=54-100%) and a specificity of 94-100% (95%CI=71-100%) (32, 34).

Marzi et al. (30) compared pretreatment IVIM values of different SCC locations: nasopharynx (n=11), oropharynx (n=15) and hypopharynx/larynx (n=8). Both D ($P \leq 0.01$) and f ($P \leq 0.05$) were significantly different between groups. D was highest in hypopharyngeal/laryngeal SCC ($D_{median}=1.07 \cdot 10^{-3}$ mm^2/s) and lowest in nasopharyngeal SCC ($D_{median}=0.83 \cdot 10^{-3}$ mm^2/s). For f, oropharyngeal SCC had the highest values ($f_{median}=22.5\%$ vs <18.6%) and also the largest range in values ($f_{range}=12.6\text{-}32.7\%$). Nasopharyngeal SCCs were the most homogeneous group ($f_{range}=7.3\text{-}16.7\%$).

Two studies assessed the difference of IVIM parameters between primary SCC and lymph nodes. Lu et al. (29) showed that lymph nodes have significantly higher D values ($P \leq 0.001$) and lower f values ($P \leq 0.001$). These results could not be confirmed by Sasaki et al. (32) who did not find any significant differences between primary SCC and SCC lymph nodes. It should be noted that in the first study included patients all had a primary tumor and a lymph node metastasis; whereas in the second study it is not specified whether primary tumors and lymph node metastasis of the same patients were assessed.

The differentiation between malignant tumor and post-chemoradiation fibrosis was assessed by Lai et al. (27) in a case-control study including pre-treatment NPC and biopsy-confirmed post-chemoradiation fibrosis. With D it was possible to separate both with a sensitivity (95%CI=93-100%) and specificity (95%CI=88-100%) of 100%.

In another study by Lai et al. (28) NPC at different disease stages were compared. For T stage, N stage and disease stage according to the American Joint Committee on Cancer (AJCC), all IVIM parameters (i.e., D, D* and f) were significantly lower in the high stage group than in the low stage group. For N staging and AJCC staging, D had the highest diagnostic accuracy (AUC=0.86 and 0.91, respectively), while f had the highest AUC in T staging (AUC=0.90). In multivariate analysis there was a non-significant trend towards a higher AUC for the combined use of all IVIM parameters.

Dikaios et al. (22) used a (conventional) non-linear (NR) and a maximum probability (MP) model to estimate IVIM-parameters for differentiating between benign and malignant lymph nodes. No significant differences between models were found. Both f and D* were significantly different ($P \leq 0.01$) between benign and malignant nodes while D was not.

Prognostic study results

Prognostic study results are mentioned in Table 6 and Appendix C. In five studies the primary tumor was assessed (23-25, 29, 35, 36) separately, in three studies lymph nodes were assessed separately (23, 26, 35) and in one both were assessed simultaneously (23).

The value of IVIM in predicting response to neo-adjuvant chemotherapy (NAC) in NPC was assessed by Xiao et al. (35) and Xiao-Ping et al. (36). In both studies pre-treatment D could significantly predict the response to NAC (P≤0.01) with a sensitivity and specificity of 64-65% and 72-81%, respectively. In both studies f was a weaker predictor than D, with only Δf being a significant predictor in the study of Xiao et al (primary tumor: *P*≤0.05; lymph node: *P* ≤0.01) (35).

Table 6 *Prognostic accuracy*

	n	Follow-up	Tumor type	Parameter	Outcome measurement	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)
Primary tumor									
Guo (24), 2016	28	3 weeks	Hypopharynx SCC	Pre-treatment D	Response to NAC	$0.85{\cdot}10^{-3}$ mm^2/s	75	89	0.81
Hauser (25), 2013	22	≥7.5 months	SCC/NPC	Pre-treatment D	Locoregional treatment outcome	...	83 (36-100)	81 (54-96)	0.79 (0.58-1.00)
Hauser (25), 2013	22	≥7.5 months	SCC/NPC	Pre-treatment f	Locoregional treatment outcome	...	67 (22-96)	94 (70-100)	0.82 (0.62-1.00)
Xiao (35), 2015	48	...	NPC	Pre-treatment D	Response to NAC	$0.91{\cdot}10^{-3}$ mm^2/s	64	81	0.71
Xiao-Ping (36), 2015	50	...	NPC	$\Delta D_{pre\text{-}post\ NAC}$	Effect of NAC[b]	26.3%	94	77	0.86 (0.72-1.00)
Xiao-Ping (36), 2015	50	...	NPC	Pre-treatment D	Residue after CRT[a]	$0.73{\cdot}10^{-3}$ mm^2/s	82	83	0.84 (0.70-0.98)
Xiao-Ping (36), 2015	50	...	NPC	Pre-treatment D	Effect of NAC[b]	$0.95{\cdot}10^{-3}$ mm^2/s	65	72	0.77 (0.61-0.92)
Xiao-Ping (36), 2015	50	...	NPC	$\Delta D_{pre\text{-}post\ NAC}$	Residue after CRT[a]	25.0%	79	83	0.76 (0.59-0.92)
Lymph node									
Hauser (26), 2014	14	≥13.5 months	SCC	Pre-treatment D	Locoregional treatment outcome	...	100 (29-100)	55 (23-83)	0.71 (0.40-1.00)
Hauser (26), 2014	14	≥13.5 months	SCC	Pre-treatment f	Locoregional treatment outcome	...	100 (29-100)	100 (72-100)	1.00 (1.00-1.00)
Xiao (35), 2015	48	...	NPC	Pre-treatment D	Response to NAC	$0.95{\cdot}10^{-3}$ mm^2/s	55	95	0.77

[a] The authors did not specify the interval between imaging and the assessment of residual disease after treatment
[b] The authors considered treatment effective if patients exhibited a complete response (CR) or partial response (PR) based on the RECIST criteria

Abbreviations: AUC = area under the curve; CRT = chemoradiotherapy; D = diffusion coefficient; f = perfusion factor; NAC = neoadjuvant chemotherapy; NPC = nasopharyngeal carcinoma; SCC = squamous cell carcinoma; 95%CI = 95% confidence interval

In the study of Xiao-Ping et al. sensitivity increased to 94% and specificity was 77% when the difference in D before and after NAC was used to assess the effect of NAC (36). It should be noted that pre-NAC D had a relatively high diagnostic accuracy in predicting the presence of residual disease after chemoradiotherapy with a sensitivity of 82% and a specificity of 83%. Ding et al. performed IVIM before and during chemoradiotherapy in HPV-positive primary HNSCC and found pre-treatment D to differ significantly between responders and non-responders (23). Guo et al. found similar results in predicting treatment response in NAC for hypopharyngeal SCC with D being the strongest predictor with a sensitivity of 75% and specificity being 89% (24).

These findings are in contrast to Hauser et al. who performed one study on primary tumors (25) and one on lymph nodes (26). The patient population consisted of HNSCC patients receiving chemoradiotherapy. In both studies, pretreatment f discriminated best between responders and non-responders ($P \leq 0.01$), whereas D was not statically significant.

To conclude it should be noted that in only one of the prognostic studies D* differed statistically significantly between responders and non-responders (24).

DISCUSSION

In studies on patients with SG tumors, all IVIM values of malignant SG tumors were between those of Warthin tumors and pleomorphic adenomas (32-34). This demonstrates that diagnostic accuracy of IVIM in distinguishing between malignant and benign head and neck lesions depends strongly on the included tumor types. With IVIM it was possible to reliably differentiate between SCCs, lymphomas, malignant SG tumors, Warthin tumors and pleomorphic adenomas (31, 33, 34, 37). It should be noted that combining IVIM parameters often yielded a higher diagnostic accuracy than using a single IVIM parameter (22, 28, 29, 31, 32, 34, 37). Future research should focus on finding the optimal combination of functional imaging parameters, for example by combining IVIM with dynamic contrast-enhanced MRI (39). Preferably only lesions should be included which are currently challenging to separate.

Detection of residual disease in irradiated tissue remains challenging (40). A recent randomized trial did reveal that ^{18}F-FDG-PET-CT can reduce the need for investigations under anesthesia after RT for laryngeal carcinoma without compromising treatment quality (41). Especially early after (chemo)radiotherapy inflammation may result in residual ^{18}F-FDG uptake. Combined with a relatively low prevalence of residual disease this results in a relatively poor positive predictive value (PPV) of PET-CT, which leaves room for improvement (40). Performing additional IVIM imaging may enhance the PPV of ^{18}F-FDG-PET-CT. Lai et al. (27) showed that pretreatment NPC could be separated from posttreatment fibrosis with an accuracy of 100%. In practice residual tumor foci may be hidden in fibrotic tissue and therefore may be more challenging to detect. Further research should assess if the resolution of IVIM imaging is appropriate to detect residual disease in the proximity to fibrosis with a high diagnostic accuracy. In order to identify the most

optimal timing of IVIM imaging after irradiation more research is necessary. Preferably by performing IVIM multiple times during and after radiation therapy. Besides individual IVIM values, it may also be valuable to look at Δ values between different time points. In the studies included in this systematic review follow-up was relatively short, with a maximum of 3 months (23, 24), in the studies where intra- or posttreatment imaging was performed.

As shown by Marzi et al. (30) nasopharyngeal tumors have lower D and f values than SCC at other locations. When we pooled the results of the individual studies reporting mean values for D, D* and f, we could confirm the results of Marzi et al. This difference in IVIM characteristics is another argument to include only one tumor site in future studies, or at least to perform subgroup analyses when both nasopharyngeal carcinoma and other HNSCC are included.

In only one of the prognostic studies D* was significantly different between responders and non-responders suggesting that this parameter has the least potential to predict prognosis (24). There is less consensus on the prognostically most promising parameter. In four studies D had the highest predictive value (23, 24, 35, 36) and in two studies f was the strongest predictor of treatment outcome (25, 26).

A proposed hypothesis is that high f is associated with a higher regional blood flow (15). High regional blood flow may be indicative of a high microvessel density which is associated with a higher likelihood of both lymph node and distant metastases (42). Therefore high f could be an unfavorable prognostic factor for survival.

D appears to be inversely correlated with cell density (43, 44). Highly cellular tumors with rapidly dividing cells are more sensitive to chemotherapy and radiotherapy and therefore associated with a more favorable prognosis (23, 24, 35, 36). An increase in D during therapy is therefore a sign of decreasing cellularity and a good response to treatment (23, 24, 35, 36).

Lai et al. (28) showed that IVIM values differed between disease stages (AJCC, T stage and N stage). Tumors with a high cell turnover with more cell division may result in a larger tumor with densely packed cells resulting in both a higher disease stage and lower D values. For more perfusion weighted parameters (D* and f) it can be hypothesized that larger tumors are more prone to intratumoral necrosis due to lacking vascularization in the central portion of the tumor and therefore lower D* and f values. This is may have implications for future prognostic studies, because it warrants the need for correcting for disease stage when assessing the prognostic value of IVIM in HNC.

Imaging protocol

The optimal combination and number of b-values remains one of the keys points which need to be addressed. Lemke et al. (45) concluded that at least 10 b-values should be used for fitting the IVIM signal in clinical settings in a simulation study on abdominal IVIM imaging while assigning more weight to low b-values. Most b-values should be in the low

range (b=0-100 s/mm^2) with only a few b-values of >450 s/mm^2 are necessary (45). This is in contrast to the findings of Gurney-Champion et al. who found seven b-values to be enough for abdominal imaging and only three b-values if only the liver is of interest in a study on 16 healthy volunteers (46).

In none of the included studies imaging was performed on both a 1.5 T and 3 T MRI system on the same patients. Therefore, we could not determine whether the field strength matters for IVIM in head and neck imaging. We do consider this to be an important issue for further research. If 1.5 T and 3 T IVIM data prove to be comparable this would create more opportunities for multicenter research. In abdominal imaging there is some evidence suggesting that D and f values of the liver are reproducible on both 1.5 T and 3 T, whereas D* values are more variable (47). However, at this moment DWI values are not considered to be robust enough to be interchangeable between institutions, regardless of field strength (48).

In the head and neck area Sasaki et al. compared the traditional least-squares method requiring 11 b-values with a geometric approach requiring only three b-values for differentiating between tumor types (32). Even though D values were significantly higher and f values significantly lower with the geometric approach compared to the least-squares method; diagnostic accuracy was comparable for differentiating SCC from lymphoma and characterizing SG tumors.

We recommend to include at least four b-values below b=200 s/mm^2 in order to appropriately fit estimated perfusion-related parameters (i.e. D*, f, P and PP) until a proposed method which requires fewer b-values is appropriately validated (45). Two of the included studies did not fulfill this criterion (22, 33). The reported D*, f and PP values of these studies should therefore be interpreted with caution.

Therefore, future research should focus on the optimal combination of b-values as well as on the optimal model for determining IVIM parameters. Especially for fitting D* and f with IVIM it is important that imaging is performed with a high signal-to-noise ratio to avoid biased parameters (22). Furthermore it should be noted that the b-value represents the strength of the diffusion pulse and is dependent of the used gradient pulse sequence, the gradient pulse duration and the gradient strength (49). It is therefore possible to create identical b-values with different imaging parameters which may result in differences in ADC values (50). More uniform definitions of b-values may improve comparability between protocols.

The least robust IVIM parameter appears to be D* with reported mean D* values of SCC ranging from $4{\cdot}10^{-3}$ mm^2/s to $49{\cdot}10^{-3}$ mm^2/s (22, 24, 29, 32) and for NPC ranging from $18{\cdot}10^{-3}$ mm^2/s to $153{\cdot}10^{-3}$ mm^2/s (27, 28, 37, 38).

Limitations

Even though this review provides an extensive overview of the use of IVIM imaging in HNC, there are some limitations. Firstly, the included studies were heterogeneous in applied

b-values, imaging protocols, tumor types and outcome measurements. This prevented us from performing any meaningful meta-analysis. The differences in selection and number of b-values in the lower range may compromise the comparability of results. Secondly, the relatively small population size comprised the analysis of confounders in the prognostic studies. The outcome parameters, mainly in prognostic studies, were also heterogeneous. It would be preferable if all prognostic studies reported at least one uniform outcome measurement. Even if this standard may not be a perfect gold standard, for example the RECIST criteria (51). Thirdly, most studies did not correct for multiple testing which may overestimate the number of significant findings. Given that most studies are relatively small and that all can be regarded as positive studies it is likely that publication bias is present. Small studies with negative results may have been regarded to be not interesting enough for publication. The heterogeneity of the included studies, especially in terms of comparisons made and outcome measurements, made it impossible to perform statistical testing for the presence of publication bias.

Conclusions

With this systematic review we provide an overview of studies on IVIM in HNC. Studies are very heterogeneous in terms of applied b-values, imaging protocols, outcome measurements and reference standards. With combinations of IVIM-parameters SCC, lymphoma, malignant SG tumors, Warthin tumors and pleiomorphic adenoma can be reliably separated from each other. Low pre-treatment D and f and an increase in D during treatment was associated with a favorable response to treatment. D* appears to be the parameter with the lowest prognostic value. Future research should focus on finding the optimal IVIM protocol. When assessing diagnostic and prognostic properties of IVIM, authors should use uniformly accepted study methods and larger patient populations.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66(1):7-30.
2. Gatta G, Botta L, Sanchez MJ, et al. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. Eur J Cancer. 2015;51(15):2130-43.
3. European crude and age adjusted incidence by cancer, years of diagnosis 2000 and 2007: Analisys based on 83 population-based cancer registries. INFORMATION NETWORK ON RARE CANCERS IN EUROPE, 2015.
4. Hashibe M, Brennan P, Benhamou S, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst. 2007;99(10):777-89.
5. Huang SH, Xu W, Waldron J, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. J Clin Oncol. 2015;33(8):836-45.
6. Brana I, Siu LL. Locally advanced head and neck squamous cell cancer: treatment choice based on risk factors and optimizing drug prescription. Ann Oncol. 2012;23 Suppl 10:x178-85.
7. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. 2011;11(1):9-22.
8. O'Sullivan B, Huang SH, Su J, 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. Tsao SW, Yip YL, Tsang CM, et al. Etiological factors of nasopharyngeal carcinoma. Oral Oncol. 2014;50(5):330-8.
10. de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol. 2009;45(11):930-5.
11. Digonnet A, Hamoir M, Andry G, et al. Post-therapeutic surveillance strategies in head and neck squamous cell carcinoma. Eur Arch Otorhinolaryngol. 2013;270(5):1569-80.
12. Srinivasan A, Mohan S, Mukherji SK. Biologic imaging of head and neck cancer: the present and the future. AJNR Am J Neuroradiol. 2012;33(4):586-94.
13. 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;168(2):497-505.
14. Burdette JH, Elster AD, Ricci PE. Acute cerebral infarction: quantification of spin-density and T2 shine-through phenomena on diffusion-weighted MR images. Radiology. 1999;212(2):333-9.
15. Lemke A, Laun FB, Simon D, Stieltjes B, Schad LR. An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. Magn Reson Med. 2010;64(6):1580-5.
16. Driessen JP, van Kempen PM, van der Heijden GJ, et al. Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. Head Neck. 2015;37(3):440-8.
17. Le Bihan D. Intravoxel incoherent motion imaging using steady-state free precession. Magn Reson Med. 1988;7(3):346-51.
18. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
20. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37.

21. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.

22. Dikaios N, Punwani S, Hamy V, et al. Noise estimation from averaged diffusion weighted images: Can unbiased quantitative decay parameters assist cancer evaluation? Magn Reson Med. 2014;71(6):2105-17.

23. Ding Y, Hazle JD, Mohamed AS, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. NMR Biomed. 2015;28(12):1645-54.

24. Guo W, Luo D, Lin M, et al. Pretreatment Intra-Voxel Incoherent Motion Diffusion-Weighted Imaging (IVIM-DWI) in Predicting Induction Chemotherapy Response in Locally Advanced Hypopharyngeal Carcinoma. Medicine (Baltimore). 2016;95(10):e3039.

25. Hauser T, Essig M, Jensen A, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. Neuroradiology. 2013;55(5):527-36.

26. Hauser T, Essig M, Jensen A, et al. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: Evaluation of lymph node metastasis. Eur J Radiol. 2014;83(5):783-7.

27. Lai V, Li X, Lee VH, Lam KO, Chan Q, Khong PL. Intravoxel incoherent motion MR imaging: comparison of diffusion and perfusion characteristics between nasopharyngeal carcinoma and post-chemoradiation fibrosis. Eur Radiol. 2013;23(10):2793-801.

28. Lai V, Li X, Lee VH, et al. Nasopharyngeal carcinoma: comparison of diffusion and perfusion characteristics between different tumour stages using intravoxel incoherent motion MR imaging. Eur Radiol. 2014;24(1):176-83.

29. Lu Y, Jansen JF, Stambuk HE, et al. Comparing primary tumors and metastatic nodes in head and neck cancer using intravoxel incoherent motion imaging: a preliminary experience. J Comput Assist Tomogr. 2013;37(3):346-52.

30. Marzi S, Piludu F, Vidiri A. Assessment of diffusion parameters by intravoxel incoherent motion MRI in head and neck squamous cell carcinoma. NMR Biomed. 2013;26(12):1806-14.

31. Sakamoto J, Imaizumi A, Sasaki Y, et al. Comparison of accuracy of intravoxel incoherent motion and apparent diffusion coefficient techniques for predicting malignancy of head and neck tumors using half-Fourier single-shot turbo spin-echo diffusion-weighted imaging. Magn Reson Imaging. 2014;32(7):860-6.

32. Sasaki M, Sumi M, Eida S, Katayama I, Hotokezaka Y, Nakamura T. Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors. PLoS One. 2014;9(11):e112866.

33. Sumi M, Nakamura T. Head and neck tumors: assessment of perfusion-related parameters and diffusion coefficients based on the intravoxel incoherent motion model. AJNR Am J Neuroradiol. 2013;34(2):410-6.

34. Sumi M, Van Cauteren M, Sumi T, Obara M, Ichikawa Y, Nakamura T. Salivary gland tumors: use of intravoxel incoherent motion MR imaging for assessment of diffusion and perfusion for the differentiation of benign from malignant tumors. Radiology. 2012;263(3):770-7.

35. Xiao Y, Pan J, Chen Y, Chen Y, He Z, Zheng X. Intravoxel Incoherent Motion-Magnetic Resonance Imaging as an Early Predictor of Treatment Response to Neoadjuvant Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma. Medicine (Baltimore). 2015;94(24):e973.

36. Xiao-ping Y, Jing H, Fei-ping L, et al. Intravoxel incoherent motion MRI for predicting early response to induction chemotherapy and chemoradiotherapy in patients with nasopharyngeal carcinoma. J Magn Reson Imaging. 2016;43(5):1179-90.

37. Yu XP, Hou J, Li FP, et al. Intravoxel Incoherent Motion Diffusion Weighted Magnetic Resonance Imaging for Differentiation Between Nasopharyngeal Carcinoma and Lymphoma at the Primary Site. J Comput Assist Tomogr. 2016;40(3):413-8.

38. Zhang SX, Jia QJ, Zhang ZP, et al. Intravoxel incoherent motion MRI: emerging applications for nasopharyngeal carcinoma at the primary site. Eur Radiol. 2014;24(8):1998-2004.

39. Sumi M, Nakamura T. Head and neck tumours: combined MRI assessment based on IVIM and TIC analyses for the differentiation of tumors of different histological types. Eur Radiol. 2014;24(1):223-31.

40. de Bree R, van der Putten L, Brouwer J, Castelijns JA, Hoekstra OS, Leemans CR. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 2009;45(4-5):386-93.

41. de Bree R, van der Putten L, van Tinteren H, et al. Effectiveness of an (18)F-FDG-PET based strategy to optimize the diagnostic trajectory of suspected recurrent laryngeal carcinoma after radiotherapy: The RELAPS multicenter randomized trial. Radiother Oncol. 2016;118(2):251-6.

42. Gasparini G, Weidner N, Maluta S, et al. Intratumoral microvessel density and p53 protein: correlation with metastasis in head-and-neck squamous-cell carcinoma. Int J Cancer. 1993;55(5):739-44.

43. Gupta RK, Cloughesy TF, Sinha U, et al. Relationships between choline magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma. J Neurooncol. 2000;50(3):215-26.

44. White ML, Zhang Y, Robinson RA. Evaluating tumors and tumorlike lesions of the nasal cavity, the paranasal sinuses, and the adjacent skull base with diffusion-weighted MRI. J Comput Assist Tomogr. 2006;30(3):490-5.

45. Lemke A, Stieltjes B, Schad LR, Laun FB. Toward an optimal distribution of b values for intravoxel incoherent motion imaging. Magn Reson Imaging. 2011;29(6):766-76.

46. Gurney-Champion OJ, Froeling M, Klaassen R, et al. Minimizing the Acquisition Time for Intravoxel Incoherent Motion Magnetic Resonance Imaging Acquisitions in the Liver and Pancreas. Invest Radiol. 2016;51(4):211-20.

47. Cui Y, Dyvorne H, Besa C, Cooper N, Taouli B. IVIM diffusion-weighted imaging of the liver at 3.0T: Comparison with 1.5T. European Journal of Radiology Open. 2015;2:123-8.

48. Kolff-Gart AS, Pouwels PJ, Noij DP, 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;36(2):384-90.

49. 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.

50. Girometti R, Furlan A, Esposito G, et al. Relevance of b-values in evaluating liver fibrosis: a study in healthy and cirrhotic subjects using two single-shot spin-echo echo-planar diffusion-weighted sequences. J Magn Reson Imaging. 2008;28(2):411-9.

51. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228-47.

CHAPTER 2.3

Functional imaging early during (chemo) radiotherapy for response prediction in head and neck squamous cell carcinoma: a systematic review and meta-analysis

Submitted

Roland M Martens
Daniel P Noij
Meedie Ali
Thomas Koopman
J Tim Marcus
Marije R Vergeer
Henrica de Vet
Marcus C de Jong
C René Leemans
Otto S Hoekstra
Remco de Bree
Pim de Graaf
Ronald Boellaard
Jonas A Castelijns

ABSTRACT

Background: Patients with advanced stage head and neck squamous cell carcinoma treated with (chemo)radiotherapy may experience locoregional failure. Non-invasive response prediction of locoregional outcome pretreatment and during treatment could allow for personalized treatment adaptation. Our aim was to assess the predictive value of functional imaging pretreatment and early after start of (chemo)radiotherapy by performing a systematic review of literature.

Methods: We searched MEDLINE/EMBASE for publications until April 1st 2018 assessing predictive performance of functional imaging (computed tomography (CT) perfusion, functional magnetic resonance imaging (fMRI), positron emission tomography (PET)/CT) performed within 4 weeks after (chemo)radiotherapy initiation. Data on study characteristics, imaging protocols and patient outcome were extracted independently by 2 reviewers. Pooled estimation and subgroup analyses were performed.

Results: We included 53 studies (CT (n=4), PET (n=25), fMRI (n=24)) with 1634 patients. Primary tumors or nodal node metastases with a high baseline perfusion and intratreatment perfusion change (increase of blood flow, blood volume and Ktrans) on CT, high baseline diffusion restriction (i.e. low ADC) and intratreatment ADC increase on MRI and high baseline ^{18}F-FDG-PET uptake followed by an intratreatment reduction of ^{18}F-FDG-PET at PET, were predictive for good treatment response and recurrence-free and overall survival. Optimal timing for treatment response assessment and long-term outcome was as early as 2-3 weeks after treatment initiation for all imaging modalities.

Conclusion: An increase of perfusion and diffusion and a decrease ^{18}F-FDG-PET uptake at 2-3 weeks during treatment were predictive for favorable treatment response, locoregional control and survival. Future studies should focus on the standardization of techniques, acquisition methods and the validation of more uniform predictive parameters in an individual patient data meta-analysis.

INTRODUCTION

Head and neck cancer (HNC) accounts for approximately 5% of cancer incidence worldwide, with head and neck squamous cell carcinoma (HNSCC) being the most common type of HNC (1). Choice of treatment depends on factors such as primary tumor location, extension into adjacent structures and the possibility of function preservation (2). Early stage disease is typically treated with single modality radiotherapy or surgery. Locally advanced tumors often require combinations of surgery, radiotherapy and/or chemotherapy (3).

Despite current treatment options, locoregional recurrence rates in the first 2 years up to 15-50% are reported in patients with advanced stage tumors (4-6). Optimization of treatment monitoring could allow for early escalation of treatment (e.g. increasing radiation dose, addition of chemotherapy or response modifiers) or an early switch to another treatment modality (i.e. primary surgery), or de-escalation of treatment, thereby reducing overtreatment and unnecessary toxicity (7-9).

Nowadays, clinical, histopathological and (conventional) imaging biomarkers are used in order to perform accurate treatment selection and response assessment (3, 5). Pretreatment conventional imaging biomarkers, on computed tomography (CT) and MRI, are mainly focused on morphologic tumor characteristics (10, 11), while functional imaging could map physiological processes (3, 12).

Change of tumor characteristics during treatment might be predictive for treatment response and long-term outcome. Changes in perfusion and metabolic activity due to cellular stress and damaged cellular membranes occur early after start of treatment and may precede changes in size (13, 14). Effects of radiation and chemotherapy start with permeability changes (15) due to reoxygenation. This is followed by the formation of interstitial edema in the first 2 weeks and progressive thickening of the connective tissue, which results in a reduction of venous and lymphatic drainage. In the end fibrosis is formed (17, 18).

These physiological changes could be captured by functional imaging (e.g. CT-perfusion, dynamic contrast-enhanced MRI (DCE-MRI), diffusion-weighted (DW-)MRI, intra-voxel incoherent motion (IVIM) MRI, 1H-MR-spectroscopy (MRS) and positron emission tomography (PET) (19).

With CT-perfusion information on angiogenesis and tumor perfusion can be obtained, which could reflect the reaction to anti-angiogenic therapy (20-22). Dynamic contrast-enhanced MRI, often referred to as MRI perfusion, depicts perfusion and permeability of tissue structures (23, 24).

With DWI the mobility of water molecules can be quantified using apparent diffusion coefficients (ADC) (25, 26). Low ADC values are generally associated with malignancy (25). Diffusion-weighted imaging can be extended by the intra-voxel incoherent motion (IVIM) technique (27), which separates the influence of tissue capillary perfusion on the diffusion (28).

^{18}F-Fluorodeoxyglucose (FDG)-PET assesses the metabolic, glycolytic activity of tissues (29). Other PET-tracers could measure hypoxia (^{18}F-fluoromisonidazol, FMISO) (30) or proliferation (3'-Deoxy-3'-^{18}F-fluorothymidine, ^{18}F-FLT) (31).

The value of functional imaging techniques and optimal timing to predict treatment response and long-term outcome early during treatment remains unknown. Therefore, a systematic review of the predictive and prognostic value of functional imaging early during treatment of HNSCC is warranted. Because treatment modification may be still possible (early) during treatment, we focused on functional imaging performed during (chemo) radiotherapy for: 1) treatment response during treatment; and 2) prediction of long-term recurrence-free and overall survival.

METHODS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews was used as guidance (32).

Search strategy and study selection

PubMed (Medline) and EMBASE were searched for articles published until April 2018 on functional imaging techniques in HNSCC performed early during (chemo)radiotherapy (within 4 weeks after initiation) (See Appendix A for the full search strategy), without language restrictions. Discrepancies were resolved by consensus. We used the following inclusion criteria: 1) Study population consisted of at least 10 patients with HNSCC; 2) functional imaging was performed with at least one of the following techniques: CT-perfusion, MR-spectroscopy, DCE-/DW-/(IVIM)-MRI or PET(CT/MRI); 3) imaging was performed within 4 weeks after the start of (chemo)radiotherapy and was used for predicting treatment response or long-term outcome; 4) histopathology, clinical and/or imaging follow-up were reference standard. Studies were excluded if 1) nasopharyngeal tumors were the main subject, due to its unique histopathology (2); 2) the article was a conference abstract or study with focus on an experimental treatment; 3) the study reported overlapping study populations with the same analysis performed.

Data extraction

Data on study and patient characteristics, imaging protocol and predictive parameters for treatment response and prognostic parameters for long-term outcome, were extracted by two reviewers (RMM, MA), independently. Discrepancies were resolved by consensus.

Early treatment response in studies were classified as "complete response" (CR), when residual tumor was absent, or as ''non-complete response" (non-CR) when residual tumor was documented within 6 months after treatment completion. Treatment response was assessed 4-8 weeks post-treatment by endoscopy and 3 months post-treatment by MRI or ^{18}F-FDG-PET-CT imaging.

Long-term outcome was described as disease free survival (DFS) and overall survival (OS). Disease free survival was divided in ''locoregional control'' (LRC) and ''locoregional failure''

(LRF). Overall survival was divided in survival and death as sub-terms. Information on predictive/prognostic outcomes and source data, (i.e. odds ratio (OR) and hazard ratio (HR) with 95% confidence intervals (CI), true positive (TP), false positive (FP), true negative (TN) and false negative (FN)) were extracted. In case of incomplete 2x2 tables, authors were contacted.

The predictive value for treatment response by CT-perfusion, functional MRI and PET was assessed. The predictive/prognostic capacity of the above-mentioned functional techniques to predict DFS and OS was assessed.

Quality assessment

We assessed the quality of eligible studies using checklists of QUADAS-2 (Quality Assessment for Diagnostic Accuracy Studies) (33) and QUIPS for long-term outcome studies (Quality in Prognostic Studies), respectively (34).

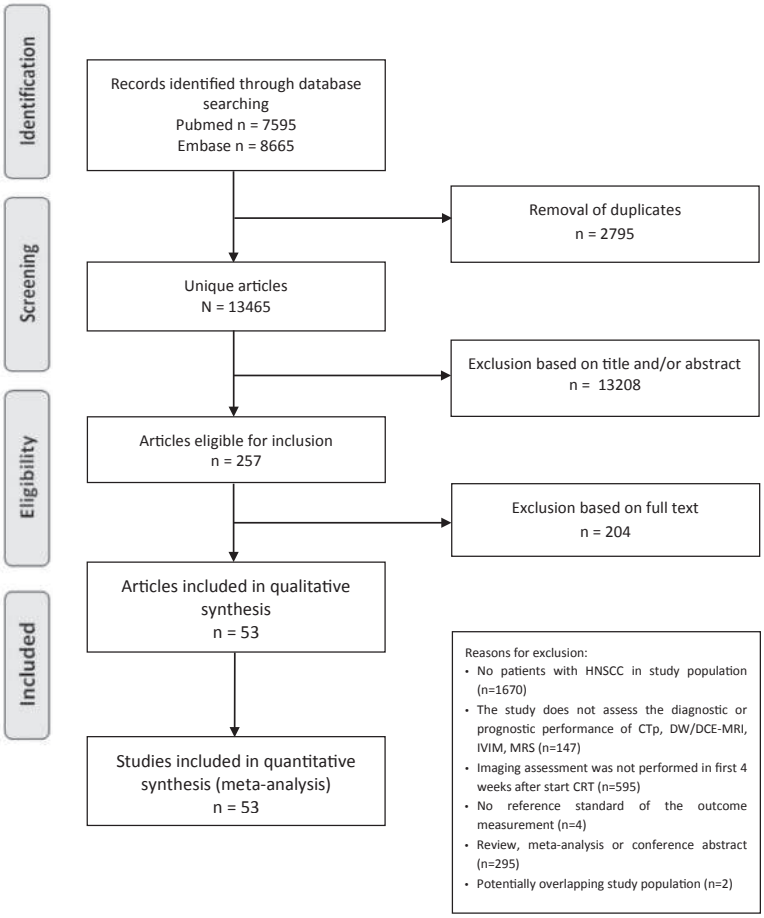


Figure 1 *PRISMA flow diagram of included studies*

Data synthesis

Early response and long-term outcome parameters were analyzed separately. Intratreatment parameters and delta (Δ, i.e. values between pre- and during treatment imaging) were evaluated with respect to the baseline value (35). Differences in outcome definitions (TP, FP, TN, FN) were aligned in order to compare results; i.e. true positive were assigned when the value was higher than the optimal threshold, which predicted an adverse response/outcome. Accuracy, odds ratio and hazard ratio (long-term outcome) were calculated based on per patient data. Variability between individual studies was evaluated by plotting the diagnostic accuracy estimates, and the proportional hazard model was pooled and presented on forest plots with 95% confidence intervals (95%CI), using RevMan 5.3 software (Cochrane collaboration, Copenhagen, Denmark). Heterogeneity was quantified using the I^2 index, which describes the percentage of variation across studies that is due to heterogeneity rather than chance. Statistical analyses were performed using SPSS (version 22, Chicago, IL, USA).

RESULTS

Study selection

The search yielded 13,465 unique studies. The full texts of 257 studies were reviewed (Figure 1). Finally we included 53 articles in which CT-perfusion (n=4) (15, 36-38), functional MRI (n=23) (10, 16, 39-59), and PET (n=26) (7, 14, 30, 31, 42, 54, 59-81) were used (Table 1). Functional MRI was applied in five DCE studies, 16 DWI studies of which three IVIM studies and one MRS study. ^{18}F-FDG-PET was applied in 15 studies, ^{18}F-FLT in four, ^{18}F-FMISO in three and ^{18}F-Hx4-PET in one study. For two studies (69, 70) we suspected overlap in study populations. However, we could not verify this and because they contained complementary information we included both studies.

Baseline characteristics

Total study population consisted of 1,670 patients, of which 75-100% was male (See Appendix B and C for extended baseline and technical details, respectively). The studies mainly consisted of T2 or T3 tumors (Appendix D1-3) among all locations (Appendix E1-3) and N2 nodal stage. The AJCC stage (7th edition) was most often III or IV. All studies were prospective, except for three MRI studies (43, 51, 82) and five PET studies (9, 68, 69, 71, 75) and in one study (76) it was not specified. In 38 out of 56 studies (70%), patients received cisplatin-based chemotherapeutic regimens. Reference standard was MR- or PET-imaging followed by (histo)pathology in case of suspicion of malignancy in all studies, except for two PET studies (42, 72) in which the reference standard was not mentioned.

Treatment response was studied in 15 studies (CT-perfusion (n=2), functional MRI (n=7), PET (n=7)), long-term outcome in 45 studies (CT-perfusion (n=2), functional MRI (n=19), PET (n=24) and five studies assessed both (Appendix F1-3 and G1-2, respectively). The prevalence of complete response was 50% in CT studies, 11-96% in MRI studies and 22-94% in PET studies. Prevalence of 2 years DFS was 50% in CT studies, 29-89% in MRI studies and 42-90% in PET studies. Prevalence of 2 years OS was 87-93% in CT studies, 29-91% in MRI studies and 32-97% in PET studies.

Table 1 *Overview of included studies*

	Studies			Patients			Follow-up	Treatment[c]						Complete response	Locoregional control	Overall survival
	Total	Treatment response[e]	Long-term outcome[e]	n	Age mean (range)	Male (%)	Follow-up (months)	RT (Gy)	Plat	EGFR	VEGF	Tax	5-FU	Range (%)	Range (%)	Range (%)
CT Perfusion[b,d]	4	2	2	79	56 (51-58)	80-96	26 (24-28)[b]	66-72	3	0	0	0	0	12-16 (50-65)	13-14(87-93)	...
MRI[b]	23	7	18	588	54 (28-83)	75-100	29 (5-76)[b]	70-72[a]	17	5	2	4	3	2-40(11-96)	5-46(29-89)	5-32 (29-91)
DCE	7	1	6													
DWI	17	3	11													
IVIM	3	2	2													
MRS	1	0	1													
PET[b]	26	3	23	1003	61 (22-118)	75-100	30 (1-83)[b]	50-78	18	6	1	0	3	6-54(22-94)	6-75(52-97)	7-80 (32-100)
FDG-PET	16	3	14													
FMISO	7	0	7													
FLT	5	0	5													
FHX4	1	0	1													

[a] Matoba(97) - Treatment 60-70Gy
[b] Follow-up was not specified in 1 CT, 3 MIR and 7 PET studies
[c] In 8 studies, a subgroup of patients were treated with neoadjuvant chemotherapy
[d] In one study patients received Mitomycin Chemotherapy
[e] 8 studies both treatment response as long term outcome

Abbreviations: 5-FU = 5-Fluoruracil; DCE = dynamic contrast-enhanced; DWI = diffusion weighted imaging; EGFR = Epidermal growth factor receptor-targeting chemotherapy; FDG-PET = Fluorodeoxyglucose positron emission tomography; FHX4 = 3-[^{18}F]fluoro-2-(4-((2-nitro-1H-imidazol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propan-1-ol; FLT = 3 -Deoxy-3 -18F-fluorothymidine; FMISO = ^{18}F-fluoromisonidazole; IVIM = intravoxel incoherent motion MRI; MRS = magnetic resonance spectroscopy; Plat = platinum-based chemotherapy; RT = radiotherapy; Tax = Taxus based chemotherapy; VEGF = Vascular endothelial growth factor-targeting chemotherapy

Study quality
The QUADAS-2 for early treatment response prediction studies (Appendix H), showed that overall risk of bias and applicability concerns were low or unclear for all studies. However, in three studies (15, 46, 49) patient selection, and flow and timing were scored as high risk for bias, due to retrospective patient selection, absence of a reference standard or exclusion of patients in the analysis. The QUIPS (Appendix I) resulted in overall low risks for bias on study participation, prognostic factor, outcome measurement and statistical analysis and reporting. However, on study attrition and confounders, one (43) and five studies (7, 14, 49, 71, 83) scored high risk for bias, respectively. Patient HPV status was reported in three MRI studies (13.6%) (57, 58) and two PET studies (7.1%) (14, 72) as possible effect-modifier.

Early treatment response prediction

Perfusion computed tomography
First-pass perfusion was assessed in two studies (15, 38). A high baseline blood flow (BF >106 ml/100g/min) (*P*=0.006) were predictive for complete response (prognostic accuracy 83.3% (95% CI, 55.2-95.3), likelihood ratio 5.0 (95% CI 1.8-13.9)). Furthermore, a combination of high BF (>106ml/100g/min), blood volume (BV ≤47ml/100g/min) and permeability surface (PS, i.e. the product between permeability and the total surface area of the capillary endothelium in a unit mass of tissue) at 3-4 weeks intratreatment were 100% (95% CI, 61.0-100) predictive for favorable treatment response (*P*=0.001, *P*=0.002, *P*=0.004) (15) (38).

Perfusion magnetic resonance imaging
One study (45) measured treatment response and found a higher pre-treatment median K^{trans} (product of blood flow, permeability and capillary surface area) combined with a K^{trans} reduction during the first week of treatment in patients with a complete response compared with patients with treatment failure.

Diffusion
The ΔADC was found to be higher in CR than non-CR, varying among the imaging acquisition moments in seven studies (46, 49, 52, 55, 57, 58): at 2 weeks, four studies (n=41, n=34, n=23 respectively) found an ADC_{mean} increase of 25, 32% and >100% in CR (*P*<0.0001-0.003) (49, 52, 55, 58). One study (84) reported an accuracy of 95% for any decreasing ADC value during treatment for predicting non-CR. For ADC_{median} at 2 weeks an increase of 24% was found in CR (n=10) (55); at 3 weeks, six studies found a 36-42% ADC_{mean} increase in CR (all p<0.003) [46, 52, 55, 57, 58] (n=10, n=34, n=15, n=31 and n=23 patients, respectively) versus 24% in non-CR (49). At week 3-4; two studies (57) (n=10, n=15) found a 104% ADC_{mean} increase in CR versus a 28% and 38% increase in non-CR (55, 57).

Positron emission tomography

The intratreatment ^{18}F-FDG-PET was found to be lower in CR than non-CR in three studies (7, 64, 85). An overview of the accuracy of treatment response prediction with Δ ^{18}F-FDG-PET uptake volume (SUV) and absolute SUV is shown in Figure 2. Lower metabolic rate measured by FDG-uptake <9 at baseline (*P*=0.01) and <16 at 3 weeks intratreatment (*P*=0.007) was associated with favorable treatment response (7).

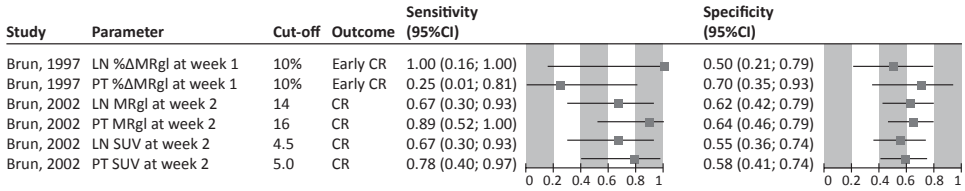


Figure 2 *An overview of FDG-PET studies on accuracy sorted by imaging time point. Sensitivity and specificity are reported with according 95% confidence interval as horizontal lines.*

Abbreviations: CR = complete response, LN = lymph node, MRgl = metabolic response in FDG-uptake, PT = primary tumor, SUV = standard uptake value

<u>*Long-term outcome prediction*</u>

Perfusion

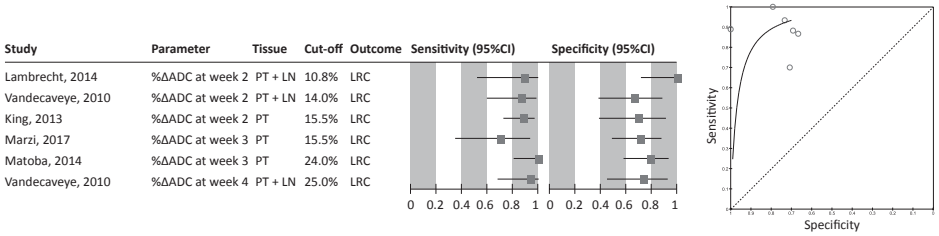
Long-term outcome was assessed in five studies using dynamic contrast-enhanced MRI (16, 41, 43-45). Two studies (16, 44) showed a significantly higher tumor-blood-volume (ΔTBV) at DCE after 2 weeks of (chemo)radiotherapy in LRC patients compared with LRF patients (*P*=0.03 and *P*=0.01). Another study found higher K^{trans} (*P*=0.012) and interstitial space volume fraction (V_e; *P*=0.012) in LRC (59).

Diffusion

Seventeen studies (46-50, 52-58, 86, 87) assessed the prognostic accuracy of DWI, of which two measured IVIM (57, 58). An optimal cut-off of minimal ΔADC_{mean} was determined per study, which was most prognostic for long-term outcome (Figure 3a). Pooled analysis showed that a ΔADC_{mean} higher than the optimal cut-off (ranging from 10.8% to 15.5%) resulted in an odds ratio of 22.38 (95%CI 7.76-64.55) at 2 weeks intratreatment for LRC during the 2-years follow-up (Figure 3b). Despite differences in patient population, imaging systems and acquisition protocols, the heterogeneity was low between pooled studies (I^2=0).

Imaging in the first week of treatment did not result in any ADC_{mean} change (48). However, at 2 weeks a ΔADC_{mean} increase of 41%, 21% and 36% was found in patients with LRC compared with a ΔADC_{mean} increase of -2%, 7% or 14% in patients with LRF, respectively (50, 51, 56). Imaging at 3 weeks intratreatment showed a ΔADC_{mean} of 22% and 51% in LRC, compared with a ΔADC_{mean} in LRF of 7% and 19%, respectively (46, 87). Imaging at 4 weeks intratreatment showed a ΔADC_{mean} in patients with a LRC of 65%, while in patients with LRF ΔADC_{mean} was -1% (56). An increasing trend of ΔADC_{mean} in LRC until week 4 was reported (55). However, 2 studies did not find significant differences in ΔADC_{mean} at 2-3 weeks between LRC and LRF patient (53, 54). No study was found in which ADC_{mean} was used for overall survival prognosis.

A



B

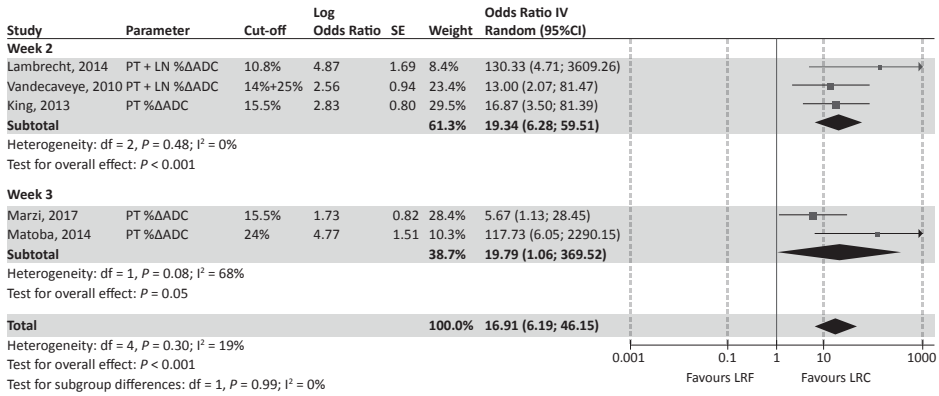
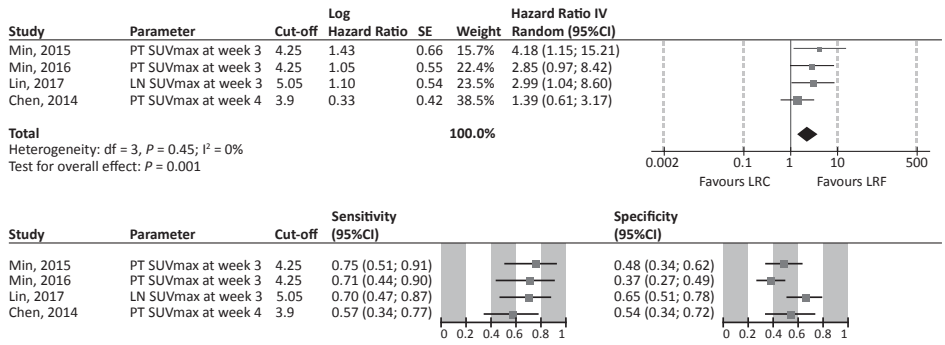


Figure 3 *A) Accuracy of MRI studies to predict locoregional control sorted by imaging time point (weeks). Sensitivity and specificity are reported with 95% confidence interval as horizontal lines. On the right, a summary receiver operating characteristic (SROC) curve is shown to represent the performance of %ΔADC to predict LRC. B) Pooled odds ratio of ΔADC to predict LRF by performing. Higher % ADC increase than the optimal cut-off value (OC) resulted in a higher odds for LRC.*

Abbreviations: ADC = apparent diffusion coefficient, df = degrees of freedom, IV = instrumental variable, LN = lymph node, LRC = locoregional control, LRF = locoregional failure, SE = standard error, PT = primary tumor

A



B

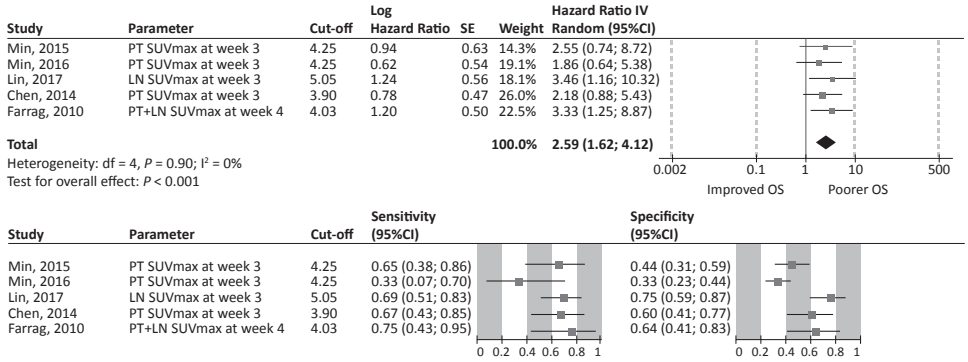


Figure 4 *A) The accuracy and hazard ratio of SUVmax for the prediction of DFS. Low to moderate accuracy is shown for the week 3-4 assessment of SUVmax. B) The accuracy and hazard ratio of SUVmax for prediction OS. Higher SUVmax than the optimal cut-off resulted in a higher hazard for death.*

Abbreviations: df = degrees of freedom, IV = instrumental variable, LN = lymph node, LRC = locoregional control, LRF = locoregional failure, OS = overall survival, SE = standard error, SUV = standard uptake value, PT = primary tumor

Positron emission tomography

In eight studies ^{18}F-FDG-PET studies SUVmax was measured (62, 65-67, 69, 74, 75, 88). An absolute SUVmax higher than the optimal cut-off defined in each study, ranging from 4.25 to 5.05, at 3-4 weeks intratreatment was found to be predictive for LRF in four studies (69, 73, 74, 88), with a hazard ratio of 2.32 (95%CI 1.39-3.87) (Figure 4a). In these pooled studies the patient population and patients outcome was quite homogeneous (I^2=0) with a LRC in 53%-75%. Although patient population and image system and acquisition protocols differed. Furthermore, glucose value, time per bed position and reconstructed matrix were not mentioned in two studies and in one study TNM-stage was not specified. Lower absolute SUVmax after 3 weeks of (chemo)radiotherapy (i.e. absolute SUVmax <4.25 g/

mL) was predictive for better DFS (*P*=0.002) (69, 88).

The accuracy of predicting overall survival with (Δ)SUVmax intratreatment is shown in Figure 4b. The absolute SUVmax resulted in a pooled hazard ratio of 2.59 (95%CI, 1.62-4.12). A SUVmax reduction ratio of ≥0.64 in the primary tumor was associated with better overall survival (HR 0.379 for death; *P*=0.035) and DFS (HR 0.429 for recurrent disease; *P*=0.045) (73).

Four studies (69, 74, 75, 88) reported that a lower (Δ) total lesion glycolysis (TLG) at 3 weeks intratreatment was moderately predictive for DFS and OS (Figure 5a and 5b, respectively). One study (75) showed that a TLG reduction of more than 5% per week was associated with improved DFS (*P*=0.04; HR= 0.37; 95%CI=0.15-0.95). Three studies (74, 88) (69) reported that an absolute TLG value of ≤9.4 or <14.0 at week 1-3 intratreatment had a significant better locoregional failure free survival of 72% and 78% compared to 35% and 41%, respectively when TLG was higher than the cut-off (*P*=0.012; *P*=0.005; HR 4.36-7.76; 95%CI=1.40-32.6).

The best reference structure to predict LRF using visual assessment of metabolic tumor response was the uptake in liver and blood pool. Complete metabolic response at 3 weeks was predictive for LRC after 2 years, with a locoregional failure and death of 89.8% in patients with intratreatment FDG-uptake more than the liver versus 71.5% in patients with uptake less than liver (*P*=0.062) (71).

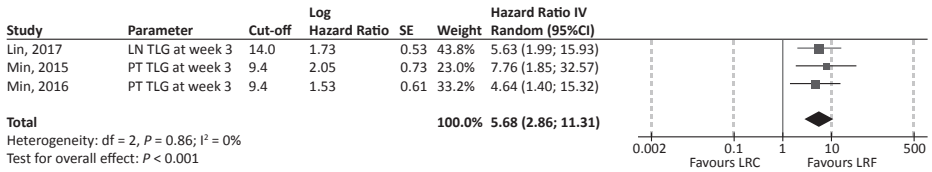
FMISO-PET uptake with tumor-to-background-ratio (TBR) (gradation of hypoxia) at 2 weeks during treatment <1.26 or <1.93 was associated with 2 year locoregional control (*P*=0.001, *P*=0.016) (30, 77). TBR_{peak} at 1 or 2 weeks intratreatment was predictive for locoregional control (*P*=0.019, *P*=0.012, respectively) (81). A tissue to blood ratio (T/Bmax) of <1.17 at 3 weeks intratreatment was predictive for good long-term outcome (*P*=0.02) (42). Delta TBR was significantly predictive for LRC (*P*<0.01) and associated with perfusion (r=0.7) (80).

A FLT-PET SUVmax decrease of ≥45% at 2 weeks (chemo)radiotherapy was associated with a better 3-year DFS (88% vs. 63%, *P*=0.035) (31).

DISCUSSION

We discuss the value and most optimal timing of performing early intratreatment functional imaging parameters regarding the effect of tumoral perfusion, diffusion and metabolic rate on treatment response, long-term disease-free survival and overall survival.

A



B

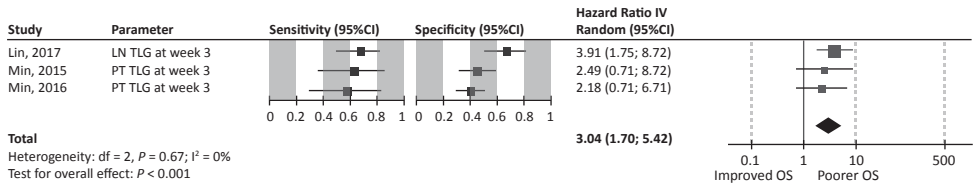


Figure 5 *A) The accuracy and hazard ratio of FDG-PET TLG was low to moderate for prediction of DFS. B) the accuracy and hazard ratio of TLG for predicting OS is shown, which resulted in a moderate accuracy.*

Abbreviations: df = degrees of freedom, IV = instrumental variable, LN = lymph node, LRC = locoregional control, LRF = locoregional failure, MRgl = metabolic response in FDG-uptake, PT = primary tumor, SE = standard error, TLG = total lesion glycolysis.

Predictive functional parameters for early outcome and optimal timing

Perfusion

Treatment response was assessed by perfusion in four studies (15, 38, 45, 59). Overall, high pretreatment blood flow and blood volume and low permeability surface (PS), followed by an intratreatment decrease of BF on perfusion CT were predictive for locoregional control (15, 38). A decrease of K^{trans} on DCE-MRI, which may reflect reductions in vascular permeability rather than perfusion, were predictive for favorable treatment response. An upregulation of angiogenic pathways and activation of different growth factors during radiotherapy in response to radiation-induced stress (89, 90) results in collateral capillary and lymphatic channels after treatment initiation, thus modulating tumor radio-sensitivity (2, 5, 37). Chemoradiotherapy-induced damage on the intratumoral microvasculature and high-resistance flow in neoplastic vessels may explain the induced decline in blood flow and volume and permeability surface (15, 38, 45, 59).

Diffusion

Studies using DWI (46, 49, 52, 55, 57, 58) suggested that higher ΔADC increase (≥25-32% at week 2, ≥36-42% at week 3 and ≥104% at week 3-4) was associated with a favorable treatment response. This was hypothetically due to an increase in diffusion capacity in the extracellular space (ECS) that occurs with cell shrinkage and death and the movement of

water from intracellular to ECS (52, 55).

Positron emission tomography
In the PET imaging studies it was reported that an intratreatment SUV_{max} higher than the optimal study-specific cut-off PET value for predicting treatment failure (7, 60, 64). This might be explained by higher proliferation rate and hypoxia. However, it could also be caused by radiation-induced inflammation (91), leading to false positive results in the early phase of treatment. A SUV_{max} reduction of more than 10% was predictive of CR, while smaller changes could occur due to random errors in patient positioning and low accuracy in tumors with only small increases in glucose metabolism compared with normal tissue metabolism. Furthermore, a decline in FDG uptake could indicate reoxygenation of tumor and thus gain in radiosensitivity (7).

Optimal timing imaging for early outcome prediction
Imaging performed 2-3 weeks intratreatment is considered optimal for treatment response assessment, in order to map important tumoral changes and to adapt treatment if necessary (46). Tumoral changes depend on early treatment-induced cellular, vascular and inflammatory reactions, which occur simultaneously during the course of (chemo) radiotherapy. Due to the overlap of these phenomena, correct timing of imaging is important to map the divergence of functional parameters in the CR and non-CR group.

Perfusion imaging, performing perfusion CT and DCE-MRI, might be effective to identify changes in tumoral vascularization and identify regions with hypoxia (19). However, only limited evidence was found for a decrease of blood flow and blood volume at 3-4 weeks on perfusion CT and 1 week during treatment on DCE-MRI, to be predictive for CR (15, 45).

Diffusion assessment showed ADC changes manifested earlier than morphological changes with an optimal time point at 3 weeks (46, 49, 52, 55, 58). The ΔADC increase of viable tumor was seen within week 2-3, after which a plateau was reached, presumably because tumor microvasculature and tumor cell environment had been effectively destroyed by concurrent chemoradiotherapy and diffusion could no longer increase (49, 55, 58).

On PET, low baseline metabolic rate (<16-20 µmol/min/100g tissue) or a high baseline tumor metabolic rate (>16-20 µmol/min/100g tissue) followed by a reduction at week 2-3 (>10%) predicted CR (7, 60). However, therapy-associated inflammation of surrounding mucosa tissue, which is ^{18}F-FDG avid due to glucose consumption by activated macrophages, limits defining correct tumor volumes mainly from the 3rd week of chemoradiotherapy (65).

The overall best time to perform DW- or PET-imaging was found to be between 2 and 3 weeks, although accuracy was hardly mentioned in the included studies. This imaging moment is still close to the start of treatment (less side-effects) and late enough to assess early changes in diffusion and metabolic rate. High ΔADC increase and ΔSUV reduction during treatment predicts a favorable response to treatment.

Predictive parameters for long-term outcome and optimal timing

Perfusion

A higher pretreatment blood flow or increase of blood flow on CTp was predictive for LRC (36, 37), hypothetically due to improved oxygenation and a greater sensitivity of tumor cells to radiation-induced free radical damage with each fraction of RT (36).

With DCE-MRI, an increase of blood volume and blood flow (16, 41-44) at 2 weeks during treatment was reported to represent sufficient oxygen provision during RT, which was associated with favorable LRC (16, 44). The persistency of some poorly perfused subvolumes with low BV or BF during treatment predicted LRF and could be targeted with local intensification of treatment (44).

Overall, small studies suggest that an increased blood flow and blood volume found on CT at week 2 during treatment and a reduction of Ktrans (product of permeability and capillary surface) predicts LRC and a favorable overall survival.

Diffusion

A higher ΔADC (increase) than the optimal cut-off (range, 10.8-25%) at 2 weeks intratreatment was predictive for locoregional control, with a pooled odds ratio of 16.91 (95%CI 6.19-46.15, Figure 3). However, ΔADC was complicated by variability (up to 15%), due to which a ΔADC of 14.6% (56) and 15.5% (50) might reflect predominantly baseline variability, whereas changes above 15% are more suggestive of true treatment-induced responses (47) Due to tumor heterogeneity (46) and presence of sub-entities (e.g. hypoxia) (58), intra-treatment response prediction, would need a multidisciplinary approach, instead of a single parameter predictor.

Low intra-voxel incoherent motion (IVIM) perfusion-free diffusion coefficient D seemed more sensitive to variation in the cellular microstructure than ΔADC, which is susceptible to effects of perfusion and diffusion. Early radiation effects, which are associated with cell damage, might be better mapped by D. A progressive increase of D throughout treatment was significantly predictive for LRC (57, 58, 92).

Positron emission tomography

The prognostic FDG-PET studies reported that a high ΔSUV_{max} (reduction; i.e. a low intratreatment absolute SUV_{max}) at week 2, 3 or 4 intratreatment was associated with LRC and OS (62, 65, 66, 69, 71). A lower absolute SUV_{max} than the optimal cut-off (range, 4.25-5.05) resulted in a hazard ratio of 2.32 (95%CI 1.39-3.87) for LRC. High intratreatment FDG-uptake is associated with increased LRF, which can be explained by radioresistant tumour parts. Tumours with intrinsic aggressiveness and high risk of distant microscopic disease are likely to have high proliferation rates prior to treatment and this rate will remain high even after a few weeks of RT. After a few weeks of effective RT, non-cancer stem cells (CSC) with limited proliferations may have been killed resulting in a larger proportion of CSCs within the tumour. In such cases, highly proliferative CSCs are likely to demonstrate high FDG-uptake and be associated with a high incidence LRF (88). Intratreatment uptake

reduction could reflect killing more radiosensitive component of tumour, which leaves the residual metabolic burden a more useful predictor. Furthermore, accumulation of FDG in peritumoral tissue could be caused by radiation-induced inflammation (62). In order to predict overall survival, a higher SUV_{max} than the optimal cut-off (range, 4.03-4.25) resulted in a pooled hazard ratio of 2.43 (95%CI, 1.45-4.05).

In primary tumor, high ΔTLG (decrease) during treatment was found predictive for LRC and OS (69, 74, 75, 88) and was described as a better reflector of the metabolic burden, compared to the highest intensity in a single voxel as measured by SUV_{max} (69). In LN metastasis, a >50% reduction in total LN TLG and MTV was the best biomarker and significantly correlated with locoregional DFS and OS (74).

A small area of hypoxia (i.e. high TBR_{max}) early during (chemo)radiotherapy was associated with a good long-term outcome due to an improved perfusion leading to better tracer delivery and wash-out of unbound tracer. Despite partial reperfusion of some regions, allowing for faster delivery and wash-out, hypoxia may still remain in regions that are at a distance from the perfused vessels (77).

^{18}F-FLT PET is expected to assess the therapeutic response much earlier than ^{18}F-FDG PET (42, 63) and may thus aid in patient-tailored treatment during an early phase of therapy (31, 42, 63, 68, 77).

Optimal imaging timing for long-term outcome prediction
The optimal timing of CT-perfusion for predicting long-term patient outcome depends on changes of blood flow and capillary permeability occurring during the first 2-3 weeks of (chemo)radiotherapy (36, 37). Increase of blood volume and reduction of poorly perfused subvolumes at week 2 predicted LRC (16, 44). High ADC increase of >10.8-15% at 2 weeks intratreatment was strongly predictive for LRC (42, 49, 50, 54, 56, 86, 87). After week 2, a reduction of Ktrans was associated with LRC and better overall survival (41), possibly due to cytotoxic effects on endothelial cells and/or overcompensating of this initial effect, which may lead to thrombosis/occlusion and/or destruction of small vessels resulting in decreased blood-flow (37).

Due to radiotherapy effects (i.e. mucositis), mainly after 2 weeks, response evaluation is advised to be done before the glucose-avid inflammatory effects would have started to dominate (66). FDG-PET at least 10 days after start of chemotherapy reduces chemotherapeutic transient FDG-PET fluctuations (metabolic flare), due to cellular stress and influx of FDG due to damaged cellular membranes (13). Low metabolic rate at week 1-3 (<5 (7, 60)), at week 3 (≤4.25 (69, 88), ≤3.25 in LN (74)) or week 4 (62, 75) were predictive of LRC. SUV_{max} reduction of >50% (65, 66) at interim PET (week 2) was associated with LRC.

In conclusion, functional imaging should be performed in the first 2-3 weeks of treatment, which could detect increased perfusion and decreased metabolic changes, which would imply LRC. In order to predict locoregional recurrence and survival accurately, FDG-PET imaging should not be performed later than 3 weeks, because of the influence of

inflammatory processes on FDG-PET uptake values.

Limitations

Even though this review provides an extensive overview of the predictive and prognostic value of intratreatment functional imaging for treatment response and long-term outcome, there are some limitations.

Firstly, some included studies were of small sample size, except for most FDG-PET and DWI studies. Furthermore, heterogeneity was found in the patient population (variability of head and neck cancer sites and (chemo)radiotherapy dose), scanning protocols, post-processing methods and statistical methods. This resulted in less comparable studies and may comprise the analysis of confounders.

Secondly, the outcome parameters were heterogeneous. It would be preferable if all response prediction and prognostic studies would report at least one uniform outcome measurement. Δ values or (percentage) change from baseline to intratreatment values are less effected by confounders of variability of single time imaging (35), which enables more accurate comparison between patients, acquisition systems and centers. On the other hand, serial imaging provides a larger logistic burden.

Thirdly, most studies examined several functional parameters without statistical correction for multiple parameters in their analysis. This may overestimate the number of significant findings. The QUADAS-2 checklist indicates a moderate (and often unclear) risk of bias. High risk bias was mainly caused by patient selection and flow and timing. The study attrition and possible confounders were often not mentioned, which might have caused bias. Most studies reported positive results, while small studies with negative results may have been regarded to be not interesting enough for publication, resulting in publication bias. In most studies, the index test (imaging) was part of the reference test during follow-up to detect presence of locoregional malignancy, which could have caused detection bias.

Conclusion

In this systematic review and meta-analysis, we conclude that intratreatment functional imaging parameters have predictive and prognostic value for treatment response, recurrence-free survival and overall survival with optimal timing of imaging at 2-3 weeks intratreatment. When performing MRI, a high pretreatment perfusion with an intratreatment decrease of blood flow or permeability (Ktrans) and high pretreatment diffusion restriction (low ADC) with a high ΔADC increase intratreatment were predictive for favorable treatment response, recurrence-free and overall survival. When performing FDG-PET, a pretreatment metabolically active tumor with a high SUV reduction of FDG-PET uptake intratreatment, was predictive for a favorable patient outcome. Future studies should focus on homogenization of techniques and acquisition methods and reporting of more uniform parameters.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66(1):7-30.
2. Arlene Forastiere WK, Andrew Trotti, David S. Idransky. Head and neck cancer. The New England journal of medicine. 2001;Vol. 345, No. 26(December 27, 2001).
3. de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol. 2009;45(11):930-5.
4. Evangelista L, Cervino AR, Chondrogiannis S, 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. Nuclear medicine communications. 2014;35(2):123-34.
5. Leusink FK, van Es RJ, de Bree R, et al. Novel diagnostic modalities for assessment of the clinically node-negative neck in oral squamous-cell carcinoma. The Lancet Oncology. 2012;13(12):e554-61.
6. Jer-Hwa Chang C-CW, Kevin Sheng-Po Yuan, Alexander T.H. Wu, Szu-Yuan Wu. Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, treatment outcomes. Oncotarget. 2017;Vol. 8, (No. 33), pp: 55600-55612(March 17, 2017).
7. Brun E, Kjellen E, Tennvall J, et al. FDG PET studies during treatment: prediction of therapy outcome in head and neck squamous cell carcinoma. Head & neck. 2002;24(2):127-35.
8. Wee JT, Anderson BO, Corry J, et al. Management of the neck after chemoradiotherapy for head and neck cancers in Asia: consensus statement from the Asian Oncology Summit 2009. The Lancet Oncology. 2009;10(11):1086-92.
9. Lee N, Schoder H, Beattie B, et al. Strategy of Using Intratreatment Hypoxia Imaging to Selectively and Safely Guide Radiation Dose De-escalation Concurrent With Chemotherapy for Locoregionally Advanced Human Papillomavirus-Related Oropharyngeal Carcinoma. International journal of radiation oncology, biology, physics. 2016;96(1):9-17.
10. Matoba M, Tuji H, Shimode Y, Kondo T, Oota K, Tonami H. Lesion regression rate based on RECIST: prediction of treatment outcome in patients with head and neck cancer treated with chemoradiotherapy compared with FDG PET-CT. Journal of radiation research. 2015;56(3):553-60.
11. Hou J, Guerrero M, Suntharalingam M, D'Souza WD. Response assessment in locally advanced head and neck cancer based on RECIST and volume measurements using cone beam CT images. Technology in cancer research & treatment. 2015;14(1):19-27.
12. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. Journal of magnetic resonance imaging : JMRI. 2010;32(1):2-16.
13. Bjurberg M, Henriksson E, Brun E, et al. Early changes in 2-deoxy-2-[18F]fluoro-D-glucose metabolism in squamous-cell carcinoma during chemotherapy in vivo and in vitro. Cancer biotherapy & radiopharmaceuticals. 2009;24(3):327-32.
14. Nyflot MJ, Kruser TJ, Traynor AM, et al. Phase 1 trial of bevacizumab with concurrent chemoradiation therapy for squamous cell carcinoma of the head and neck with exploratory functional imaging of tumor hypoxia, proliferation, and perfusion. International journal of radiation oncology, biology, physics. 2015;91(5):942-51.
15. Rana L, Sharma S, Sood S, et al. Volumetric CT perfusion assessment of treatment response in head and neck squamous cell carcinoma: Comparison of CT perfusion parameters before and after chemoradiation therapy. European Journal of Radiology Open. 2015;2:46-54.
16. Cao Y, Popovtzer A, Li D, et al. Early prediction of outcome in advanced head-and-neck cancer based on tumor blood volume alterations during therapy: a prospective study. International journal of radiation oncology, biology, physics. 2008;72(5):1287-90.

17. Zahra MA, Hollingsworth KG, Sala E, Lomas DJ, Tan LT. Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy. The Lancet Oncology. 2007;8(1):63-74.

18. Hermans R. Diffusion-weighted MRI in head and neck cancer. Current opinion in otolaryngology & head and neck surgery. 2010;18(2):72-8.

19. Quon H, Brizel DM. Predictive and prognostic role of functional imaging of head and neck squamous cell carcinomas. Seminars in radiation oncology. 2012;22(3):220-32.

20. Miles KA. Perfusion CT for the assessment of tumour vascularity: which protocol? The British journal of radiology. 2003;76 Spec No 1:S36-42.

21. Srinivasan A, Mohan S, Mukherji SK. Biologic imaging of head and neck cancer: the present and the future. AJNR American journal of neuroradiology. 2012;33(4):586-94.

22. Schmitt P, Kotas M, Tobermann A, Haase A, Flentje M. Quantitative tissue perfusion measurements in head and neck carcinoma patients before and during radiation therapy with a non-invasive MR imaging spin-labeling technique. Radiotherapy and Oncology. 2003;67(1):27-34.

23. Jansen JF, Parra C, Lu Y, Shukla-Dave A. Evaluation of Head and Neck Tumors with Functional MR Imaging. Magnetic resonance imaging clinics of North America. 2016;24(1):123-33.

24. Chikui T, Kitamoto E, Kawano S, et al. Pharmacokinetic analysis based on dynamic contrast-enhanced MRI for evaluating tumor response to preoperative therapy for oral cancer. Journal of magnetic resonance imaging : JMRI. 2012;36(3):589-97.

25. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR Imaging in the Head and Neck. Radiology. 2012;263(1):19-32.

26. D Le Bihan RT, P Douek and N Patronas Diffusion MR imaging: clinical applications. American Journal of Roentgenology. 1992;159: 591-599. 10.2214.

27. Hauser T, Essig M, Jensen A, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. Neuroradiology. 2013;55(5):527-36.

28. Noij DP, Martens RM, Marcus JT, et al. Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value. Oral oncology. 2017;68:81-91.

29. Strauss LG, Conti PS. The applications of PET in clinical oncology. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 1991;32(4):623-48; discussion 49-50.

30. Wiedenmann NE, Bucher S, Hentschel M, et al. Serial [18F]-fluoromisonidazole PET during radiochemotherapy for locally advanced head and neck cancer and its correlation with outcome. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2015;117(1):113-7.

31. Hoeben BA, Troost EG, Span PN, et al. 18F-FLT PET during radiotherapy or chemoradiotherapy in head and neck squamous cell carcinoma is an early predictor of outcome. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2013;54(4):532-40.

32. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.

33. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.

34. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing Bias in Studies of Prognostic Factors. Annals of Internal Medicine. 2013;158(4):280-6.

35. Quarles van Ufford HM, van Tinteren H, Stroobants SG, Riphagen, II, Hoekstra OS. Added value of baseline 18F-FDG uptake in serial 18F-FDG PET for evaluation of response of solid extracerebral tumors to systemic cytotoxic neoadjuvant treatment: a meta-analysis. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2010;51(10):1507-16.

36. Truong MT, Saito N, Ozonoff A, et al. Prediction of locoregional control in head and neck squamous cell carcinoma with serial CT perfusion during radiotherapy. AJNR American journal of neuroradiology. 2011;32(7):1195-201.

37. Abramyuk A, Hietschold V, Appold S, von Kummer R, Abolmaali N. Radiochemotherapy-induced changes of tumour vascularity and blood supply estimated by dynamic contrast-enhanced CT and fractal analysis in malignant head and neck tumours. The British journal of radiology. 2015;88(1045):20140412.

38. Ursino S, Faggioni L, Guidoccio F, et al. Role of perfusion CT in the evaluation of functional primary tumour response after radiochemotherapy in head and neck cancer: preliminary findings. The British journal of radiology. 2016;89(1065):20151070.

39. Bhatia KS, King AD, Yu KH, et al. Does primary tumour volumetry performed early in the course of definitive concomitant chemoradiotherapy for head and neck squamous cell carcinoma improve prediction of primary site outcome? The British journal of radiology. 2010;83(995):964-70.

40. King AD, Yeung DK, Yu KH, et al. Pretreatment and early intratreatment prediction of clinicopathologic response of head and neck cancer to chemoradiotherapy using 1H-MRS. Journal of magnetic resonance imaging : JMRI. 2010;32(1):199-203.

41. Baer AH, Hoff BA, Srinivasan A, Galban CJ, Mukherji SK. Feasibility analysis of the parametric response map as an early predictor of treatment efficacy in head and neck cancer. AJNR American journal of neuroradiology. 2015;36(4):757-62.

42. Dirix P, Vandecaveye V, De Keyzer F, Stroobants S, Hermans R, Nuyts S. 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. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2009;50(7):1020-7.

43. Jansen JF, Lu Y, Gupta G, et al. Texture analysis on parametric maps derived from dynamic contrast-enhanced magnetic resonance imaging in head and neck cancer. World journal of radiology. 2016;8(1):90-7.

44. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. Medical physics. 2012;39(8):5277-85.

45. Yoo DS, Kirkpatrick JP, Craciunescu O, et al. Prospective trial of synchronous bevacizumab, erlotinib, and concurrent chemoradiation in locally advanced head and neck cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18(5):1404-14.

46. Galban CJ, Mukherji SK, Chenevert TL, et al. A feasibility study of parametric response map analysis of diffusion-weighted magnetic resonance imaging scans of head and neck cancer patients for providing early detection of therapeutic efficacy. Translational oncology. 2009;2(3):184-90.

47. Hoang JK, Choudhury KR, Chang J, Craciunescu OI, Yoo DS, Brizel DM. Diffusion-weighted imaging for head and neck squamous cell carcinoma: quantifying repeatability to understand early treatment-induced change. AJR American journal of roentgenology. 2014;203(5):1104-8.

48. Kim S, Loevner L, Quon H, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(3):986-94.

49. King AD, Mo FKF, Yu KH, et al. Squamous cell carcinoma of the head and neck: Diffusion-weighted MR imaging for prediction and monitoring of treatment response. European radiology. 2010;20(9):2213-20.

50. King AD, Chow KK, Yu KH, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology. 2013;266(2):531-8.

51. Lambrecht M, Van Herck H, De Keyzer F, et al. Redefining the target early during treatment. Can we visualize regional differences within the target volume using sequential diffusion weighted MRI? Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2014;110(2):329-34.

52. Martins EB, Chojniak R, Kowalski LP, Nicolau UR, Lima EN, Bitencourt AG. Diffusion-Weighted MRI in the Assessment of Early Treatment Response in Patients with Squamous-Cell Carcinoma of the Head and Neck: Comparison with Morphological and PET/CT Findings. PloS one. 2015;10(11):e0140009.

53. Scalco E, Marzi S, Sanguineti G, Vidiri A, Rizzo G. Characterization of cervical lymph-nodes using a multi-parametric and multi-modal approach for an early prediction of tumor response to chemo-radiotherapy. Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB). 2016;32(12):1672-80.

54. Schouten CS, de Bree R, van der Putten L, et al. Diffusion-weighted EPI- and HASTE-MRI and 18F-FDG-PET-CT early during chemoradiotherapy in advanced head and neck cancer. Quantitative imaging in medicine and surgery. 2014;4(4):239-50.

55. Tyagi N, Riaz N, Hunt M, et al. Weekly response assessment of involved lymph nodes to radiotherapy using diffusion-weighted MRI in oropharynx squamous cell carcinoma. Medical physics. 2016;43(1):137.

56. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. European radiology. 2010;20(7):1703-14.

57. Ding Y, Hazle JD, Mohamed AS, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. NMR in biomedicine. 2015;28(12):1645-54.

58. Paudyal R, Oh JH, Riaz N, 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. Journal of magnetic resonance imaging : JMRI. 2017;45(4):1013-23.

59. Wong KH, Panek R, Dunlop A, et al. Changes in multimodality functional imaging parameters early during chemoradiation predict treatment response in patients with locally advanced head and neck cancer. European journal of nuclear medicine and molecular imaging. 2017.

60. Brun E, Ohlsson T, Erlandsson K, et al. Early prediction of treatment outcome in head and neck cancer with 2- 18FDG PET. Acta Oncologica. 1997;36(7):741-7.

61. Lowe VJ, Dunphy FR, Varvares M, et al. Evaluation of chemotherapy response in patients with advanced head and neck cancer using [F-18]fluorodeoxyglucose positron emission tomography. Head & neck. 1997;19(8):666-74.

62. Farrag A, Ceulemans G, Voordeckers M, Everaert H, Storme G. Can 18F-FDG-PET response during radiotherapy be used as a predictive factor for the outcome of head and neck cancer patients? Nuclear medicine communications. 2010;31(6):495-501.

63. Troost EG, Bussink J, Hoffmann AL, Boerman OC, Oyen WJ, Kaanders JH. 18F-FLT PET/CT for early response monitoring and dose escalation in oropharyngeal tumors. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2010;51(6):866-74.

64. Ceulemans G, Voordeckers M, Farrag A, Verdries D, Storme G, Everaert H. Can 18-FDG-PET during radiotherapy replace post-therapy scanning for detection/demonstration of tumor response in head-and-neck cancer? International journal of radiation oncology, biology, physics. 2011;81(4):938-42.

65. Hentschel M, Appold S, Schreiber A, et al. 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. European journal of nuclear medicine and molecular imaging. 2011;38(7):1203-11.

66. Castaldi P, Rufini V, Bussu F, et al. Can “early” and “late”18F-FDG PET-CT be used as prognostic factors for the clinical outcome of patients with locally advanced head and neck cancer treated with radio-chemotherapy? Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2012;103(1):63-8.

67. Kishino T, Hoshikawa H, Nishiyama Y, Yamamoto Y, Mori N. Usefulness of 3'-deoxy-3'-18F-fluorothymidine PET for predicting early response to chemoradiotherapy in head and neck cancer. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2012;53(10):1521-7.

68. Hoshikawa H, Mori T, Kishino T, et al. Changes in (18)F-fluorothymidine and (18)F-fluorodeoxyglucose positron emission tomography imaging in patients with head and neck cancer treated with chemoradiotherapy. Annals of nuclear medicine. 2013;27(4):363-70.

69. Min M, Lin P, Lee MT, et al. Prognostic role of metabolic parameters of (18)F-FDG PET-CT scan performed during radiation therapy in locally advanced head and neck squamous cell carcinoma. European journal of nuclear medicine and molecular imaging. 2015;42(13):1984-94.

70. Min M, Lin P, Lee M, et al. 18F-FDG PET-CT performed before and during radiation therapy of head and neck squamous cell carcinoma: Are they independent or complementary to each other? J Med Imag Radiat On. 2016;60(3):433-40.

71. Min M, Lin P, Lee M, et al. Prognostic Value of 2-[18F] Fluoro-2-deoxy-D-glucose Positron Emission Tomography-Computed Tomography Scan Carried out During and After Radiation Therapy for Head and Neck Cancer Using Visual Therapy Response Interpretation Criteria. Clinical Oncology. 2016;28(6):393-401.

72. Zegers CML, van Elmpt W, Szardenings K, et al. Repeatability of hypoxia PET imaging using [18F]HX4 in lung and head and neck cancer patients: a prospective multicenter trial. European journal of nuclear medicine and molecular imaging. 2015;42(12):1840-9.

73. Chen SW, Hsieh TC, Yen KY, et al. Interim FDG PET/CT for predicting the outcome in patients with head and neck cancer. The Laryngoscope. 2014;124(12):2732-8.

74. Lin P, Min M, Lee M, et al. Nodal parameters of FDG PET/CT performed during radiotherapy for locally advanced mucosal primary head and neck squamous cell carcinoma can predict treatment outcomes: SUVmean and response rate are useful imaging biomarkers. European journal of nuclear medicine and molecular imaging. 2016;44(5):801-11.

75. Pollom EL, Song J, Durkee BY, et al. Prognostic value of midtreatment FDG-PET in oropharyngeal cancer. Head & neck. 2016;38(10):1472-8.

76. Thorwarth D, Eschmann SM, Paulsen F, Alber M. A model of reoxygenation dynamics of head-and-neck tumors based on serial 18F-fluoromisonidazole positron emission tomography investigations. International journal of radiation oncology, biology, physics. 2007;68(2):515-21.

77. Zips D, Zophel K, Abolmaali N, et al. Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2012;105(1):21-8.

78. Zschaeck S, Haase R, Abolmaali N, et al. Spatial distribution of FMISO in head and neck squamous cell carcinomas during radio-chemotherapy and its correlation to pattern of failure. Acta oncologica (Stockholm, Sweden). 2015;54(9):1355-63.

79. Lee YH, Song JH, Choi HS, et al. Using primary tumor volumetry to predict treatment outcome for patients with oropharyngeal cancer who were treated with definitive chemoradiotherapy. Asia-Pacific journal of clinical oncology. 2018;14(2):e21-e8.

80. Grkovski M, Lee NY, Schoder H, et al. Monitoring early response to chemoradiotherapy with (18)F-FMISO dynamic PET in head and neck cancer. European journal of nuclear medicine and molecular imaging. 2017;44(10):1682-91.

81. Lock S, Perrin R, Seidlitz A, et al. Residual tumour hypoxia in head-and-neck cancer patients undergoing primary radiochemotherapy, final results of a prospective trial on repeat FMISO-PET imaging. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2017;124(3):533-40.

82. Bhatia KS, King AD, Yeung DK, et al. Can diffusion-weighted imaging distinguish between normal and squamous cell carcinoma of the palatine tonsil? The British journal of radiology. 2010;83(993):753-8.

83. Hoeben BA, Troost EG, Bussink J, van Herpen CM, Oyen WJ, Kaanders JH. 18F-FLT PET changes during radiotherapy combined with cetuximab in head and neck squamous cell carcinoma patients. Nuklearmedizin Nuclear medicine. 2014;53(2):60-6.

84. King AD, Mo FK, Yu KH, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. European radiology. 2010;20(9):2213-20.

85. Brun E, Ohlsson T, Erlandsson K, et al. Early prediction of treatment outcome in head and neck cancer with 2-18FDG PET. Acta oncologica (Stockholm, Sweden). 1997;36(7):741-7.

86. Lambrecht M, Van Calster B, Vandecaveye V, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2014;110(3):429-34.

87. Matoba M, Tuji H, Shimode Y, et al. Fractional change in apparent diffusion coefficient as an imaging biomarker for predicting treatment response in head and neck cancer treated with chemoradiotherapy. AJNR American journal of neuroradiology. 2014;35(2):379-85.

88. Min M, Lin P, Lee M, et al. 18F-FDG PET-CT performed before and during radiation therapy of head and neck squamous cell carcinoma: Are they independent or complementary to each other? Journal of medical imaging and radiation oncology. 2016;60(3):433-40.

89. Folkman J, Camphausen K. Cancer. What does radiotherapy do to endothelial cells? Science. 2001;293(5528):227-8.

90. Mitsuko Furuya MN, Yoshitoshi Kasuya, Sadao Kimura, Hiroshi Ishikura. Pathophysiology of tumor neovascularization. Vascular health and risk management. 2005;1(4) 277–290.

91. Cheung PK, Chin RY, Eslick GD. Detecting Residual/Recurrent Head Neck Squamous Cell Carcinomas Using PET or PET/CT: Systematic Review and Meta-analysis. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery. 2016;154(3):421-32.

92. Marzi S, Piludu F, Sanguineti G, et al. The prediction of the treatment response of cervical nodes using intravoxel incoherent motion diffusion-weighted imaging. European journal of radiology. 2017;92:93-102.

93. Rana L, Sharma S, Sood S, et al. Volumetric CT perfusion assessment of treatment response in head and neck squamous cell carcinoma: Comparison of CT perfusion parameters before and after chemoradiation therapy. European Journal of Radiology Open. 2015;2:46-54.

CHAPTER 3

REPRODUCIBILITY OF DIFFUSION-WEIGHTED IMAGING

CHAPTER 3.1

Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions

American Journal of Neuroradiology 2015;36:384-90

Nominated for 2014 Lucien Levy Best Research Article Award

Anna S Kolff-Gart
Petra J Pouwels
Daniel P Noij
Redina Ljumanovic
Vincent Vandecaveye
Frederik de Keyzer
Remco de Bree
Pim de Graaf
Dirk L Knol
Jonas A Castelijns

ABSTRACT

Background and purpose: Diffusion-weighted imaging (DWI) is typically performed with echo-planar imaging (EPI) sequences in single center studies. The purpose of this study was to determine the reproducibility of apparent diffusion coefficient (ADC) values in the head and neck region in healthy subjects. In addition, reproducibility of ADC values in different tissues was assessed to identify the most suitable reference tissue.

Materials and Methods: We prospectively studied seven healthy subjects, with EPI and turbo spin-echo (TSE) sequences, on five MRI systems, at three time points in two institutes. ADC maps of EPI (with 2 b-values and 6 b-values) and TSE sequences were compared. Mean ADC values for different tissues (submandibular gland, sternocleidomastoid muscle, spinal cord, subdigastric lymph node and tonsil) were used to evaluate intra- and intersubject-, intersystem- and intersequence-variability using a linear mixed model.

Results: On 97% of images an ROI could be placed on the spinal cord, compared to 87% in the tonsil. ADC values derived from EPI-DWI-2b and calculated EPI-DWI-2b extracted from EPI-DWI-6b did not differ significantly. Standard error of ADC measurement (SEM) was the smallest for tonsil and spinal cord (SEM=$151.2 \cdot 10^{-6}$ mm/s^2 and $190.1 \cdot 10^{-6}$ mm/s^2, respectively). Intersystem difference for mean ADC values and the influence of MRI system on ADC values between the subjects were statistically significant (P<0.001). The mean difference between examinations was negligible (i.e. less than $10 \cdot 10^{-6}$ mm/s^2).

Conclusions: In this study, the spinal cord was the most appropriate reference tissue and EPI-DWI-6b was the most reproducible sequence. ADC values are more precise if subjects are measured on the same MRI system and with the same sequence. ADC values differ significantly between MRI systems and sequences.

INTRODUCTION

Almost 3% of all malignancies are head and neck cancer (HNC), ninety-five percent of which are squamous cell carcinomas (1). Magnetic resonance imaging (MRI) is one of the imaging modalities used in the workup of HNC patients (2). Diffusion-weighted imaging (DWI) is an MRI technique by which diffusion properties of water can be quantified as an apparent diffusion coefficient (ADC) (3). Changes in ADC are inversely correlated with changes in cellularity (4). In tissues with high cellularity, diffusion of extracellular water in particular is limited by cell membranes, which gives low ADC values. In tissues with low cellularity, when diffusion is facilitated (e.g. in edematous or necrotic tissue), ADC values are high.

Indications for DWI in HNC include tissue characterization of primary tumors and nodal metastases, prediction and monitoring of treatment response after (chemo)radiotherapy, and differentiation between radiation changes and residual or recurrent disease (5).

Neither the optimal DWI sequence for assessment of the head and neck region nor its reproducibility has been clearly established. Diffusion-weighted imaging can be performed with either echo-planar imaging (EPI) or turbo spin-echo (TSE) sequences, of which the EPI sequence is most commonly used in the head and neck area (6, 7). On EPI-DWI more malignant lesions can be detected and lesion delineation is facilitated. However, the interobserver agreement of ADC values is reported to be higher on TSE-DWI, probably due to the frequent occurrence of artifacts and geometric distortions in EPI-DWI (8).

Currently the use of DWI in head and neck imaging is mostly confined to research protocols and advanced academic centers. Before DWI can be used in multicenter studies, its reproducibility across different centers and MRI systems should be validated (9). Apparent diffusion coefficient values may be affected by the selected technique and MRI system, e.g. due to differences in gradient systems, coils, pulse sequence designs, imaging parameters, and artifacts related to susceptibility effects or eddy currents (10). Information on variance is needed (11). Furthermore, the use of reference tissues might help ascertain variability between different MRI systems and could potentially help to correct for differences in ADC values between MRI systems.

The purpose of this prospective study was to determine the reproducibility of ADC values in the head and neck region obtained from DWI based on both EPI and TSE sequences in repeated measurement on different MRI systems in healthy subjects. In addition, we assessed which tissue shows the highest reproducibility in ADC values, such that it could function as a reference tissue in future studies.

Table 1 *Specification of DWI sequences obtained at each MRI system: '+' indicates the sequence is performed; '-' sequence not performed/ not available; 'o' data extracted from 6b*

	I	II	III	IV	V
Manufacturer	Siemens	Siemens	GE	Siemens	Philips
Model	Avanto	Sonata	Signa HDxt	Aera	Achieva
Center	Amsterdam	Amsterdam	Amsterdam	Leuven	Leuven
Field strength	1.5T	1.5T	1.5T	1.5T	3.0T
Conventional T2	+	+	+	+	+
EPI-DWI-2b	+	+	+	o	o
EPI-DWI-6b	+	+	-	+	+
TSE-DWI-2b	+	+	+	-	-

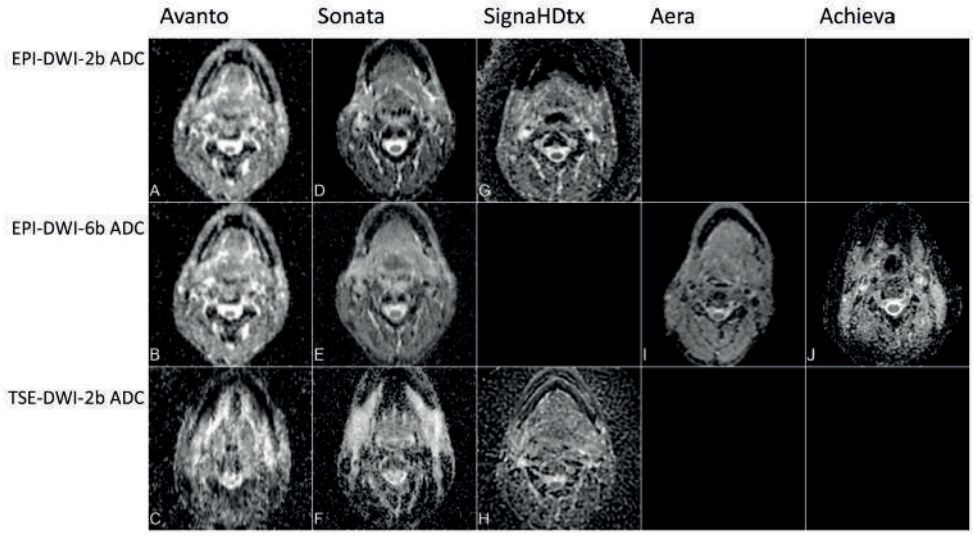


Figure 1 *ADC maps of all DWI sequences on all MRI systems. On Signa HDtx EPI-DWI-6b was not performed. On Aera and Achieva EPI-DWI-2b ADC was extracted from EPI-DWI-6b and TSE-DWI-2b was not performed.*

MATERIALS AND METHODS

Subjects

The study population consisted of seven healthy subjects, five men and two women (age range, 27-54 years; median age, 30 years). The subjects were examined in two institutions: 1) VU University Medical Center; 2) University Hospitals Leuven. All examinations were performed in 2011, after obtaining approval from the relevant institutional review boards and written informed consent from all subjects. The following MRI systems were used: I) Siemens Sonata, II) Siemens Avanto and III) Siemens Aera (Erlangen, Germany), and IV) GE Signa Excite HDxt (Milwaukee, WI, USA), all at 1.5 T, as well as V) Philips Achieva (Best,

the Netherlands) at 3 T. All examinations were performed with a dedicated head and neck radiofrequency coil in combination with a spine-array coil.

All subjects were examined on all MRI systems, at three time points per MRI system, yielding a total of 15 sessions per subject. Two examinations were performed on the same day (between examinations the subject was removed from the MRI system) and one examination at least one month later.

Imaging protocol
Each session included an anatomical T2-weighted sequence through the neck and up to three DWI sequences, with acquisition parameters as similar as possible among the MRI systems. Due to technical limitations, no EPI-DWI-6b was performed on one MRI system (Signa HDxt), and on two MRI systems (Aera and Achieva) no separate EPI-DWI-2b were performed. The sequences used per MRI system are shown in Table 1 and Figure 1.

All imaging was acquired with 21 transverse sections centered on the epiglottis (section thickness, 4 mm; intersection gap, 0.4 mm). The imaging protocol consisted of both conventional T2-weighted (TR/TE = at least 3700/ 90-110 ms, in-plane pixel size of 0.95x 0.95 mm) and EPI-DWI (TR/TE = at least 4300/ 59-98 ms, in-plane pixel size of 1.5-1.9x1.5-1.9 mm, interpolated in-plane pixel size of 0.75-0.95 mm) or TSE-DWI (TR/TE = 900-3000/ 84-113 ms, in-plane pixel size of 1.3x1.3 mm). B-values for the three DWI series were as follows: (1) EPI-DWI obtained with 6 b-values of 0, 50, 100, 500, 750 and 1000 s/mm^2, (2) EPI-DWI obtained with 2 b-values of 0 and 1000 s/mm^2 and (3) TSE-DWI obtained with 2 b-values also of 0 and 1000 s/mm^2.

Data analysis
All ADC maps were calculated online or off-line using MRI system software of the respective vendor. EPI-DWI-6b was analyzed assuming a mono-exponential ADC. Apparent diffusion coefficient values for EPI-DWI-2b on the two MRI systems without EPI-DWI-2b were derived from EPI-DWI-6b by selecting only the images acquired using b=0 s/mm^2 and b=1000 s/mm^2 (12). This 'generated' EPI-DWI-2b data was compared with the other EPI-DWI-2b data. Data was transferred to a DICOM-viewer (Centricity Radiology RA 650, version 6.1: GE medical System Milwaukee WI, USA).

For each examination, one ellipse-shaped region of interest (ROI) per tissue was manually drawn on the slice which contained the bulk of the tissue of interest by one observer (RL) with seven years of experience in head and neck imaging. For each of the following five tissues in the head and neck ADC values were determined: 1) submandibular gland, 2) sternocleidomastoid muscle, 3) spinal cord, 4) subdigastric lymph node and 5) tonsil. For the selection of a subdigastric lymph node, either the left or right one was selected, consistently within each subject. The size (range, 20-50 mm^2) and position of the ROI were identified on T2-weighted images. ROIs were drawn on corresponding b0-images by visual comparison with the anatomical T2WI. ROIs drawn on the b0-images were copied to the corresponding ADC maps.

Statistical Analysis

Firstly, it was determined whether ADC values of the EPI-DWI-2b sequences can be substituted by ADC maps obtained by selecting only the b=0 and b=1000 s/mm²-images from the EPI-DWI-6b (for MRI systems IV and V), because they are theoretically equivalent. We used a linear mixed model, with fixed effects for subjects, MRI systems, sequences, and a 'MRI system × sequences'-interaction' (13, 14). Random effects were all possible interactions with the subjects (Appendix A). This was tested using data from MRI system I and II, being the only MRI systems on which both sequences had been performed.

For the main variance analysis, five MRI systems and three sequences were compared by using the same statistical modeling approach and reasoning as used for the linear mixed model and by incorporating tissues as fixed effects (Appendix A). All three examinations of each subject were assumed to be pure replications, and were nested within 'subject × MRI system'-combinations. Models with sequence-specific error variances were compared using Akaike's Information Criterion (15). The standard error of measurement (SEM) for ADC values per tissue was expressed as the square root of the sum of residual variance (σ^2_E) and the variance expressing the interaction between replication and subjects at different MRI systems

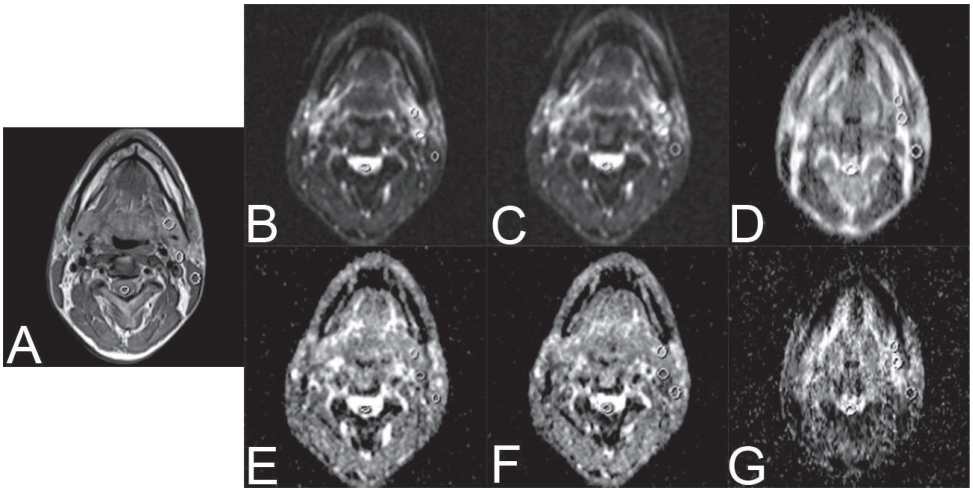


Figure 2 *Example of regions of interest (ROIs) drawn on T2 (A), EPI-DWI-2b b0 (B), EPI-DWI-6b b0 (C), TSE-DWI-2b b0 (D), EPI-DWI-2b ADC (E), EPI-DWI-6b ADC (F) and TSE-DWI-2b ADC. The tonsils are not visible at this level. Images were acquired with Siemens Avanto.*

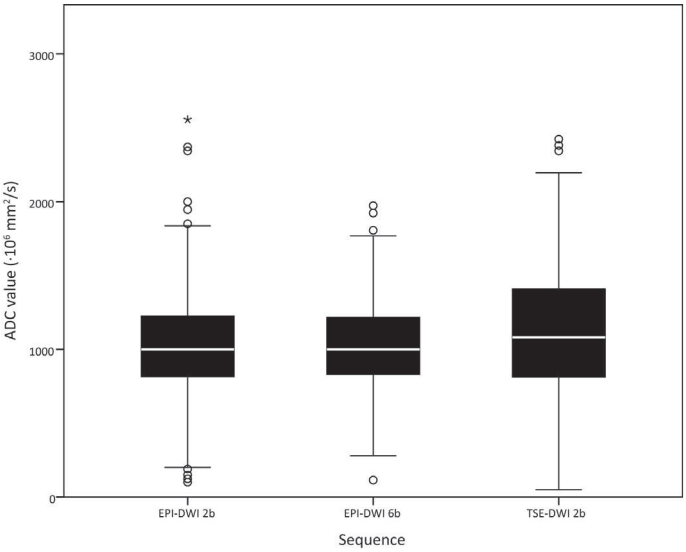


Figure 3 *Box-plots showing the distribution of ADC values (·10^{-6} mm/s^{2}) per sequence. The points are outliers (i.e.>1.5 IQR away from the 25th or 75th percentile). The asterisk is an extreme outlier (i.e.>3 IQR away from the 25th or 75th percentile).*

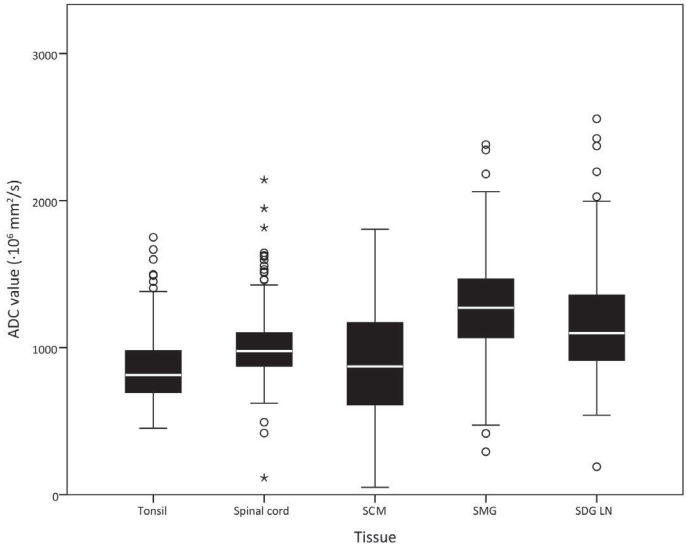


Figure 4 *Box-plots showing the distribution of ADC values (·10^{-6} mm/s^{2}) per tissue. The points are outliers (i.e.>1.5 IQR away from the 25th or 75th percentile). The asterisks are extreme outlier (i.e.>3 IQR away from the 25th or 75th percentile).*

($\sigma^2_{R:IM}$), sequences ($\sigma^2_{SR:IM}$), and tissues ($\sigma^2_{TR:IM}$), (Appendix A):

$$SEM = \sqrt{\sigma^2_{R:IM} + \sigma^2_{SR:IM} + \sigma^2_{TR:IM} + \sigma^2_E}$$

Differences in mean ADC values for all systems, and the between subjects- effects, were tested using a Levene's test of equality of error variances, and α-level of 0.05 was used for statistical significance (16). All missing data or images with poor quality of DWI were specifically labeled for statistical analysis. Box plots were created using SPSS (version 20.0; Chicago, IL, USA). All other analyses were performed with SAS version 9.2 (Proc NLMIXED; SAS Inc, Cary, NC, USA).

RESULTS

Diffusion-weighted imaging

All subjects underwent multiple DWI sessions, with multiple sequences, on all MRI systems. For MRI system III EPI-DWI-6b was unavailable; for MRI systems IV and V, ADC maps for EPI-DWI-2b were constructed using only the b=0 and b=1000 s/mm^2 images from the EPI-DWI-6b, yielding a total of 12 DWI sequences per subject (Table 1). Two subjects underwent two instead of three replications. One subject had prior bilateral tonsillectomies. Therefore, the maximum number of possible ROIs was 1104. For a detailed overview of the number of possible ROIs we refer to Appendix B. Further elimination was due to technically failed images and image specific poor quality, and in 37 cases it was impossible to place a ROI: in 95% of tissues ROI placement was possible on TSE-DWI-2b, on EPI-DWI-2b in 96% and on EPI-DWI-6b in 97% (Table 2). Examples of ADC maps on different MRI systems and sequences are shown in Figure 1. An example of drawn ROIs is shown in Figure 2.

When combining the results of the three DWI sequences, ROI placement was possible in 96% of tissues (Table 2). However, in only 87% (range, 83-90%) of images a ROI could be placed on the tonsil. In the other regions ROIs could be placed in 97% to 98% of the cases. These data indicate that the tonsil is probably not a good reference tissue for future evaluations.

A variance component analysis was carried out for MRI system I and II to test for potential differences between ADC values derived from the EPI-DWI-2b sequence and the calculated EPI-DWI-2b extracted from EPI-DWI-6b (Table 3). The lowest bias was found in the subdigastric lymph node ($0.7 \cdot 10^{-6}$mm^2/s) and the highest bias was found in the tonsil ($-23.2 \cdot 10^{-6}$mm^2/s). Furthermore, this analysis showed a small range of limits of agreement (LoA) (range, $-307.0 \cdot 10^{-6}$ mm^2/s; $302.4 \cdot 10^{-6}$ mm^2/s) for all tissues combined. This implies that both ADC values are not significantly different. Therefore, we used calculated EPI-DWI-2b-ADC values extracted from EPI-DWI-6b on systems if EPI-DWI-2b was not available

Table 2 *Number of placed ROIs per tissue and per sequence. The percentage of the maximum number of possible ROIsis displayed in parentheses. Elimination is due to poor image quality or artifacts.*

	Tonsil	Spinal cord	SCM	SMG	SDG LN	Total
EPI-DWI-2b, n (%)	58 (90)	74 (96)	75 (97)	76 (99)	74 (99)	357 (96)
EPI-DWI-6b, n (%)	57 (89)	76 (99)	76 (99)	76 (99)	74 (99)	359 (97)
TSE-DWI-2b, n (%)	50 (83)	76 (97)	77 (97)	76 (96)	72 (97)	351 (95)
Total	165 (87)	226 (97)	228 (98)	228 (98)	220 (98)	1067 (96)

Table 3 *Comparison of ADC values derived from calculated EPI-DWI-2b extracted from EPI-DWI-6b and EPI-DWI-2b for MRI system I and II. In parentheses the bias is displayed as a percentage of the mean ADC from EPI-DWI-2b for MRI system I and II*

	Bias ($\cdot 10^{-6}$ mm^2/s)	LoA ($\cdot 10^{-6}$ mm^2/s)
Tonsil	-23.2 (-2.9)	-307.0; 260.7
Spinal cord	-12.7 (-1.2)	-296.6; 271.1
SCM	10.8 (1.1)	-273.1; 294.6
SMG	18.6 (1.3)	-265.3; 302.4
SDG LN	0.7 (0.1)	-283.2; 284.5

Table 4 *Actual ADC values ($\cdot 10^{-6}$ mm^2/s) and standard error of ADC measurement ($\cdot 10^{-6}$ mm^2/s) for all subjects and MRI systems*

	EPI-DWI-2b		EPI-DWI-6b		TSE-DWI-2b		Total per tissue	
	median (IQR)	SEM	median (IQR)	SEM	median (IQR)	SEM	median (IQR)	SEM
Tonsil	791 (675; 876)	134.2	746 (674; 857)	119.6	1089 (839; 1272)	203.0	813 (694;980)	151.2
Spinal cord	950 (868; 1053)	194.4	950 (865; 1016)	170.6	1076 (908; 1303)	204.2	976 (873;1100)	190.1
SCM	990 (782; 1276)	221.6	1084 (810; 1317)	210.5	534 (286; 822)	285.0	872 (611;1171)	237.8
SMG	1257 (1090; 1462)	247.0	1233 (1066; 1362)	222.5	1392 (1030; 1638)	431.2	1271 (1066;1468)	295.5
SDG LN	1042 (809; 1211)	307.9	1027 (870; 1174)	242.9	1393 (1124; 1709)	322.0	1099 (910; 1360)	291.0
Total per sequence	1000 (815; 1226)	216.6	1000 (830; 1217)	190.3	1082 (812; 1414)	284.5	1020 (819; 1273)	238.3

for further analysis. The intersystem difference between the MRI systems, with mean ADC values as dependent variable was statistically significant (P<0.001). The influence of the sequence, the MRI system and the interaction between these two parameters was significant (P=0.011*)*. The influence of MRI system on the ADC values between the subjects (P<0.001) was also significant.

Main variance analysis

For the main analysis the actual median ADC values and the results of the main variance components analysis per sequence and per tissue are shown in Table 4. The three used DWI sequences showed some differences (Figure 3); the EPI-DWI-6b sequence demonstrated

the smallest interquartile range (IQR) values (830-1217·10^{-6} mm^2/s) and lowest SEM (190.3) in ADC for all tissues. The TSE-DWI-2b sequence demonstrated the broadest IQR (812-1414·10^{-6} mm^2/s) and largest SEM (284.5·10^{-6} mm^2/s) for all tissues, while EPI-DWI-2b and EPI-DWI-6b showed a more narrow IQR (815-1226·10^{-6} mm^2/s and 830-1217·10^{-6} mm^2/s, respectively) and smaller SEM (216·10^{-6} mm^2/s and 190.3·10^{-6} mm^2/s, respectively). Therefore, measurements on EPI-DWI-2b and EPI-DWI-6b are more precise. Note that with TSE-DWI-2b the lowest number of ADCs was available for analysis (95%, see Table 2).

The spinal cord and tonsil show the smallest IQR (873-1100·10^{-6} mm/s^2 and 694-980·10^{-6} mm/s^2, respectively) and lowest SEM (151.2·10^{-6} mm/s^2 and 190.1·10^{-6} mm/s^2, respectively) (Table 4, Figure 4). These tissues have the lowest SEM, indicating that ADC measurements in these tissues are the most precise and the best reproducible. However, even though the SEM is low for the spinal cord (SEM=190·10^{-6} mm^2/s), with a median ADC of 976·10^{-6} mm^2/s, the range of normal values is still broad (IQR=873-1100·10^{-6} mm^2/s).

Variance caused by time is limited (Figure 5). The mean difference in ADC values of the second examination compared to the first, which were on the same day, was 6·10^{-6} mm/s^2 (standard deviation (SD)=310·10^{-6} mm/s^2). For the third examination, one month after the first, the mean difference in ADC values was -5·10^{-6} mm/s^2 (SD=310·10^{-6} mm/s^2) compared to the first measurement.

DISCUSSION

Before quantitative DWI can be applied in a multicenter study, knowledge is required about reproducibility of ADC values within a subject, between different MRI systems and between sequences (10). This study is a first step to obtain that knowledge.

In this study we assessed the reproducibility of ADC values for different DWI sequences, MRI systems, and different tissues in the head and neck. As expected, the variance in ADC values per subject per tissue is the smallest if the subject is measured on the same MRI system with the same sequence. The EPI-DWI-6b sequence showed the best reproducibility for all compared tissues, although it must be stressed that this sequence was not available on all MRI systems. The EPI-DWI-2b sequence had a slightly lower reproducibility than the EPI-DWI-6b. Advantages of EPI-DWI-2b are a shorter acquisition time and that the sequence is more widely clinically available. Apparent diffusion coefficient measurements in the spinal cord and tonsil were the most precise and reproducible. Since the spinal cord is almost always present in the field of view during a head and neck study, this tissue can potentially be used as a reference. It also has the advantage that it is rarely affected by malignancy; this in contrast to the tonsils, which are absent in case of tonsillectomy and frequently prove to be the location of an initially unknown primary tumor (17). Therefore, the spinal cord seems to be the best suitable to serve as reference tissue.

Diffusion-weighted imaging is frequently used in oncologic imaging (18, 19). Previous studies have shown the potential of DWI in diagnosing malignancies in the head and neck area, response prediction and differentiation between treatment-induced tissue changes,

and residual or recurrent disease (6, 20, 21). However, these studies were conducted in a single institution, without variance in MRI system and protocol. It is well known that quantitative MRI parameters (e.g. ADC) can differ substantially between MRI systems and imaging protocols (22), which is also confirmed in the present study. We obtained three examinations on five MRI systems on healthy subjects. This study validates that differences in ADC values are statistically significant for sequences, MRI systems and also for the interaction between MRI systems and sequences.

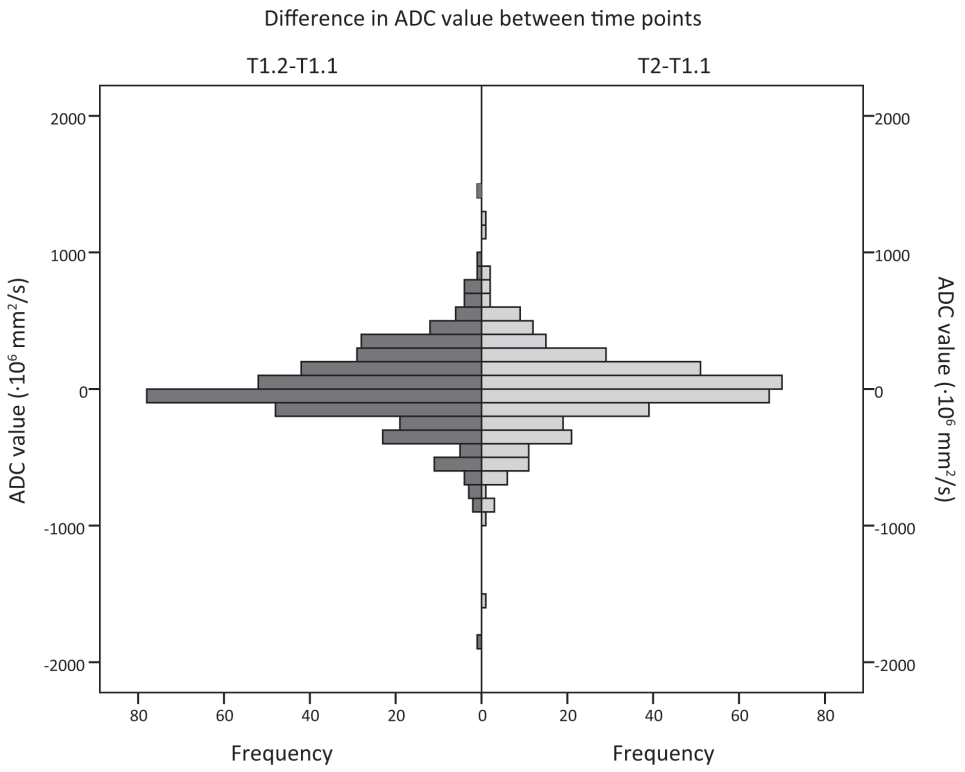


Figure 5 *Histograms showing the difference in ADC values (·10^{-6} mm/s^2) compared to the first scan (T1.1). T1.1 and T1.2 were on the same day. T2 was one month later. T1.2 – T1.1, mean=6·10^{-6} mm/s^2; standard deviation=310·10^{-6} mm/s^2. T2 – T1.1, mean=-5·10^{-6} mm/s^2; standard deviation=310·10^{-6} mm/s^2).*

Verhappen and colleagues found TSE-DWI to be more reproducible between observers than EPI-DWI in a single-center, single-system study on primary tumors and lymph nodes of 12 patients with HNC (8). In the current multicenter, multi-system study, ADC values derived with the EPI-DWI-6b sequence turned out to be the most reproducible in healthy subjects over time, followed by EPI-DWI-2b and TSE-DWI-2b was the least reproducible sequence.

These different findings may be attributed to the included subjects: healthy volunteers in the current study, and patients with head and neck malignancies which display diffusion restriction in the study by Verhappen et al. (8). Turbo spin-echo DWI has an inherently lower signal-to-noise ratio (SNR) (23), which limits the reproducibility in healthy tissue, whereas it does not suffer from geometrical distortion and is apparently sensitive enough to detect diffusion restriction. In the current study, ROIs were drawn on b=0 s/mm^2 images in visual correlation with anatomical T2 images. Because EPI-DWI has a higher SNR, small structures (e.g. benign lymph nodes) are more easily visualized. Therefore EPI-DWI may be more appropriate for the evaluation of small structures. In a study by Vandecaveye et al. 57% of malignant lymph nodes had a diameter of less than 1 cm: therefore appropriate evaluation of small (apparently benign) structures is vital (20). Verhappen et al. drew ROIs on ADC maps of malignant tissue that showed diffusion restriction (8). Especially DWI of primary tumors in the head and neck area may suffer from geometric distortion, due to the tumor location at the air-tissue interface. In that case, geometric distortion of EPI-techniques may reduce reproducibility between observers.

There is also a difference in reproducibility among different tissues in the head and neck area. On all MRI systems and sequences, ADC values of the submandibular gland were the least precise (Table 4). An explanation for the relatively poor reproducibility might be the intrinsic physiological change in salivary glands during the time of day. In subdigastric lymph nodes ADC values have a relatively poor reproducibility (Table 4). Subdigastric lymph nodes are often too small for drawing reliable ROIs, particularly in healthy subjects. Moreover, lymph nodes are prone to changes in time (e.g. due to frequently occurring inflammation in the head and neck area). In contrast, ADC values of the spinal cord and the tonsil are the most reproducible within subjects. In 87% of the images a ROI could be drawn on the tonsils, which is lower than the other tissues (range, 97-98%) (Table 2). In healthy subjects the tonsils are sometimes too small to reliably draw a ROI on DWI. However, if the tonsils are large enough to allow for the assessment of ADC values, these values appear to be relatively stable over time within a subject resulting in relatively high precision and reproducibility of ADC measurements. The sternocleidomastoid muscle has intermediate reproducibility. Small changes in ADC values of muscle tissue may be explained by small differences in muscle tone in time.

Sasaki and colleagues previously assessed the reproducibility of ADC measurements in the brain between MRI systems, imaging protocols on different time points and in different institutions. It was concluded that there was significant variability in ADC values depending on the coil systems, imagers, vendors and field strengths (10). However, only 3 out of 10 patients were imaged more than once on the same MRI system. In our study all patients were imaged multiple times on the same MRI system, in different institutions and with a time interval of at least a month between imaging. We found significant differences between MRI systems and sequences.

The present study shows that, although physiology of healthy subjects may change over time, ADC values obtained within one person and with the same MRI system, protocol and sequences immediately after the first scan and with an interval of at least 1 month have a

low variance (i.e. the intra-subject variance is small) (Figure 5). This finding indicates that ADC measurements are reproducible and independent of time. The spinal cord and tonsil are the tissues with the lowest ADC variability when different MRI systems, protocols and sequences are used.

This study had some limitations. We only included healthy subjects with a broad age-range for whom a stable physiological status over time for all normal tissues can only be assumed. Based on Figure 5, influence of time appears to be limited with mean ADC differences being less than $10 \cdot 10^{-6}$ mm/s^2 between measurements. Stability of used MRI systems and sequences also needs to be assumed. Furthermore, the study population was too small to calculate a conversion factor for different MRI systems. In order to calculate such a conversion factor for different MRI systems, a group size of 50 subjects or more is needed (13).

Conclusion

The smallest range of ADC values can be obtained by imaging a subject on the same MRI system with an EPI-DWI with 6 b-values. Of the investigated tissues, the spinal cord shows the least variance and therefore is a candidate to serve as reference tissue in the head and neck region.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69-90.
2. de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol. 2009;45(11):930-5.
3. 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;168(2):497-505.
4. Chenevert TL, Meyer CR, Moffat BA, et al. Diffusion MRI: a new strategy for assessment of cancer therapeutic efficacy. Mol Imaging. 2002;1(4):336-43.
5. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
6. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010;20(7):1703-14.
7. Kim S, Loevner L, Quon H, et al. 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. 2009;15(3):986-94.
8. Verhappen MH, Pouwels PJ, Ljumanovic R, et al. Diffusion-weighted MR imaging in head and neck cancer: comparison between half-fourier acquired single-shot turbo spin-echo and EPI techniques. AJNR Am J Neuroradiol. 2012;33(7):1239-46.
9. Padhani AR, Liu G, Mu-Koh D, et al. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. Neoplasia. 2009;11(2):102-25.
10. Sasaki M, Yamada K, Watanabe Y, et al. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology. 2008;249(2):624-30.
11. Do RK, Chandarana H, Felker E, et al. Diagnosis of liver fibrosis and cirrhosis with diffusion-weighted imaging: value of normalized apparent diffusion coefficient using the spleen as reference organ. AJR Am J Roentgenol. 2010;195(3):671-6.
12. Burdette JH, Elster AD, Ricci PE. Calculation of apparent diffusion coefficients (ADCs) in brain using two-point and six-point methods. J Comput Assist Tomogr. 1998;22(5):792-4.
13. Carstensen B. Comparing Clinical Measurement Methods: A practical guide. Hoboken: John Wiley & Sons Ltd, 2010: 59-67.
14. Carstensen B, Simpson J, Gurrin LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. Int J Biostat. 2008;4(1):Article 16.
15. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716-23.
16. Bryman A, Cramer D. Quantitative Data Analysis with SPSS Release 12 and 13: A Guide for Social Scientist. Quantitative Data Analysis with SPSS Release 12 and 13: A Guide for Social Scientist. New York: Routledge, 2005: 147-177
17. Cianchetti M, Mancuso AA, Amdur RJ, et al. Diagnostic evaluation of squamous cell carcinoma metastatic to cervical lymph nodes from an unknown head and neck primary site. Laryngoscope. 2009;119(12):2348-54.
18. Sinkus R, Van Beers BE, Vilgrain V, DeSouza N, Waterton JC. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. Eur J Cancer. 2012;48(4):425-31.

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

20. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46.

21. Vandecaveye V, De Keyzer F, Nuyts S, 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;67(4):960-71.

22. Pagani E, Hirsch JG, Pouwels PJ, et al. Intercenter differences in diffusion tensor MRI acquisition. J Magn Reson Imaging. 2010;31(6):1458-68.

23. Reiser M, Faber SC. Recent and future advances in high-speed imaging. Eur Radiol. 1997;7 Suppl 5:166-73.

CHAPTER 4

DIAGNOSTIC CAPACITY OF DIFFUSION-WEIGHTED IMAGING

CHAPTER 4.1

Whole-body-MR imaging including DWIBS in the work-up of patients with head and neck squamous cell carcinoma: a feasibility study

European Journal of Radiology 2014; 83:1144-51

Daniel P Noij
Els J Boerhout
Indra C Pieters-van den Bos
Emile F Comans
Daniela Oprea-Lager
Rinze Reinhard
Otto S Hoekstra
Remco de Bree
Pim de Graaf
Jonas A Castelijns

ABSTRACT

Objectives: To assess the feasibility of whole-body magnetic resonance imaging (WB-MRI) including diffusion-weighted whole-body imaging with background-body-signal-suppression (DWIBS) for the evaluation of distant malignancies in head and neck squamous cell carcinoma (HNSCC); and to compare WB-MRI findings with ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) and chest-CT.

Methods: Thirty-three patients with high risk for metastatic spread (26 males; range, 48-79 years; mean age, 63 ± 7.9 years (mean ± standard deviation) years) were prospectively included with a follow-up of six months. WB-MRI protocol included short-TI inversion recovery and T1-weighted sequences in the coronal plane and half-fourier acquisition single-shot turbo spin-echo T2 and contrast-enhanced-T1-weighted sequences in the axial plane. Axial DWIBS was reformatted in the coronal plane. Interobserver variability was assessed using weighted kappa and the proportion specific agreement (PSA).

Results: Two second primary tumors and one metastasis were detected on WB-MRI. WB-MRI yielded seven clinically indeterminate lesions which did not progress at follow-up. The metastasis and one second primary tumor were found when combining ^{18}F-FDG-PET-CT and chest-CT findings. Interobserver variability for WB-MRI was κ=0.91 with PA ranging from 0.82 to 1.00. For ^{18}F-FDG-PET-CT κ could not be calculated due to a constant variable in the table and PA ranged from 0.40 to 0.99.

Conclusions: Our WB-MRI protocol with DWIBS is feasible in the work-up of HNSCC patients for detection and characterization of distant pathology. WB-MRI can be complementary to ^{18}F-FDG-PET-CT, especially in the detection of non ^{18}F-FDG avid second primary tumors.

INTRODUCTION

In head and neck squamous cell carcinoma (HNSCC) patients, 2-18% present with clinically identified distant spread of disease, while autopsy incidences have been reported to be up to 57% (1). Only palliative treatment remains when distant metastases are present in patients with HNSCC. Therefore, efforts should be made to detect distant metastases and avoid futile treatment.

Screening for distant metastases is currently done on a routine basis by means of ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) in combination with a diagnostic chest-CT in patients at high risk of developing distant metastases. Most metastases or second primary tumors (SPT) develop within 15 months after the end of treatment with curative intent, despite negative screening on ^{18}F-FDG-PET-CT. Since false negative rates are up to 50%, room for improvement remains (1-5).

Due to several technical improvements, it is now clinically feasible to perform high-resolution whole-body magnetic resonance imaging (WB-MRI) protocols in less than one hour. In HNSCC patients, WB-MRI showed a promising role for the evaluation of metastatic spread of disease despite variations in diagnostic accuracy of WB-MRI versus ^{18}F-FDG-PET-CT (6-8).

In addition to conventional WB-MRI, diffusion-weighted imaging (DWI) has shown potential. In order to deal with motion artifacts, Takahara et al. developed diffusion-weighted whole-body imaging with background body signal suppression (DWIBS). This sequence allows for the acquisition of DWI under free-breathing (9). The addition of DWIBS might improve the accuracy of WB-MRI to detect distant metastases (10-12).

The reported imaging sequences as well as imaging planes are quite variable (3, 4, 6-8, 10-18). As the addition of either DWI or contrast-enhanced (CE) sequences may improve the outcome of diagnostic interpretation, the value of these modalities needs to be clarified further. In addition, the choice of the imaging plane (e.g. axial versus coronal) has considerable effect on the duration of the scan, and potentially on the interpretation of the images as well. Therefore, tailor-made imaging protocols may optimize the performance of WB-MRI.

The purpose of our study therefore was to prospectively assess the feasibility of WB-MRI including DWIBS for the evaluation of distant malignancies in HNSCC patients with high risk factors for the presence of metastatic disease; and to compare MRI findings with ^{18}F-FDG-PET-CT and chest-CT.

MATERIALS AND METHODS

Study population

This prospective study was performed in a tertiary referral center for HNSCC between August 2009 and June 2012. Inclusion criteria comprised histopathologically proven HNSCC; planned extensive treatment with curative intent (surgery and/or radiotherapy with or without chemotherapy); planned routine screening for the presence of distant metastases by means of ^{18}F-FDG-PET-CT and chest-CT, i.e., at least one of the high risk factors for the development of distant metastases, as previously defined by De Bree et al. (clinically three or more lymph node metastases; bilateral lymph node metastases; lymph node metastases of 6 cm or larger; low jugular lymph node metastases; locoregional recurrence or second primary tumor) (5); and an age of 18-80 years. Exclusion criteria were pregnancy and contraindications for MRI. After approval of the local institutional review board and informed consent, 33 patients were included. For more detailed patient characteristics, we refer to Table 1. Whole-body ^{18}F-FDG-PET-CT and WB-MRI were performed at random order as dictated by logistics (mean time difference, 15.8 ± 11.3 days).

Whole-body-MRI

MR imaging was performed on a 1.5 T system (Magnetom Avanto; Siemens, Erlangen, Germany), using a total imaging matrix (TIM) coil system combined with dedicated coils. Whole-body-MRI up to the upper femora was performed with the acquisition of a T1-weighted sequence in the coronal plane; a short-tau inversion recovery (STIR) sequence in the coronal plane; an axial T2-weighted sequence covering the entire body; dedicated axial liver sequences covering the upper abdomen in the axial plane, including in- and opposed phase T1 gradient-echo (GRE).

DWIBS was acquired using with a 2D EPI sequence in the axial plane and reformatted in the coronal plane and presented with inverted signal intensity (b-value, 1000 s/mm^2; number of averages, 2; fat saturation, SPAIR; parallel imaging: GRAPPA).

After administration of 0.2 mmol/kg gadoteric acid in 17 patients (Dotarem; Guerbet, Roissy, France) and of 0.15 mmol/kg gadobutrol in 15 patients (Gadovist; Bayer Schering AG, Berlin, Germany) dynamic contrast-enhanced fat-suppressed volumetric interpolated breath-hold (VIBE) T1-weighted sequences in the arterial and delayed venous phases and a T1-weighted sequence covering the entire body were acquired in the axial plane. One patient did not receive MR-contrast due to renal failure. An overview of the scanned anatomic regions and sequences is provided in Table 2 and 3. Total examination time approximated 60 minutes, with scan time being 35 minutes.

^{18}F-FDG-PET-CT

Thirty-one patients underwent a PET/low-dose CT (LD-CT) scan after a 6 hour fast period and adequate hydration. The examination was performed from mid-thigh to skull vertex, 60 minutes after intravenous administration of 250-370 MBq ^{18}F-FDG. Scans were acquired on a Gemini TOF-64 PET-CT scanner (Philips Medical Systems, Best, The Netherlands) with an axial field of view of 18 cm. Time of flight (TOF) information was used during reconstruction. Reconstructed images had an image matrix size of 144x144, a pixel size of

4x4 mm and a slice thickness of 5 mm. Low-dose-CT was collected using a beam current of 30 to 50 mAs at 120 keV. CT-scans were reconstructed using an image matrix size of 512x512 resulting in pixel sizes of 1.17x1.17 mm and a slice thickness of 5 mm.

In two patients, examinations were performed at other institutions using Gemini TOF-64 and TOF-16 PET-CT scanners (Philips Medical Systems, Best, The Netherlands), respectively.

Table 1 *Patient characteristics of 33 HNSCC patients at whole-body MR imaging*

n	Sex	Age	Location	Recurrence	TNM	Previous treatment
1	Male	58	Hypopharynx	Locoregional recurrence	T4N3	Chemoradiation
2	Male	68	Hypopharynx	Second primary tumor	T4N2	CO2 laser excision
3	Male	57	Larynx	Primary tumor	T4N0	-
4	Male	57	Oropharynx	Primary tumor	T2N2	-
5	Female	69	Oropharynx	Locoregional recurrence	T3N2	Chemoradiation
6	Female	73	Larynx	Locoregional recurrence	T4N2	Excision + neck dissection + chemoradiation
7	Male	58	Oropharynx + oropharynx	Primary tumors	T2N1 + T2N1	-
8	Male	74	Nasopharynx	Second primary tumor	T1N2	Radiotherapy + excision + neck dissection + postoperative radiotherapy
9	Male	48	Oropharynx	Primary tumor	T4N2	-
10	Male	55	Oropharynx	Second primary tumor	T3N1	Neck dissection + radiotherapy
11	Male	63	Oropharynx	Primary tumor	T3N2	-
12	Male	65	Oral cavity	Locoregional recurrence	T1N0	Excision + neck dissection + radiotherapy
13	Male	70	Oropharynx	Locoregional recurrence	T2N1	Radiotherapy
14	Male	64	Hypopharynx	Primary tumor	T2N2	-
15	Male	59	Larynx	Locoregional recurrence	T2N2	Chemoradiation + excision
16	Male	62	Hypopharynx + oropharynx	Locoregional recurrence	T3N2 + T2N2	Chemoradiation
17	Female	67	Oral cavity	Second primary tumor	T4N0	Excision + neck dissection
18	Male	59	Tongue	Locoregional recurrence	T3N0	Radiotherapy
19	Male	75	Hypopharynx	Second primary tumor	T3N0	Radiotherapy
20	Male	57	Oral cavity	Second primary tumor	T2N0	Excision + neck dissection + radiotherapy
21	Male	50	Oropharynx	Primary tumor	T4N2	-
22	Female	51	Larynx	Locoregional recurrence	T2N0	Radiotherapy
23	Female	69	Oropharynx	Locoregional recurrence	T2N0	Chemoradiation
24	Male	61	Oropharynx	Primary tumor	T1N2	Radiotherapy + excision + neck dissection
25	Male	63	Oropharynx	Third primary	T3N0	-
26	Female	52	Larynx	Locoregional recurrence	T2N2	Chemoradiation
27	Male	59	Oropharynx	Primary tumor	T3N2	-
28	Male	66	Oropharynx	Primary tumor	T2N2	-
29	Male	59	Oropharynx	Primary tumor	T1N2	-
30	Male	66	Oral cavity	Recurrence	T2N0	Radiotherapy + neck dissection + excision
31	Male	79	Oropharynx	Primary tumor	T4N2	-
32	Female	60	Oropharynx	Third primary	T2N0	Excision + neck dissection
33	Male	78	Oral cavity	Locoregional recurrence	T2N0	Neck dissection + postoperative radiotherapy

Table 2 *MR Imaging protocol at 1.5T used in HNSCC patients*

	Sequence	Region	TR (ms)	TE (ms)	Matrix	FOV (mm)	Slices	Thickness (mm)	Flip angle	Scan time (min:sec)
Pre-contrast	Cor STIR	Whole body	6000	62	320 x 224	500	31	4	150	10:00
	Cor T1 TSE	Whole body	520	9.1	320 x 256	500	31	4	150	9:00
	Ax DW-MRI-EPI	Whole body	8200	66	128 x 88	500	60	4	90	9:00
	Ax T2 TSE	Head and Neck	4750	108	448 x 252	250	28	5	180	1:41
	Ax T1 GRE	Liver	100	2.38/4.76	256 x 154	350	20	6	70	0:32
	Ax HASTE-T2	Thorax- Pelvis	1000	65	256 x 165	500	20	8	150	0:22
	Ax VIBE FS	Liver	5.46	2.38	256 x 135	450	64	3	10	0:21
Postcontrast	Ax VIBE FS	Liver	5.46	2.38	256 x 135	450	64	3	10	1:30
	Ax T1 TSE	Head and Neck	755	9.5	320 x 256	250	28	5	150	1:31
	Ax FLASH 2D FS	Thorax- Pelvis	202	4.76	256 x166	500	35	6	70	0:58
									Total	34:55

Abbreviations: DW = diffusion-weighted; EPI = echo-planar imaging; FLASH = fast low angle shot; FS = fat saturation; GRE = gradient echo; HASTE = half-Fourier acquisition single-shot turbo spin-echo; STIR = short-TI inversion recovery; TSE = turbo spin echo; VIBE = volumetric interpolated breath-hold

Chest-CT

Chest-CT-scans were performed in 32 patients in the early arterial phase on a fourth-generation CT-scanner (Somaton Plus; Siemens, Erlangen, Germany) after intravenous contrast administration (Ultravist, Bayer Schering AG, Berlin, Germany) with a reconstructed slice thickness of 5 mm. In one patient only LD-CT was performed.

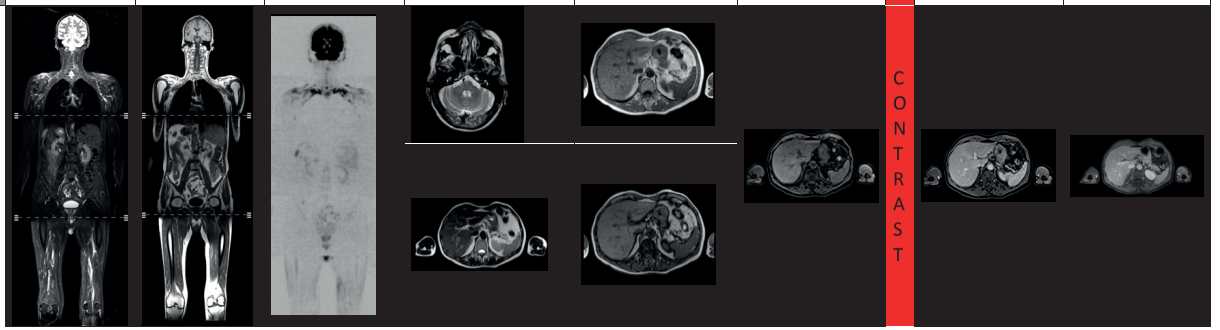
Image analysis

All readers were aware of the HNSCC diagnosis, but blinded to all other information, including the other imaging test results.

Whole-body-MRI images were analyzed for distant metastasis and SPT by two independent reviewers, with four and two years' experience in WB-MRI. After separate analysis the final decision was made in consensus. The analysis consisted of two parts: 1. evaluation of all conventional sequences without DWIBS; and 2. evaluation after DWIBS was added to the conventional sequences. Overall image quality and artifacts were assessed, per sequence, on a four-point Likert scale. For image quality: 1=inadequate, 2=adequate, 3=good, 4=excellent. For artifacts: 1=none present, 2=irrelevant, 3=diagnostically relevant, 4=marked. To complete the assessment of image quality the sequences that best depicted the pathology were selected. Although the primary goal was screening for distant metastases, SPT and incidental findings were also registered. Based on all MRI findings the likelihood of metastasis and/or SPT was scored on a three-point Likert scale: 1=yes, 2=clinically indeterminate, 3= no. The presence of malignancy was suspected on conventional WB-MRI in focal lesions with different signal intensities compared to the surrounding tissue. On the DWIBS malignancy was suspected in case of abnormal signal intensity in focal lesions.

Table 3 *Schematic MR imaging protocol at 1.5T used in HNSCC patients*

TIM coil system										
Head coil *4 elements*		Whole-body	Whole-body	Whole-body	Head and neck	Liver	Liver		Liver	Whole-body
Neck coil *2 elements*	Spine coil *24 elements*	STIR *coronal*	SE-T1 *coronal*	DWIBS *coronal*	TSE-T2 *axial*	In-phase GRE-T1 *axial*	Pre-contrast VIBE *axial*	CONTRAST	3-phase VIBE *axial*	CE-T1 *axial*
Torso coil *4 elements*					Thorax - abdomen - pelvis HASTE-T2 *axial*	Liver Opposed-phase GRE-T1 *axial*				
Abdominal coil *4 elements*										



Abbreviations: CE = contrast-enhanced; DWIBS = diffusion-weighted whole-body imaging with background-body-signal-suppression; GRE = gradient echo; HASTE = half-Fourier acquisition single-shot turbo spin-echo; SE = spin echo; STIR = short-TI inversion recovery; TSE = turbo spin echo; VIBE = volumetric interpolated breath-hold

At first, ^{18}F-FDG-PET-CT images were analyzed independently for distant metastasis and SPT by two reviewers with 12 and four years' experience in PET analysis. Again, the final decision was made in consensus. The likelihood of metastasis and/or SPT was scored on a three-point Likert scale: 1=yes, 2=clinically indeterminate, 3=no. Again, the primary goal was screening on distant metastases, but SPT and other abnormalities were also registered. Lesions were characterized as suspicious for malignancy based on increased ^{18}F-FDG uptake, incompatible with physiological ^{18}F-FDG distribution, within structures with an anatomical substrate on the (LD-)CT. Chest-CT was analyzed for distant metastasis and SPT by a radiologist with seven years of experience. The likelihood of metastasis and/or SPT was scored on a three-point Likert scale: 1=yes, 2=clinically indeterminate, 3=no.

When the presence of metastasis and/or SPT based on imaging was classified as 'yes' or 'clinically indeterminate', the final diagnosis regarding the presence of malignancy was based on histopathology or progression at six months of clinical follow-up (i.e. clinical assessment in the outpatient clinic every two months).

Statistical analysis

Interobserver variability for WB-MRI and ^{18}F-FDG-PET-CT was calculated with weighted kappa using Stata (version 11.2; College Station, TX, USA) and with proportion specific agreement using Microsoft Excel (Microsoft Office 2010, Microsoft, Redmond, WA, USA) (19). For the interpretation of weighted kappa, the following cut-off values are used: ≤0.20=poor; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=substantial; 0.81-1.00=very good. The proportion specific agreement consists of two parts: positive agreement (PA) and negative agreement (NA) (20). These two numbers express the agreement on positive and negative ratings respectively. Two sets of positive and negative ratings are calculated to deal with the 'clinically indeterminate' category regarding the presence of malignancy. In the first set clinically indeterminate is recoded into 'yes': $PA_{\text{clinically indeterminate=yes}}$ and $NA_{\text{clinically indeterminate=yes}}$. In the second set 'clinically indeterminate' is recoded into 'no': $PA_{\text{clinically indeterminate=no}}$ and $NA_{\text{clinically indeterminate=no}}$.

RESULTS

MRI quality

One patient did not receive MR-contrast due to renal failure. All other patients completed the entire protocol. Median image quality scores of the MR sequences were: 4 (range, 3-4) for coronal T1; 4 (range, 2-4) for coronal STIR; 3 (range, 3-4) for axial T2; 3 (range, 1-4) for axial T1 with contrast and 4 (range, 2-4) for DWIBS, and for artifacts: 1 (range, 1-2) for coronal T1; 2 (range, 1-3) for coronal STIR; 1 (range, 1-2) for axial T2; 1 (range, 1-4) for axial T1 with contrast and 1 (range, 1-3) for DWIBS. The coronal STIR was indicated to be most informative in 19 patients, coronal DWIBS in 24 patients and axial T2 in 17 patients by both reviewers.

Comparison between WB-MRI, ^{18}F-FDG-PET-CT and chest-CT

One patient had a distant HNSCC metastasis (lung; maximum axial diameter: 8 mm) (Figure 1) and two had SPT (renal cell carcinoma (RCC); maximum axial diameter: 80 mm, and a neuroendocrine tumor with liver metastases; maximum axial diameter: 20 mm) (Figure 2). On WB-MRI, without DWIBS, this metastasis was suspected and both SPTs were found. On DWIBS these three lesions all showed diffusion restriction, this confirmed the presence of malignancy. In another patient DWIBS aided in favor of the correct final diagnosis: the addition of DWIBS changed the conclusion regarding the presence of malignancy from 'clinically indeterminate' to 'no' in a benign cervical bone lesion. Seven lesions on WB-MRI including DWIBS were classified as clinically indeterminate: two vertebral lesions (one had a negative biopsy and both did not progress at follow-up), four thoracic lesions (all regressed at follow-up) and one pancreatic lesion (did not progress at follow-up). An adrenal lesion was correctly qualified as benign, whereas diagnostic chest-CT was equivocal. The lesion did not show ^{18}F-FDG uptake on ^{18}F-FDG-PET-CT and did not progress at follow-up. Other relevant incidental findings detected on WB-MRI were bone infarction, cholelithiasis, (old) brain infarction, scoliosis, hemochromatosis and atelectasis. WB-MRI was correctly negative in 25 patients, after the addition of DWIBS in 24 patients.

On ^{18}F-FDG-PET-CT the HNSCC lung metastasis was also detected, but not the SPTs. Two lesions, a focal lung lesion and a vertebral bone lesion, were classified as clinically indeterminate. ^{18}F-FDG-PET-CT was correctly negative in 30 patients. On chest-CT the HNSCC lung metastasis and the RCC were identified. Eight lesions were classified as clinically indeterminate: four focal lung lesions, two lymph nodes, one bone lesion and a liver lesion. None of these lesions did progress at follow-up. Chest-CT was correctly negative in 25 patients.

The clinical standard of practice (^{18}F-FDG-PET-CT and chest-CT) yielded metastasized HNSCC in one patient and RCC in another. One vertebral bone lesion remained clinically indeterminate using the clinical standard of practice. This lesion did not progress at follow-up.

The interobserver agreement for WB-MRI was very good (κ=0.91, $PA_{clinically\ indeterminate=yes}$=0.82; $NA_{clinically\ indeterminate=yes}$=0.96; $PA_{clinically\ indeterminate=no}$=1.00 and $NA_{clinically\ indeterminate=no}$=1.00). For ^{18}F-FDG-PET-CT weighted kappa could not be calculated. This is because the table contained a constant variable, which made it impossible to calculate weighted kappa. Proportion specific agreement was: $PA_{clinically\ indeterminate=yes}$=0.40; $NA_{clinically\ indeterminate=yes}$=0.98; $PA_{clinically\ indeterminate=no}$=0.67 and $NA_{clinically\ indeterminate=no}$=0.99.

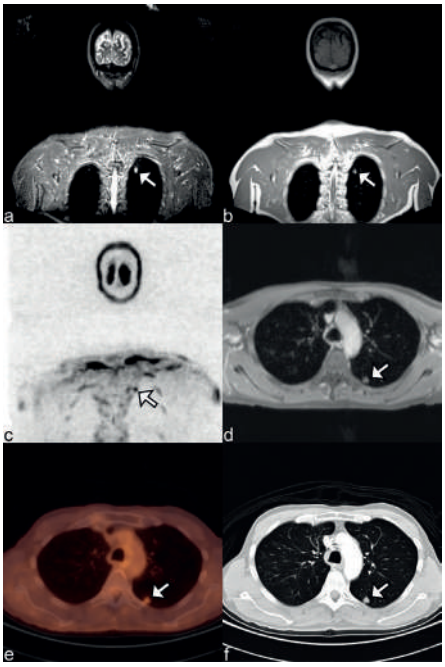


Figure 1 *Coronal and axial images in a 62-year old male with a lung metastasis in the apex of the left lower lobe (**arrow**). A) Coronal STIR, B) coronal T1, C) DWIBS, D) axial contrast-enhanced T1, E) axial fused ^{18}F-FDG-PET-CT and F) axial chest-CT. Mainly due to diffusion-restriction on the coronal DWIBS this lesion is suspected to be malignant. The lesion demonstrates ^{18}F-FDG uptake and on chest-CT a solitary non-calcified nodule is seen.*

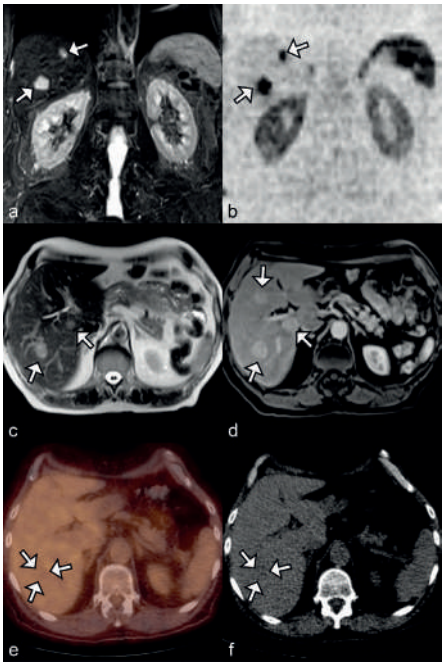


Figure 2 *Images of a focal liver lesion (**arrows**) in a 68-year old male (multiple other lesions with identical characteristics not shown). A) Coronal STIR, B) DWIBS, C) axial HASTE-T2, D) axial CE-T1 VIBE in the arterial phase, E) axial fused ^{18}F-FDG-PET-CT and F) axial LD-CT. Based on MR imaging findings this lesion is suspicious of malignancy, with neuroendocrine liver metastases as first differential option. This has been confirmed after biopsy. The high signal of the spleen on DWIBS can be considered physiological. The lesion is outside the field of view of the diagnostic chest-CT. No uptake of ^{18}F-FDG is seen. On low-dose CT a minimally hypodense lesion is seen only after visual correlation with MR-images. Therefore this lesion is regarded as undetected in further data-analysis.*

DISCUSSION

Various WB-MRI protocols have been compared to ^{18}F-FDG-PET-CT in the work-up of patients with (suspicion of) distant metastases. In patients with colorectal and breast cancer Schmidt et al. used an imaging protocol containing coronal STIR, coronal T1 and axial CE-T1. Radiological follow-up of at least five months served as a standard of reference (13, 14). In both studies WB-MRI and ^{18}F-FDG-PET-CT had comparable diagnostic accuracy in detecting distant metastases. However, sensitivity and specificity of WB-MRI to detect metastatic disease were variable: 95% and 92% in breast cancer and 78% and 95% in colorectal cancer respectively. This suggests that the value of WB-MRI may depend on the type of malignancy and its metastatic pattern. Ohno et al. found the combination of conventional WB-MRI and DWI to have a diagnostic accuracy comparable to ^{18}F-FDG-PET-CT for M-stage assessment in non-small cell lung cancer, using a combination of imaging, biopsy and at least 12 months of clinical follow-up as the reference standard. Sensitivity seemed to improve after the addition of DWI to conventional MRI (from 60% to 70%) (18). Heusner et al. demonstrated high sensitivity (91%), but low specificity (72%) of whole-body DWI alone in the detecting breast cancer metastases. Specificity was especially compromised in lymph nodes and bone lesions (16).

Taken together, these data suggest that imaging protocols containing more MR-sequences than DWI alone may be preferable. Compared to conventional DWI, DWIBS has the advantage that it allows for DWI during free breathing (9). Due to background suppression small lesions are more easily detected on DWIBS (21).

In our study WB-MRI including DWIBS was superior to ^{18}F-FDG-PET-CT in the detection of SPT. In general SPTs in HNSCC mainly emerging in the head and neck area and the lungs (22). The level of ^{18}F-FDG uptake of RCC and neuroendocrine tumors is variable. Populations have been described where only 31% of the RCCs showed increased ^{18}F-FDG uptake (23, 24). Ng et al. performed two studies in patients with advanced HNSCC. In a study of 79 patients with advanced HNSCC, Ng et al. reported that ^{18}F-FDG-PET-CT showed a (non-significant) trend towards higher diagnostic capability than conventional WB-MRI in detecting SPT below the clavicles (4/5 vs 2/5) (8). In another study in 150 patients with advanced HNSCC both modalities were comparable. On WB-MRI a bronchoalveolar cell carcinoma was detected, due to low ^{18}F-FDG-uptake this was interpreted as inflammation on ^{18}F-FDG-PET-CT. Using ^{18}F-FDG-PET-CT colon carcinoma was found, which was missed on WB-MRI. On both modalities another SPT in the lung was detected (25).

Whole-body-MRI is considerably less expensive than ^{18}F-FDG-PET-CT. If WB-MRI can replace ^{18}F-FDG-PET-CT, a substantial reduction of health costs seems to be possible. Moreover, patients undergoing WB-MRI are not exposed to radiation as by ^{18}F-FDG-PET-CT. To replace ^{18}F-FDG-PET-CT, WB-MRI needs to have at least comparable diagnostic accuracy(26). However, the biological information provided by the level of ^{18}F-FDG uptake may carry prognostic relevance, and serial uptake measurements may serve as a predictive biomarker (27). Hence, if the chest-CT information proves to be redundant, PET-MRI might become the method of choice for personalized therapy.

In this pilot study we used a combination of STIR and T1 in the coronal plane combined with T2 and dedicated liver sequences in the axial plane. DWIBS was acquired in the axial plane and reformatted in the coronal plane. By using this combination, we demonstrated the feasibility of WB-MRI not only to detect benign and malignant lesions, but also characterize them (e.g. hemangiomas, renal cysts, bone infarction and hemochromatosis). In our study population WB-MRI allowed for the detection of two non-^{18}F-FDG avid malignancies. Whole-body MRI yielded seven clinically indeterminate lesions. In one of these lesions biopsy was performed. None of the clinically indeterminate lesions did progress at follow-up. The addition of DWIBS aided in making the correct final diagnosis of a HNSCC lung metastasis and a benign cervical bone lesion. Particular in bone and thoracic lesions WB-MRI including DWIBS remained clinically inconclusive. On ^{18}F-FDG-PET-CT small thoracic lesions were difficult to deal with and for chest-CT lung nodules and mediastinal lymph nodes were challenging to characterize.

We believe that there is a learning curve in the evaluation of WB-MRI including DWIBS. The addition of DWIBS to WB-MRI protocols allows for fast image interpretation since it enables distinguishment of malignant from benign tissue “at-a-glance” (28). The use of coronal images requires additional training as most radiologists are more familiar with axial images. Incidental findings are more frequently present than on ^{18}F-FDG-PET-CT due to the higher soft tissue detail on WB-MRI. Therefore, some experience in WB-MRI is necessary to deal with them properly.

Our study had some limitations. First, the incidence of distant metastases was lower than would be expected according to our inclusion criteria as defined by de Bree et al. (5); only one patient had a distant metastasis and two patients demonstrated SPTs. Some patients with high suspicion of distant metastases visualized on ^{18}F-FDG-PET-CT refrained from WB-MRI. In the future this could be prevented by performing all imaging on the same day. Other patients refrained from WB-MRI due to claustrophobia. This limited the possibilities for statistical analysis. Second, because this is a pilot study, the number of patients was limited. Therefore, it is necessary to prospectively validate this MR-protocol in a larger population.

Conclusions

The presented WB-MRI protocol with DWIBS is feasible in the work-up of patients with advanced HNSCC for the detection and characterization of distant pathology; it allowed for the detection of non ^{18}F-FDG avid malignancies and can therefore be complementary to ^{18}F-FDG-PET-CT.

REFERENCES

1. de Bree R, Haigentz M, Jr., Silver CE, et al. Distant metastases from head and neck squamous cell carcinoma. Part II. Diagnosis. Oral Oncol. 2012;48(9):780-6.
2. Senft A, de Bree R, Hoekstra OS, et al. Screening for distant metastases in head and neck cancer patients by chest CT or whole body FDG-PET: a prospective multicenter trial. Radiother Oncol. 2008;87(2):221-9.
3. Ng SH, Chan SC, Liao CT, et al. Distant metastases and synchronous second primary tumors in patients with newly diagnosed oropharyngeal and hypopharyngeal carcinomas: Evaluation of 18F-FDG PET and extended-field multi-detector row CT. Neuroradiology. 2008;50(11):969-79.
4. Kim SY, Roh JL, Yeo NK, et al. Combined 18F-fluorodeoxyglucose-positron emission tomography and computed tomography as a primary screening method for detecting second primary cancers and distant metastases in patients with head and neck cancer. Ann Oncol. 2007;18(10):1698-703.
5. de Bree R, Deurloo EE, Snow GB, Leemans CR. Screening for distant metastases in patients with head and neck cancer. Laryngoscope. 2000;110(3 Pt 1):397-401.
6. Fruehwald-Pallamar J, Czerny C, Mayerhoefer ME, 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;38(6):1009-19.
7. Chan SC, Wang HM, Yen TC, et al. 18F-FDG PET/CT and 3.0-T whole-body MRI for the detection of distant metastases and second primary tumours in patients with untreated oropharyngeal/hypopharyngeal carcinoma: A comparative study. Eur J Nucl Med Mol Imaging. 2011;38(9):1607-19.
8. Ng SH, Chan SC, Yen TC, et al. Pretreatment evaluation of distant-site status in patients with nasopharyngeal carcinoma: accuracy of whole-body MRI at 3-Tesla and FDG-PET-CT. Eur Radiol. 2009;19(12):2965-76.
9. Takahara T, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M. Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. Radiat Med. 2004;22(4):275-82.
10. Murtz P, Krautmacher C, Traber F, Gieseke J, Schild HH, Willinek WA. Diffusion-weighted whole-body MR imaging with background body signal suppression: a feasibility study at 3.0 Tesla. Eur Radiol. 2007;17(12):3031-7.
11. Komori T, Narabayashi I, Matsumura K, et al. 2-[Fluorine-18]-fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography versus whole-body diffusion-weighted MRI for detection of malignant lesions: initial experience. Ann Nucl Med. 2007;21(4):209-15.
12. Sommer G, Klarhofer M, Lenz C, Scheffler K, Bongartz G, Winter L. Signal characteristics of focal bone marrow lesions in patients with multiple myeloma using whole body T1w-TSE, T2w-STIR and diffusion-weighted imaging with background suppression. Eur Radiol. 2011;21(4):857-62.
13. Schmidt GP, Baur-Melnyk A, Haug A, et al. Comprehensive imaging of tumor recurrence in breast cancer patients using whole-body MRI at 1.5 and 3 T compared to FDG-PET-CT. Eur J Radiol. 2008;65(1):47-58.
14. Schmidt GP, Baur-Melnyk A, Haug A, et al. Whole-body MRI at 1.5 T and 3 T compared with FDG-PET-CT for the detection of tumour recurrence in patients with colorectal cancer. Eur Radiol. 2009;19(6):1366-78.
15. Fischer MA, Nanz D, Hany T, et al. Diagnostic accuracy of whole-body MRI/DWI image fusion for detection of malignant tumours: a comparison with PET/CT. Eur Radiol. 2011;21(2):246-55.
16. Heusner TA, Kuemmel S, Koeninger A, et al. Diagnostic value of diffusion-weighted magnetic resonance imaging (DWI) compared to FDG PET/CT for whole-body breast cancer staging. Eur J Nucl Med Mol Imaging. 2010;37(6):1077-86.
17. Brauck K, Zenge MO, Vogt FM, et al. Feasibility of whole-body MR with T2- and T1-weighted real-time steady-state free precession sequences during continuous table movement to depict metastases. Radiology. 2008;246(3):910-6.

18. Ohno Y, Koyama H, Onishi Y, et al. Non-small cell lung cancer: whole-body MR examination for M-stage assessment--utility for whole-body diffusion-weighted imaging compared with integrated FDG PET/CT. Radiology. 2008;248(2):643-54.

19. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.

20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

21. Kwee TC, Takahara T, Ochiai R, Nievelstein RA, Luijten PR. Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology. Eur Radiol. 2008;18(9):1937-52.

22. Morris LG, Sikora AG, Patel SG, Hayes RB, Ganly I. Second primary cancers after an index head and neck cancer: subsite-specific trends in the era of human papillomavirus-associated oropharyngeal cancer. J Clin Oncol. 2011;29(6):739-46.

23. Oyama N, Ito H, Takahara N, et al. Diagnosis of complex renal cystic masses and solid renal lesions using PET imaging: comparison of 11C-acetate and 18F-FDG PET imaging. Clin Nucl Med. 2014;39(3):e208-14.

24. Jadvar H. Hepatocellular carcinoma and gastroenteropancreatic neuroendocrine tumors: potential role of other positron emission tomography radiotracers. Semin Nucl Med. 2012;42(4):247-54.

25. Ng SH, Chan SC, Yen TC, et al. PET/CT and 3-T whole-body MRI in the detection of malignancy in treated oropharyngeal and hypopharyngeal carcinoma. Eur J Nucl Med Mol Imaging. 2011;38(6):996-1008.

26. Plathow C, Walz M, Lichy MP, et al. [Cost considerations for whole-body MRI and PET/CT as part of oncologic staging]. Radiologe. 2008;48(4):384-96.

27. Quarles van Ufford HM, van Tinteren H, Stroobants SG, Riphagen, II, Hoekstra OS. Added value of baseline 18F-FDG uptake in serial 18F-FDG PET for evaluation of response of solid extracerebral tumors to systemic cytotoxic neoadjuvant treatment: a meta-analysis. J Nucl Med. 2010;51(10):1507-16.

28. Padhani AR, Koh DM, Collins DJ. Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology. 2011;261(3):700-18.

CHAPTER 4.2

Diagnostic value of diffusion-weighted imaging and ^{18}F-FDG-PET/CT for the detection of unknown primary head and neck cancer in patients presenting with cervical metastasis

European Journal of Radiology 2018;107:20-25

Daniel P Noij
Roland M Martens
P de Graaf
B Zwezerijnen
Thomas Koopman
Remco de Bree
Otto S Hoekstra
Jonas A Castelijns

ABSTRACT

Background and purpose: Head and neck squamous cell carcinoma (HNSCC) may present with cervical metastases without an apparent primary tumor. Detecting the primary tumor results in more targeted treatment. Acquisition of diffusion-weighted imaging (DWI) is improving with less artifacts and image distortion. We assessed the diagnostic value of DWI and ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) for detecting primary tumors in patients presenting with nodal metastasis of an unknown primary HNSCC.

Materials and methods: For this retrospective study we included 31 patients (male/female ratio=23/8; median age=66 years; age range, 40-80 years) who presented with a pathologically proven cervical nodal metastasis from HNSCC without overt primary tumor location between January 2013 and November 2016 and underwent both DWI and ^{18}F-FDG-PET-CT. Both modalities were assessed qualitatively and quantitatively. With receiver operating characteristic (ROC) analysis we determined the optimal cut-off for imaging parameters in separating occult malignancy from benign tissue.

Results: Qualitative analysis of magnetic resonance imaging including DWI resulted in a sensitivity of 81.3% (95%CI)=53.7-95.0) and specificity of 73.3% (95%CI=44.8-91.1). With qualitative scoring of ^{18}F-FDG-PET-CT a sensitivity and specificity of 93.8% (95%CI=67.8-99.7) and 73.3% (95%CI=44.8-91.1) were found. With quantitative analysis sensitivity and specificity of maximum standardized uptake value (SUV_{max}) were 81.3% (95%CI=53.6-95.0) and 93.3% (95%CI=66.0-99.7), respectively. Combining DWI and ^{18}F-FDG-PET-CT resulted in a sensitivity of 93.8% (95%CI=67.7-99.7%) and specificity of 60.0% (95%CI=32.9-82.5%).

Conclusion: In this study on HNSCC patients presenting with clinically unknown primary lesions the diagnostic accuracy of qualitative analysis with DWI and ^{18}F-FDG-PET-CT and quantitative analysis of ^{18}F-FDG-PET-CT using SUV_{max} were high. Adding DWI did not improve the accuracy of ^{18}F-FDG-PET-CT.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) may present with cervical lymphadenopathy (1, 2). Diagnostic testing usually consists of cross-sectional imaging (e.g. magnetic resonance imaging (MRI), computed tomography (CT), and/or ^{18}F-Fluorodeoxyglucose positron emission tomography (^{18}F-FDG-PET)-CT), and an examination under anesthesia (EUA) with taking of biopsies and/or diagnostic tonsillectomy (1). Recently, promising results have been achieved with transoral robotic surgery to identify small primary lesions (3). The presence of human papillomavirus (HPV)-DNA suggests a primary lesion in the oropharynx (4), whereas the Epstein-Barr virus (EBV) is suggestive for a nasopharyngeal lesion (5). However, in 2-9% of patients presenting with cervical metastases from HNSCC no primary tumor is detected despite extensive diagnostic testing (2). Initially unknown primary (UP) tumors are most frequently located in the tonsils, followed by the base of the tongue (6).

In UP HNSCC the whole mucosal lining of the upper aerodigestive tract -where the UP tumor is expected to be located- is treated with radiotherapy (7). This extensive radiation field is associated with increased morbidity. Therefore, efforts should be made to detect occult primary tumors. The use of imaging may result in more directed biopsies and a higher yield of EUAs.

^{18}F-FDG-PET-CT is generally considered to be the most valuable imaging technique for this indication (6, 8-12). However, the presence of small lesions may result in false-negative findings on ^{18}F-FDG-PET-CT. Moreover, ^{18}F-FDG-uptake in the head and neck area can have other causes than malignancy (e.g. tonsillitis or tongue movement). Some ^{18}F-FDG-uptake in tonsils and the base of the tongue is physiological (13). False positive rates up to 20% have been reported (1).

Diffusion-weighted imaging (DWI) has been introduced as a potential technique for detecting occult primary HNSCC (14, 15). Diffusion can be quantified in an apparent diffusion coefficient (ADC) (16). In general, HNSCC is associated with lower ADC values than benign tissues (17-20). However, Choi et al. found clinically UP tonsillar carcinoma to have higher ADC values than normal tonsillar tissue and lower than overt tonsillar carcinoma (9). Due to improving image quality and reduction of artifacts and image distortion lesions as small as 4 mm can be assessed with DWI (18). Therefore, DWI may be a valuable modality in detecting UP HNSCC. Since DWI and ^{18}F-FDG-PET-CT are based on different properties, both techniques may be complementary.

In an attempt to increase the accuracy of diagnostic imaging, which may result in more directed biopsies and less extensive treatment, we firstly assessed the diagnostic value of DWI and ^{18}F-FDG-PET-CT, either as stand-alone or in combination, to detect primary tumors in patients presenting with nodal metastases of UP HNSCC. Secondly, different quantitative measurements were performed in order to find the most suitable for detecting the UP.

METHODS AND MATERIALS

This retrospective study was performed in a tertiary referral center for head and neck cancer. Patients treated between January 2013 and November 2016 were included. The institutional review board waived need for informed consent. Inclusion criteria were: 1) pathologically proven cervical nodal metastasis from HNSCC without overt primary tumor location after a complete head and neck examination including flexible endoscopy at the outpatient clinic; 2) MRI including DWI of at least oropharynx and hypopharynx; 3) ^{18}F-FDG-PET-CT. Exclusion criteria were: 1) history of previous HNSCC; 2) history of malignancy requiring systemic therapy; 3) mucosal biopsy performed before ^{18}F-FDG-PET-CT.

We used the final diagnosis as defined by the multi-disciplinary team with access to all available diagnostic results, including EUA with taking of biopsies, as the reference standard. If the primary location remained unknown, T-stage was defined as Tx (21).

We also included a control group presenting with overt primary HNSCC and nodal metastases to compare quantitative imaging parameters between UP HNSCC and overt HNSCC. The other in- and exclusion criteria for the control group were identical to the UP group. This group was chosen as a control group, because these patients most closely represent the UP group. We only included the controle group in the analysis if mentioned specifically.

The clinically occult primary group consisted of 31 consecutive patients (male/female ratio=23/8; median age, 66 years; age range, 40-80 years). In 16 of these patients a primary tumor was found after EUA (Table 1). See Figure 1 for a flowchart of patient inclusion. The control group consisted of 20 patients (male/female ratio=16/4; median age, 62 years; age range, 48-88 years).

Diffusion-weighted MRI

We used a 1.5T MRI system (Signa Excite HDxt, GE Healthcare, Milwaukee, WI, USA) with a standard head-neck coil. Conventional MRI protocols consisted of at least 4 mm short tau inversion recovery (STIR) and spin-echo T1 in the axial plane.

Diffusion-weighted imaging was acquired using echo-planar imaging (EPI) (TR/TE=5600/76.2 ms; matrix=128x86) with two b-values: b=0 (1 average) and 1000 (6 averages) s/mm^2. ADC maps were calculated using a monoexponential formula with software present on the MRI system.

Positron emission tomography–computed tomography

For ^{18}F-FDG-PET-CT we used two systems (Ingenuity or Gemini TF, Philips Medical Systems, Best, the Netherlands), both EARL accredited (22), depending on logistics after a fasting period of six hours in accordance with the EANM guidelines (22). One hour after injection of 2.5 MBq/kg (±10%) of ^{18}F-FDG, whole body ^{18}F-FDG-PET-CT with the arms down was performed in 13 patients (two minutes per bed position, image matrix size=144x144, voxel size=4x4x4 mm) and, after a change in the institutional protocol, dedicated head and neck ^{18}F-FDG-PET-CT in 38 patients (four minutes per bed position, image matrix size=288x288,

voxel size=2x2x2 mm). In both protocols low-dose CT with 120 kV and 50 mAS was used for anatomic correlation of ^{18}F-FDG uptake and attenuation correction. In the whole body protocol post reconstruction image slice thickness was 7 mm full width at half maximum (FWHM), and 5 mm FWHM in the head and neck protocol.

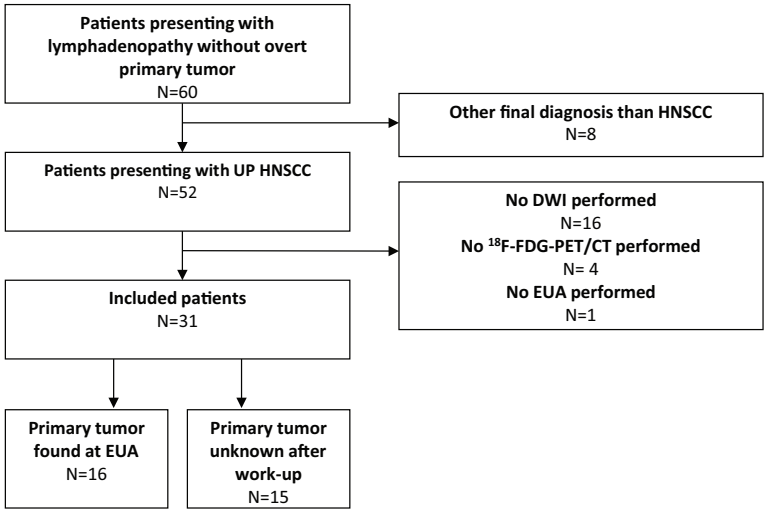


Figure 1 *Flow chart of patient inclusion*

Image analysis

On both modalities observers routinely assessed the following regions in each patient: tonsils, base of the tongue and hypopharynx in each patient. Lesions were qualitatively divided in three categories: benign (score=1), equivocal (score=2) or malignant (score=3) (Table 2) (13, 17). In the absence of a lesion the region was classified as benign, and consequently only included in qualitative analysis. Anonymized images were presented in a random order to the observers, who were aware of the presence of a control group. Observers were blinded to the final diagnosis and results of other observers and modalities.

Two radiologists with 33 (JCA) and 9 (PGR) years of experience in head and neck imaging assessed MRI including DWI. Consensus values were used for the final analysis. Conventional sequences were used for anatomical localization. Malignancy was suspected in focal lesions with asymmetry or ingrowth into surrounding structures, with high signal intensity on b1000 images and concurrent low ADC value (17). Restricted diffusion was not considered indicative of malignancy in: parotid and submandibular salivary glands, thyroid and palatine tonsils in the absence of other imaging characteristics suggestive for malignancy (9, 23) (Table 2). ROIs were placed on areas with high signal intensity on b1000 images and copied to the ADC map for histogram analysis of ADC resulting in the following parameters: ADC_{volume}, ADC_{mean}, standard deviation (ADC_{SD}), ADC_{median}, minimal value (ADC_{min}), maximal value (ADC_{max}), $ADC_{skewness}$, $ADC_{kurtosis}$.

A nuclear medicine resident (BZ) under supervision of a nuclear medicine physician with 21 years of experience (OSH) assessed ^{18}F-FDG-PET-CT imaging. Malignancy was suspected in case of abnormal focal uptake higher than the mediastinal blood pool and not attributable to normal ^{18}F-FDG distribution (Table 2) (13). We used the maximal standardized uptake value (SUV_{max}) normalized for body weight (with and without correction for blood pool uptake) for quantitative analysis.

Statistical analysis

Three analyses were performed: 1) In the per-patient analysis we used the highest score per imaging modality in that patient for analysis; 2) in the per-lesion analysis all lesions were analyzed separately; 3) in the tonsil-only analysis we included only the results of tonsils (i.e. between 0-2 per patient) as this is the most frequent location of initially UP HNSCC.

Chi-square tests were used to compare qualitative image results and p16 or HPV status. Mann-Whitney U tests were used to compare ADC and SUV_{max} values of malignant and benign tissue; and of p16 or HPV positive and negative patients. For parameters that were significantly different between malignant and benign tissue we used receiver operating characteristic (ROC) analysis to determine diagnostic accuracy. Two cut-offs were determined: 1) the highest Youden Index (24), and 2) a sensitivity of at least 90%. With McNemar tests we compared diagnostic accuracy between modalities. For qualitative DWI analysis, interobserver agreement was assessed with weighted κ and proportion specific agreements. Proportion specific agreements expresses the level of agreement separately for positive and negative ratings resulting in a positive agreement and negative agreement (25). To calculate the proportion specific agreement, we used the sensitive cut-off value to create a 2x2 table. Two-way mixed intraclass correlation coefficient (ICC) was used for quantitative parameters. Interobserver agreement measures were interpreted as follows: ≤0.20=poor; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=substantial; 0.81-1.00=very good (26). Statistical analyses were done with SPSS (version 22.0; Chicago, IL, USA). *P* values <0.05 were considered statistically significant.

RESULTS

Qualitative analysis

All 31 patients were included in qualitative analysis. Due to local artifacts caused by dental artifacts or patient movement (e.g. swallowing), there were 170/186 evaluable regions in the per-lesion analysis of DWI, compared to 186/186 regions on ^{18}F-FDG-PET-CT. One T1 hypopharyngeal tumor could not be included in the per-lesion analysis of DWI due to local image artifacts (Table 3). There was no significant relation between p16 or HPV status and qualitative image results (data not shown).

Table 1 *Patient characteristics*

	Clinically occult primary group (n=31)	Control group (n=20)
Age, median, range	66, 40-80 years	62, 48-88 years
Male/Female ratio	23/8	16/5
Final tumor location	Tonsil=8	Tonsil=11
	Base of the tongue=6	Base of the tongue=5
	Hypopharynx=2	Hypopharynx=2
	Unknown=15	Pharyngeal wall=2
T-stage, n	Tx=15	T1 =1
	T1=7	T2=18
	T2=8	T3=1
	T3=1	
N-stage, n	N1=8	N1=0
	N2a=5	N2a=1
	N2b=12	N2b=16
	N2c=3	N2c=3
	N3=3	N3=0
AJCC-stage, n	III=8	III=0
	IV=23	IV=20
p16, n	Neg=6	Neg=6
	Pos=17	Pos=13
	Unknown=8	Unknown=1
HPV DNA, n	Neg=4	Neg=1
	Pos=16	Pos=11
	Unknown=11	Unknown=8

Best diagnostic accuracy with DWI was achieved with a sensitive approach (i.e. a score of ≥2 is considered a positive read) (Table 3); resulting in a sensitivity and specificity of 81.3% and 73.3% in the per-patient analysis. Sensitivity and specificity were 86.7% and 96.1% in the per-lesion approach. When including only tonsils sensitivity and specificity were 75.0% and 92.6%, respectively.

Table 2 *Qualitative imaging criteria for DWI and ^{18}F-FDG-PET-CT, respectively*

Likert scale	DWI (17)	^{18}F-FDG-PET-CT (13)
1. (probably) benign	No diffusion restriction/ ADC higher than surrounding structures	No or physiological ^{18}F-FDG-uptake
2. Probably malignant	Focal lesion with minimally elevated signal on b1000 and slightly lower ADC value compared to surrounding structure	Minimal higher ^{18}F-FDG-uptake than the mediastinal blood pool and not attributable to normal ^{18}F-FDG distribution
3. Suggestive of malignancy	Focal lesion with high signal on b1000 and low ADC value compared to surrounding structure	Abnormal focal ^{18}F-FDG-uptake higher than the mediastinal blood pool and not attributable to normal ^{18}F-FDG distribution

With ^{18}F-FDG-PET-CT best results were achieved with a conservative approach (i.e. a score of 3 is considered a positive read) (Table 3). In the per-patient analysis sensitivity was 93.8% and specificity was 73.3%. In three patients the nuclear medicine physician located the tumor in the tonsil. We considered this as false-positive, because the reference standard located the tumor in the ipsilateral base of the tongue. This resulted in a sensitivity of 75.0% and specificity was 96.5% in the per-lesion analysis. When including only tonsils sensitivity and specificity were 87.5% and 88.9%, respectively.

There were no significant differences between diagnostic accuracy of DWI (sensitive approach) and ^{18}F-FDG-PET-CT (conservative approach) in the per-patient, per-lesion or tonsil-only analysis. Interobserver agreement for DWI was substantial on the per-patient (weighted κ=0.61 (95%CI=0.42-0.80)) and per-lesion level (weighted κ=0.77 (95%CI=0.67-0.87)). On a per-patient level positive agreement was very good (0.81) and negative agreement was substantial (0.72). On a per-lesion level positive agreement was substantial (0.76) and negative agreement was very good (0.97).

The sensitive approach for DWI and conservative approach for ^{18}F-FDG-PET-CT were combined as follows. Firstly, we considered positive reads on both modalities to be an overall positive read. This resulted in a sensitivity of 81.3% (95%CI=53.7-95.0%) and specificity of 86.7% (95%CI=58.4-97.7%) in the per-patient analysis. In the per-lesion approach sensitivity was 66.7% (95%CI=38.7-87.0%) and specificity was 98.7% (95%CI=94.9-99.8%). In the tonsil-only analysis sensitivity was 75.0% (95%CI=35.6-95.5%) and specificity was 96.3% (95%CI=86.2-99.4%). Secondly, we considered positive reads on either modality to be an overall positive read. In the per-patient analysis this resulted in a sensitivity of 93.8% (95%CI=67.7-99.7%) and specificity of 60.0% (95%CI=32.9-82.5%). In a per-lesion analysis sensitivity was 93.3% (95%CI=67.8-99.7%) and specificity was 93.5% (95%CI=88.1-96.7%). In the tonsil-only analysis sensitivity was 87.5% (95%CI=46.7-99.3%) and specificity was 85.2% (95%CI=72.3-92.9%). Other combinations did not further improve sensitivity. See Figure 2 for a representative case of an initial unknown primary tumor.

Quantitative analysis

In three patients there were no focal lesions on DWI to include in quantitative analysis. One of these patients did have a T1 hypopharyngeal tumor. Therefore, we performed quantitative analysis in 28 patients with 104 evaluable lesions in the UP group and 20 patients with 69 evaluable lesions in the control group. On ^{18}F-FDG-PET-CT all 31 patients with 128 evaluable lesions in the UP group were included in quantitative analysis, and 20 patients with 85 evaluable lesions in the control group

Except for a larger lesion volume of malignant tissue (P<0.001), no ADC parameter differed significantly between malignant and benign lesions. Receiver-operation-characteristic analysis of lesion volume resulted in a sensitivity and specificity of 66.7% and 90.0%, respectively (Appendix A). There were no differences in DWI parameters between p16 or HPV positive and negative patients (data not shown).

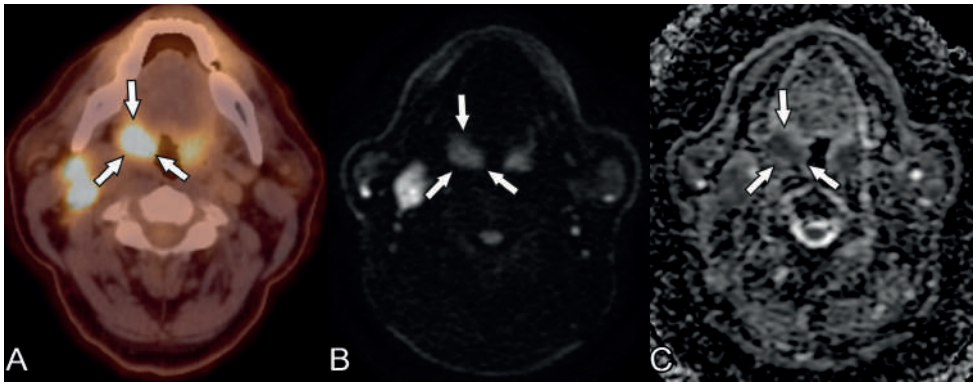


Figure 2 *Imaging of a 56-year old male patient who presented with a progressive nodal swelling on the right side and initial unknown primary tumor and who was finally diagnosed with a T1N2b tonsillar carcinoma. A) ^{18}F-FDG-PET-CT shows higher ^{18}F-FDG uptake of the right tonsil compared to the left tonsil. B) On b1000 imaging and the C) ADC map the right tonsil has a slightly lower ADC value compared to the left tonsil ($1.05 \cdot 10^{-3}$ mm^2/sec vs $1.16 \cdot 10^{-3}$ mm^2/sec) combined with a slightly larger right tonsil compared to the left tonsil.*

Table 3 *Qualitative image results of DWI and ^{18}F-FDG-PET-CT in the unknown primary group*

		Sensitive (cut-off 1-2)			Conservative (cut-off 2-3)		
	AUC (95%CI)	Sensitivity (%, 95%CI, ratio)	Specificity (%, 95%CI, ratio)	YI	Sensitivity (%, 95%CI, ratio)	Specificity (%, 95%CI, ratio)	YI
Per-patient DWI (n=31)	0.813 (0.655-0.970)	81.3 (53.7-95.0, 13/16)	73.3 (44.8-91.1, 11/15)	0.483	50.0 (25.5-74.5, 8/16)	93.3 (66.0-99.7, 14/15)	0.433
Per-patient ^{18}F-FDG-PET-CT (n=31)	0.829 (0.672-0.986)	93.8 (67.8-99.7, 15/16)	53.3 (27.4-77.8, 8/15)	0.471	93.8 (67.8-99.7, 15/16)	73.3 (44.8-91.1, 11/15)	0.671
Per-lesion DWI (n=170)	0.922 (0.819-1.000)	86.7 (58.4-97.8, 13/15[a])	96.1 (91.4-98.4, 149/155)	0.761	53.3 (27.4-77.7, 8/15[a])	99.4 (95.9-100.0, 154/155)	0.527
Per-lesion ^{18}F-FDG-PET-CT (n=186)	0.855 (0.726-0.984)	75.0 (47.4-91.7, 12/16)	94.7 (89.9-97.4, 161/170)	0.697	75.0 (47.4-91.7, 12/16)	96.5 (92.1-98.6, 164/170)	0.715
Only tonsil DWI (n=62)	0.845 (0.661-1.000)	75.0 (35.6-95.5, 6/8)	92.6 (81.3-97.6, 50/54)	0.676	37.5 (10.2-74.1, 3/8)	98.1 (88.8-99.9, 53/54)	0.356
Only tonsil ^{18}F-FDG-PET-CT (n=62)	0.880 (0.735-1.000)	87.5 (46.7-99.3, 7/8)	85.2 (72.3-92.9, 46/54)	0.727	87.5 (46.7-99.3, 7/8)	88.9 (76.6-95.4, 48/54)	0.764

[a] In one patient with a hypopharyngeal tumor we could not assess this subsite due to local image artifacts. We therefore excluded this lesion from the per-lesion analysis. If all artifacts would have been included as negative findings, then sensitivity would have been 81.3% (95%CI=53.7-95.0, ratio=13/16) with specificity being 96.5% (95%CI=92.1-98.6, ratio=164/170) with the sensitive approach. With the conservative approach sensitivity and specificity would have been 50.0% (95%CI=25.5-74.5, ratio=8/16) and 99.4% (95%CI=96.3-100, ratio=169/170), respectively.

When comparing DWI parameters between clinically UP tumors (n=15) and overt malignancies in the control group (n=20), we only found clinically UP tumors to be smaller than overt malignancies (*P*=0.033).

When only tonsils were included in the analysis we found lower ADC_{mean} (*P*=0.05), ADC_{min} (*P*=0.005), ADC_{median} (*P*=0.039), and a larger volume (*P*=0.001) of UP tonsillar tumors compared to benign tonsils resulting in AUCs ranging from 0.720 to 0.845 (Appendix A). When comparing UP tonsillar carcinoma to overt tonsillar tumors, we only found a significantly larger volume for overt tonsillar malignancy (*P*=0.016).

Interobserver agreement was very good for lesion volume (ICC=0.937) and substantial for other DWI parameters (ICC_{range}=0.605-0.793) (Table 4).

Based on SUV_{max} diagnostic accuracy was relatively high with AUC ranging from 0.811 to 0.906 depending on the approach, per-lesion or per-patient (Appendix A). Correcting SUV_{max} for blood pool uptake did not significantly change the results. SUV_{max} was significantly higher in overt primary tumors compared to UP tumors (per-

patient: median SUV_{max}=11.56 vs 6.43 (*P*=0.002); per-lesion: median SUV_{max}=10.88 vs 6.43 (*P*=0.05)). There was no difference in SUV_{max} between p16 or HPV positive and negative patients (data not shown).

Table 4 *Results of intraclass correlation coefficient and concurrent interpretation*

Parameter	ICC	Interpretation
ADC_{Mean}	0.781 (0.720-0.830)	Substantial
ADC_{SD}	0.629 (0.536-0.706)	Substantial
ADC_{Min}	0.737 (0.667-0.795)	Substantial
ADC_{Max}	0.656 (0.569-0.729)	Substantial
$ADC_{Skewness}$	0.647 (0.558-0.721)	Substantial
$ADC_{Kurtosis}$	0.605 (0.509-0.687)	Substantial
ADC_{median}	0.793 (0.735-0.840)	Substantial
ADC_{Volume}	0.937 (0.917-0.952)	Very good

DISCUSSION

We assessed the diagnostic value of DWI and ^{18}F-FDG-PET-CT for detecting primary tumors in patients presenting with nodal metastases of UP HNSCC. Reducing toxicity of extensive field radiotherapy outweighs the risk of an extra biopsy during EUA. Therefore, we considered high sensitivity of diagnostic imaging more important than high specificity. In qualitative analysis we found a non-significant trend towards higher sensitivity of ^{18}F-FDG-PET-CT compared to DWI (93.8% vs 81.3%). Because DWI and ^{18}F-FDG-PET-CT are based on different properties we hypothesized that combining modalities would improve diagnostic accuracy. However, we could not confirm this hypothesis: adding DWI did not improve the accuracy of ^{18}F-FDG-PET-CT. Single modality diagnostic accuracy was already high, leaving hardly any room for improvement.

For quantitative analysis ^{18}F-FDG-PET-CT outperformed DWI. This indicates that other factors than the quantified degree of diffusion restriction need to be taken into account (e.g. lesion size or asymmetry). Sensitivity of DWI and ^{18}F-FDG-PET-CT were comparable when only tonsils were assessed quantitatively, with ^{18}F-FDG-PET-CT being more specific.

Choi et al. retrospectively determined the added value of ADC histogram analysis to conventional MRI and ^{18}F-FDG-PET-CT in detecting occult tonsillar carcinoma (9). Occult tonsillar carcinoma was defined as occult after flexible endoscopy and contrast-enhanced CT, but pathologically proven after tonsillectomy or biopsy. We did not include CT findings because contrast-enhanced CT scans are not routinely performed in our institution for this indication. The authors found higher heterogeneity in occult tonsillar carcinoma, displayed by a higher ADC_{SD} than normal tonsils, with a sensitivity of 78.9% and specificity of 60.0%. Sensitivity increased to 94.7% after combining conventional MRI with ADC histogram analysis and ^{18}F-FDG-PET-CT. The authors found higher ADC values in malignant compared

to benign tonsils. This is in line with another study by Bhatia et al. who compared ADC values of normal tonsils to tonsillar carcinoma (27). In our population ADC values of tonsillar carcinoma were lower than benign tonsillar tissue. Our findings are in line with the general finding that HNSCC is associated with lower ADC values than benign tissue, because HNSCC consists of more densely packed cells than most benign tissues (18-20). Benign tonsils contain small lymphocytes which results in some degree of diffusion restriction. However more or less the same applies to lymph nodes where low ADC values are indicative of malignant infiltration (18). Choi et al. did report final ^{18}F-FDG-PET results; however no extensive analysis of ^{18}F-FDG-PET was performed. We used a more extensive analysis of ^{18}F-FDG-PET-CT by including SUV_{max} and correcting for blood pool activity. We also included the base of the tongue in routine image assessment.

In another study on the use of MRI including DWI and ^{18}F-FDG-PET-CT in UP HNSCC comparable results for both modalities were found with sensitivities of 88.2% and 94.4%, respectively (28). These findings are in line with the findings in this study. The authors used a comparable definition of UP HNSCC as in this study: biopsy proven nodal metastasis without an overt primary lesion after a complete head and neck physical examination. Unfortunately, the authors did not specify how images were analyzed except for that both modalities were assessed by two observers in consensus. We performed a more extensive qualitative and quantitative analysis and included interobserver statistics.

When deciding which modality to use, arguments for DWI are: more general availability of MRI compared to ^{18}F-FDG-PET-CT; no requirement of ionizing radiation; and lower costs of DWI. Argument to use ^{18}F-FDG-PET-CT are: more experience with ^{18}F-FDG-PET-CT for this indication; and that the intrinsic whole-body nature of ^{18}F-FDG-PET-CT allows for the detection of distant metastases. Because patients with unknown primary HNSCC present with nodal disease, they have a relatively high risk of developing distant metastases (29). Currently, we consider ^{18}F-FDG-PET-CT to be the preferred imaging technique to perform in the work-up of UP HNSCC. However, if ^{18}F-FDG-PET-CT is unavailable, DWI can be used as an alternative technique with high diagnostic accuracy.

Ideally imaging is performed before an EUA. Taking biopsies results in ^{18}F-FDG uptake or changes in diffusion parameters by local inflammation that may cause false-positive results (6). Secondly, more directed biopsies can be performed using imaging results, which may increase the yield of EUA. Therefore, we only included patients if imaging was performed before EUA.

Positive HPV status is associated with lower ADC and SUV_{max} values in overt HNSCC (30, 31). In occult HNSCC we could not confirm these results. The smaller size of occult HNSCC may result in a higher influence of partial volume effects and relatively large contributions of stromal tissue to the final ADC value of the lesion. This might have reduced the discriminatory effect of HPV status. Because SUV_{max} is volume dependent, assessing small lesions like occult HNSCC results in lower SUV_{max} values independent of HPV status. This makes ADC and SUV_{max} less valuable to predict HPV status in occult HNSCC.

This study had some limitations. Using the combination of all diagnostic procedures as reference standard has some limitations: 1) due to sampling errors with biopsies during EUA occult primary tumors may be missed; 2) missed occult primary tumors may not become clinically evident due to extensive field irradiation including potential primary tumor sites; and 3) because of the low incidence of UP HNSCC the patient population is relatively small. This resulted in wide confidence intervals for diagnostic accuracy. Our imaging protocols are continuously being improved, therefore a longer inclusion period would limit the comparability between patients. We did not assess the nasopharynx because it was often outside the field of view on DWI. However, none of the included patients was finally diagnosed with nasopharyngeal carcinoma. Qualitative image assessment always contains a degree of subjectivity, especially for determining the degree of diffusion restriction on DWI. It should be noted that a learning curve may be present for DWI analysis for this indication, and more experience may improve its results. We could not include the hypopharynx in routine quantitative analysis because the normal hypopharynx consists of a single layer of non-keratinizing squamous cell epithelium. The side of lymphadenopathy was known to the observers which may have biased observations of the potential primary tumor site towards the ipsilateral side. However, in clinical practice this information is also known. Finally, because ^{18}F-FDG-PET-CT is the current standard technique, ^{18}F-FDG-PET-CT was assessed by one resident under supervision of a nuclear medicine physician, which prevented us from performing interobserver analyses.

Conclusion

In this study on HNSCC patients presenting with clinically UP lesions the diagnostic accuracy of qualitative analysis with DWI and ^{18}F-FDG-PET-CT and quantitative analysis of ^{18}F-FDG-PET-CT using SUV_{max} were high. Adding DWI did not improve the accuracy of ^{18}F-FDG-PET-CT.

REFERENCES

1. Galloway TJ, Ridge JA. Management of Squamous Cancer Metastatic to Cervical Nodes With an Unknown Primary Site. J Clin Oncol. 2015;33(29):3328-37.
2. Calabrese L, Jereczek-Fossa BA, Jassem J, et al. Diagnosis and management of neck metastases from an unknown primary. Acta Otorhinolaryngol Ital. 2005;25(1):2-12.
3. Geltzeiler M, Doerfler S, Turner M, et al. Transoral robotic surgery for management of cervical unknown primary squamous cell carcinoma: Updates on efficacy, surgical technique and margin status. Oral Oncol. 2017;66:9-13.
4. Huang SH, Xu W, Waldron J, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. J Clin Oncol. 2015;33(8):836-45.
5. Pathmanathan R, Prasad U, Sadler R, Flynn K, Raab-Traub N. Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. N Engl J Med. 1995;333(11):693-8.
6. Johansen J, Buus S, Loft A, et al. Prospective study of 18FDG-PET in the detection and management of patients with lymph node metastases to the neck from an unknown primary tumor. Results from the DAHANCA-13 study. Head Neck. 2008;30(4):471-8.
7. Shoushtari A, Saylor D, Kerr KL, et al. Outcomes of patients with head-and-neck cancer of unknown primary origin treated with intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys. 2011;81(3):e83-91.
8. Lee HJ, Kim J, Woo HY, Kang WJ, Lee JH, Koh YW. 18F-FDG PET-CT as a supplement to CT/MRI for detection of nodal metastasis in hypopharyngeal SCC with palpably negative neck. Laryngoscope. 2015;125(7):1607-12.
9. Choi YJ, Lee JH, Kim HO, et al. Histogram Analysis of Apparent Diffusion Coefficients for Occult Tonsil Cancer in Patients with Cervical Nodal Metastasis from an Unknown Primary Site at Presentation. Radiology. 2016;278(1):146-55.
10. Lee JR, Kim JS, Roh JL, et al. Detection of occult primary tumors in patients with cervical metastases of unknown primary tumors: comparison of (18)F FDG PET/CT with contrast-enhanced CT or CT/MR imaging-prospective study. Radiology. 2015;274(3):764-71.
11. Regelink G, Brouwer J, de Bree R, et al. Detection of unknown primary tumours and distant metastases in patients with cervical metastases: value of FDG-PET versus conventional modalities. Eur J Nucl Med Mol Imaging. 2002;29(8):1024-30.
12. Rusthoven KE, Koshy M, Paulino AC. The role of fluorodeoxyglucose positron emission tomography in cervical lymph node metastases from an unknown primary tumor. Cancer. 2004;101(11):2641-9.
13. Nakamoto Y, Tatsumi M, Hammoud D, Cohade C, Osman MM, Wahl RL. Normal FDG distribution patterns in the head and neck: PET/CT evaluation. Radiology. 2005;234(3):879-85.
14. Vandecaveye V, De Keyzer F, Nuyts S, 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;67(4):960-71.
15. Razek AA, Megahed AS, Denewer A, Motamed A, Tawfik A, Nada N. Role of diffusion-weighted magnetic resonance imaging in differentiation between the viable and necrotic parts of head and neck tumors. Acta Radiol. 2008;49(3):364-70.
16. 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;168(2):497-505.
17. 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;23(2):562-9.
18. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of

diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46.

19. Srinivasan A, Dvorak R, Perni K, Rohrer S, Mukherji SK. Differentiation of benign and malignant pathology in the head and neck using 3T apparent diffusion coefficient values: early experience. AJNR Am J Neuroradiol. 2008;29(1):40-4.
20. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
21. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(2):122-37.
22. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-54.
23. Vandecaveye V, De Keyzer F, Dirix P, Lambrecht M, Nuyts S, Hermans R. Applications of diffusion-weighted magnetic resonance imaging in head and neck squamous cell carcinoma. Neuroradiology. 2010;52(9):773-84.
24. Bohning D, Holling H, Patilea V. A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. Stat Methods Med Res. 2011;20(5):541-50.
25. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
27. Bhatia KS, King AD, Yeung DK, et al. Can diffusion-weighted imaging distinguish between normal and squamous cell carcinoma of the palatine tonsil? Br J Radiol. 2010;83(993):753-8.
28. Godeny M, Lengyel Z, Polony G, et al. Impact of 3T multiparametric MRI and FDG-PET-CT in the evaluation of occult primary cancer with cervical node metastasis. Cancer Imaging. 2016;16(1):38.
29. Senft A, Hoekstra OS, Witte BI, Leemans CR, de Bree R. Screening for distant metastases in head and neck cancer patients using FDG-PET and chest CT: validation of an algorithm. Eur Arch Otorhinolaryngol. 2015.
30. Schouten CS, Hakim S, Boellaard R, et al. Interaction of quantitative (18)F-FDG-PET-CT imaging parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. Head Neck. 2016;38(4):529-35.
31. Driessen JP, van Bemmel AJ, van Kempen PM, et al. Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. Head Neck. 2015.

CHAPTER 4.3

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 Surgery, Oral Medicine, Oral Pathology and Oral Radiology 2017;124:296-305

Daniel P Noij
Viresh A Jagesar
Pim de Graaf
Marcus C de Jong
Otto S Hoekstra
Remco de Bree
Jonas A Castelijns

ABSTRACT

Objective: Diagnosing residual malignancy after (chemo)radiotherapy presents a diagnostic challenge due to overlapping symptoms and imaging characteristics. We assessed the added diagnostic value of diffusion-weighted imaging (DWI) to fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) in head and neck squamous cell carcinoma (HNSCC) patients with residual ^{18}F-FDG uptake at the primary tumor site three months after (chemo)radiotherapy.

Study Design: For this retrospective study from January 2010 to June 2012, 22 patients (median age, 61 years; range, 41-77 years) were included for analysis. Both ^{18}F-FDG-PET-CT and MRI including DWI were performed as a part of the institutional protocol and were qualitatively assessed for the presence of residual malignancy at the primary tumor site.

Results: Sensitivity and specificity of ^{18}F-FDG-PET-CT were 100% and 47%, respectively. For DWI, sensitivity and specificity were 80% and 82%, respectively. When DWI was added to ^{18}F-FDG-PET-CT with residual ^{18}F-FDG uptake and only a positive read on both ^{18}F-FDG-PET-CT and DWI was considered to be overall positive, sensitivity remained 80% (95%CI: 28-99%), and specificity was 88% (95%CI: 64-99%).

Conclusions: In this pilot study of the selected patients with residual ^{18}F-FDG -uptake at the primary tumor site 3 months after (chemo)radiotherapy, we demonstrated that the addition of DWI to ^{18}F-FDG-PET-CT has the potential to increase the specificity of the response evaluation with limited decrease of sensitivity.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) accounts for 4% of malignancies worldwide (1). Approximately 30% of patients present with advanced disease (2). Treatment with curative intent consists of surgery, radiotherapy, and chemotherapy alone or combined. To decrease morbidity, non-surgical treatments are increasingly applied, with reported locoregional control rates of 43-96% for (chemo)radiotherapy in locally advanced tumors (3). Local control and survival rates rapidly decrease when salvage surgery is delayed. For example, in patients treated by radiotherapy for resectable oropharyngeal carcinoma, salvage surgery at 1-2 months was successful in 70% of residues detected at the time of response evaluation but in 33% of the later detected recurrences (4).

Differentiation between post-radiation changes and residual malignancy can be challenging due to overlapping symptoms such as hoarseness, pain, and swallowing complaints. Taking repeated biopsies confers risks of infection, chondritis, and edema, thereby exacerbating radiotherapy effects (5). Therefore, non-invasive techniques are warranted to reliably select patients for examinations under anesthesia (EUA) to avoid unnecessary examinations and reduce risks of complications, patient burden, and cost.

Positron emission tomography combined with computed tomography (PET-CT) has often been used to detect residual HNSCC, mostly using fluorodeoxyglucose (^{18}F-FDG). Current consensus is that a negative ^{18}F-FDG-PET-CT is highly reliable, with reported negative predictive values (NPV) for the presence of malignancy of 92-99% (6-10). However, inflammatory post-irradiation effects can compromise its diagnostic specificity; together with the typically low prevalence of local residual disease, the positive predictive value of ^{18}F-FDG-PET-CT is suboptimal (6-10). Hence, residual ^{18}F-FDG uptake after (chemo) radiotherapy warrants further investigation. Improving specificity without compromising sensitivity might reduce the number of unnecessary (i.e., tumor-negative) biopsies (5).

Diffusion-weighted imaging (DWI) has shown promising results for detecting residual locoregional disease after (chemo)radiotherapy in HNSCC with high sensitivity (80-100%) and specificity (90-100%) (11-15).

As ^{18}F-FDG-PET-CT and DWI are based on different biochemical properties, combining both modalities may be synergistic in detecting local residue after (chemo)radiotherapy. Because of the high false-positive rate (1-specificity) and high NPV of ^{18}F-FDG-PET-CT, further analysis of lesions with ^{18}F-FDG uptake may be an effective way to reduce unnecessary EUAs. Vandecaveye et al. (11) found a specificity of 95% for DWI in the differentiation between recurrent disease and complete remission in HNSCC patients three weeks after chemoradiotherapy. DWI appears to be able to detect lesions with a diameter of at least 4 mm, whereas with ^{18}F-FDG-PET-CT it is difficult to characterize sub-centimeter lesions (16, 17).

In response evaluation of advanced nodal disease after chemoradiotherapy, ^{18}F-FDG-PET-CT showed a sensitivity of 100% and a specificity of 84% for the detection of residual neck disease, while these figures for DWI were 60% and 93%, respectively (18). Adding DWI to ^{18}F-FDG-PET-CT increased the specificity of ^{18}F-FDG-PET-CT alone: sensitivity was 100% and specificity was 95% (18). This increase in specificity may ensure that fewer patients are exposed to unnecessary neck dissections in clinical practice.

The primary aim of this pilot study was to assess the potential added diagnostic value of DWI to ^{18}F-FDG-PET-CT in HNSCC patients with residual ^{18}F-FDG uptake at the primary tumor site three months after (chemo)radiotherapy. The hypothesis was that combining the specific DWI and the sensitive ^{18}F-FDG-PET-CT would result in higher specificity than with ^{18}F-FDG-PET-CT alone, without compromising sensitivity.

METHODS & MATERIALS

We used the Standards for Reporting of Diagnostic Accuracy (STARD) statement as a guideline for the study methods (19).

Patients

This retrospective study was conducted at a tertiary referral center (VU University Medical Center, Amsterdam, the Netherlands) for HNSCC and approved by the institutional review board, with a waiver of informed consent. We consecutively included patients who received (chemo)radiotherapy for HNSCC from January 2010 until June 2012.

Inclusion criteria were: 1) previously untreated, histopathologically proven HNSCC treated with primary (chemo)radiotherapy with curative intent; 2) residual ^{18}F-FDG-uptake at the primary tumor site on ^{18}F-FDG-PET-CT at three months after the end of (chemo) radiotherapy as reported by the attending nuclear medicine physician; 3) magnetic resonance imaging (MRI) with DWI series at three months follow-up; 4) interval between ^{18}F-FDG-PET-CT and MRI less than one month; and 5) biopsies taken during EUA at three months follow-up. Histopathological evidence of residual disease during clinical follow-up was used as the reference standard. In our institution, follow-up procedures at three months after (chemo)radiotherapy consisted of routine clinical examination, ^{18}F-FDG-PET-CT, and MRI including DWI and EUA. Thereafter, patients received clinical follow-up with two- to three-monthly clinical examinations by a head and neck surgeon for the first two years. Additional diagnostic procedures (e.g., additional imaging with either MRI, ^{18}F-FDG-PET-CT, CT or ultrasound, and EUA) were performed when deemed necessary by the attending physician.

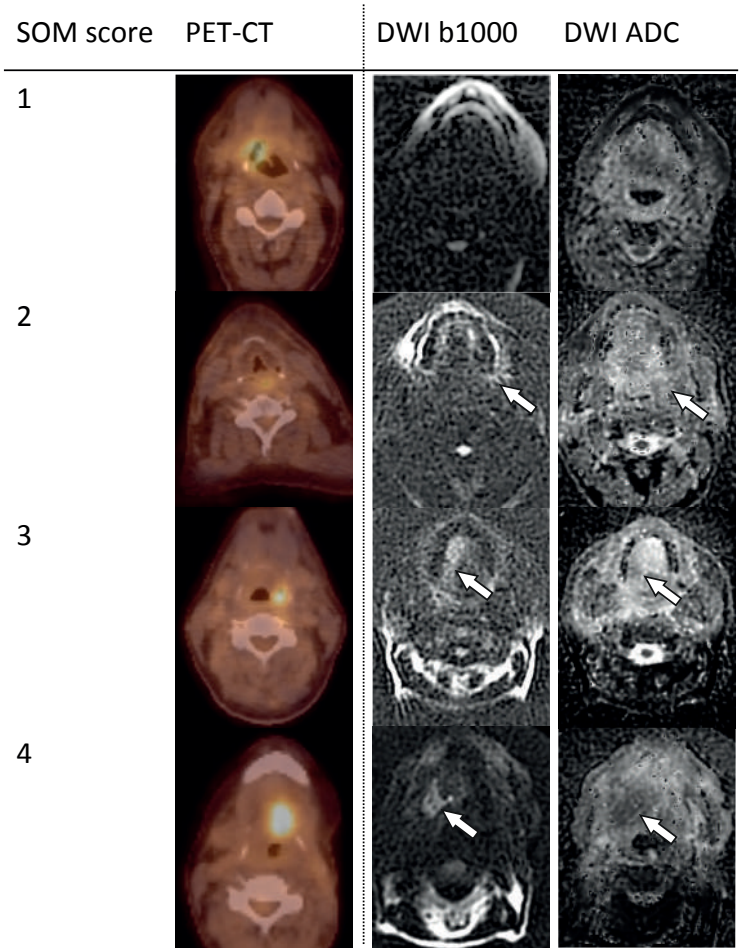


Figure 1 *Axial images for each SOM score for PET-CT and DWI. As both modalities were assessed independently, PET-CT and DWI images with identical SOM scores are not from the same patients. The arrows mark the area of diffusion restriction.*

Table 1 *Patients characteristics*

No	Gender	Age	Tumor location	TNM	Treatment (dose (Gy))	Chemotherapy	PET-CT (SOM score)	DWI (SOM score)	Local outcome (duration of follow-up (months))
1	Female	64	Oropharynx	T2N2b	CRT (70)	Cisplatin	Negative (1)	Negative (1)	CR (51)
2	Male	49	Hypopharynx	T2N2a	CRT (70)	Cisplatin	Negative (1)	Negative (1)	CR (31)
3	Male	63	Oropharynx	T4aN0	CRT (70)[a]	Cisplatin	Negative (1)	Negative (1)	CR (30)
4	Male	59	Oropharynx	T3N2c	CRT (70)	Cisplatin	Positive (3)	Negative (1)	CR (29)
5	Male	53	Oropharynx	T3N2b	CRT (70)	Cisplatin	Negative (1)	Negative (1)	CR (27)
6	Female	50	Oropharynx	T4bN2b	CRT (70)	Cisplatin	Negative (1)	Negative (1)	CR (24)
7	Female	55	Oropharynx	T3N2b	CRT (70)	Cisplatin	Positive (4)	Negative (1)	CR (50)
8	Female	63	Oropharynx	T4N0	CRT (70)[b]	Cetuximab	Negative (1)	Negative (1)	CR (48)
9	Female	64	Hypopharynx	T3N0	CRT (70)	Cisplatin	Positive (3)	Negative (1)	CR (17)
10	Male	63	Larynx	T3N0	CRT (70)	Cetuximab	Positive (3)	Positive (4)	CR (4)
11	Female	77	Hypopharynx	T4bN0	CRT (70)	Cetuximab	Positive (3)	Inadequate quality	Residue (4)
12	Male	61	Oropharynx	T1N2b	CRT (70)	Cisplatin	Negative (1)	Negative (1)	CR (34)
13	Male	73	Oropharynx	T4aN0	CRT (70)	Cetuximab	Positive (4)	Positive (4)	Residue (5)
14	Male	41	Larynx	T2N2c	CRT (70)	Cisplatin	Positive (3)	Negative (1)	CR (30)
15	Female	44	Oral cavity	T4aN2c	CRT (70)	Cisplatin	Positive (2)	Positive (3)	Residue (3)
16	Male	62	Oropharynx	T3N0	CRT (70)	Cisplatin	Positive (4)	Positive (3)	CR (29)
17	Male	61	Oropharynx	T1N2b	CRT (70)	Cisplatin	Positive (3)	Negative (1)	CR (29)
18	Male	54	Hypopharynx	T3N2c	CRT (70)[c]	Cisplatin	Positive (2)	Negative (1)	CR (4)
19	Male	60	Hypopharynx	T4aN2c	CRT (70)[a]	Cisplatin	Negative (1)	Positive (2)	CR (26)
20	Male	60	Oropharynx	T4aN0	RT (70)[c]	-	Positive (4)	Positive (4)	Residue (3)
21	Male	49	Oropharynx	T4bN1	CRT (70)[d]	Cisplatin	Positive (4)	Positive (2)	Residue (11)
22	Male	64	Larynx	T2N0	RT (70) [c]	-	Negative (1)	Inadequate quality	CR (5)
23	Male	65	Oropharynx	T3N3	CRT (70)	Cetuximab	Positive (2)	Negative (1)	Residue (1)
24	Male	64	Larynx	T2N1	RT (70)[c]	-	Positive (2)	Negative (1)	CR (26)

[a] Also received induction chemotherapy (2x TPF) and weekly cisplatin 40 mg/m^2
[b] Accelerated radiotherapy with 6 radiation treatments per week
[c] Received weekly cisplatin 40 mg/m2
[d] Also received induction chemotherapy (4x TPF)

Abbreviations: CR = Complete remission; CRT = Chemoradiotherapy; DWI = Diffusion-weighted imaging; PET-CT = Positron emission tomography computed tomography; RT = Radiotherapy; SOM = Suspicion of malignancy

From 208 patients with HNSCC treated with (chemo)radiotherapy with curative intent, we identified 24 eligible patients (median age, 61 years; range, 41-77 years; male/female ratio=17/7). The reasons for exclusion were: 1) no ^{18}F-FDG-uptake at the primary tumor site; 2) no ^{18}F-FDG-PET-CT or DWI available; or 3) an interval between ^{18}F-FDG-PET-CT and MRI of greater than one month. Both ^{18}F-FDG-PET-CT and MRI were performed within a median interval of 6 days (interquartile range (IQR), 2-27 days). In 13 patients ^{18}F-FDG-PET-CT was performed first, which was dictated by logistics. An EUA was performed within a median of five days after the last imaging test (IQR, 3-12 days). Patient characteristics are shown in Table I.

Imaging

^{18}F-FDG-PET-CT was performed as described previously (20). In short, PET and low-dose CT were performed from the mid-thigh to the skull vertex with the arms elevated over the head after a six hour fasting period and adequate hydration. Procedures and imaging reconstruction were compatible with European Association of Nuclear Medicine (EANM) guidelines (21).

Magnetic resonance imaging was performed at 1.5 T with three MRI systems as dictated by logistics (Sonata (n=3) and Avanto (n=3); Siemens, Erlangen, Germany and Signa HDxt (n=18); GE Healthcare, Milwaukee, WI, United States), using a head coil combined with a phased-array spine and neck coil. Axial images (22 sections of 4 mm section thickness, 0.4 mm gap, in-plane pixel size of 0.9×0.9 mm) were obtained with Short TI Inversion Recovery (STIR), DWI, and T1-weighted imaging before and after the administration of contrast material (0.2 ml/kg gadobutrol (Gadovist; Bayer Schering AG, Berlin, Germany) (n=2) or 0.4 ml/kg gadoteric acid (Dotarem; Guerbet, Roissy, France) (n=22)).

Diffusion-weighted imaging was acquired with either echo-planar imaging (EPI) or turbo spin-echo (TSE) (i.e., half-Fourier acquisition single-shot turbo spin-echo (HASTE) and periodically rotated overlapping parallel lines with enhanced reconstruction (PROPELLER)) sequences. On the Sonata and Avanto units EPI-DWI and HASTE-DWI were performed (n=6), on the Signa HDxt unit PROPELLER-DWI was performed (n=18; Table 2).

Table 2 *DWI imaging protocol for each MRI system*

Sequence	Avanto (n=3)	Sonata (n=3)	Signa HDxt (n=18)
EPI-DWI (TR/TE, no of averages, b-values)	5000/111 ms, 3 averages, b = 0/500/1000 s/mm^2	5000/95 ms, 3 averages, b = 0/500/1000 s/mm^2	-
TSE-DWI (sequence, TR/TE, no of averages, b-values)	HASTE, 729/113 ms, 2 averages, b = 0/750/1000 s/mm^2	HASTE, 900/110 ms, 3 averages, b = 0/750/1000 s/mm^2	PROPELLER, 3500/83.87, 2 averages, b= 0/750/1000 s/mm^2

Abbreviations: CE-T1 = Contrast-enhanced T1; EPI-DWI = Echo planar imaging diffusion-weighted imaging; HASTE = Half-Fourier Acquisition Single-Shot Turbo Spin-Echo; PROPELLER = Periodically Rotated Overlapping Parallel Lines with Enhanced Reconstruction; SE-T1 = Spin-echo T1; STIR = Short TI Inversion Recovery; TE = Excitation time; TR = Relaxation time; TSE-DWI = Turbo spin-echo diffusion weighted imaging

Image analysis
All observers were aware that all included patients had residual ^{18}F-FDG uptake at the primary tumor site as reported by the attending nuclear medicine physician. Observers had access to basic patient information (age, gender, and treatment), tumor location, pretreatment tumor stage, and pretreatment imaging. Observers were blinded to findings of the other observers, EUA results, and treatment outcome. However, the surgeon who performed the EUA was aware of imaging findings.

A nuclear medicine physician with 26 years of experience qualitatively re-evaluated ^{18}F-FDG-PET-CT images using a 4-point Likert scale; the suspicion of malignancy (SOM) was scored as follows: 1=low suspicion, 2=moderate suspicion, 3=substantial suspicion, 4=high suspicion. The level of suspicion was deduced from the localization of the abnormal ^{18}F-FDG uptake, its aspect (focal uptake increasing the level of suspicion), and level of intensity compared to the surrounding background and contralateral physiological uptake (higher uptake increasing the level of suspicion).

Two radiologists with 30 (referred to as radiologist 1) and seven years (referred to as radiologist 2) of experience in head and neck radiology independently qualitatively re-evaluated DWI to determine the SOM score using the same 4-point Likert scale as for ^{18}F-FDG-PET-CT. Pretreatment and posttreatment anatomical MR sequences and pretreatment ^{18}F-FDG-PET-CT images were used only for tumor localization to ensure that the same lesions were assessed with both modalities. Malignancy was suspected in focal lesions with high signal intensity on high b-value imaging combined with low signal intensity on the apparent diffusion coefficient (ADC) map. In six patients both EPI and HASTE had been acquired; after scoring each series separately, a single SOM score was given for DWI findings in these patients. Discrepancies between the radiologists were resolved by consensus. Representative images of cases SOM scores of 1 through 4 are shown in Figure 1.

Statistical analysis
Receiver operating characteristic (ROC) analysis was used to determine the optimal cut-off of the SOM score for PET-CT and DWI to calculate diagnostic accuracy with a high weight on sensitivity to prevent failure to detect residual disease. Local progression-free survival was analyzed with the Kaplan-Meier product-limit method and log-rank tests to compare subgroups. To assess potential bias introduced by using different MRI systems, we performed a subgroup analysis for the most frequently used MRI system. Interobserver variability was assessed using weighted Kappa analysis and the proportion specific agreement (22). Kappa was interpreted as follows: 0.00-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=substantial; and 0.81-1.00=almost perfect agreement (23). The proportion specific agreement consisted of two parts: the agreement on positive ratings (PA) (i.e., presence of residual disease) and the agreement on negative ratings (NA) (i.e., absence of residual disease). Point estimates and distribution were reported as mean ± SD, unless stated otherwise. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) (version 20.0; Chicago, IL, USA). Two-sided *P*-values <0.05 were considered significant.

RESULTS

Two patients (patient 11 and 22) were excluded from further analysis due to significant artifacts and image distortion on DWI. From the remaining 22 evaluable patients, five (23%) were diagnosed with histopathologically proven residue at the primary tumor site. One patient with tumor residue had received only accelerated radiotherapy (patient 20). Another patient with residual disease had received induction chemotherapy followed by concurrent chemoradiotherapy (patient 21) (Table 1). Median follow-up after treatment was 29 months (IQR=25-33 months) in patients with complete remission. Residual disease was discovered at a median follow-up of three months (IQR, 2-8 months).

Positron emission tomography

Of the finally included patients, SOM scores were 1 in eight patients, 2 in four patients, 3 in five patients and 4 in five patients (Table 1). The two earlier mentioned patients who were excluded had SOM scores of 1 and 3. ROC analysis resulted in an area under the curve (AUC) of 0.78 (95% confidence interval (95%CI)=0.55-1.00). Sensitivity decreased from 100% to 60% when the threshold of test positivity was increased from a SOM score of 1 vs 2-4 to 1-3 vs 4, whereas specificity increased from 47% to 88% (Table 3). To optimize sensitivity, we considered SOM scores of 1 (n=8) as negative and SOM scores ≥2 (n=14) as positive, resulting in a sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the detection of malignancy of 100% (95%CI=48-100%), 47% (95%CI=23-72%), 36% (95%CI=13-65%) and 100% (95%CI=63-100%), respectively (Tables 3 and 4). Patients with a negative ^{18}F-FDG-PET-CT result (SOM score 1) tended to have better local progression-free survival than those with a positive one (*P*=0.054) (Figure 2a).

Table 3 *Operational characteristics of PET-CT, DWI and combined PET-CT and DWI accuracy in the 22 patients*

	Threshold between positive and negative findings, % (95%CI)		
	1-2	2-3	3-4
PET-CT (n=22)			
Sensitivity	100 (48-100)	60 (15-95)	60 (15-95)
Specificity	47 (23-72)	59 (33-82)	88 (64-99)
PPV	36 (13-65)	30 (7-65)	60 (15-95)
NPV	100 (63-100)	83 (52-98)	88 (64-99)
DWI (n=22)			
Sensitivity	80 (28-99)	60 (15-95)	40 (5-85)
Specificity	82 (57-96)	88 (64-99)	94 (71-100)
PPV	57 (18-90)	60 (15-95)	67 (9-99)
NPV	93 (68-100)	88 (64-99)	84 (60-97)
PET-CT and DWI combined (n=22)[a]			
Sensitivity	80 (28-99)	60 (15-95)	40 (5-85)
Specificity	88 (64-99)	88 (64-99)	94 (71-100)
PPV	67 (22-96)	60 (15-95)	67 (9-99)
NPV	94 (70-100)	88 (64-99)	84 (60-97)

[a] Only the lowest threshold for SOM in PET-CT had a sensitivity of 100%. In order not to miss residual disease we did not use other SOM score thresholds for PET-CT.

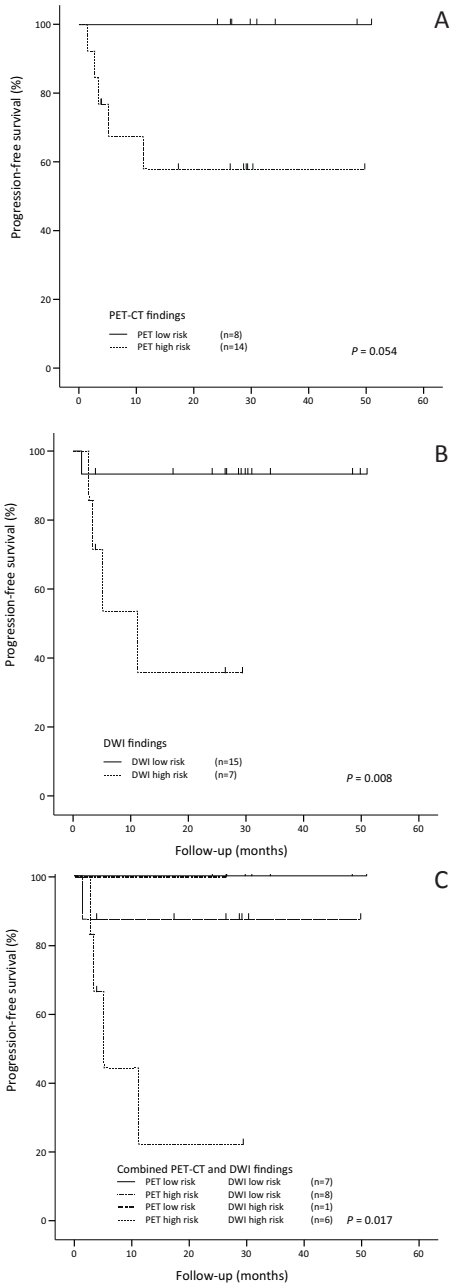


Figure 2 *Kaplan-Meier curves of local progression-free survival for patients: A) considered low and high risk based on PET-CT findings (**P**=0.054); B) considered low and high risk based on DWI findings (**P**=0.008); C) for combined PET-CT and DWI findings (**P**=0.017)*

Diffusion-weighted imaging

SOM-scores after consensus were 1 in 15 patients, 2 in two patients, 3 in two patients and 4 in three patients. One patient with residual disease had an SOM score of 1 (patient 23). This patient had a small tumor residue at the uvula without diffusion restriction (Table 1). ROC analysis resulted in an AUC of 0.82 (95%CI=0.59-1.00). Sensitivity decreased from 80% to 40% while increasing the threshold of test positivity from SOM scores of 1 vs 2-4 to 1-3 vs 4. Specificity increased from 82% to 94% (Table 3). To optimize sensitivity we considered SOM scores of 1 (n=15) as negative and SOM scores ≥2 (n=7) as positive, which yielded a sensitivity, specificity, PPV and NPV of 80% (95%CI=28-99%), 82% (95%CI=57-96%), 57% (95%CI=18-90%) and 93% (95%CI=68-100%), respectively (Tables 3 and 4). Patients with a negative DWI test result (SOM score 1) had significantly better local progression-free survival than those with a positive DWI (*P*=0.008) (Figure 2b). There was substantial agreement between the DWI readers (weighted kappa=0.65 (95%CI=0.18-0.75)), with lower proportion specific agreement for positive ratings than for negative ones (agreement for positive ratings=0.63 (95%CI=0.35-0.85), vs agreement for negative ratings=0.79 (95%CI=0.59-0.92); Appendix A).

In a subgroup analysis of the most frequently used MRI system (Signa HDxt (n=16)), we found similar results (sensitivity, specificity, PPV and NPV were 80% (95%CI=28-99%), 72% (95%CI=39-94%), 57% (95%CI=18-90%) and 73% (95%CI=39-94%), respectively). There were no discrepancies in SOM scores between EPI and HASTE in the six patients in whom both sequences were performed.

Imaging modalities combined
Only the lowest threshold for SOM in ^{18}F-FDG-PET-CT had a sensitivity of 100%. In order not to miss residual disease we did not use other SOM score thresholds for ^{18}F-FDG-PET-CT. The use of other of thresholds of positivity for DWI did improve combined test specificity from 88% (95%CI=64-99%) to 94% (95%CI=71-100%), however at the expense of sensitivity, which decreased from 80% (95%CI=28-99%) to 40% (95%CI=5-85%) (Table 3).

When determining the combined score of ^{18}F-FDG-PET-CT and DWI, we only considered a positive read on both ^{18}F-FDG-PET-CT and DWI (i.e., SOM score >1) to be a positive read in the combined score. Remaining score combinations were considered to be negative for malignancy. This resulted in a sensitivity, specificity, PPV and NPV of 80% (95%CI=28-99%), 88% (95%CI=64-99%), 67% (95%CI=22-96%) and 94% (95%CI=70-100%), respectively (Tables 3 and 4). Imaging of two representative cases in which ^{18}F-FDG-PET-CT was positive for malignancy are shown in Figure 3 (negative results on DWI) and Figure 4 (positive results on DWI).

Local progression-free survival was best in patients with negative findings on both modalities. If both DWI and ^{18}F-FDG-PET-CT findings were positive for malignancy, local progression-free survival was the lowest (*P*=0.017) (Figure 2c).

Decreasing the number of EUAs
When EUA would have been performed only on the basis of positive scores on ^{18}F-FDG-PET-CT, 64% (9/14) of EUAs would be unnecessary and all cases of residue would have been detected. If ^{18}F-FDG-PET-CT and DWI were combined and only patients with positive test findings on both techniques would undergo EUA, 33% (2/6) of EUAs would have been unnecessary, however at the expense of missing one patient with tumor residue (Appendix B).

DISCUSSION

We assessed the added value of DWI to ^{18}F-FDG-PET-CT in patients with residual ^{18}F-FDG-uptake at the primary tumor site for response evaluation three months after (chemo) radiotherapy. When only ^{18}F-FDG-PET-CT in this cohort was used, sensitivity was 100% and specificity was 47%. The combination of DWI and ^{18}F-FDG-PET-CT, considering only patients with positive scores on both techniques to be overall positive, resulted in a sensitivity and specificity of 80% and 88%, suggesting that DWI and ^{18}F-FDG-PET-CT may be complementary. Diffusion-weighted imaging also had additional prognostic value, as positive findings on both DWI and ^{18}F-FDG-PET-CT resulted in the worst local progression-free survival (*P*=0.017). The combination of DWI and ^{18}F-FDG-PET-CT in response evaluation three months after (chemo)radiotherapy may therefore be a valuable application for PET-MRI.

In our institution patients receive routine EUAs three months after (chemo)radiotherapy. The most sensitive strategy (i.e., missing the fewest residues, with the least imaging and with the lowest number of unnecessary EUAs) for selecting patients for EUA would have been performing EUA based only on ^{18}F-FDG-PET-CT findings, since no tumor residues

were missed. Then in 64% (9/14) of patients with residual ^{18}F-FDG-uptake at the primary tumor site an unnecessary EUA would have been performed (Table 4). If only patients with positive test results on both ^{18}F-FDG-PET-CT and DWI would have received an EUA, unnecessary EUAs would have been performed in only 33% (2/6) of patients. However, this reduction of EUAs would have been at the expense of missing one tumor residue.

Table 4 *Diagnostic accuracy of PET-CT, DWI and combined PET-CT and DWI for the detection of tumor residue at the primary tumor site For the individual imaging tests, we classified a SOM score of 1 as a negative test results. When combining PET-CT and DWI only patients with positive findings on both techniques were considered overall positive.*

	PET-CT			**DWI**			**PET-CT and DWI combined**		
	Residue	Complete remission	Total	Residue	Complete remission	Total	Residue	Complete remission	Total
Positive	5	9	14	4	3	7	4	2	6
Negative	0	8	8	1	14	15	1	15	16
Total	5	17	22	5	17	22	5	17	22

Abbreviations: DWI = Diffusion-weighted imaging; PET-CT = Positron emission tomography computed tomography

Currently, ADC is not considered interchangeable between MRI systems and sequences (24). Moreover, consensus is lacking on sequences, b-values, and software (17). This is reflected by conflicting study findings regarding the presence (8, 25) or absence (26, 27) of a significant correlation between ADC and standardized uptake value (SUV) in HNSCC. When the ^{18}F-FDG-PET-CT guidelines as proposed by Boellaard et al. (21) are followed, standardized uptake values (SUVs) can be regarded as interchangeable between ^{18}F-FDG-PET-CT-scanners.

In this study, bias may be introduced by using multiple MRI systems and sequences; however, conclusions were identical in all 6 patients where both EPI and HASTE were performed. Furthermore, we performed a subgroup analysis by including only the most commonly used MRI system with its specific sequence, which revealed findings comparable to those of the whole group. This suggests that the specific MRI system and diffusion sequence are of limited importance when image analysis is performed in a qualitative manner.

For DWI, interobserver variability was substantial (i.e., weighted kappa=0.65) and PA was lower than NA (0.63 and 0.79, respectively). The presence of restricted diffusion seems to be more susceptible to variable scoring than its absence.

King et al. (12) performed DWI six weeks after (chemo)radiotherapy in 20 patients with HNSCC. ROC analysis was used to determine the optimal ADC threshold for locoregional failure with the highest specificity. Sensitivity, specificity, PPV and NPV were 45%, 100%, 100% and 63%, respectively. ROC analysis is mostly used to find the combination of the highest sensitivity and highest specificity. However, since missing residual disease outweighs an unnecessary EAU with biopsies, sensitivity is more important than specificity in this clinical problem. King et al. stated that DWI may rival the results of 18F-FDG-PET-CT.

In another study performed by Yu et al. (15), 41 patients with oropharyngeal SCC were included. DWI was performed eight weeks after treatment and 18F-FDG-PET-CT 12 weeks after treatment. For DWI, the optimal ADC threshold resulted in sensitivity, specificity, PPV, and NPV values of 100%, 92%, 50%, and 100%, respectively and 100%, 71%, 23%, and 100% for PET-CT, respectively. In this study the authors suggested that DWI may rival the results of 18F-FDG-PET-CT. In our study, DWI had higher specificity than 18F-FDG-PET-CT (82% vs 47%) in patients with residual 18F-FDG-uptake at the primary tumor site after (chemo)radiotherapy. However, 18F-FDG-PET-CT had higher sensitivity than DWI (100% vs 80%). As mentioned earlier, our findings suggest complementary instead of rivaling roles for 18F-FDG-PET-CT and DWI. Furthermore, the standardization issues of ADC need to be overcome before ADC-based decisions can be incorporated into clinical practice.

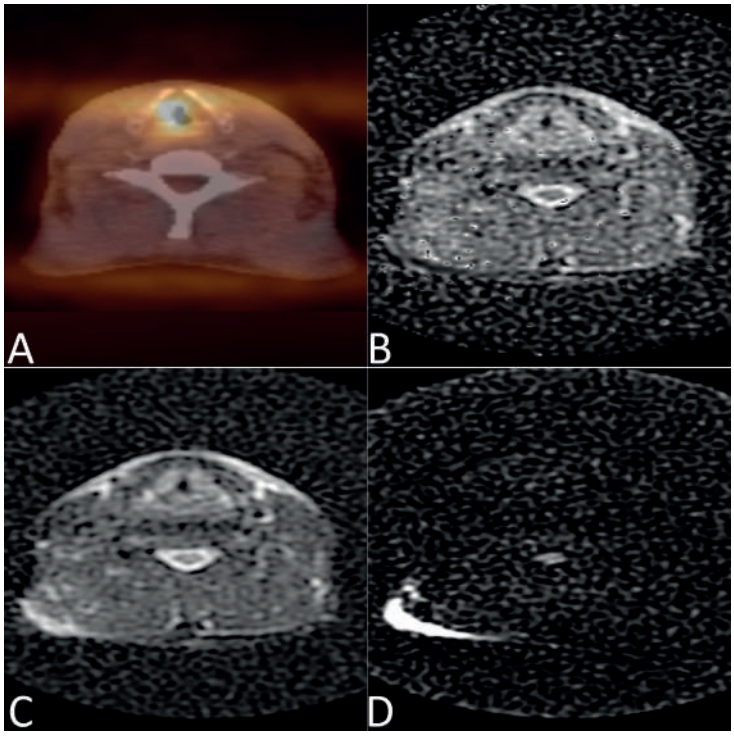


Figure 3 *Representative images of a patient with negative DWI findings: Axial images of a 41-year old male 3 months after chemoradiotherapy for a T2N2c supraglottic laryngeal carcinoma (patient 14). On (A) PET-CT there was ^{18}F-FDG-uptake at the right supraglottic level which was substantially suspicious of malignancy with a SOM score of 3. On the PROPELLER DWI series ((B) ADC map, (C) b=0 s/mm^2 and (D) b=1000 s/mm^2) there was no suspicion of malignancy (i.e., there is neither high signal intensity on (D) b=1000 s/mm^2 nor low signal intensity on the (C) ADC map) with a SOM score of 1. Biopsy did not reveal malignancy and at 30 months of follow-up the patient is still in complete remission. In this patient PET-CT findings were therefore considered false-positive and DWI findings true-negative.*

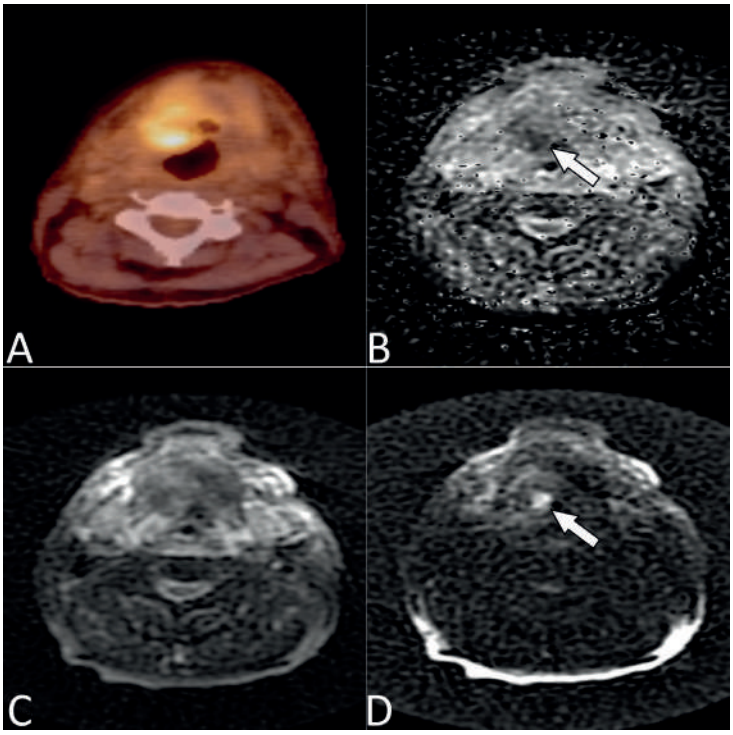


Figure 4 *Representative images of a patient with positive DWI findings: Axial images of a 73-year old male 3 months after chemoradiotherapy for a T4N0 oropharyngeal carcinoma (patient 13). On (A) PET-CT there was ^{18}F-FDG-uptake at the right vallecula which was highly suspicious of malignancy with a SOM score of 4. On the PROPELLER DWI series ((B) ADC map, (C) b=0 s/mm^2 and (D) b=750 s/mm^2) there was also high risk of malignancy (i.e., there is high signal intensity on (D) b=1000 s/mm^2 and low signal intensity on the (C) ADC map (arrow)) with a SOM score of 4. Biopsy confirmed the presence of malignancy. In this patient both PET-CT and DWI findings are true-positive.*

This study had some limitations. Firstly, we only assessed patients with residual ^{18}F-FDG-uptake three months after (chemo)radiotherapy. This leads to selection bias and makes it difficult to compare our study results to those of other groups. However, we considered it valid to assess only patients with residual ^{18}F-FDG-uptake, because diagnostic problems arise in these patients in clinical practice. Secondly, we only assessed the primary tumor site and not lymph nodes or distant sites. Disease progression to lymph nodes or distant sites has major implications for patient prognosis. Thirdly, we used threshold values for positive test results with the highest sensitivity in this cohort of patients. The validity of these threshold values should ideally be verified in future prospective studies. Moreover, the acceptable amount of decrease in sensitivity (risk of missing of residual tumor) to increase the specificity (diminishing unnecessary EUA) should be debated.

Conclusion

In this pilot study of a selected population of patients with residual ^{18}F-FDG -uptake at the primary tumor site three months after (chemo)radiotherapy, we demonstrated that the addition of DWI to ^{18}F-FDG-PET-CT has the potential to substantially increase the specificity of response evaluation by imaging with limited decrease of sensitivity. The most sensitive strategy for the selection of patients for EUA was using only PET-CT. To limit unnecessary EUAs, DWI can be added to increase the efficacy of selecting patients for EUA by diagnostic imaging. These results have to be confirmed by larger prospective (multicenter) studies. Combining DWI and PET in detecting residual disease after (chemo)radiotherapy may be a valuable application of PET-MRI.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015;65(1):5-29.
2. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Bethesda, MD2014.
3. Alkureishi LW, de Bree R, Ross GL. RADPLAT: an alternative to surgery? Oncologist. 2006;11(5):469-80.
4. Yom SS, Machtay M, Biel MA, et al. Survival impact of planned restaging and early surgical salvage following definitive chemoradiation for locally advanced squamous cell carcinomas of the oropharynx and hypopharynx. Am J Clin Oncol. 2005;28(4):385-92.
5. de Bree R, Putten Lvd, Brouwer J, Castelijns JA, Hoekstra OS, René Leemans C. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 2009;45(4-5):386-93.
6. Ryan WR, Fee WE, Jr., Le QT, Pinto HA. Positron-emission tomography for surveillance of head and neck cancer. Laryngoscope. 2005;115(4):645-50.
7. Malone JP, Gerberi MA, Vasireddy S, et al. Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography. Arch Otolaryngol Head Neck Surg. 2009;135(11):1119-25.
8. Nakajo M, Nakajo M, Kajiya Y, et al. FDG PET/CT and Diffusion-Weighted Imaging of Head and Neck Squamous Cell Carcinoma Comparison of Prognostic Significance Between Primary Tumor Standardized Uptake Value and Apparent Diffusion Coefficient. Clin Nucl Med. 2012;37(5):475-80.
9. Ho AS, Tsao GJ, Chen FW, et al. Impact of positron emission tomography/computed tomography surveillance at 12 and 24 months for detecting head and neck cancer recurrence. Cancer. 2013;119(7):1349-56.
10. van der Putten L, Hoekstra OS, de Bree R, et al. 2-Deoxy-2[F-18]FDG-PET for detection of recurrent laryngeal carcinoma after radiotherapy: interobserver variability in reporting. Mol Imaging Biol. 2008;10(5):294-303.
11. Vandecaveye V, Dirix P, De Keyzer F, et al. Diffusion-weighted magnetic resonance imaging early after chemoradiotherapy to monitor treatment response in head-and-neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys. 2012;82(3):1098-107.
12. King AD, Mo FK, Yu KH, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010;20(9):2213-20.
13. Abdel Razek AA, Kandeel AY, Soliman 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. 2007;28(6):1146-52.
14. 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;23(2):562-9.
15. Yu Y, Mabray M, Silveira W, et al. Earlier and more specific detection of persistent neck disease with diffusion-weighted MRI versus subsequent PET/CT after definitive chemoradiation for oropharyngeal squamous cell carcinoma. Head Neck. 2017;39(3):432-8.
16. Fukui MB, Blodgett TM, Snyderman CH, et al. Combined PET-CT in the head and neck: part 2. Diagnostic uses and pitfalls of oncologic imaging. Radiographics. 2005;25(4):913-30.
17. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
18. Schouten C, De Graaf P, Alberts F, et al. Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDGPET-CT. Head Neck. 2015;37:E49.

19. Bossuyt. Towards complete and accurate reporting of studies of diagnostic accuracy the STARD initiative. BMJ. 2003.

20. Noij DP, Boerhout EJ, Pieters-van den Bos IC, et al. 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. 2014;83(7):1144-51.

21. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging. 2010;37(1):181-200.

22. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

24. Kolff-Gart AS, Pouwels PJ, Noij DP, 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;36(2):384-90.

25. Choi SH, Paeng JC, Sohn CH, et al. Correlation of 18F-FDG uptake with apparent diffusion coefficient ratio measured on standard and high b value diffusion MRI in head and neck cancer. J Nucl Med. 2011;52(7):1056-62.

26. Fruehwald-Pallamar J, Czerny C, Mayerhoefer ME, 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;38(6):1009-19.

27. Varoquaux A, Rager O, Lovblad KO, 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;40(6):842-52.

CHAPTER 4.4

Use of diffusion-weighted imaging and ^{18}F-FDG-PET/CT in the response assessment for (chemo) radiotherapy in head and neck squamous cell carcinoma

Accepted for publiction in Clinical Oncology

Daniel P Noij
Roland M Martens
Thomas Koopman
Otto S Hoekstra
Emile F Comans
Ben Zwezerijnen
Remco de Bree
Pim de Graaf
Jonas A Castelijns

ABSTRACT

Purpose: Our purpose was to assess the diagnostic accuracy and prognostic value of diffusion-weighted imaging (DWI) and ^{18}F-fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) performed 3-6 months after (chemo)radiotherapy in head and neck squamous cell carcinoma (HNSCC).

Materials and Methods: For this retrospective cohort study we included 82 patients with advanced stage HNSCC treated between 2012 and 2015. Primary tumors and lymph nodes were assessed separately. DWI was analyzed qualitatively and quantitatively. ^{18}F-FDG-PET-CT was evaluated using the Hopkins criteria. Dichotomous qualitative analysis was performed for both modalities. Cox regression analysis was used for univariate analysis of recurrence-free survival (RFS). Significant univariate parameters were included in multivariate analysis.

Results: In 12 patients locoregional recurrence occurred. With all imaging strategies, either single-modality or multi-modality, a high NPV was achieved (94.3-100%). Best results were obtained with a sequential approach only including the second modality in positive reads of the first modality. It did not matter which modality was assessed first. For primary tumor assessment this resulted in a sensitivity, specificity, PPV and NPV of 57.1%, 97.3%, 66.7%, 96.0%, respectively. For lymph node analysis sensitivity, specificity, PPV and NPV were 83.3%, 95.6%, 62.5%, and 98.5%, respectively. After correction for received treatment and HPV-status, primary tumor (P=0.009) or lymph node (P<0.001) Hopkins score ≥4 on ^{18}F-FDG-PET-CT remained significant predictors of RFS.

Conclusion: A sequential approach including both DWI and ^{18}F-FDG-PET-CT resulted in the best diagnostic accuracy for follow-up after (chemo)radiotherapy. Qualitative analysis of ^{18}F-FDG-PET-CT is a stronger predictor of RFS than DWI analysis.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) has a worldwide incidence of approximately 500.000 (1). Approximately 75% patients presents with advance-staged disease (2). Overall 5-year survival is approximately 50%, depending on tumor location, stage and HPV-status (3, 4). For these patients treatment generally consists of combinations of surgery, radiotherapy and chemotherapy. In the last decade primary non-surgical treatments, i.e. radiotherapy with or without chemotherapy, with salvage surgery in reserve are more frequently applied. Stringent follow-up is important to detect recurrent tumor early since salvage options may then still be feasible (5-7).

For treatment monitoring ^{18}F-fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) is frequently used. With a sensitivity and negative predictive value (NPV) of 74-85%, and 93-97% ^{18}F-FDG-PET-CT is a cost-effective screening tool for the presence of residual or recurrent disease (8, 9). However, because of the relatively low incidence of residual and recurrent malignancy combined with a specificity of 82-90%, the positive predictive value (PPV) of ^{18}F-FDG-PET-CT is suboptimal (47-65%) (8). Therefore, additional diagnostic procedures are warranted in patients with a positive ^{18}F-FDG-PET-CT read. Additional diagnostic testing may reduce the need for investigations under anesthesia with taking of biopsies. Unnecessary taking of biopsies should be avoided in irradiated tissue because of the high risk of complications and associated morbidity (10).

Diffusion-weighted imaging (DWI) is a functional magnetic resonance imaging (MRI) sequence based on Brownian motion of water in tissue (11). Reported PPVs of DWI in the detection of recurrent HNSCC are 75-100% (12). Because both techniques are based on different properties, combining DWI and ^{18}F-FDG-PET-CT may be an effective non-surgical screening strategy. With histogram analysis lesion heterogeneity can be detected which may result in better lesion characterization than solely by an ADC_{mean} (13-16).

The purpose of this study was to assess the diagnostic accuracy of DWI and ^{18}F-FDG-PET-CT performed 3-6 months after curative (chemo)radiotherapy in HNSCC. Secondary aims were to assess the prognostic value and interobserver agreement.

METHODS AND MATERIALS

This retrospective study was approved by the institutional review board with a waiver of informed consent. Inclusion criteria were: 1) histopathologically proven HNSCC treated with (chemo)radiotherapy; 2) diffusion-weighted imaging and ^{18}F-FDG-PET-CT performed 3-6 months after treatment. Exclusion criteria were: 1) earlier episodes of malignancy requiring systemic therapy or radiotherapy in the head and neck area; 2) inadequate image quality. The reference standard consisted of clinical follow-up with clinical assessment by a head and neck surgeon every 6-8 weeks during the first year and every 2-3 months during the second year. Additional diagnostic procedures (e.g., additional imaging and invasive procedures) were performed when deemed necessary by the attending physician.

We finally included 82 patients treated between March 2012 and October 2015 (Figure 1, Table 1). Tumor HPV-status was assessed using an earlier described algorithm (17): firstly, p16-status was determined immunohistochemically on formalin-fixed and paraffin-embedded tumor tissue; if high-risk HPV-DNA was also detected in p16-positive tumors, then these tumors were classified as HPV-positive.

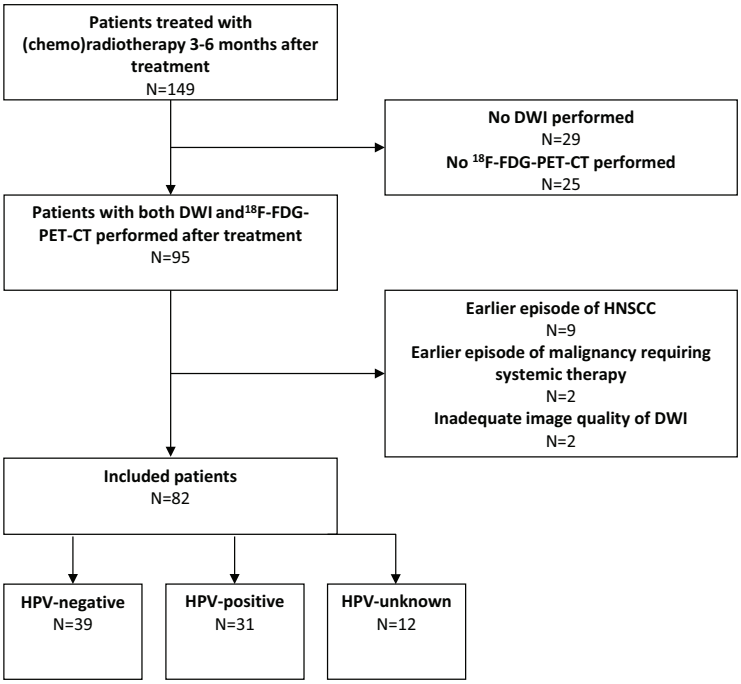


Figure 1 *Flowchart of patient inclusion*

Magnetic resonance imaging

MRI was acquired on a 1.5 T MRI system (Signa Excite HDxt, GE Healthcare, Milwaukee, WI, USA) (n=70) or a 3 T MRI system (Achieva, Philips Medical Systems, Best, the Netherlands) (n=12). Examinations were performed with a standard head and neck coil. The protocol consisted of at least 7 mm axial short tau inversion recovery (STIR) covering the entire neck; and 4 mm axial STIR and spin-echo T1 covering the primary tumor and lymph nodes. Spin-echo T1 was performed before and after administering gadoterate meglumine (Dotarem; Guerbet, Roissy, France). Diffusion-weighted imaging was acquired on the 1.5 T system using echo-planar imaging (TR/TE=5600/76.2 ms; matrix=128x86) with three b-values: b=0 (1 average), 750 (3 averages) and 1000 (6 averages) s/mm^2. On the 3 T system STIR-DWI (TR/TE=4692/64 ms; matrix=120x116) with at least four b-values was acquired: b=0, 50, 500 and 1000 s/mm^2 (3 averages). ADC maps were calculated using a mono-exponential formula with software present on the MRI system.

Positron emission tomography/computed tomography

We used an EARL accredited Ingenuity or Gemini TF PET-CT-system depending on logistics (Philips Medical Systems, Best, the Netherlands). Patients fasted for at least 6 hours. Blood glucose levels were determined prior to administration of 2.5 MBq/kg body weight of ^{18}F-FDG. Imaging started 60 minutes after ^{18}F-FDG administration. Low-dose CT (120 kV; 50 mAs) was used for attenuation correction and anatomic correlation of ^{18}F-FDG uptake. Whole-body ^{18}F-FDG-PET-CT was performed from mid-thigh to the skull vertex with the arms down in 32 patients (two minutes per bed position, image matrix size=144x144, voxel size=4x4x4 mm). After a change in imaging protocol dedicated head and neck ^{18}F-FDG-PET-CT was performed in 50 patients (four minutes per bed position, image matrix size=288x288, voxel size=2x2x2 mm). Post-reconstruction image resolution was 7 mm full width at half maximum in the whole-body protocol and 5 mm in the head and neck protocol. Protocols were in accordance with the EANM guidelines (18).

Table 1 *Patient characteristics*

Age	
Median	59.2
Range	43-81
Male/female	61/21
Tumor location	
Oropharynx	63
Hypopharynx	12
Larynx	5
Oral cavity	1
Unknown primary	1
T-stage	
Tx	1
T1	11
T2	26
T3	22
T4	22
N-stage	
N0	8
N1	15
N2a	6
N2b	37
N2c	15
N3	1
AJCC	
III	16
IVa	65
IVb	1
HPV	
Positive	39
Negative	31
Unknown	12
Treatment	
Cisplatin + radiotherapy	61[a]
Cetuximab + radiotherapy	9[b]
Radiotherapy alone	12[c]
Time between initiation of treatment and imaging (months, median; interquartile range)	
DWI	4.42; 4.21-4.57
^{18}F-FDG-PET-CT	4.43; 4.37-4.70

[a] 59 patients received cisplatin 100 mg/m^2 on day 1, 22 and 43; 2 patients received cisplatin in a weekly dose of 40 mg/m^2. Radiotherapy consisted of 70 Gy in 35 fractions of 2 Gy

[b] Cetuximab was administered 400 mg/m^2 one week before the start of radiotherapy followed by 250 mg/m^2 every week during radiotherapy. Radiotherapy consisted of 70 Gy in 35 fractions of 2 Gy

[c] 8 patients received accelerated radiotherapy

Table 2 *Hopkins criteria (19) with representative examples. Pretreatment tumor stage: Hopkins 1: T2N2b vallecula carcinoma left; Hopkins 2: T1N2b supraglottic laryngeal carcinoma left; Hopkins 3: T1N2b tonsillar carcinoma right; Hopkins 4: T3N0 glottic carcinoma right; Hopkins 5: T4aN1 base of the tongue carcinoma right*

Score	FDG uptake pattern and intensity	Response category
1	FDG uptake at the primary site and nodes, less than background blood pool (IJV).	Complete metabolic response
2	Focal FDG uptake at the primary site and nodes greater than blood pool (IJV), but less than liver.	Likely complete metabolic response
3	Diffuse FDG uptake at the primary site or nodes is greater than blood pool (IJV) or liver	Likely postradiation inflammation
4	Focal FDG uptake at the primary site or nodes, greater than liver.	Likely residual tumor
5	Focal and intense FDG uptake at the primary site and nodes.	Residual tumor

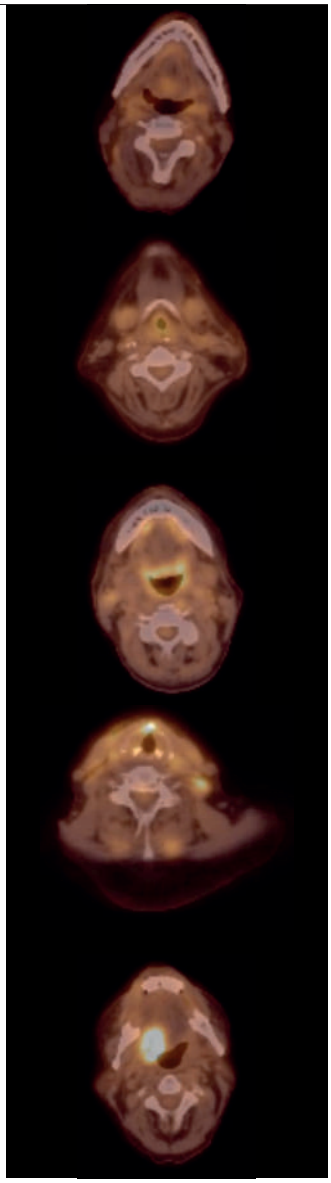


Image analysis
Observers had access to pre-treatment disease stage and pre-treatment imaging, but were blinded to patient outcome and findings of the other observers. In the results section consensus values are reported. The primary tumor and lymph nodes were assessed independently. Per patient the most suspicious lymph node was included in the analysis.

Diffusion-weighted imaging was independently assessed qualitatively by two radiologists with 33 (JCA) and 9 (PG) years of experience in head and neck imaging and quantitatively by a researcher with five years of experience in head and neck imaging (DPN) and a radiologist (JCA). Morphologic sequences were used for anatomic localization. Residual malignancy was suspected in tissue with diffusion restriction (i.e. high signal intensity on b1000 images and low ADC value) at the site of a pre-treatment tumor or lymph node and classified on a three-point Likert scale: 1. Probably benign; 2. Equivocal; 3. Probably malignant. A whole lesion region of interest (ROI) was drawn for quantitative ADC histogram analysis when a focal lesion was present (i.e. volume, mean, standard deviation (ADC_{SD}), median, minimal value (ADC_{min}), maximal value (ADC_{max}), $ADC_{skewness}$, $ADC_{kurtosis}$). Furthermore, ROIs were drawn on the myelum on the three most central slices to compare between MRI systems.

A nuclear medicine physician with 15 (EFC) and a nuclear medicine resident with 5 (BZ) years of experience independently reviewed ^{18}F-FDG-PET-CT images by using the Hopkins criteria (Table 2) (19). Generally, a score of ≥4 is considered suspicious of malignancy. For comparison we also determined the diagnostic accuracy with a cut-off of ≥3 as suspicious of malignancy.

Statistical analysis
Receiver operating characteristics (ROC) curves and the Youden index were used to determine cut-off values of qualitative and quantitative parameters (20). Cox-regression analysis of patient and imaging characteristics was used to predict recurrence-free survival (RFS) and overall survival (OS). Significant parameters in univariate analysis were combined in multivariate analysis. We used the last day of radiotherapy as t0. With Mann-Whitney U tests we compared quantitative DWI data between responders and non-responders, and ADC_{myelum} values between MRI systems. McNemar tests were used to compare diagnostic accuracy between modalities. Weighted κ and proportion specific agreements were used for interobserver agreement for categorical variables (21). To calculate proportion specific agreements, we used the cut-off value with the highest Youden Index to create a 2x2 table. Two-way mixed intraclass correlation coefficient (ICC) was used for quantitative parameters. Statistical analyses were done with SPSS (version 22.0; Chicago, IL, USA). *P*-values of <0.05 were considered statistically significant.

RESULTS

Patient population
Median follow-up was 26 months (interquartile range=9-39 months). Seven patients had primary tumor failure, in six patients nodal failure occurred, and one of these patients

had both primary tumor and nodal failure. Nineteen patients died during follow-up, all with locoregional or distant disease progression. Positive HPV-status was positively associated with OS (*P*=0.010; hazard ratio (HR)=0.140; 95%CI=0.032-0.620), but not with RFS (*P*=0.128; HR=0.294; 95%CI=0.061-1.423). Patients treated with cisplatin had better RFS than patients treated with radiotherapy alone or with concurrent cetuximab (*P*=0.008; HR=4.744; 95%CI=1.501-14.995); however, OS was not significantly different (*P*=0.057; HR=2.423; 95%CI=0.974-6.033). Treatment type was not associated with HPV-status. None of the following parameters were predictive of either RFS or OS: gender, age, T stage (T1-2 vs T3-4), N stage (N0-1 vs N 2-3), AJCC stage (III vs IV) (Table 3).

Single modality assessment

At qualitative DWI analysis best diagnostic accuracy (i.e. the highest Youden Index) was achieved with a sensitive approach (i.e. a score ≥2 is considered a positive read) (Table 4). For primary tumor analysis sensitivity, specificity, PPV and NPV were 57.1% (95%CI=20.2-88.2), 91.9% (95%CI=82.7-96.7), 40.0% (95%CI=13.7-72.6%), and 95.8% (95%CI=87.3-98.9), respectively. For lymph node analysis sensitivity, specificity, PPV and NPV were 100% (95%CI=51.7-100), 72.1% (95%CI=59.7-81.9), 24.0% (95%CI=10.2-45.5%), and 100% (95%CI=90.9-100%), respectively. A DWI score ≥2 was associated with worse RFS, for primary tumor (*P*=0.003; HR=7.486; 95%CI=2.007-27.926) and lymph node (*P*=0.009; HR=5.389; 95%CI=1.1518-19.140). There was no significant association between qualitative DWI parameters and OS (Table 3).

For ^{18}F-FDG-PET-CT, the conventional Hopkins criteria cut-off (considering scores ≥4 as a positive read), resulted in best diagnostic accuracy of ^{18}F-FDG-PET-CT (Table 4): primary tumor sensitivity, specificity, PPV and NPV were 85.7% (95%CI=42.0-99.2%), 86.5% (95%CI=76.1-93.0), 37.5% (95%CI=16.3-64.1%), and 98.5% (95%CI=90.6-99.9), respectively. For lymph node assessment sensitivity, specificity, PPV and NPV were 83.3% (95%CI=36.5-99.1), 92.6% (95%CI=83.0-97.3), 50.0% (95%CI=20.1-79.9), and 98.4% (95%CI=90.5-99.9), respectively. RFS was significantly worse in patients with a Hopkins score ≥4 for primary tumor (*P*<0.001; HR=10.873; 95%CI=3.245-36.435) and lymph node (*P*<0.001; HR=21.570; 95%CI=5.526-84.203). Overall survival was significantly worse in patients with a primary tumor (*P*=0.015; HR=3.205; 95%CI=1.258-8.164) or lymph node (*P*=0.027; HR=3.248; 95%CI=1.141-9.251) Hopkins score ≥4 (Table 3). See Figure 2 for a representative case.

Both modalities compared

Specificity of nodal assessment with ^{18}F-FDG-PET-CT was significantly higher than DWI (92.6% vs 72.1%; *P*<0.001), with comparable sensitivity (83.3% vs 100%) (Table 4). Survival analysis, including qualitative DWI and ^{18}F-FDG-PET-CT results, corrected for treatment (cisplatin vs non-cisplatin) and HPV-status, revealed that primary tumor (*P*=0.009) and lymph node (*P*<0.001) assessment with ^{18}F-FDG-PET-CT remained a significant predictor of RFS. Nodal assessment with ^{18}F-FDG-PET-CT and DWI were independent significant predictors of OS (*P*<0.001 and *P*=0.042, respectively).

Both modalities combined
We used the cut-off with the highest Youden index when combining results of DWI and ^{18}F-FDG-PET-CT (i.e. DWI cut-off=1-2; ^{18}F-FDG-PET-CT cut-off=3-4). See appendix A for the diagnostic accuracy of different combinations of imaging.

For primary tumor analysis NPVs ranged from 94.7% to 100%, depending on the imaging strategy. Highest NPV was achieved with a sensitive approach (considering a positive read on one of the modalities as an overall positive read) resulting in a sensitivity of 100% (95%CI=56.1-100.0%), specificity of 79.7% (95%CI=68.5-87.8%), PPV of 31.8% (95%CI=14.7-54.9%) and NPV of 100% (95%CI=92.3-100%) for primary tumor analysis. Highest PPV was achieved with a sequential approach only including the second modality in positive reads of the first modality. It did not matter which modality was assessed first. This resulted in a sensitivity, specificity, PPV and NPV of 57.1% (95%CI=20.2-88.2%), 97.3% (95%CI=89.7-99.5%), 66.7% (95%CI=24.1-94.0%), 96.0 (95%CI=87.8-99.0%), respectively.

Table 3 *Results of univariate Cox regression analysis*

	P-value	HR (95%CI)	P-value	HR (95%CI)
Gender	0.963	0.970 (0.262-3.584)	0.109	0.301 (0.069-1.308)
Age	0.109	1.067 (0.986-1.156)	0.340	1.032 (0.967-1.101)
T stage (T1-2 vs T3-4)	0.308	1.866 (0.562-6.200)	0.121	2.152 (0.817-5.668)
N stage (N0-1 vs N2-3)	0.055	0.329 (0.106-1.023)	0.839	0.899 (0.322-2.510)
AJCC	0.530	0.658 (0.178-2.436)	0.407	1.862 (0.428-8.104)
Treatment (cisplatin vs non-cisplatin)	**0.008**	4.744 (1.501-14.995)	0.057	2.423 (0.974-6.033)
HPV-status	0.128	0.294 (0.061-1.423)	**0.010**	0.140 (0.032-0.620)
DWI PT (1 vs 2-3)	**0.003**	7.486 (2.007-27.926)	0.179	3.461 (0.629-11.985)
^{18}F-FDG-PET-CT PT (1-3 vs 4-5)	**<0.001**	10.873 (3.245-36.435)	**0.015**	3.205 (1.258-8.164)
DWI LN (1 vs 2-3)	**0.009**	5.389 (1.1518-19.140)	0.421	1.672 (0.478-5.850)
^{18}F-FDG-PET-CT LN (1-3 vs 4-5)	**<0.001**	21.570 (5.526-84.203)	**0.027**	3.248 (1.141-9.251)

Abbreviations: LN = lymph node; PT = primary tumor

For lymph node analysis NPVs ranged from 94.3% to 100% depending on the imaging strate gy. Highest NPV was achieved by only including DWI findings, resulting in a sensitivity, specificity, PPV and NPV of 100% (95%CI=51.7-100%) 72.1% (95%CI=59.7-81.9%), 24.0% (95%CI=10.2-45.5%), and 100% (95%CI=90.9-100), respectively. Highest PPV was achieved with a sequential approach only including the second modality in positive reads of the first modality resulting in a sensitivity, specificity, PPV and NPV of 83.3% (95%CI=36.5-99.1%), 95.6% (95%CI=86.8-98.9%), 62.5% (95%CI=25.9-89.8%), and 98.5% (95%CI=90.7-99.9), respectively. It did not matter which modality was assessed first.

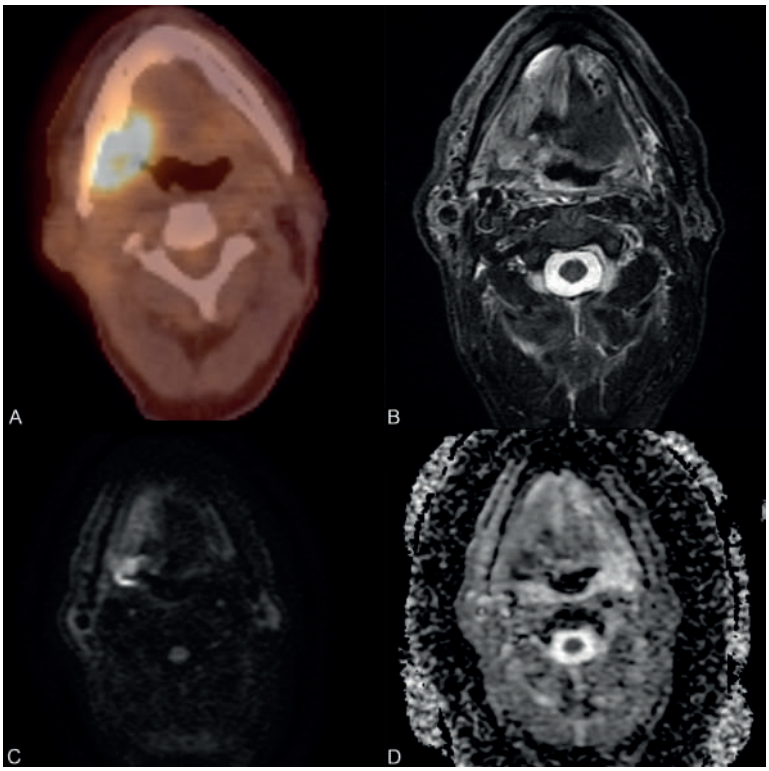


Figure 2 *Imaging of 64-year old with a T4N2b base of the tongue carcinoma on the right side with residual disease 3 months after (chemo)radiotherapy. The residual disease was detected both with A) ^{18}F-FDG-PET-CT (Hopkins=5) and C/D) DWI (Qualitative score=3). Note the difference in estimated size of the residual lesion between modalities that appears larger on ^{18}F-FDG-PET-CT. B) STIR imaging is included for anatomic correlation.*

Table 4 *Diagnostic accuracy of qualitative image analysis of DWI and ^{18}F-FDG-PET-CT, respectively*

			Sensitive approach[a]					**Conservative approach**[b]				
Tissue	Parameter	AUC (95%CI)	Sensitivity (%, 95%CI; ratio)	Specificity (%, 95%CI; ratio)	Positive predictive value (%, 95%CI; ratio)	Negative predictive value (%, 95%CI; ratio)	YI	Sensitivity (%, 95%CI; ratio)	Specificity (%, 95%CI; ratio)	Positive predictive value (%, 95%CI; ratio)	Negative predictive value (%, 95%CI; ratio)	YI
Primary tumor (n=81)	DWI	0.759 (0.524-0.994)	57.1 (20.2-88.2; 4/7)	91.9 (82.7-96.7; 68/74)	40.0 (13.7-72.6; 4/10)	95.8 (87.3-98.9; 68/71)	0.490	42.9 (11.8-79.8; 3/7)	98.6 (91.7-99.9; 73/74)	75.0 (21.9-98.7; 3/4)	94.8 (86.5-98.3; 73/77)	0.415
	^{18}F-FDG-PET-CT	0.934 (0.853-1.000)	100.0 (56.1-100.0; 7/7)	56.8 (44.8-68.1; 42/74)	17.9 (8.1-34.1; 7/39)	100 (89.6-100; 42/42)	0.568	85.7 (42.0-99.2; 6/7)	86.5 (76.1-93.0; 64/74)	37.5 (16.3-64.1; 6/16)	98.5 (90.6-99.9; 64/65)	0.722
Lymph node (n=74)	DWI	0.855 (0.761-0.949)	100 (51.7-100; 6/6)	72.1 (59.7-81.9; 49/68)	24.0 (10.2-45.5; 6/25)	100 (90.9-100; 49/49)	0.721	33.3 (6.0-75.9; 2/6)	89.7 (79.3-95.4; 61/68)	22.2 (3.9-59.8; 2/9)	93.8 (84.2-98.0; 61/65)	0.230
	^{18}F-FDG-PET-CT	0.952 (0.899-1.000)	100.0 (51.7-100.0; 6/6)	79.4 (67.5-87.9; 54/68)	30.0 (12.8-54.3; 6/20)	100 (91.7-100; 54/54)	0.794	83.3 (36.5-99.1; 5/6)	92.6 (83.0-97.3; 63/68)	50.0 (20.1-79.9; 5/10)	98.4 (90.5-99.9; 63/64)	0.759

[a] For DWI a sensitive cut-off was between 1-2. For ^{18}F-FDG-PET-CT a sensitive cut-off was between 2-3.
[b] For DWI a sensitive cut-off was between 2-3. For ^{18}F-FDG-PET-CT a sensitive cut-off was between 4-5.

Abbreviations: AUC = area under the curve; YI = Youden index

Table 5 *Quantitative DWI results*

ADC parameter	Primary tumor (n=42)				Lymph node (n=63)			
	Control	Failure	P value	AUC (95%CI)	Control	Failure	P value	AUC
Mean ($\cdot 10^{-3}$ mm^2/s)	1.88	1.62	0.099	0.702 (0.460-0.944)	1.55	1.61	0.480	0.600 (0.423-0.777)
Standard deviation ($\cdot 10^{-3}$ mm^2/s)	0.24	0.28	0.287	0.633 (0.389-0.876)	0.23	0.26	0.326	0.638 (0.418-0.858)
Median ($\cdot 10^{-3}$ mm^2/s)	1.87	1.60	0.099	0.702 (0.453-0.951)	1.55	1.61	0.450	0.607 (0.439-0.774)
Min ($\cdot 10^{-3}$ mm^2/s)	1.39	0.92	0.114	0.694 (0.410-0.977)	1.10	0.99	0.612	0.572 (0.337-0.808)
Max ($\cdot 10^{-3}$ mm^2/s)	2.37	2.32	0.817	0.531 (0.293-0.768)	2.04	2.29	0.053	0.762 (0.637-0.887)
Skewness ($\cdot 10^{-3}$ mm^2/s)	0.03	0.10	0.741	0.543 (0.337-0.749)	0.10	0.22	0.931	0.514 (0.259-0.769)
Kurtosis ($\cdot 10^{-3}$ mm^2/s)	2.56	2.46	0.792	0.535 (0.292-0.777)	2.67	2.97	0.364	0.628 (0.388-0.867)
Volume (cm^3)	0.48	2.41	0.335	0.618 (0.301-0.935)	0.35	2.62	0.004	0.872 (0.735-1.000)

Abbreviation: AUC = area under the curve

Quantitative diffusion-weighted imaging analysis

Quantitative DWI analysis was possible in 42 patients with focal residual lesions at the primary tumor site (seven patients with primary tumor failure) and in 63 patients with focal residual nodal lesions (six patients with nodal failure) (Table 5). Residual nodal volume was significantly smaller in nodal control (cut-off=0.26 cm^3; *P*=0.004; sensitivity=100% (95%CI=46.3-100%); specificity=62.1%; 95%CI=48.3-74.2%). Primary tumor ADC_{mean} was non-significantly higher in primary tumor control (*P*=0.099; cut-off=1.70·10^{-3} mm^2/s; sensitivity=71.4% (95%CI=30.3-94.9%); specificity=77.1%; 95%CI=59.4-89.0%). Nodal ADC_{max} was non-significantly lower in patients with nodal control (*P*=0.053; cut-off=2.13·10^{-3} mm^2/s; sensitivity=80% (95%CI=29.9-98.9%); specificity=86.2%; 95%CI=74.1-93.4%). Residual nodal volume was predictive of RFS (*P*=0.001), also after correction for treatment (cisplatin vs non-cisplatin) and HPV-status (*P*=0.045). Nodal volume was not predictive of OS (*P*=0.650).

ADC_{myelum} values on the 1.5T MRI system (n=70) were significantly lower than on the 3T MRI system (n=12) (ADC_{myelum}= 1.02·10^{-3} mm^2/s vs 1.21 ·10^{-3} mm^2/s, *P*=0.039).

Interobserver agreement

For DWI weighted κ was 0.516-0.618 with a three-point Likert scale. Positive agreement was 0.56-0.77. Negative agreement was 0.88-0.92. ICC for determining ADC was 0.916-0.996. For ^{18}F-FDG-PET-CT weighted κ was 0.406-0.533 with a five-point Likert scale. Positive agreement was 0.55-0.62. Negative agreement was 0.92. See Appendix C for a detailed overview over interobserver agreement.

DISCUSSION

We determined the diagnostic and prognostic accuracy of routinely performed DWI and ^{18}F-FDG-PET-CT 3-6 months after treatment of HNSCC with (chemo)radiotherapy. High NPV was achieved with all imaging strategies (Appendix A, NPV=94.3-100%). Highest PPV was found with a sequential approach considering only the results of the second modality in those patients with positive findings on the first modality. It did not matter which modality was performed first. Reasons to perform DWI first are that DWI is cheaper and generally more readily available (22). The main reason to perform ^{18}F-FDG-PET-CT first is that there is more experience with ^{18}F-FDG-PET-CT for this indication. Moreover, with ^{18}F-FDG-PET-CT there can be screened for distant metastases, which may avoid futile salvage surgery. ^{18}F-FDG-PET-CT was an independent significant predictor of RFS, while DWI was not. For OS prediction nodal assessment with both modalities had independent predictive value.

In a recent meta-analysis it was stated that DWI may be superior to ^{18}F-FDG-PET-CT in the detection of recurrent HNSCC (12). The included studies only contained patients with a clinical suspicion of recurrent disease. The authors found a PPV of DWI of 91-100%. In another meta-analysis the PPV of ^{18}F-FDG-PET-CT was 47-65% for the detection of recurrent HNSCC was found (8). In the second meta-analysis both routinely performed imaging and imaging performed in patients with clinical suspicion of recurrent disease were included. This limits the comparability between both meta-analyses. In a recent study DWI at eight weeks after chemoradiotherapy was compared to ^{18}F-FDG-PET-CT at 14 weeks for neck failure (23). Using nodal ADC_{mean} values the authors found a sensitivity of 100% and specificity of 92% for DWI compared to a sensitivity of 100% and specificity of 71% for ^{18}F-FDG-PET-CT.

Vogel et al. compared qualitative and quantitative DWI assessment in 46 patients with laryngeal or hypopharyngeal cancer treated with (chemo)radiotherapy and clinical suspicion of recurrent disease (24). Post-treatment MRI was performed two to 108 months after treatment. Qualitative DWI analysis sensitivity, specificity and accuracy were 94%, 100%, and 98%, respectively. Accuracies of quantitative analysis ranged from 74% to 80% and were inferior to qualitative image analysis. This is in line with our findings. We found limited additional value for histogram analysis of DWI. Only primary tumor ADC_{mean}, lymph node ADC_{max} and lymph node volume on DWI performed more or less equally well as qualitative analysis. Pre-treatment, high ADC_{mean} is associated with an adverse prognosis and an increase of ADC during treatment is prognostically favorable (25-27). In a posttreatment setting we found a trend towards a higher primary tumor ADC_{mean} in patients with locoregional control. Malignant tissue is associated with low ADC values and necrosis/fibrosis with high ADC values (28). Low pretreatment ADC may be indicative of a tumor which is more responsive for (chemo)radiotherapy. After treatment ADC values should increase if all malignant tissue is eradicated by the (chemo)radiotherapy. However, we also found higher lymph node ADC_{max} to be associated with an adverse prognosis. A benign lymph node contains multiple small densely packed lymphocytes which results in diffusion restriction on DWI. For tonsils there is some evidence that benign tonsils have lower ADC values than tonsils harboring HNSCC (14, 29).

We found better interobserver agreement for quantitative analysis compared to qualitative analysis. High agreement of quantitative analysis was expected, because with DWI lesions are relatively easy to detect on high b-value images. Qualitative image analysis is more subjective in nature which results in less interobserver agreement. Also a larger learning curve may be present for qualitative analysis, especially for post-treatment imaging which is considered one the most challenging parts of head and neck radiology (30). Often small areas of residual diffusion restriction are present and it remains difficult to appreciate the clinical suspicion of residual malignancy of these areas resulting in a relatively low positive agreement (0.56-0.77).

With the Hopkins criteria we attempted to limit subjectivity of qualitative image analysis with ^{18}F-FDG-PET-CT. However, we found only moderate agreement for ^{18}F-FDG-PET-CT (κ=0.41-0.53), which was lower than the κ values reported by Marcus et al. (0.69-0.89) (19). The largest discrepancy was between a Hopkins score of either 1 or 3, which are both considered indicative of the absence of residual malignancy. Again, the positive agreement was relatively low (0.55-0.62), due to disagreement between observers whether uptake should be regarded as either focal of diffuse. It should be noted that for DWI a 3-point Likert scale was used compared to a 5-point Likert scale for ^{18}F-FDG-PET-CT.

This study had some limitations. Firstly, DWI was performed on two MRI systems as dictated by logistics. ADC values may vary between MRI systems (31). Therefore, quantitative data should be interpreted with caution because we did find significant differences in ADC_{myelum} values between MRI systems. We used the ADC_{myelum} to compare between MRI systems because this tissue is not affected by malignancy and is always present in the field of view (31). For qualitative image assessment the used MRI system may be of less significance. Secondly, because a significant amount of the patients did not have a residual lesion, we could only perform quantitative image analysis in 42/81 primary tumors and 63/74 lymph nodes.

Conclusion

A sequential approach including both modalities resulted in the best the diagnostic accuracy for follow-up after (chemo)radiotherapy. Qualitative analysis of ^{18}F-FDG-PET-CT is a stronger predictor of RFS than DWI analysis. Quantitative analysis of DWI did not add value to qualitative analysis.

REFERENCES

1. Ferlay J, Soerjomataram IE, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F.;, . GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. 2013.
2. Brana I, Siu LL. Locally advanced head and neck squamous cell cancer: treatment choice based on risk factors and optimizing drug prescription. Ann Oncol. 2012;23 Suppl 10:x178-85.
3. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. The Lancet Oncology. 2010;11(1):21-8.
4. Adelstein DJ, Saxton JP, Rybicki LA, et al. Multiagent concurrent chemoradiotherapy for locoregionally advanced squamous cell head and neck cancer: mature results from a single institution. J Clin Oncol. 2006;24(7):1064-71.
5. Kano S, Homma A, Hayashi R, et al. Salvage surgery for recurrent oropharyngeal cancer after chemoradiotherapy. Int J Clin Oncol. 2013;18(5):817-23.
6. Roosli C, Studer G, Stoeckli SJ. Salvage treatment for recurrent oropharyngeal squamous cell carcinoma. Head Neck. 2010;32(8):989-96.
7. Zafereo ME, Hanasono MM, Rosenthal DI, et al. The role of salvage surgery in patients with recurrent squamous cell carcinoma of the oropharynx. Cancer. 2009;115(24):5723-33.
8. Gupta T, Master Z, Kannan S, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2011;38(11):2083-95.
9. Uyl-de Groot CA, Senft A, de Bree R, Leemans CR, Hoekstra OS. Chest CT and whole-body 18F-FDG PET are cost-effective in screening for distant metastases in head and neck cancer patients. J Nucl Med. 2010;51(2):176-82.
10. de Bree R, van der Putten L, Brouwer J, Castelijns JA, Hoekstra OS, Leemans CR. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 2009;45(4-5):386-93.
11. 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;168(2):497-505.
12. Driessen JP, van Kempen PM, van der Heijden GJ, et al. Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. Head Neck. 2015;37(3):440-8.
13. Galban CJ, Mukherji SK, Chenevert TL, et al. A feasibility study of parametric response map analysis of diffusion-weighted magnetic resonance imaging scans of head and neck cancer patients for providing early detection of therapeutic efficacy. Transl Oncol. 2009;2(3):184-90.
14. Choi YJ, Lee JH, Kim HO, et al. Histogram Analysis of Apparent Diffusion Coefficients for Occult Tonsil Cancer in Patients with Cervical Nodal Metastasis from an Unknown Primary Site at Presentation. Radiology. 2016;278(1):146-55.
15. King AD, Chow KK, Yu KH, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology. 2013;266(2):531-8.
16. Tyagi N, Riaz N, Hunt M, et al. Weekly response assessment of involved lymph nodes to radiotherapy using diffusion-weighted MRI in oropharynx squamous cell carcinoma. Med Phys. 2016;43(1):137.
17. Smeets SJ, Hesselink AT, Speel EJ, et al. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. Int J Cancer. 2007;121(11):2465-72.
18. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-54.

19. Marcus C, Ciarallo A, Tahari AK, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes. J Nucl Med. 2014;55(9):1411-6.

20. Bohning D, Holling H, Patilea V. A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. Stat Methods Med Res. 2011;20(5):541-50.

21. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.

22. Greuter MJ, Schouten CS, Castelijns JA, et al. Cost-effectiveness of response evaluation after chemoradiation in patients with advanced oropharyngeal cancer using (18)F-FDG-PET-CT and/or diffusion-weighted MRI. BMC Cancer. 2017;17(1):256.

23. Yu Y, Mabray M, Silveira W, et al. Earlier and more specific detection of persistent neck disease with diffusion-weighted MRI versus subsequent PET/CT after definitive chemoradiation for oropharyngeal squamous cell carcinoma. Head Neck. 2017;39(3):432-8.

24. 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;23(2):562-9.

25. Kim S, Loevner L, Quon H, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(3):986-94.

26. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010;20(7):1703-14.

27. King AD, Mo FKF, Yu KH, et al. Squamous cell carcinoma of the head and neck: Diffusion-weighted MR imaging for prediction and monitoring of treatment response. European Radiology. 2010;20(9):2213-20.

28. Padhani AR, Koh DM, Collins DJ. Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology. 2011;261(3):700-18.

29. Bhatia KS, King AD, Yeung DK, et al. Can diffusion-weighted imaging distinguish between normal and squamous cell carcinoma of the palatine tonsil? Br J Radiol. 2010;83(993):753-8.

30. Saito N, Nadgir RN, Nakahira M, et al. Posttreatment CT and MR imaging in head and neck cancer: what the radiologist needs to know. Radiographics. 2012;32(5):1261-82; discussion 82-4.

31. Kolff-Gart AS, Pouwels PJ, Noij DP, 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;36(2):384-90.

4

CHAPTER 5

PROGNOSTIC CAPACITY OF DIFFUSION-WEIGHTED IMAGING

CHAPTER 5.1

Predictive value of diffusion-weighted imaging without and with including contrast-enhanced magnetic resonance imaging in image analysis of head and neck squamous cell carcinoma

European Journal of Radiology 2015;84:108-16

Daniel P Noij
Petra J Pouwels
Redina Ljumanovic
Dirk L Knol
Patricia Doornaert
Remco de Bree
Jonas A Castelijns
Pim de Graaf

ABSTRACT

Objectives: To assess disease-free survival (DFS) in head and neck squamous cell carcinoma (HNSCC) treated with (chemo)radiotherapy.

Methods: Pretreatment magnetic resonance images of 78 patients were retrospectively studied. Apparent diffusion coefficients (ADC) were calculated with two sets of two b-values: 0-750 s/mm² (ADC_{750}) and 0-1000 s/mm² (ADC_{1000}). One observer assessed tumor volume on T1-weighted imaging (T1-WI). Two independent observers assessed ADC values of primary tumor and largest lymph node in two sessions (i.e. without and with including CE-T1WI in image analysis). Interobserver and intersession agreement were assessed with intraclass correlation coefficients (ICC), separately for ADC_{750} and ADC_{1000}. Lesion volumes and ADC values were related to DFS using Cox regression analysis.

Results: Median follow-up was 18 months. Interobserver ICC was better without than with CE-T1WI (primary tumor=0.92 and 0.75-0.83, respectively; lymph node=0.81-0.83 and 0.61-0.64, respectively). Intersession ICC ranged from 0.84 to 0.89. With CE-T1WI, mean ADC values of primary tumor and lymph node were higher at both b-values than without CE-T1WI ($P<0.001$). Tumor volume (sensitivity=73%; specificity=57%) and lymph node ADC_{1000} (sensitivity=71-79%; specificity=77-79%) were independent significant predictors of DFS without and with including CE-T1WI ($P<0.05$).

Conclusions: Pretreatment primary tumor volume and lymph node ADC_{1000} were significant independent predictors of DFS in HNSCC treated with (chemo)radiotherapy. Disease-free survival could be predicted from ADC values acquired without and with including CE-T1WI in image analysis. The inclusion of CE-T1WI did not result in significant improvements in the predictive value of DWI. Diffusion-weighted imaging without including CE-T1WI was highly reproducible.

INTRODUCTION

Head and neck cancer accounts for approximately 3% of all malignancies (1). Treatment selection is based on the best tradeoff between cure rate and quality of life, and consists of (a combination) of surgery, chemotherapy and radiotherapy depending on disease stage (2).

With better treatment selection, patients with a high probability of an unfavorable treatment outcome after (chemo)radiotherapy could undergo primary surgical treatment. The same applies when treatment response to (chemo)radiotherapy can be monitored in an early stage; then (chemo)radiotherapy might be terminated prematurely. After a full (chemo)radiotherapy treatment, salvage surgery with curative intent is still possible to perform. However this is not preferred because of a higher risk of complications like impaired wound healing. Moreover, salvage treatment is not always possible because of extension of the residual or recurrent tumor outside its original location. Therefore a minority of patients (21-31%) receives salvage surgery after local failure (3-5).

Diffusion-weighted imaging (DWI) is an emerging magnetic resonance imaging (MRI) technique in response prediction in HNSCC patients treated with (chemo)radiotherapy (6) including in head and neck radiology. The main indications for performing DW imaging in this relatively small but challenging region of the body are tissue characterization, nodal staging, therapy monitoring, and early detection of treatment failure by differentiating recurrence from posttherapeutic changes. Lower apparent diffusion coefficients (ADCs.

Diffusion-weighted imaging is based on differences in water mobility in different tissues which can be quantified into an apparent diffusion coefficient (ADC) (7). The extent of diffusion weighting depends on the timing and the strength of the gradient and is expressed as a b-value. In order to reconstruct an ADC at least two different b-values are needed, typically a low (e.g. <150 s/mm^2) and a high b-value (e.g. >700 s/mm^2) are used. In hypercellular tissue (e.g. malignant tissue) with a small amount of extracellular space diffusion is restricted, which gives a low ADC value. In contrary, in hypocellular tissue where diffusion in the extracellular space is facilitated, ADC values are high. Necrosis and inflammation generally meet these criteria (8, 9). There is still no consensus on the optimal combination of b-values, though a combination of b=0 s/mm^2 and b=1000 s/mm^2 is commonly used (9-15).

Diffusion-weighted imaging has shown potential in the prediction of prognosis in patients with head and neck squamous cell carcinoma (HNSCC) treated with (chemo)radiotherapy and to monitor therapy in a very early stage. Higher pretreatment ADC values are associated with adverse prognosis (8, 12, 13, 16). Furthermore, DWI has shown potential to detect central necrosis and (subcentimeter) metastatic lymph nodes (9, 15).

Contrast-enhanced imaging is often used to exclude necrosis, which allows that the ADC value only of the solid part of lesions can be determined (9, 15). To our knowledge there has not been a study that assessed the clinical relevance of using contrast-enhanced imaging for excluding necrosis on DWI. Hatakenaka et al. did suggest that pretreatment

ADC would be superior to contrast-enhanced MRI to predict local failure (10). Since DWI and contrast-enhanced imaging are based on different properties, both techniques may be synergistic in predicting the outcome of treatment.

The purpose of this study was to assess the prediction of disease-free survival (DFS) and interobserver agreement of DWI without and with including contrast-enhanced T1-weighted imaging (CE-T1WI) in image analysis of HNSCC treated with (chemo)radiotherapy.

METHODS & MATERIALS

Study population

This retrospective study was approved by the local ethics committee. The requirement for informed consent was waived.

Inclusion criteria were: histologically proven HNSCC treated with (chemo)radiotherapy in the oral cavity, oropharynx; hypopharynx or larynx; and turbo spin-echo (TSE)-DWI of adequate diagnostic quality for the primary tumor or the lymph node on at least one b-value image. Exclusion criteria were previous malignancies in the head and neck area and distant metastases at the start of therapy. All patients were clinically assessed by a head and neck surgeon who performed a physical examination and endoscopic evaluation of the primary tumor. The nodal stage was assessed using ultrasound-guided fine-needle aspiration cytology. A total of 111 consecutive patients received pre-treatment DWI and (chemo) radiotherapy of the head and neck between August 2009 and December 2011.

To allow for optimally comparable data we selected the largest

Table 1 *Patient, tumor and treatment characteristics*

Characteristics		Total (n=78)
Age (year), mean (SD)		62 (10)
Gender, n (%)	male	45 (58)
	female	33 (42)
T stage, n (%)	T1	8 (10)
	T2	28 (36)
	T3	28 (36)
	T4	14 (18)
N stage, n (%)	N0	38 (49)
	N1	13 (17)
	N2	27 (35)
AJCC tumor stage, n (%)	I	4 (5)
	II	13 (17)
	III	27 (35)
	IV	34 (44)
Tumor location, n (%)	Oral cavity	8 (10)
	Oropharynx	40 (51)
	Larynx	22 (28)
	Hypopharynx	8 (10)
Primary treatment, n (%)	Radiotherapy	40 (51)
	Chemoradiotherapy	38 (49)
Radiation dose (Gy) , n (%)	<70	8 (10)
	70	70 (90)

patient group that was scanned on the same MR-system, therefore 18 patients were excluded due to the use of another MR-system. One patient was excluded because no CE-T1WI was acquired. Fourteen patients were excluded because neither the primary tumor nor the largest lymph node was visible on DWI due to small tumor size (n=8) or poor image quality (n=4). Finally, the study consisted of 78 patients. In all 78 patients b1000 images were acquired. In 64 of these patients b750 images were also acquired. See Figure 1 for a detailed flow-chart of patient inclusion.

Radiotherapy was delivered to the primary tumor and affected lymph nodes in a total dose of 70 Gy in 35 fractions of 2 Gy in 70 patients. Three patients received a total dose of 69 Gy in 30 fractions of 2.3 Gy. All these 73 patients received an elective dose to the lymph nodes at risk for microscopic tumor. In four patients a total dose of 52 Gy was delivered in 16 fractions of 3.25 Gy. One patient received 60 Gy in 25 fractions of 2.4 Gy. In these last five patients, no elective dose to the lymph node regions was given. Thirty-eight patients received additional chemotherapy (i.e. 100 mg/m^2 cisplatin in the first, fourth and seventh week after the start of radiotherapy (n=24) or 400 mg/m^2 cetuximab one week before the start of radiotherapy followed by 250 mg/m^2 every week during radiotherapy (n=14)). Median time between MRI examinations and the start of treatment was 25 days (range, 7-63 days). Patient, tumor and treatment characteristics are summarized in Table 1.

Follow-up consisted of clinical assessment every 6-8 weeks during the first year, every 2-3 months during the second year and every 3-4 months in the third year. Additional imaging and diagnostic procedures were performed in case of clinical suspicion of recurrent disease, residual disease or distant metastases. Positive biopsy or locoregional disease progression within six months after the end of treatment was considered to be residual disease; after six months it was considered to be a locoregional recurrence.

Magnetic resonance imaging

Imaging was performed on a 1.5 T system (Signa HDxt; GE Healthcare, Milwaukee, WI, United States), using a standard head and neck coil with 29 elements. For all sequences the field of view was 250 mm. DWI was acquired using two PROPELLER sequences with two sets of two b-values: b=0 and 750 s/mm^2 and b=0 and 1000 s/mm^2 respectively. Apparent diffusion coefficient maps were calculated by using two sets of b-values: b=0 and 750 s/mm^2 (ADC_{750}) and b=0 and 1000 s/mm^2 (ADC_{1000}). After the administration of 0.4 ml/kg gadoteric acid (Dotarem; Guerbet, Roissy, France) (n=72) and 0.2 ml/kg gadobutrol (Gadovist; Bayer Schering AG, Berlin, Germany) (n=6), CE-T1WI without fat saturation was acquired. An overview of our imaging protocol is provided in Table 2. Because of differences in resolution and to correct for patient movement, CE-T1WI and DWI were co-registered using the linear registration software tool FLIRT from the FSL package (FMRIB Centre, Oxford, United Kingdom).

Image analysis

Images were evaluated with Centricity Radiology RA 600 (version 6.1, GE Healthcare, Milwaukee, WI, USA). Volume of the primary tumor and largest lymph node were assessed on T1-weighted images by one reader (JCA) by drawing manual ROIs on each slice

containing the lesion. The same reader also assessed the short axis diameter of the largest lymph node (17).

Table 2 *Imaging protocol at 1.5 T in HNSCC patients*

	Sequence	TR (ms)	TE (ms)	TI (ms)	Matrix	Slice thickness (mm)	Intersection gap (mm)	Slices	b-value (s/mm^2)
Pre-contrast	Ax T1 SE	460-500	12	-	256x256	3.0-4.0	0.3-0.4	22	-
	Ax STIR	4150	60	160	352x224 or 352x192	7	2.1	25	-
	Ax TSE DWI	3500	84	-	128x128	3.0-5.0	0-0.4	16	0, 750, 1000
Postcontrast	Ax CE-T1	560	14	-	256x256	7	2.1	25	-

Abbreviations: Ax = axial; CE-T1 = contrast-enhanced T1; DWI = diffusion-weighted imaging; SE = spin-echo; STIR = short-TI inversion recovery; TE = time to echo; TI = time for inversion; TR = repetition time; TSE = turbo spin-echo

Image analysis without and with including CE-T1WI in image analysis was done in two sessions by two radiologists (JCA and PGR) with 21 and six years of experience in head and neck radiology. In both sessions observers had access to conventional MR-sequences for anatomical correlation, and patient information regarding age, gender, global tumor location (i.e. oral cavity, oropharynx, larynx and hypopharynx) and TNM-stage, but were blinded to treatment outcome and the results of the other observer.

In the first session observers had access to all diffusion sequences (i.e. b0, b750, b1000 and corresponding ADC maps) but not to the CE-T1WI. Free-hand regions of interest (ROIs) were drawn on the high b-value images to delineate the solid parts of the tumor and largest lymph node on the slide that contained the core of the lesion, avoiding areas of necrosis. Solid parts were characterized by a high signal intensity on the high b-value images and low signal intensity on the ADC map. Regions of interest were copied from the high b-value images to the ADC map. Mean ADC value and ROI volume were recorded. Image quality of DWI was assessed separately for the primary tumor and largest lymph node using a five-point Likert scale: 1=inadequate; 2=moderate; 3=fair; 4=good; 5=excellent.

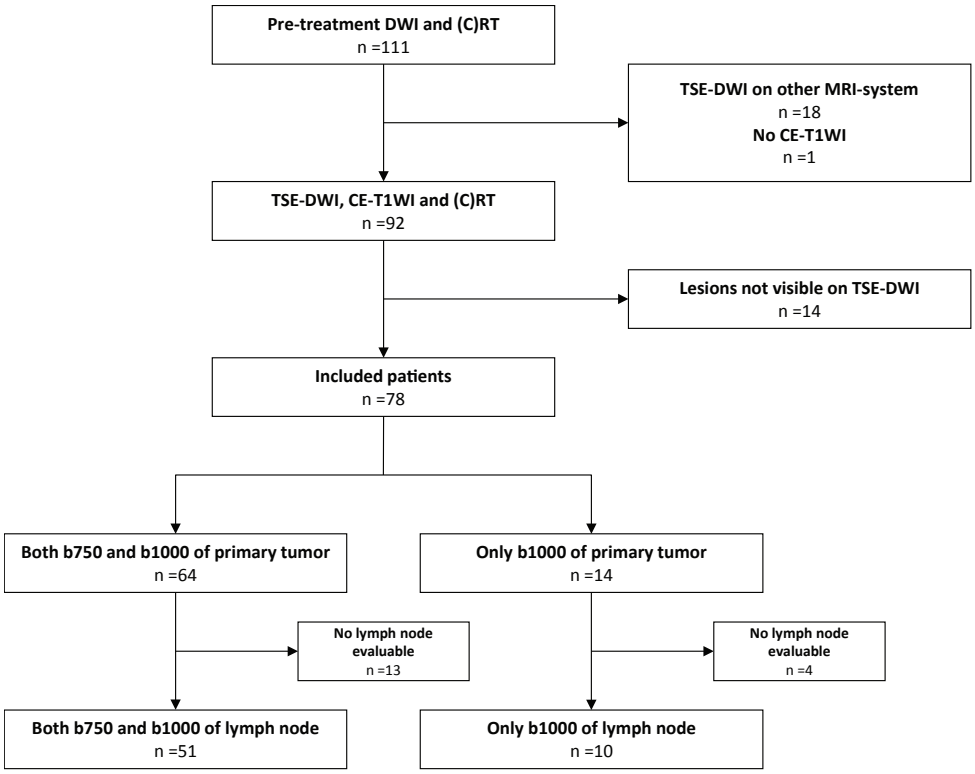


Figure 1 *Flow chart of patient inclusion*

Abbreviations: CE-T1WI = contrast-enhanced T1-weighted imaging; (C)RT = (chemo)radiotherapy; DWI = diffusion-weighted imaging; HNSCC = head and neck squamous cell carcinoma; MRI = magnetic resonance imaging; TSE = turbo spin-echo

Table 3 *Mean ADC-values, image quality and interobserver agreement for both observers without and with including CE-T1WI in image analysis. Image quality was only assessed in the first session. Data are expressed as mean ± standard deviation.*

	Variable	Observer 1 ADC ($\cdot 10^{-3}$ mm^2/s)	Median ROI volume (cm^3)	Image quality	Observer 2 ADC ($\cdot 10^{-3}$ mm^2/s)	Median ROI volume (cm^3)	Image quality	Average ADC ($\cdot 10^{-3}$ mm^2/s)	Median ROI volume (cm^3)	ICC (95%CI)
Without CE-T1WI	Primary tumor ADC_{750} (n=56)	1.68 ± 0.29	0.65	3.5[a]	1.66 ± 0.29	0.65	2.8[a]	1.67 ± 0.29	0.62	0.92 (0.86-0.95)
	Primary tumor ADC_{1000} (n=62)	1.46 ± 0.26	0.64	3.4	1.47 ± 0.26	0.72	2.7	1.47 ± 0.25	0.68[b]	0.92 (0.87-0.95)
	Lymph node ADC_{750} (n=49)	1.62 ± 0.25	0.29	3.6	1.62 ± 0.33	0.31	2.8	1.62 ± 0.27	0.29[b]	0.75 (0.60-0.85)
	Lymph node ADC_{1000} (n=57)	1.43 ± 0.23	0.27	3.5	1.41 ± 0.27	0.31	2.7	1.42 ± 0.24	0.29[b]	0.83 (0.72-0.89)
With CE-T1WI	Primary tumor ADC_{750} (n=56)	1.86 ± 0.35	0.89	-	1.83 ± 0.37	0.67	-	1.85 ± 0.34	0.82	0.81 (0.69-0.82)
	Primary tumor ADC_{1000} (n=62)	1.63 ± 0.29	1.01	-	1.59 ± 0.31	0.69	-	1.61 ± 0.29	0.85	0.83 (0.73-0.89)
	Lymph node ADC_{750} (n=49)	1.75 ± 0.28	0.25	-	1.80 ± 0.37	0.37	-	1.77 ± 0.30	0.30	0.64 (0.42-0.78)
	Lymph node ADC_{1000} (n=57)	1.58 ± 0.28	0.30	-	1.56 ± 0.30	0.42	-	1.57 ± 0.26	0.38	0.61 (0.42-0.75)

[a] Resembles a significant difference between ROI volumes without and with CE-T1WI (P =0.002)
[b] Resembles a significant difference between b-values (P <0.05)

Abbreviations: ADC = apparent diffusion coefficient; CE-T1WI = contrast-enhanced T1-weighted imaging; ICC = intraclass correlation coefficient; ROI = region of interest

In the second session observers had access to CE-T1WI weighted images, b0 images and ADC maps, but not to high b-value images. On CE-T1WI single slice free-hand ROIs were placed on the contrast-enhancing part of the tumor and largest lymph node, again areas of necrosis were avoided. The ROI was first copied to the b0 image to verify the anatomical position and subsequently to the ADC map. Again mean ADC value and ROI volume were recorded. To minimize recall bias, the second session was at least four weeks after the first. To ensure that the same lesions were assessed in both sessions, observers had access to the slice position of the ROI in the first session.

Statistical analysis

Statistical analyses were performed using SPSS (version 20.0; Chicago, IL, USA). Wilcoxon signed rank tests were used to compare the image quality of both b-values for both observers separately and to compare ROI volumes acquired without and with including CE-T1WI in image analysis. Interobserver agreement and intersession agreement (i.e. between ADC values acquired without and with including CE-T1WI in image analysis) were assessed by calculating the two-way mixed model intraclass correlation coefficient (ICC) (18). Intraclass correlation coefficients can be interpreted according to Nunnally (19): Techniques with an ICC>0.80 are reliable for basic research, to be clinically applicable ICCs>0.90 are necessary.

Mean ADC values of both observers were used for subsequent analyses. We compared ADC values derived without and with including CE-T1WI in image analysis by using paired sample t-tests and Bland-Altman plots. Paired sample t-tests were also used to compare ADC_{750} with ADC_{1000}.

Disease-free survival was assessed for various predictors using univariable Cox regression analysis. Significant predictors were tested further with multivariable Cox regression analysis. Receiver operating characteristic (ROC) analysis was performed to determine the optimal cut-off value with the highest Youden Index for lesion volume and ADC values in predicting DFS. This optimal cut-off was used to create Kaplan-Meier curves of these continuous variables.

RESULTS

Treatment outcome

Median follow-up was 18 months (interquartile range (IQR), 9-25 months). Five patients were censored because of a second primary tumor. One patient died due to euthanasia. This patient was censored because we did not consider this to be death due to disease progression. Sixty-nine percent (54/78) of the patients remained disease-free during follow-up. During follow-up, biopsies were positive for malignancy in 23% (18/78) of the patients. Five patients had residual disease, eight developed a locoregional recurrence and five patients were diagnosed with distant metastasis.

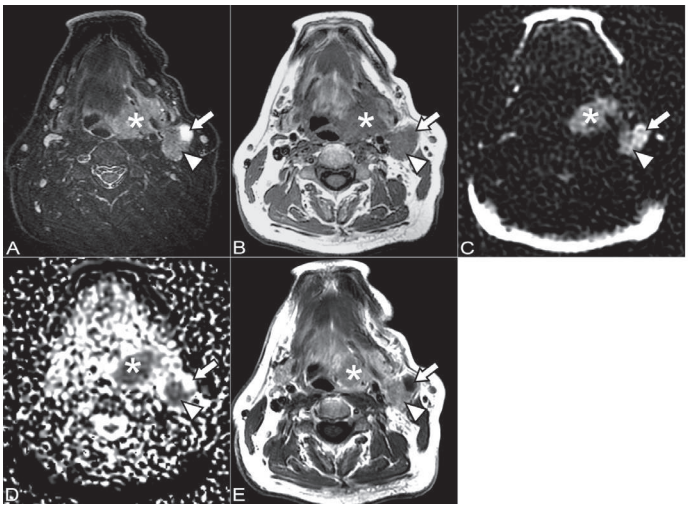


Figure 2 *MR images in a 59-year old male patient with a T3N2c oropharyngeal carcinoma (*). Necrosis (arrow) in the level II lymph node (arrowhead) is detected on A) STIR (hyperintensity), on B) T1 minimal hypo-intensity is seen in the necrotic area. The findings of C) b750 (high signal) and D) ADC_{750} (high signal) are also indicative of necrosis. On E) CE-T1WI necrosis is seen due to low contrast-enhancement in the necrotic part of the lymph node (arrow). After 16 months of follow-up this patient remained disease free.*

Image analysis

In 88% (56/64) of the patients the primary tumor was visible on b750 images, on b1000 images in 79% (62/78). Lymph nodes could be evaluated on b750 images in 96% (49/51) of the patients and b1000 images in 93% (57/61) (Figure 1, Table 3). According to both observers the primary tumor (P<0.05) and largest lymph node (P>0.05) were better depicted on the b750 images than the b1000 images (Table 3). Median primary tumor volume was 7.3 cm^3 (IQR=2.7-14.6 cm^3) with median lymph node volume being 1.3 cm^3 (IQR, 0.5-4.2 cm^3) and median minimal axial diameter of the largest node being 8.2 mm (IQR, 5.5-15.4 mm). Representative images are shown in Figure 2.

Interobserver agreement

Without including CE-T1WI in image analysis ICC for ADC values of the primary tumor was 0.92 for both b-values. For the largest lymph node ICC was 0.75 for the ADC_{750} and 0.83 for the ADC_{1000}. Including CE-T1WI in image analysis resulted in lower ICCs being 0.81 and 0.83 for the primary tumor for the ADC_{750} and ADC_{1000} respectively. In lymph nodes these values were also lower being 0.64 for ADC_{750} and 0.61 for ADC_{1000} (Table 3).

Comparison between ADC values and ROI volumes without and with including CE-T1WI in image analysis

Regardless of including CE-T1WI in image analysis ADC_{750} was higher than ADC_{1000} in both primary tumor and lymph node (P<0.001). Bland-Altman plots are provided in Figure 2.

With inclusion of CE-T1WI, mean ADC values of primary tumor and lymph node were higher at both b-values than without CE-T1WI (P<0.001). Also ROIs were larger when including CE-T1WI. This difference in ROI volume was significant (P=0.002), except for the ADC_{750} of the primary tumor (P>0.05). Biases (i.e. mean difference without and with including CE-T1WI) ranged from -0.14·10^{-3} mm²/s to -0.18·10^{-3} mm²/s with ICC ranging from 0.84 to 0.89 (Figure 3, Table 4).

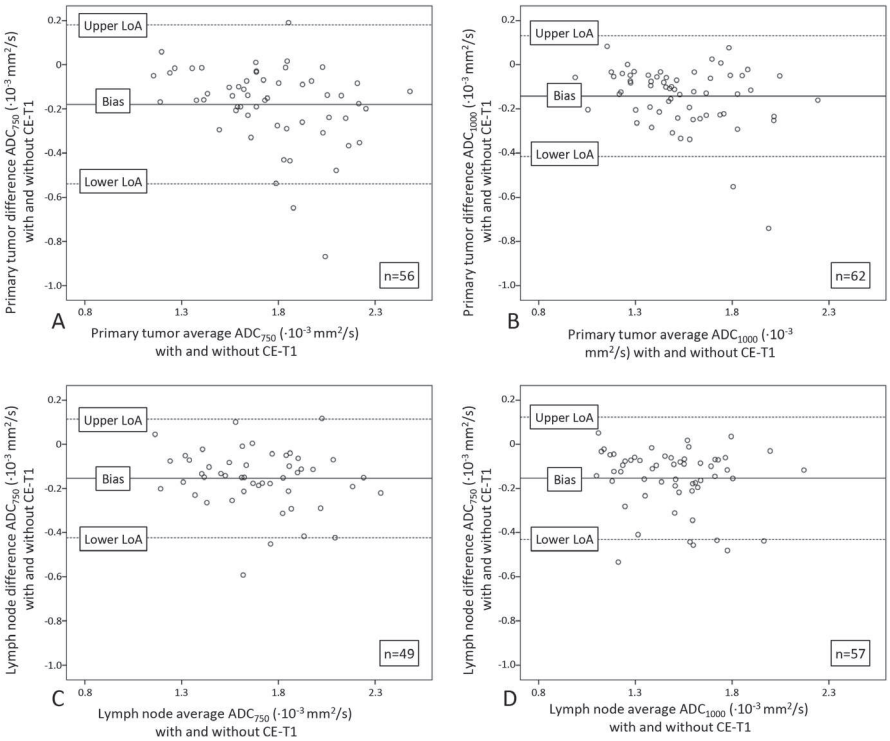


Figure 3 *Bland-Altman plots representing the agreement regarding ADC-measurements without and with including CE-T1WI in image analysis. Positive values indicate a higher ADC-value without including CE-T1WI in image analysis. A) Primary tumor ADC_{750}; B) primary tumor ADC_{1000}; C) lymph node ADC_{750}; D) lymph node ADC_{1000}.*

Survival analysis

Results of ROC analysis are shown in Table 5. Area under the curve ranged from 0.54 to 0.82. Without including CE-T1WI in image analysis, significant prognostic factors in univariable Cox regression of DFS were large primary tumor volume on T1 (P=0.001), and a high lymph node ADC_{1000} (P=0.001). Primary tumor ADC_{750}, primary tumor ADC_{1000}, lymph node volume on T1, minimal axial diameter on T1 and lymph node ADC_{750} were not significant parameters (P>0.05). Other variables included in univariable Cox regression

were: age, gender, treatment, adjuvant treatment, and radiation dose. None of these variables was a significant predictor of DFS (P>0.05). Both significant variables remained significant when adding both primary tumor volume on T1 (P=0.009) and ADC_{1000} of the lymph node (P=0.014) to a multivariable Cox regression model (Table 6).

Table 4 *Agreement between ADC-values without and with including CE-T1WI in image analysis*

Variable	Bias ($\cdot 10^{-3}$ mm^2/s)	LoA ($\cdot 10^{-3}$ mm^2/s)	ICC (95%CI)
Primary tumor ADC_{750}	-0.18	-0.54; 0.18	0.84 (0.74-0.90)
Primary tumor ADC_{1000}	-0.14	-0.42; 0.13	0.88 (0.80-0.92)
Lymph node ADC_{750}	-0.15	-0.42; 0.14	0.89 (0.81-0.94)
Lymph node ADC_{1000}	-0.15	-0.43; 0.12	0.85 (0.75-0.91)

Abbreviations: ADC = Apparent diffusion coefficient; ICC = Intraclass correlation coefficient; LoA = Limits of agreement

When including CE-T1WI in image analysis, high lymph node ADC_{1000} (P<0.001) was also a significant predictor of DFS in univariable Cox regression. In a multivariable Cox regression model both primary tumor volume on T1 (P=0.011) and lymph node ADC_{1000} (P=0.002) remained significant predictors (Table 6). Kaplan-Meier curves of the significant predictors are shown in Figure 4.

DISCUSSION

To our knowledge this is the first study that assesses the predictive value of DWI without and with including CE-T1WI in image analysis of HNSCC. Hatakenaka et al. suggested that pretreatment ADC would be superior to CE-MRI to predict local failure (10). We did not find any significant differences in the predictive value of DWI with or without including CE-T1WI in image analysis. The intersession agreement was high (ICC=0.84-0.89). Differences in ADC values might be explained by a systematic error due to larger ROI volume when including CE-T1WI in image analysis. This may be caused by peritumoral contrast-enhancement. In both settings high lymph node ADC_{1000} and primary tumor volume were independent significant predictors of DFS. These findings suggest that DWI analysis without CE-T1WI is non-inferior to DWI including CE-T1WI in image analysis for predicting DFS. The inclusion of CE-T1WI did not result in significant improvements in the predictive value of DWI. This suggests that DWI can be used to detect necrosis at a comparable level as CE-T1WI, or at least without clinically significant differences. An advantage of DWI compared to CE-T1WI is that it can be used in patients with renal failure.

In our study high pretreatment lymph node ADC_{1000} was a significant predictor of treatment response. Kim et al. also found high pretreatment ADC of metastatic lymph nodes in HNSCC to be a significant predictor of local failure in a study on 33 patients with a median follow-up of 12 months. Sensitivity and specificity were 65% and 86%, respectively. When the change in ADC value between pre-treatment DWI and DWI one week after the start of treatment was used, sensitivity increased to 86% and specificity was 83% (12).

However, in a study on 37 HNSCC patients with a follow-up of at least 2 years performed by King et al., pretreatment ADC was not a significant predictor of local failure (20). Only ADC changes between DWI examinations before and during treatment showed a significant correlation with treatment outcome. A (large) increase of ADC values in early follow-up was predictive of local control. Treatment induced cell death may lead to reduced diffusion restriction and therefore a rise in ADC values. These findings suggest that DWI before and during treatment (e.g. two weeks after the start of (chemo)radiotherapy) provides the highest diagnostic accuracy in response prediction. However, this is not yet implemented in clinical practice because early follow-up MRI findings are not yet incorporated in treatment protocols. Besides there is a logistic challenge, because for reliable repeated ADC measurements patients need to be scanned on the same scanner in the same hospital, ideally in the same position (21).

Malignant tissue is characterized by low ADC value implying high cellularity compared to benign tissue (22). In order to treat HNSCC with (chemo)radiotherapy high cell turnover (i.e. low ADC) is required as (chemo)radiotherapy mainly targets dividing cells (23, 24). Therefore relatively high pre-treatment ADC may result in adverse prognosis for patients treated with (chemo)radiotherapy (12). For surgery the relation between ADC and prognosis may be different as more slowly dividing malignancies may be more easy to remove radically, however this is beyond the scope of this article.

It should be noted that abscesses are also characterized by high signal intensity on high b-value imaging combined with a low ADC value and may therefore be difficult to distinguish from malignant tissue. In a study of Koç et al. on patients with necrotic and cystic head and neck lesions abscesses could be differentiated from (necrotic) tumors with a sensitivity and specificity of 100%. Abscesses were characterized by even lower ADC values than malignancies (25). The lower ADC value of abscesses is attributed to the higher cell density in an abscess combined with the presence of proteins and other macromolecules in abscesses (26). Therefore lesions with high intensity on high b-value imaging and low ADC values cannot always considered to be malignant. Other sequences and clinical parameters (e.g. fever and tender lymphadenopathy) may further aid in differentiating abscesses from malignancy.

Table 5 *Results of ROC-analysis. The highest Youden Index was used to determine the optimal cut-off value.*

	Parameters	Cut-off	Sensitivity	Specificity	AUC
	Tumor volume on T1	7.3 cm^3	73%	57%	0.66
	Lymph node volume on T1	0.8 cm^3	81%	40%	0.56
	Minimal axial diameter	7.6 mm	75%	54%	0.58
Without CE-T1WI	Primary tumor ADC_{750}	1.63 $\cdot 10^{-3}$ mm^2/s	73%	53%	0.59
	Primary tumor ADC_{1000}	1.73 $\cdot 10^{-3}$ mm^2/s	39%	88%	0.59
	Lymph node ADC_{750}	1.63 $\cdot 10^{-3}$ mm^2/s	78%	60%	0.66
	Lymph node ADC_{1000}	1.51 $\cdot 10^{-3}$ mm^2/s	71%	74%	0.75
With CE-T1WI	Primary tumor ADC_{750}	1.55 $\cdot 10^{-3}$ mm^2/s	100%	22%	0.55
	Primary tumor ADC_{1000}	1.44 $\cdot 10^{-3}$ mm^2/s	85%	33%	0.54
	Lymph node ADC_{750}	1.83 $\cdot 10^{-3}$ mm^2/s	78%	63%	0.62
	Lymph node ADC_{1000}	1.68 $\cdot 10^{-3}$ mm^2/s	79%	77%	0.82

Abbreviations: ADC = apparent diffusion coefficient; AUC = area under the curve; CE-T1WI = contrast-enhanced T1-weighted imaging

We used two sets of two b-values (0-750 s/mm^2 and 0-1000 s/mm^2). In most clinical studies a maximum b-value of 1000 s/mm^2 is used (9-15). Only King et al. used a maximum b-value of 500 s/mm^2 to limit signal loss and image distortion (20). We used a TSE sequence for DWI instead of the more commonly used echo planar imaging (EPI) sequence. TSE sequences suffer less from geometric distortion, susceptibility artifacts and motion artifacts, but the signal-to-noise-ratio is lower (27). Therefore the use of a maximum b-value of 1000 s/mm^2 could result in a too low signal-to-noise ratio to allow for proper image interpretation. This is supported by our data; the image quality of b750 images was rated slightly higher than b1000 images. Further the primary tumor and lymph node were more frequently visualized on b750 images (88% and 96%, respectively) than on b1000 images (79% and 93%, respectively).

Table 6 *Results of univariable and multivariable Cox regression without and with including CE-T1WI in image analysis. Significant predictors in univariable Cox regression were tested further with multivariable Cox regression analysis.*

Parameter		**Univariable Cox regression, *P* value**	**Multivariable Cox regression, *P* value**
Without CE-T1WI	Primary tumor volume on T1 (cm^3)	0.001	0.009
	Primary tumor ADC_{750} ($\cdot 10^{-3}$ mm^2/s)	0.571	-
	Primary tumor ADC_{1000} ($\cdot 10^{-3}$ mm^2/s)	0.226	-
	Lymph node volume on T1	0.763	
	Minimal axial lymph node diameter on T1	0.414	
	Lymph node ADC_{750} ($\cdot 10^{-3}$ mm^2/s)	0.202	-
	Lymph node ADC_{1000} ($\cdot 10^{-3}$ mm^2/s)	0.001	0.014
With CE-T1WI	Primary tumor volume on T1 (cm^3)	0.001	0.011
	Primary tumor ADC_{750} ($\cdot 10^{-3}$ mm^2/s)	0.572	-
	Primary tumor ADC_{1000} ($\cdot 10^{-3}$ mm^2/s)	0.471	-
	Lymph node volume on T1	0.763	
	Minimal axial lymph node diameter on T1	0.414	
	Lymph node ADC_{750} ($\cdot 10^{-3}$ mm^2/s)	0.240	-
	Lymph node ADC_{1000} ($\cdot 10^{-3}$ mm^2/s)	<0.001	0.002

Abbreviations: ADC = apparent diffusion coefficient; CE-T1WI = contrast-enhanced T1-weighted imaging

In both primary tumor and lymph node ADC_{750} was higher than ADC_{1000}. This may be explained our assumption of a monoexponential model. In this model ADC values are lower at higher b-values due to perfusion effects at low b-values due to a non-linear relation between b-values and signal intensity. At higher b-values a linear relation exists between b-value and signal intensity. ADC values may be better represented with a biexponential model which accounts for the perfusion effects at low b-values (28). It has also been shown that the high b-value component (i.e. an ADC value obtained exclusively from b-values above 500 s/mm^2) has a stronger correlation with outcome (29).

Verhappen et al. compared primary tumor and lymph node delineation between TSE-DWI and EPI-DWI in 12 patients with HNSCC. They concluded that lesions, in particular small lymph nodes, are more easily visualized with EPI-DWI (30). This may be explained by a lower signal-to-noise ratio of TSE-DWI compared to EPI-DWI when using a maximum b-value of 1000 s/mm^2. However the results of TSE-DWI were more reproducible between observers (ICC=0.79 for EPI-DWI vs ICC=0.92 for TSE-DWI). In our study ICC was 0.92 in the primary tumor when only DWI is used. According to the criteria of Nunnally (19) TSE-DWI would be clinically useful only for primary tumor assessment.

In this manuscript we have used mean ADC values per ROI. Standard deviations were also acquired, however these values did not show any significant relations and were therefore not included in the manuscript. Histogram analysis of ADC has been used as marker of tumor heterogeneity (e.g. skewness, kurtosis, ADC_{min} or ADC_{max}) with promising results (31).

This study had some limitations. In the first place, all events in disease-free survival analysis were considered equal, however not all events had the same clinical consequences. We did this because we expect recurrent and residual disease to occur more frequently in the higher tumor stages, regardless of the severity of the event. We also did not assess overall survival, because patients are sometimes referred to other institutions for palliative care. We therefore could not reliably determine the time and cause of death. Secondly, since patients were treated non-surgically it is not fully clear if the largest lymph nodes really contained metastatic tissue. In our institution ultrasound-guided fine-needle aspiration cytology performed by experienced investigators is used for N-staging, which confers the risk of sampling errors. In reviews by de Bree et al. (32) and de Bondt et al. (33) ultrasound guided fine needle aspiration cytology appears to be the best minimally invasive alternative to the gold standard (i.e. histological examination after an elective neck dissection). Thirdly, 14 patients were excluded because neither tumor nor lymph node was visible at both b-values. These patients mainly had small lesions. Observers only had access to global tumor location, but not to the exact tumor location, the outcome of other diagnostic procedures nor patient symptoms, which makes it more difficult to identify small lesions.

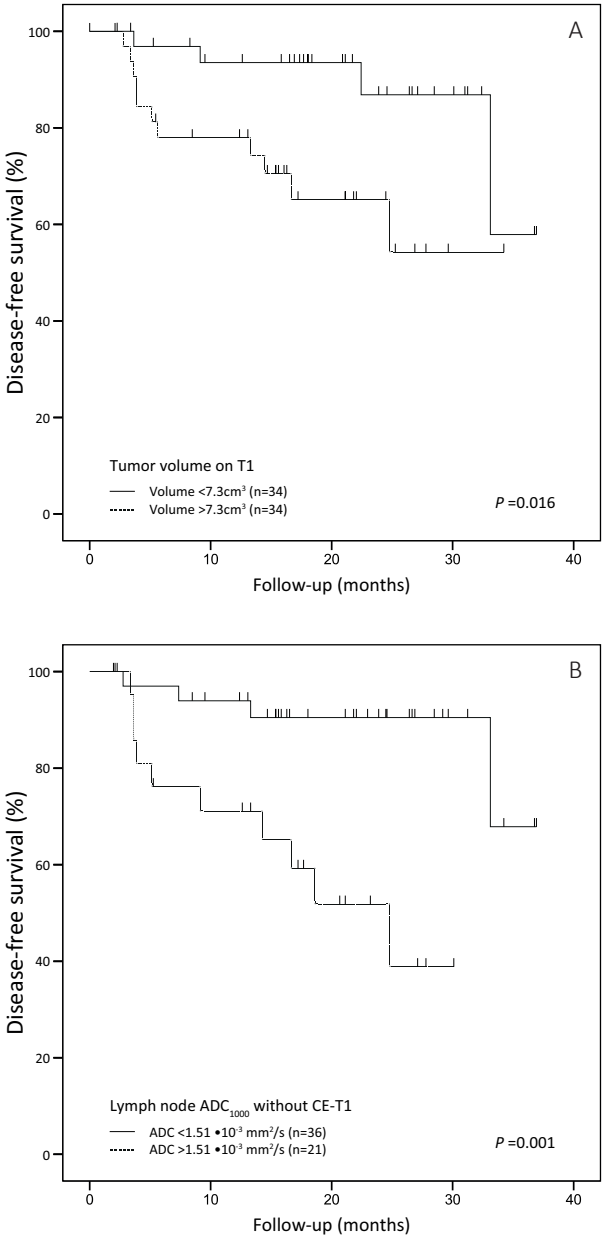


Figure 4 *Kaplan-Meier curves of A) tumor volume on T1, B) lymph node ADC_{1000} without including CE-T1WI in image analysis and C) lymph node ADC_{1000} with including CE-T1WI in image analysis.*

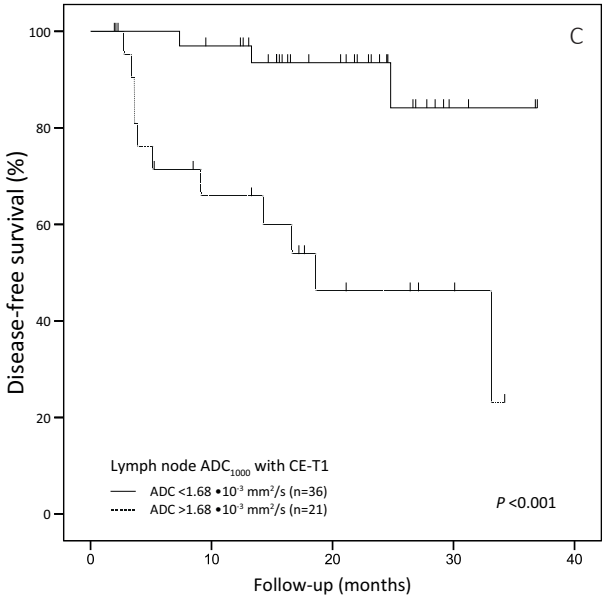


Figure 4 *continued*

Conclusion

In conclusion, pretreatment primary tumor volume and the lymph node ADC_{1000} are independent significant predictors of DFS in patients with HNSCC treated with (chemo) radiotherapy. In addition, lymph node ADC_{1000} is a significant predictor with and without including CE-T1WI in image analysis. Diffusion-weighted imaging-analysis without CE-T1WI is highly reproducible, demonstrated by good interobserver agreement. ADC values were lower without than with including CE-T1W1 in image analysis. The inclusion of CE-T1WI results in a lower interobserver agreement in measuring ADC on DWI. Therefore pretreatment DWI may be an additional tool to determine patient prognosis. As injection of any contrast agents is not necessary to perform DWI, using DWI without CE-T1WI may result in lower imaging costs with an equal predictive value. Further research is necessary to validate the value of TSE-DWI in response prediction in comparison to EPI-DWI.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69-90.
2. Pignon JP, le Maitre A, Maillard E, Bourhis J, Group M-NC. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol. 2009;92(1):4-14.
3. Kano S, Homma A, Hayashi R, et al. Salvage surgery for recurrent oropharyngeal cancer after chemoradiotherapy. Int J Clin Oncol. 2013;18(5):817-23.
4. Roosli C, Studer G, Stoeckli SJ. Salvage treatment for recurrent oropharyngeal squamous cell carcinoma. Head Neck. 2010;32(8):989-96.
5. Zafereo ME, Hanasono MM, Rosenthal DI, et al. The role of salvage surgery in patients with recurrent squamous cell carcinoma of the oropharynx. Cancer. 2009;115(24):5723-33.
6. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
7. 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;168(2):497-505.
8. Vandecaveye V, De Keyzer F, Nuyts S, 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;67(4):960-71.
9. Abdel Razek AAK, Kandeel AY, Soliman N, et al. Role of diffusion-weighted Echo-Planar MR imaging in differentiation of residual or recurrent head and neck tumors and posttreatment changes. American Journal of Neuroradiology. 2007;28(6):1146-52.
10. Hatakenaka M, Nakamura K, Yabuuchi H, 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;81(2):339-45.
11. Kato H, Kanematsu M, Tanaka O, et al. Head and neck squamous cell carcinoma: usefulness of diffusion-weighted MR imaging in the prediction of a neoadjuvant therapeutic effect. Eur Radiol. 2009;19(1):103-9.
12. Kim S, Loevner L, Quon H, et al. 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. 2009;15(3):986-94.
13. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010;20(7):1703-14.
14. Dirix P, Vandecaveye V, De Keyzer F, et al. Diffusion-weighted MRI for nodal staging of head and neck squamous cell carcinoma: impact on radiotherapy planning. Int J Radiat Oncol Biol Phys. 2010;76(3):761-6.
15. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46.
16. King AD, Mo FKF, Yu KH, et al. Squamous cell carcinoma of the head and neck: Diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010;20(9):2213-20.
17. Van Den Brekel MWM, Stel HV, Castelijns JA, et al. Cervical lymph node metastasis: Assessment of radiologic criteria. Radiology. 1990;177(2):379-84.
18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30-46.
19. Nunnally JC. Psychometric Theory. New York: McGraw Hill, 1967.

20. King AD, Chow KK, Yu KH, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology. 2013;266(2):531-8.

21. Sasaki M, Yamada K, Watanabe Y, et al. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology. 2008;249(2):624-30.

22. Gupta RK, Cloughesy TF, Sinha U, et al. Relationships between choline magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma. J Neurooncol. 2000;50(3):215-26.

23. Marcu LG, Bezak E. Neoadjuvant cisplatin for head and neck cancer: Simulation of a novel schedule for improved therapeutic ratio. J Theor Biol. 2012;297:41-7.

24. Pawlik TM, Keyomarsi K. Role of cell cycle in mediating sensitivity to radiotherapy. Int J Radiat Oncol Biol Phys. 2004;59(4):928-42.

25. Koc O, Paksoy Y, Erayman I, Kivrak AS, Arbag H. Role of diffusion weighted MR in the discrimination diagnosis of the cystic and/or necrotic head and neck lesions. Eur J Radiol. 2007;62(2):205-13.

26. Mishra AM, Gupta RK, Saksena S, et al. Biological correlates of diffusivity in brain abscess. Magn Reson Med. 2005;54(4):878-85.

27. Pipe JG, Farthing VG, Forbes KP. Multishot diffusion-weighted FSE using PROPELLER MRI. Magn Reson Med. 2002;47(1):42-52.

28. Chandarana H, Kang SK, Wong S, et al. Diffusion-weighted intravoxel incoherent motion imaging of renal tumors with histopathologic correlation. Invest Radiol. 2012;47(12):688-96.

29. Lambrecht M, Van Calster B, Vandecaveye V, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiother Oncol. 2014;110(3):429-34.

30. Verhappen MH, Pouwels PJ, Ljumanovic R, et al. Diffusion-weighted MR imaging in head and neck cancer: comparison between half-fourier acquired single-shot turbo spin-echo and EPI techniques. AJNR Am J Neuroradiol. 2012;33(7):1239-46.

31. Ahn SJ, Choi SH, Kim YJ, et al. Histogram analysis of apparent diffusion coefficient map of standard and high B-value diffusion MR imaging in head and neck squamous cell carcinoma: a correlation study with histological grade. Acad Radiol. 2012;19(10):1233-40.

32. de Bree R, Takes RP, Castelijns JA, et al. Advances in diagnostic modalities to detect occult lymph node metastases in head and neck squamous cell carcinoma. Head Neck. 2015;37(12):1829-39.

33. de Bondt RB, Nelemans PJ, Hofman PA, 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;64(2):266-72.

CHAPTER 5.2

Predictive and prognostic value of quantitative ^{18}F-FDG-PET and diffusion-weighted imaging in head and neck squamous cell cancer

Submitted

Roland M. Martens
Daniel P. Noij
T. Koopman
Ben Zwezerijnen
Martijn Heymans
Otto S. HoekstraMarije R. Vergeer
Remco de Bree
C. René Leemans
Pim de Graaf
Ronald Boellaard
Jonas A. Castelijns

ABSTRACT

Background and Purpose: In head and neck squamous cell carcinoma (HNSCC) (chemo) radiotherapy is increasingly being used to preserve organ functionality. The purpose was identifying predictive diffusion weighted imaging (DWI) and ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography (^{18}F-FDG-PET-CT) parameters in a large cohort of resectable HNSCC patients.

Materials and Methods: We retrospectively included 134 histologically proven HNSCC patients treated with (chemo)radiotherapy between 2012-2017. In 58 patients pre-treatment DWI and ^{18}F-FDG-PET-CT were performed, in 31 only DWI and in 45 patients only ^{18}F-FDG-PET-CT. Primary tumor ($_{PT}$) and largest lymphnode ($_{LN}$) metastasis were quantitatively assessed for treatment failure, recurrence-free survival (RFS) and overall survival (OS). Multivariate analysis was performed for ^{18}F-FDG-PET-CT and DWI separately. Thereafter, in patients who underwent both imaging modalities positive and negative predictive value and differences in survival (log-rank test) at optimal cut-offs were assessed.

Results: Mean follow-up was 25.6 months (interquartile range, 14.0-37.1 months). Predictors of treatment failure, corrected for treatment, TNM-stage and HPV-status, were $SUV_{max\text{-}PT}$, $ADC_{max\text{-}PT}$, $ADC_{p20\text{-}LN}$, total lesion glycolysis ($TLG_{\text{-}LN}$) (P=0.049, P=0.024, P=0.047, P=0.031, respectively). $TLG_{\text{-}PT}$ was predictive for RFS (P=0.003); metabolic active tumor volume ($MATV_{\text{-}PT}$) and $ADC_{GTV\text{-}PT}$ were significant predictors for OS (P=0.003, P<0.001, respectively). In patients with both imaging modalities $SUV_{max\text{-}PT}$ remained predictive for treatment failure (P=0.042), $TLG\text{-}_{LN}$ for RFS (P=0.021) and $ADC_{GTV\text{-}PT}$ for OS (P=0.009). Higher predictive value for treatment response was found when including both $SUV_{max\text{-}PT}$ and $ADC_{max\text{-}PT}$, compared to either one separately.

Conclusion: Both DWI- and ^{18}F-FDG-PET-CT-parameters appear to have predictive value for treatment response and long-term outcome. Combining $SUV_{max\text{-}PT}$ and $ADC_{max\text{-}PT}$ resulted in better prediction of treatment response compared to single parameter assessment.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the most common head and neck malignancy with 550,000 cases annually worldwide (1). Most patients presenting with locally advanced disease are treated with (chemo)radiotherapy in the context of organ preservation. Complete response isn't achieved in approximately 25-30% (1). Surgical salvage treatment can be difficult and functional outcome may be uncertain (2). It is essential to identify tumor characteristics predictive of response to (chemo)radiotherapy to increase treatment efficacy and long-term outcome by treatment intensification or by offering alternative treatment options as early as possible to these patients (e.g. surgery or best supportive care).

Besides clinical and histopathological factors (3), imaging parameters may provide important prognostic biomarkers (4, 5). Functional imaging, such as diffusion-weighted imaging (DWI) and ^{18}F-Fluorodeoxyglucose positron-emission computer tomography (^{18}F-FDG-PET-CT), could provide additional information on pathophysiology (6, 7). Diffusion-weighted imaging measures cell density and diffusion capacity, which can be quantified with the apparent diffusion coefficient (ADC). Low ADC values are associated with hypercellular tissues (e.g. malignant tumors) (8). With ^{18}F-FDG-PET-CT metabolic characteristics associated with active proliferation and invasion can be quantified, e.g. by assessing the standardized uptake value (SUV) (9).

Tumor characteristics associated with treatment failure or adverse long-term outcome, might be identified with functional imaging (5, 7, 10, 11). Qualitative assessment is inadequate to measure this tumor heterogeneity and often carries a certain extent of subjectivity, resulting in high interobserver variability. Quantitative analysis could provide more reproducible information of tumor heterogeneity by utilizing whole lesion assessment (12, 13).

Earlier studies showed that (combinations of) DWI and ^{18}F-FDG-PET-CT predict treatment response (10) and long-term outcome (5). However, only small patient populations were analyzed without correction for tumor-stage or HPV-status and different image acquisition systems, resulting in limited robustness of predictive and prognostic value (5, 11, 14, 15). Furthermore, limited use of higher-order statistics or the clinically relevant predictive assessment of combined use of DWI and ^{18}F-FDG-PET-CT (16) was performed. The purpose of this study was identifying DWI- and ^{18}F-FDG-PET-CT-parameters corrected for important confounders in a large cohort of resectable head and neck squamous cell carcinoma (HNSCC) patients treated with (chemo)radiotherapy.

METHODS

Patients

We retrospectively enrolled 134 consecutive patients treated from 2012-2017. The local ethics committee waived informed consent. This study was performed according to the STARD-criteria (17). Inclusion criteria were: histopathologically proven HNSCC; pretreatment DWI and/or ^{18}F-FDG-PET-CT and planned (chemo)radiotherapy. Exclusion criteria were: previous locoregional treatment for HNSCC and insufficient image quality for lesion segmentation. Treatment response was defined as: absence of residual malignant tissue within the first six months post-treatment. For response assessment we used a standardized physical examination and imaging with MRI and ^{18}F-FDG-PET-CT which was performed between 3-6 months after treatment, depending on logistics. In case of suspected malignancy, histological confirmation was acquired. Locoregional recurrence-free survival (RFS) was defined as absence of locoregional (primary tumor ($_{PT}$) or lymph node ($_{LN}$)) recurrence in 2-years follow-up period after end of treatment, which overlapped with the treatment response group.

MRI acquisitions

MR imaging was performed using a 1.5 T MR system (Signa HDxt, GE Healthcare, Milwaukee, WI) with a 12-channel neurovascular head-and-neck-array coil. The protocol consisted of at least conventional T1-weighted and short-tau-inversion-recovery (STIR). Echo-planar imaging (EPI)-DWI with TR/TE: 4300-5600/ 59-98 ms; inversion time 160 ms, averages: 3; flip-angle: 90°; matrix size: 256x256, voxel size 1x1x4 mm^3, 21-28 slices. ADC was measured using 2 b-values (0 and 1000 s/mm^2).

^{18}F-FDG-PET-CT acquisitions

For ^{18}F-FDG-PET/low-dose-CT we used a Gemini TOF-64 PET-CT (Philips Medical Systems, Best, The Netherlands) with EARL accreditation and the imaging protocol was in accordance with the EANM guidelines 2.0 on (18). Low-dose non-contrast CT (120 kV; 50 mAs) was used for attenuation correction and anatomic correlation of ^{18}F-FDG uptake. Whole-body ^{18}F-FDG-PET-CT was performed in arms-down-position, with 18 cm axial field of view, from mid-thigh to skull vertex, 60 minutes after intravenous administration of 2.5 MBq/kg ^{18}F-FDG, two minutes per bed position. The ^{18}F-FDG-PET images were reconstructed using vendor-provided and EARL-compliant reconstruction protocol with photon-attenuation correction. Reconstructed images had an image matrix size of 144x144 and voxel size of 4x4x4 mm. Post-reconstruction image resolution was 7 mm full-width at half maximum.

Image analysis

Delineation on MRI was performed manually by two independent radiologists (JAC, PG) with 34 and 10 years of experience in head-and-neck radiology, respectively. A resident (BZ) supervised by a nuclear medicine physician (OH) with 5 and 30 years of experience in head-and-neck PET ^{18}F-FDG-PET-CT performed semi-automated delineations (Figure 1), based on a background-corrected 50% of SUV_{peak} isocontour (19). Observers were aware of the HNSCC diagnosis, TNM-stage (7th edition) and primary tumor location for delineation of proven malignant lesions.

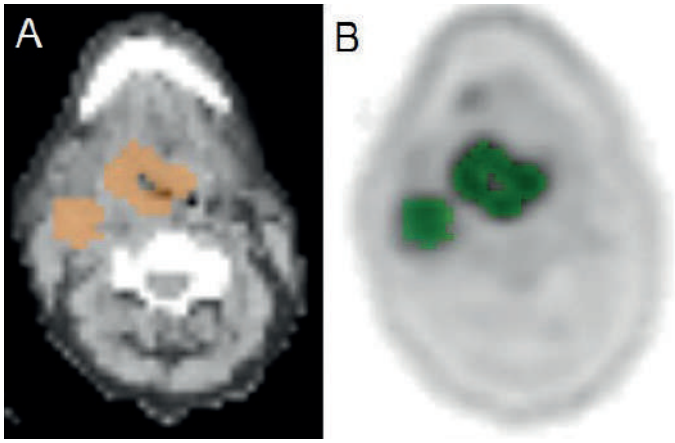


Figure 1 *Example of a ^{18}F-FDG-PET-CT semi-automated delineation based on a background-corrected 50% of SUV_{peak} isocontour in a patients with both a primary tumor and lymph node. A) shows the low-dose CT for anatomic correlation, B) shows the attenuation-corrected ^{18}F-FDG-PET images*

Whole-lesion segmentation was performed on MRI on the T1 and ADC map in accordance with other sequences (Figure 2) using VelocityTM software (Varian Medical Systems, Inc, Palo Alto, USA). Histograms were generated with Matlab (MathWorks Inc, MA, USA), based on included voxels. The following MRI parameters were extracted: gross tumor volume on T1 ($T1_{GTV}$) and ADC (ADC_{GTV}), ADC_{mean}, ADC_{max}, ADC_{min}, ADC standard deviation (ADC_{SD}), ADC percentiles ($ADCp_{10}$ to $ADCp_{90}$), $ADC_{kurtosis}$, and $ADC_{skewness}$. Consensus values were used for final analyses.

With ^{18}F-FDG-PET-CT, segmentation was performed using a background-corrected 50% isocontour of tumor SUV_{peak} (20). The SUV was normalized to body weight. SUV_{peak} was defined as the highest uptake in 1 mL spherical volume of interest across all tumor voxel locations. Uptake parameters were derived using in-house developed Accurate software, (18) to quantify lesions on ^{18}F-FDG-PET-CT were: metabolically active tumor volume (MATV), SUV_{max}, SUV_{peak}, SUV_{mean}. Total lesion glycolysis (TLG) was calculated by the tumor volume multiplied by the SUV_{mean} of the included voxels (12).

Statistical analysis

Logistic regression was used to predict treatment response. We used Cox regression to assess locoregional RFS and OS. P-values of <0.05 were considered significant in the univariate analysis. Significant univariate parameters of DWI and ^{18}F-FDG-PET-CT were combined in multivariate analysis for DWI ^{18}F-FDG-PET/CT, separately, using a backward Wald test. The remaining significant parameters for DWI or ^{18}F-FDG-PET-CT, for tumor and

for lymph node metastasis were combined and corrected for TNM-stage and HPV status. Multivariable regression analysis was performed according to the TRIPOD-statement, accepting *P*-values up to 0.157 to enhance the model applicability to other patient groups (21). For survival analyses, the end of treatment was used as start of follow-up. For multivariate analysis only patients were included of whom all data was available (i.e. DWI and ^{18}F-FDG-PET-CT). Subgroup analyses was performed for tumors with a MATV >4.2ml (22), to avoid bias from partial volume effect in small tumors.

With receiver operating characteristic (ROC) analysis three optimal cut-offs (highest Youden index (YI)) of significant multivariate parameters were determined. The positive and negative predictive value (post-test risk and 1-post-test risk, respectively) at each cut-off value was calculated for treatment failure using a prevalence ranging from 10 to 50%. A log-rank test of remaining significant multivariable parameters (divided by the optimal cut-off) was presented as Kaplan Meier survival curves. Analyses were performed using SPSS (version 18.0; SPSS Inc., Chicago, III, USA).

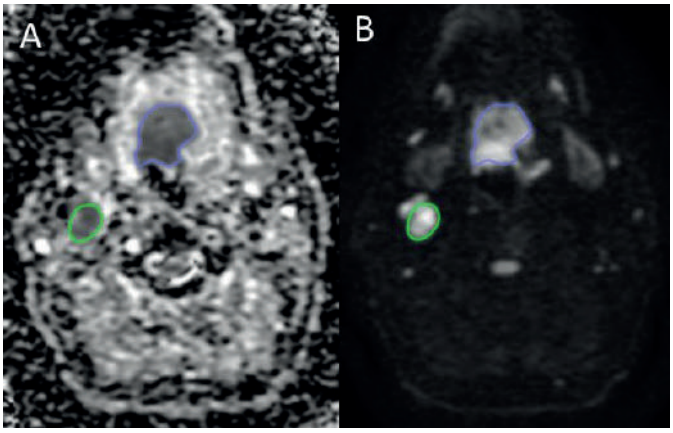


Figure 2 *Example of whole-lesion segmentation DWI on the A) ADC map and B) b1000 imaging. Delineation was performed in areas with low signal intensitiy on the ADC map and high signal intensity on the b1000 images in accordance with anatomic sequences*

RESULTS

Patients characteristics

The study population consisted of 134 patients (97 males; mean age, 64.7 years; range, 41-91 years; Table 1). Fifty-eight patients underwent both DWI and ^{18}F-FDG PET-CT (within a mean interval of six days (range, 0-13), mean of 16 days (range, 7-25 days) before treatment initiation), 31 patients underwent DWI only, 45 patients ^{18}F-FDG-PET-CT only (Figure 3). All 134 patients completed radiotherapy with 70 Gy in total, of which 18 accelerated radiotherapy. Eighty-seven patients (58.2%) received concomitant cisplatin and 18 patients received cetuximab).

Table 1 *Patient characteristics*

Characteristics		Total (n=134)
Sex, n (%)	Male	97 (72.4)
	Female	37 (27.6)
Age (year)	Mean age	66.4
	Range	41.0 - 90.1
Site, n (%)	Oral cavity	3 (2.2)
	Orofarynx	96 (71.6)
	Hypofarynx	24 (17.9)
	Larynx	11 (8.2)
T-stage, n (%)	T1	16 (11.9)
	T2	44 (32.8)
	T3	33 (24.6)
	T4	41 (30.6)
N-stage, n (%)	N0	20 (14.9
	N1	25 (18.7)
	N2	88 (65.7)
	N3	1 (0.7)
HPV status, n (%)	Positive	47 (35.1)
	Negative	67 (50.0)
	Unknown	20 (14.9)
Treatment, n (%)	Radiotherapy	134 (100)
	Chemotherapy total	105 (78.4)
	Cisplatin	87 (82.9)
	Cetuximab	18 (17.1)
Outcome, n (%)[a]	Residual tumors	17 (12.7)
	Locoregional recurrence	28 (20.9)
	Deaths	94 (70.1)

[a] Multiple outcomes were possible, therefore the total exceeds 100%

Treatment outcome

Median follow-up was 25.6 months (interquartile range, 14.0-37.1 months). Seventeen (12.5%) patients had locoregional treatment failure. The failure rate was similar in the DWI only and ^{18}F-FDG-PET-CT only group for residual disease (13.5% versus 10.7%), recurrences (20.2% versus 20.4%) and death (29.2% versus 32.0%). Mean follow-up of patients without residual or recurrent disease was 27.3 months (interquartile range, 16.3-37.8 months). In 28 patients (20.9%) a recurrence occurred. During follow-up 42 patients (31.3%) died.

Histogram analysis of voxels in the VOI is illustrated in Figure 4. A wider distribution (higher ADC_{SD}) and higher ADC of voxels (higher ADC_{max}) were seen in the treatment failure group.

Correlation with HPV-status

Differences in functional imaging parameters between HPV-negative and HPV-positive tumours (Appendix A) showed that low-to-moderate correlation with HPV-status was found for $ADC_{mean\text{-}PT}$ (r=0.452, *P*<0.01), $ADC_{max\text{-}PT}$ (r=0.319, *P*=0.002), $ADC_{p10\text{-}PT}$ to $ADC_{p90\text{-}PT}$ (r=0.4-0.48, *P*<0.01), $SUV_{peak\text{-}PT}$ (r=0.326, P<0.01), $SUV_{mean\text{-}PT}$ (r=0.325, *P*<0.01).

Univariate analysis of treatment response

Diffusion-weighted imaging and ^{18}F-FDG-PET-CT parameters predictors of treatment response (Appendix B) were higher in patients with treatment failure. Subgroup analysis in tumors with MATV larger than 4.2 ml (i.e., not effected by partial volume effects), did not show any significant predictors.

Univariate survival analysis of long-term outcome (RFS, OS)

The evaluation of recurrence-free survival (Appendix B) showed that $ADC_{max\text{-}PT}$ (*P*=0.012), $MATV_{PT}$ (*P*=0.001), and TLG_{PT} (*P*=0.039) were higher in patients with a tumor recurrence. Subgroup analysis in tumors with MATV >4.2 ml, showed that $MATV_{PT}$ and TLG_{PT} remained significant predictors (*P*=0.001, *P*=0.003, respectively).

Corrected for TNM-stage and HPV-status, $ADC_{gtv\text{-}PT}$ (*P*<0.001), $ADC_{SD\text{-}PT}$ (*P*=0.031), $ADC_{max\text{-}PT}$ (*P*=0.007), $SUV_{peak\text{-}PT}$ (*P*=0.025), $SUV_{mean\text{-}PT}$ (*P*=0.039) and TLG_{PT} (*P*=0.021) were significantly higher in patients with poor OS. Subgroup analysis showed that $MATV_{PT}$ and TLG_{PT} remained significant (*P*<0.001, P=0.009, respectively) in tumors >4.2ml.

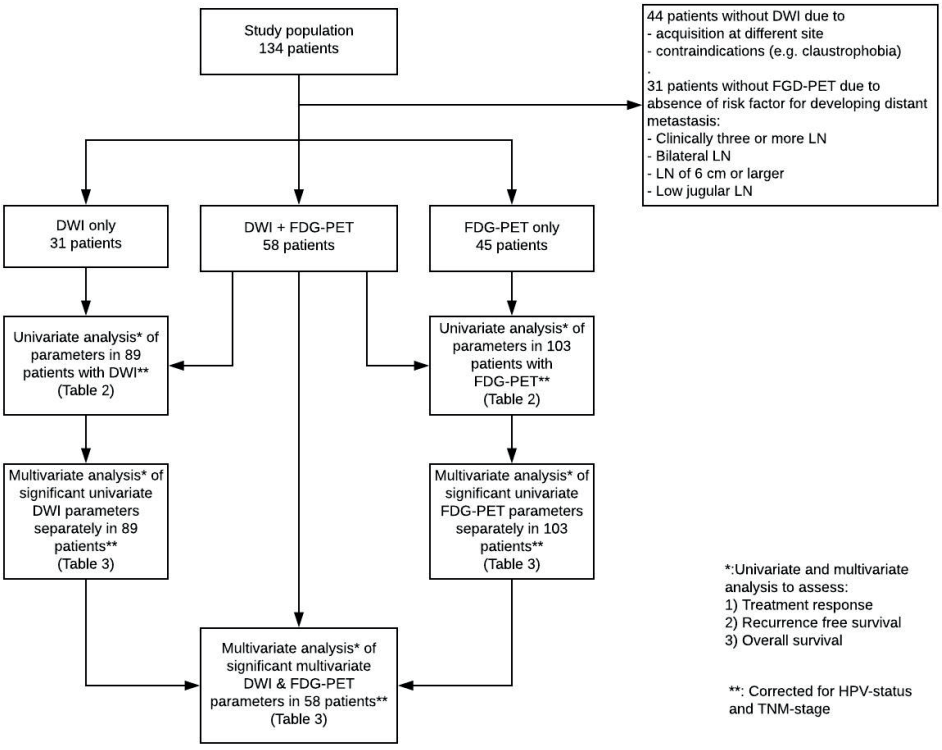


Figure 3 *Flowchart of patient inclusion*

Table 2 *Multivariate analysis of significant univariate parameters assessed per DWI or PET separately, corrected for TNM-stage and HPV-status. In the 'modalities together' section significant parameters of the single modality section are combined.*

			Treatment response						Recurrence free survival						Overall survival					
			Single modality			Modalities together			Single modality			Modalities together			Single modality			Modalities together		
			P value	Odds ratio	95%CI	P value	Odds ratio	95%CI	P value	Odds ratio	95%CI	P value	Odds ratio	95%CI	P value	Odds ratio	95%CI	P value	Odds ratio	95%CI
Primary tumor	PET	MATV	NS						NS						0.003	1.041	1.014-1.069	NS		
		SUV_{max}	0.049	1.188	1.000-1.411	0.049	1.302	1.010-1.679												
		SUV_{peak}	NS												NS					
		SUV_{mean}	NS												NS					
		TLG	NS						0.003	1.006	1.002-1.011	0.003	1.006	1.010-1.035	NS					
	DWI	ADC_{GTV}													<0.001	1.063	1.031-1.095	<0.001	1.051	1.013-1.090
		ADC_{SD}	NS												NS					
		ADC_{max}	0.024	5.586	1.250-24.972	NS									NS					
		ADC_{p70}							NS											
		ADC							NS											
Lymph Node	PET	TLG	0.031	1.012	1.001-1.024	NS														
	DWI	ADC_{p20}	0.047	22.985	0.987-535.043	NS														
		ADC_{p30}	NS																	

Abbreviations: NS = not significant

Multivariate analysis of treatment response

The significant univariate parameters for treatment response prediction in a multivariate analysis were assessed in DWI and ^{18}F-FDG-PET-CT separately, and for PT and LN separately (Table 2), corrected for TNM-stage and HPV-status. Primary tumor $SUV_{max\text{-}PT}$ (*P*=0.049) and $ADC_{max\text{-}PT}$ (*P*=0.024), and nodal $TLG_{\text{-}LN}$ (*P*=0.031) and $ADC_{p20\text{-}LN}$ (*P*=0.047) were significant predictors of treatment response.

Finally, these remaining significant parameters per modality were combined in patients with both imaging acquisitions for a head-to-head comparison (Table 2), which revealed $SUV_{max\text{-}PT}$ as significant predictive factor (*P*=0.042) in multivariate analysis with a hazard ratio of 1.204 (95%CI, 0.991-1.462). Subgroup analysis in tumors with a MATV larger than 4.2 ml, resulted in similar predictive parameter.

In order to determine the additional value of predicting treatment response using a combination of both modalities, the remaining significant multivariate parameters of both cohorts (ADC_{max} of DWI-cohort and SUV_{max} of ^{18}F-FDG-PET-CT-cohort) were analysed in patients with MATV>4.2 ml (Table 3). First, optimal cut-offs for $SUV_{max\text{-}PT}$ were 7.13, 11.3, 13.58; and for ADC_{max} were 2.236, 2.528, 2.671, respectively. The positive predictive value (PPV) and negative predictive value (NPV) were calculated for ADC_{max} and SUV_{max} at the cut-off values described above, showing additional value of combining ADC_{max} and SUV_{max}. The PPV for treatment failure was highest when both a high ADC_{max} (>2.528) and high SUV_{max} (>13.58) were measured. In our calculation, regardless of ADC_{max} and SUV_{max} <7.13 ruled out treatment failure. At higher SUV_{max} the NPV was lower as values of ADC_{max} and SUV_{max} decreased. The combination of significant multivariate nodal parameters $ADC_{p20\text{-}LN}$ and TLG_{LN} resulted in only a slightly better prediction of treatment failure (Appendix C).

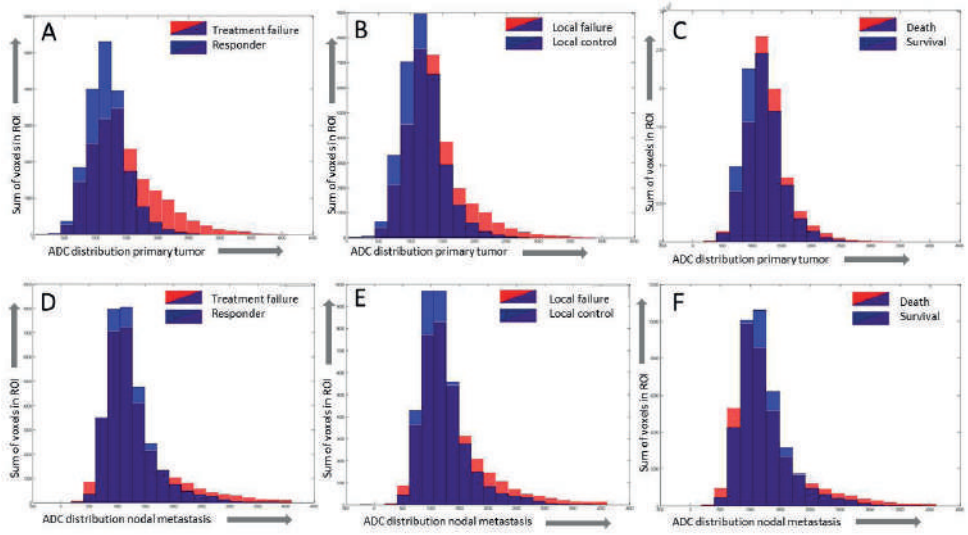


Figure 4 *Histogram analysis of in ROI included ADC voxels of primary tumors and lymph node metastasis with adverse outcome (red), good outcome (blue) and with overlap of voxels (purple). A) In treatment response assessment is shown, in which a wider distribution of the non-responders (ADC_{SD}) and higher maximum values of the ADC in non-responders (ADC_{max}). B) The locoregional failure of PT in the RFS showed a significant higher ADC_{max}. C) The overall survival of patients showed for PT a wider distribution and higher ADC_{max} values in non-responders. D) The histogram of the LN shows non-responders have slightly more of the lowest ADC voxels, lower kurtosis and higher ADC values. E) The LN ADC histogram of RFS shows higher kurtosis and a less divided distribution in the local control group. F) The overall survival showed for LN ADC parameters no significant predictor, which can be seen in the almost similar distribution of the twee compared group.*

Table 3 *Positive A) and negative B) predictive value for treatment failure using both ADCmax (horizontal) and SUVmax (vertical) for each prevalence using any ADCmax and SUVmax value and using the most optimal cut-off values.*

A. Positive predictive value for treatment failure

Prevalence	SUV_{max} cut-off	Any ADC_{max}	ADC_{max} > 1.927	ADC_{max} > 2.236	ADC_{max} > 2.528
0.10	Any SUV_{max}	0.10	0.20	0.36	0.43
	> 7.13	0,13	0.26	0.44	0.50
	> 11.30	0.19	0.34	0.54	0.61
	> 13.58	0.28	0.47	0.66	0.72
0.20	Any SUV_{max}	0.20	0.36	0.56	0.63
	> 7.13	0.25	0.44	0.64	0.70
	> 11.30	0.34	0.54	0.73	0.78
	> 13.58	0.46	0.66	0.82	0.85
0.30	Any SUV_{max}	0.30	0.49	0.69	0.74
	> 7.13	0.37	0.57	0.75	0.80
	> 11.30	0.47	0.67	0.82	0.86
	> 13.58	0.60	0.77	0.88	0.91
0.40	Any SUV_{max}	0.40	0.60	0.77	0.82
	> 7.13	0.48	0.67	0.82	0.86
	> 11.30	0.58	0.76	0.88	0.90
	> 13.58	0.70	0.84	0.92	0.94
0.50	Any SUV_{max}	0.50	0.69	0.84	0.87
	> 7.13	0.58	0.76	0.87	0.90
	> 11.30	0.67	0.82	0.91	0.93
	> 13.58	0.78	0.89	0.95	0.96

B. Negative predictive value for treatment failure

Prevalence	SUV_{max} cut-off	Any ADC_{max}	ADC_{max} < 2.528	ADC_{max} < 2.236	ADC_{max} < 1.927
0.10	Any SUV_{max}	0.90	0.93	0.94	0.95
	< 13.58	0.93	0.95	0.96	0.97
	< 11.30	0.94	0.96	0.97	0.97
	< 7.13	1.00	1.00	1.00	1.00
0.20	Any SUV_{max}	0.80	0.85	0.88	0.89
	< 13.58	0.86	0.90	0.92	0.93
	< 11.30	0.88	0.92	0.93	0.94
	< 7.13	1.00	1.00	1.00	1.00
0.30	Any SUV_{max}	0.70	0.77	0.81	0.83
	< 13.58	0.79	0.84	0.87	0.89
	< 11.30	0.82	0.86	0.89	0.90
	< 7.13	1.00	1.00	1.00	1.00
0.40	Any SUV_{max}	0.60	0.68	0.73	0.76
	< 13.58	0.70	0.77	0.81	0.84
	< 11.30	0.74	0.80	0.84	0.86
	< 7.13	1.00	1.00	1.00	1.00
0.50	Any SUV_{max}	0.50	0.59	0.64	0.68
	< 13.58	0.61	0.69	0.74	0.77
	< 11.30	0.65	0.73	0.77	0.80
	< 7.13	1.00	1.00	1.00	1.00

Multivariate analysis of long-term outcome (RFS, OS)

The significant univariate parameters for RFS prediction, were assessed per modality in a multivariate analysis, corrected for TNM-stage and HPV-status. This resulted in only TLG_{PT} as significant (P=0.003) prognostic multivariate ^{18}F-FDG-PET-CT parameter (Table 2). None of the DWI metrics remained significant. In multivariate analysis, TLG_{PT} (*P*=0.002) remained a significant predictor of RFS. Subgroup analysis (MATV >4.2 ml) resulted in TLG_{PT} (*P*=0.039) as predictive parameter for RFS. Using TLG_{-PT}, the optimal cut-off value of 36.2 resulted in a significant log-rank test for RFS prediction (Figure 5A).

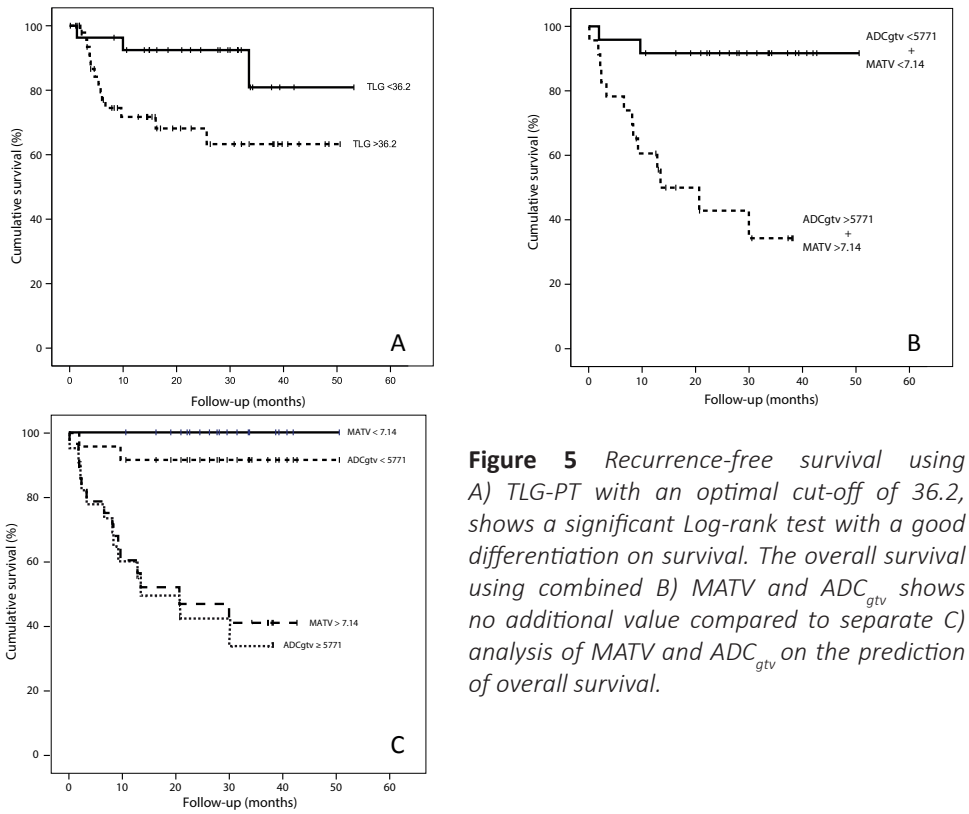


Figure 5 *Recurrence-free survival using A) TLG-PT with an optimal cut-off of 36.2, shows a significant Log-rank test with a good differentiation on survival. The overall survival using combined B) MATV and ADC_{gtv} shows no additional value compared to separate C) analysis of MATV and ADC_{gtv} on the prediction of overall survival.*

The univariate ^{18}F-FDG-PET-CT-parameters for prognosis of OS ($MATV_{PT}$, SUV_{max-PT}, $SUV_{peak-PT}$, and TLG_{PT}) combined in a multivariate analysis, resulted in $MATV_{PT}$ (*P*=0.003) as remaining significant prognosticator (single modality, Table 2).

The univariate significant DWI parameters for OS (ADC_{GTV-PT}, ADC_{SD-PT} and ADC_{max-PT}) combined in a multivariate analysis (modalities together, Table 2) resulted in ADC_{GTV-PT} (*P*<0.001) and ADC_{SD-PT} (*P*=0.05) as significant prognosticators. ADC_{GTV-PT} (*P*=0.009) remained a significant predictor in subgroup analysis (MATV >4.2 ml) who underwent both imaging modalities.

The use of both $MATV_{PT}$ and ADC_{GTV-PT} (Figure 5C) resulted in a similar survival curve as for single parameter assessment (Figure 5B), which showed no additional value for the use of both modalities for the OS prognosis.

DISCUSSION

Pretreatment DWI and ^{18}F-FDG-PET-CT were evaluated for the predictive value for treatment response, recurrence-free survival and overall survival. The assessment of both $SUV_{max\text{-}PT}$ and $ADC_{max\text{-}PT}$ showed additional value for treatment response prediction, compared with single parameter assessment. Furthermore, a high TLG_{PT} was prognostic for locoregional recurrence, and high $ADC_{GTV\text{-}PT}$, $MATV_{PT}$, and $ADC_{SD\text{-}PT}$ were prognostic for OS. Below, we will discuss the aetiology of the applied functional parameters, the predictive and prognostic value, and finally the implications for clinical practice.

In order to identify tumor characteristics with quantitative analyses, whole-lesion delineation could capture tumor heterogeneity and ignores subjective exclusion of necrosis or other potential predictive characteristics (15, 23, 24). This heterogeneity may be caused by areas with high cellularity, necrosis, stroma and areas with increased or decreased vascularity. The mean value of an imaging parameter may be insufficient, because when areas with low and high ADC values are included in the ROI, heterogeneity is flattened out.

Tumor heterogeneity (Figure 3) may be depicted by ADC histogram analysis (15, 25). A wide distribution, as sign of lesion heterogeneity, might be reflected by the ADC_{SD} (25). Furthermore, low-cellular tumor parts (e.g. necrosis) might be measured with high ADC values (high ADC_{max} or ADC percentiles), while high-cellular solid viable areas are reflected by low ADC values. In our study primary tumor ADC_{SD} and ADC_{max} parameters were higher in non-responders, which might be due to limited efficacy of radiotherapy in structures with a low cellularity (e.g. low diffusion restriction such as necrosis and fibrosis) (26). Lymph node $ADC_{p20\text{-}LN}$ was found higher in non-responders. This might be caused by the high density of metastatic cell besides the nodal lymphocytes, insufficient vascularization, a higher degree of necrosis (7, 14, 27).

An HPV-negative tumor status is correlated with a high ADC value (r=0.452). This is reflected by a higher ADC_{SD} and higher ADC_{max}, and characterized by a keratinizing morphology and large areas of central necrosis with cystic changes, small stromal volume and insufficient vascularization (23). Contrary, HPV-positive tumors have a typical non-keratinizing morphology, with small central necrosis and large amounts of infiltrating lymphocytes, reduced extracellular space (i.e increased ADC), sufficient vascularization and, therefore, responsive to (chemo)radiotherapy (10, 26).

In HPV-negative tumors HIF1α-induced glucose metabolism is increased as evidenced by increased glycolysis and proliferation, production of lactate and hypoxic microenvironment. In contrast, HPV-positive cells effectively utilize mitochondrial respiration as evidenced by increased oxygen consumption (28-30). We confirmed this hypothesis by finding significant higher ^{18}F-FDG-PET parameters ($MATV_{PT}$, $SUV_{max\text{-}PT}$, $SUV_{peak\text{-}PT}$, $SUV_{mean\text{-}PT}$ and TLG_{PT} in univariate analysis) in HPV-negative than HPV-positive patients.

The main predictor for treatment response, when both imaging techniques were analyzed together, was $SUV_{max\text{-}PT}$, which was significant higher in non-responders. This is in line with literature, where highly metabolic 'aggressive' tumor activity is associated with larger

tumor size, high proliferation rate, necrosis and HPV-negative status leading to a poor treatment response (24, 28). However, higher odds ratios for treatment failure were found for $ADC_{max\text{-}PT}$ (OR=5.59; 95%CI 1.25-24.97) and $ADC_{p20\text{-}LN}$ (OR=22.98; 95%CI, 0.99-535.04) than $SUV_{max\text{-}PT}$ (OR=1.19; 95%CI 1.00-1.41).

We found $TLG\text{-}_{PT}$ to be prognostic for RFS. This is in line with other studies, in which identification of necrosis or hypoxia and a higher glycolytic and increased tumor heterogeneity in HPV-negative tumors was predictive for an adverse outcome (11, 15). Total lesion glycolysis has been suggested to reflect global metabolic activity in whole tumors better than MATV, as TLG represents both functional tumor burden and biological aggressiveness (31). Regarding OS, high $MATV_{PT}$ and $ADC_{GTV\text{-}PT}$ were found to be predictive for poor OS, which was described in previous studies (4, 32).

The identification of predictive tumor characteristics for response to (chemo)radiotherapy, such as a high intratumoral heterogeneity (ADC_{SD}, ADC_{max}, $ADC_{P20\text{-}LN}$), tumor aggressiveness (SUV_{max}) and a negative HPV-status, could help stratify risk groups, which might implicate tailored treatment (e.g., intensifying concurrent chemotherapy or offering alternative treatment options). Patients with high TLG_{PT} and $ADC_{GTV\text{-}PT}$ may be considered at high risk of adverse outcome and may benefit from intensifying post-treatment monitoring.

Our findings suggested that pre-treatment response prediction benefits from combining DWI with ^{18}F-FDG-PET-CT imaging, by measuring the $ADC_{max\text{-}PT}$ and $SUV_{max\text{-}PT}$ (Table 3). However, the cons of performing both imaging must be considered, e.g. more frequent artifacts with DWI, radiation and higher costs with ^{18}F-FDG-PET and CT. Optimal cut-off values of $ADC_{max\text{-}PT}$ and $SUV_{max\text{-}PT}$ showed their diagnostic potential to considerably modify the a priori risk of treatment failure. In contrast, no additional value of combining imaging modalities was found for RFS and OS.

Limitations

Some limitations must be acknowledged. First, ^{18}F-FDG-PET-CT delineation bias could have occurred using a 50% SUV_{peak} threshold, which excludes necrotic areas with low SUV (10). This results in overall higher SUV and lower MATV, although it is more reproducible. Second, the assessment of the largest LNs only could falsely ignore the adverse effect of having several small LN metastases. However, small LNs (<5 mm) cannot be assessed with sufficient accuracy on ^{18}F-FDG-PET-CT (33). Moreover, it is likely that the chance of eradication of small lymph node metastases is higher than larger LN. Therefore, the prognosis might be determined by the largest lymph node metastases.

Conclusion

Both DWI and ^{18}F-FDG PET-CT parameters appear to have predictive value for treatment response, recurrence-free survival and overall survival. Combining $SUV_{max\text{-}PT}$ with $ADC_{max\text{-}PT}$ improved treatment failure prediction compared to single parameter assessment. Being able to stratify patients with favorable and unfavorable outcomes might help determining patient tailored treatment.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66(1):7-30.
2. Semrau S, Waldfahrer F, Lell M, et al. Feasibility, toxicity, and efficacy of short induction chemotherapy of docetaxel plus cisplatin or carboplatin (TP) followed by concurrent chemoradio-therapy for organ preservation in advanced cancer of the hypopharynx, larynx, and base of tongue. Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft [et al]. 2011;187(1):15-22.
3. Chiesa F. Surfing prognostic factors in head and neck cancer at the millenium.
4. Xie P, Li M, Zhao H, Sun X, Fu Z, Yu J. 18F-FDG PET or PET-CT to evaluate prognosis for head and neck cancer: a meta-analysis. J Cancer Res Clin Oncol. 2011;137(7):1085-93.
5. Nakajo M, Nakajo M, Kajiya Y, et al. FDG PET/CT and diffusion-weighted imaging of head and neck squamous cell carcinoma: Comparison of prognostic significance between primary tumor standardized uptake value and apparent diffusion coefficient. Clinical nuclear medicine. 2012;37(5):475-80.
6. Harriet C. Thoeny M, Frederik De Keyzer, MSc, Ann D. King, MD. Diffusion-weighted MR Imaging in the Head and Neck. 2012.
7. Plaxton NA, Brandon DC, Corey AS, et al. Characteristics and Limitations of FDG PET/CT for Imaging of Squamous Cell Carcinoma of the Head and Neck: A Comprehensive Review of Anatomy, Metastatic Pathways, and Image Findings. AJR American journal of roentgenology. 2015;205(5):W519-31.
8. Hermans R. Diffusion-weighted MRI in head and neck cancer. Curr Opin Otolaryngol Head Neck Surg. 2010;18(2):72-8.
9. Machtay M, Natwa M, Andrel J, et al. Pretreatment FDG-PET standardized uptake value as a prognostic factor for outcome in head and neck cancer. Head & neck. 2009;31(2):195-201.
10. Schouten CS, de Graaf P, Bloemena E, et al. Quantitative diffusion-weighted MRI parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. AJNR American journal of neuroradiology. 2015;36(4):763-7.
11. Chawla S, Kim S, Dougherty L, 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;200(1):35-43.
12. Manca G, Vanzi E, Rubello D, et al. 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.
13. Krabbe CA, Pruim J, Scholtens AM, et al. 18F-FDG PET in Squamous Cell Carcinoma of the Oral Cavity and Oropharynx: A Study on Inter- and Intraobserver Agreement. J Oral Maxillofac Surg. 2010;68(1):21-7.
14. Kim S, Loevner L, Quon H, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(3):986-94.
15. Srinivasan A, Chenevert TL, Dwamena BA, et al. Utility of pretreatment mean apparent diffusion coefficient and apparent diffusion coefficient histograms in prediction of outcome to chemoradiation in head and neck squamous cell carcinoma. Journal of computer assisted tomography. 2012;36(1):131-7.
16. Sheng-Chieh Chan N-MC, Chia-Hsun Hsieh, Shu-Hang Ng, Chien-Yu Lin, Tzu-Chen Yen, Cheng-Lung Hsu, Hung-Ming Wan, Chun-Ta Liao, Kai-Ping Chang, Jiun-Jie Wang. Multiparametric imaging using 18 parameters and functional MRI F-FDG PET/CT heterogeneity techniques: prognostic significance in patients with primary advanced oropharyngeal or hypopharyngeal squamous cell carcinoma treated withchemoradiotherapy. Oncotarget. 2017.
17. Patrick M. Bossuyt JBR, David E. Bruns, Constantine A. Gatsonis, Paul P. Glasziou, Les Irwig, Jeroen G. Lijmer, David Moher, Drummond Rennie, Henrica C.W. de Vet, Herbert Y. Kressel, Nader Rifai, Robert M. Golub, Douglas G. Altman, Lotty Hooft, Daniël A. Korevaar, MD Jérémie F. Cohen. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. 2015.

18. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-54.

19. Virginie Frings FHPvV, Linda M. Velasquez, Wendy Hayes, Peter M. van de Ven, Otto S. Hoekstra, Ronald Boellaard. Repeatability of Metabolically Active Tumor Volume Measurements with FDG PET/CT in Advanced Gastrointestinal Malignancies: A Multicenter Study.

20. Schinagl DA, Vogel WV, Hoffmann AL, van Dalen JA, Oyen WJ, Kaanders JH. Comparison of five segmentation tools for 18F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer. International journal of radiation oncology, biology, physics. 2007;69(4):1282-9.

21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63.

22. Cysouw MCF, Kramer GM, Schoonmade LJ, Boellaard R, de Vet HCW, Hoekstra OS. Impact of partial-volume correction in oncological PET studies: a systematic review and meta-analysis. European journal of nuclear medicine and molecular imaging. 2017;44(12):2105-16.

23. Driessen JP, van Bemmel AJ, van Kempen PM, et al. Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. Head Neck. 2015.

24. Paidpally V, Chirindel A, Lam S, Agrawal N, Quon H, Subramaniam RM. FDG-PET/CT imaging biomarkers in head and neck squamous cell carcinoma. Imaging Med. 2012;4(6):633-47.

25. Choi YJ, Lee JH, Kim HO, et al. Histogram Analysis of Apparent Diffusion Coefficients for Occult Tonsil Cancer in Patients with Cervical Nodal Metastasis from an Unknown Primary Site at Presentation. Radiology. 2016;278(1):146-55.

26. Foo SS, Abbott DF, Lawrentschuk N, Scott AM. Functional imaging of intratumoral hypoxia. Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging. 2004;6(5):291-305.

27. Chawla S, Kim S, Wang S, Poptani H. Diffusion-weighted imaging in head and neck cancers. Future Oncol. 2009;5(7):959-75.

28. Young-Suk Jung1, *, Abdo J. Najy1,*, Wei Huang1, Seema Sethi1, Michael Snyder2,3,, Wael Sakr1 GD, Maik Hüttemann4, Icksoo Lee4,8, Rouba Ali-Fehmi1, Silvia, Franceschi5 LS, Harold E. Kim2,3, Ikuko Kato2,1 and Hyeong-Reh Choi Kim1. HPV-associated differential regulation of tumor metabolism in oropharyngeal head and neck cancer.

29. Schouten CS, Hakim S, Boellaard R, et al. Interaction of quantitative (18)F-FDG-PET-CT imaging parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. Head Neck. 2016;38(4):529-35.

30. Paul Burgman JAOD, John L. Humm, and C. Clifton Ling. Hypoxia-Induced Increase in FDG Uptake in MCF7 Cells. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2001;42:170–175.

31. Steven M. Larson YE, Timothy Akhurst, Madhu Mazumdar,, Homer A. Macapinlac RDF, Cecille Casilla, Melissa Fazzari, MS,, Neil Srivastava HWDY, John L. Humm, Jose Guillem,, Robert Downey MK, Alfred E. Cohen, Robert Ginsberg. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging: The Visual Response Score and the Change in Total Lesion Glycolysis. Clinical Positron Imaging, Volume 2, Number 3. 1999.

32. Chen SW, Hsieh TC, Yen KY, Liang JA, Kao CH. Pretreatment (18)F-FDG PET/CT in whole-body total lesion glycolysis to predict survival in patients with pharyngeal cancer treated with definitive radiotherapy. Clinical nuclear medicine. 2014;39(5):e296-300.

33. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2009;50 Suppl 1:122S-50S.

CHAPTER 6

SUMMARY, GENERAL DISCUSSION AND FUTURE PERSPECTIVES

SUMMARY, GENERAL DISCUSSION AND FUTURE PERSPECTIVES

In this thesis we assessed the role of functional imaging in head and neck cancer. With the use of functional imaging we aimed to improve the diagnostic workup of: 1) patients presenting with an unknown primary tumor; 2) patients with a high risk of developing distant metastases; and 3) patients with potential residual disease after (chemo) radiotherapy. Further we assessed the prognostic capacity of functional imaging before the start of (chemo)radiotherapy and we assessed the reproducibility of diffusion-weighted (DWI) magnetic resonance imaging (MRI). We will discuss our findings below starting with discussing the findings of three systematic reviews on the use of DWI, contrast-enhanced perfusion MRI, and intra-treatment functional imaging, respectively.

Potential applications of functional imaging

In **Chapter 2** we discuss the potential applications of functional imaging in three systematic reviews.

In **Chapter 2.1** we focused on the value of contrast-enhanced perfusion magnetic resonance imaging (MRI). The diagnostic and prognostic capacity were assessed separately. With contrast-enhanced perfusion MRI, serial MRI images are made in order to estimate contrast-enhancement. Tissue perfusion can be estimated by using pharmacokinetic modelling. We included 33 studies. In 22 of these studies the diagnostic potential was assessed, and in 11 studies the prognostic capacity was assessed. The most frequently used technique was dynamic contrast enhanced (DCE) MRI. Several studies assessed the correlation between DCE parameters and positron emission tomography (PET) parameters and found a low, but significant, correlation between both modalities. This indicates that both modalities offer complementary information, which was also expected because both techniques are based on different properties. Therefore, the combination of DCE and PET may be an application of PET-MRI (1-4).

One of the most promising results of diagnostic DCE-MRI was the potential to detect heterogeneity within lesions. A correlation between the standard deviation of the rate constant between interstitial space and plasma (k_{ep}) and vascular endothelial growth factor (VEGF) staining was found (2). Lesion heterogeneity may be a sign of tissue hypoxia. Hypoxia is a sign of inadequate perfusion of the tumor (5). Radiotherapy requires the presence of oxygen in order to form free radicals and induce tumor cell death. Moreover, adequate tumor perfusion is needed for chemotherapy to reach the tumor. Therefore, lesion heterogeneity on imaging may be indicative of an unfavorable response to chemoradiotherapy. This hypothesis was confirmed by some of the prognostic studies where high skewness of the volume transfer constant between plasma and interstitial space (K^{trans}) (i.e. heterogeneous tumor perfusion) was associated with a poor prognosis (6). Moreover, several studies found low pre-treatment K^{trans} (i.e. low permeability) to be indicative of poor treatment outcome (7-9).

We concluded that perfusion-weighted MRI shows great potential in various aspects of diagnosing head and neck squamous cell carcinoma (HNSCC) and in the prediction of short-term prognosis. However, at this moment perfusion-weighted MRI is not considered

to be reproducible enough to be used in clinical practice for HNSCC. More research with uniform study methods and with larger sample sizes is needed.

In **Chapter 2.2** we assessed the diagnostic and prognostic value of intravoxel incoherent motion (IVIM) MRI. With IVIM, tissue perfusion and diffusion can be estimated without using intravenous contrast. Diffusion restriction is considered to be indicative of malignancy (10-14). For this review we included 17 studies (10 diagnostic, five prognostic and two assessing both). Five diagnostic studies focused on differentiating between different malignant and benign lesions (15-19). By combining IVIM parameters squamous cell carcinomas, lymphomas, malignant salivary gland tumors, Warthin's tumors and pleomorphic adenomas could be differentiated with a sensitivity of 85-87% and specificity of 80-100% (15, 17, 19). Three studies found malignant salivary gland tumors to have intermediate IVIM values compared to the benign Warthin's tumors (lower values) and pleomorphic adenoma (higher value). Combining the diffusion coefficient (D) and pseudodiffusion coefficient (D*) resulted in a sensitivity of 100% (95%CI, 54-100%) and a specificity of 94-100% (95%CI, 71-100%) in differentiating between benign and malignant salivary gland tumors (16-18). Two studies assessed the prognostic value of IVIM in neoadjuvant chemotherapy (NAC) in nasopharyngeal carcinoma (20, 21). Pretreatment D was the single best parameter with a sensitivity and specificity of 64-65% and 72-81% in predicting response to NAC. When using the difference in D before and after NAC sensitivity increased to 94% and specificity was 77% (21). In another study the response of hypopharyngeal carcinoma to NAC was predicted with pretreatment D, resulting in a sensitivity of 75% and specificity being 89% (22). Contrary to these results Hauser et al. found the perfusion factor (f) to be the strongest predictor in HNSCC patients receiving chemoradiotherapy (23, 24). Low values of D and f were associated with a more favorable prognosis (20-24).

The finding that low f (i.e. low perfusion) was associated with favorable prognosis appears to be contradictory to the observation that low K^{trans}, also a sign of low perfusion, was associated with a poor prognosis (7-9, 23, 24). As mentioned before, a well perfused tumor should be more sensitive for chemoradiotherapy which should result in a more favorable outcome (6). On the other hand, it can also be argued that a well-perfused tumor with a high microvessel density is more aggressive and has more metastatic potential, thereby resulting in an adverse prognosis (25, 26). Preferably a direct comparison between DCE and IVIM parameters is performed in the same population to deal with this apparently contradictory finding. Ideally then also potential confounders are included in the analysis (e.g. tumor volume, disease stage, smoking status, HPV status and treatment).

In conclusion, we found that combinations of IVIM parameters made it possible to reliably differentiate between various types of tumors. Low pre-treatment D and f and an increase in D during treatment were associated with a favorable response to treatment.

In **Chapter 2.3** we focused on the use of functional imaging in early follow-up after the start of (chemo)radiotherapy in HNSCC. We did choose for intra-treatment imaging because we expected functional changes to precede changes is size. Moreover during treatment, it is still possible to make adjustments based on the response (e.g. either de-

escalation or intensification of treatment or convert to surgical (salvage) treatment with less complications). Functional computed tomography (CT), MRI and PET techniques were included. Finally, we included 53 studies (four on CT, 23 on MRI and 26 on PET). Studies were divided in short-term response prediction up to three months after treatment and long-term response prediction two years after treatment.

For short-term response prediction most data were available for DWI and PET imaging. We found between two and three weeks after the start of treatment to be the most optimal timing of imaging. At this time apparent diffusion coefficient (ADC) increase and standardized uptake value (SUV) reduction were predictive for a favorable response to treatment (27-33). A plateau in ADC change was observed three weeks after treatment (28, 29, 31). For ^{18}F-Fluorodeoxyglucose (^{18}F-FDG) PET-CT, (chemo)radiotherapy results in ^{18}F-FDG uptake by causing an inflammatory response which may last until as long as three months after treatment (34, 35). However an early decrease in ^{18}F-FDG after 2-3 weeks is a sign of a favorable response (32, 33).

For long-term response prediction it was also possible to predict the prognosis by performing imaging in the first three weeks of treatment. An increase of blood volume on perfusion CT (36, 37), an increase in ADC (10, 31, 38-42), a reduction in K^{trans} (43) and reductions in metabolic PET-parameters (32, 33, 35, 44-49) were all associated with locoregional control.

To conclude, we found various functional imaging parameters to have a role in prediction response to treatment and locoregional control. However, although many of the studies included in this systematic review show predictive value, cut-off points which can be used in daily clinical practice for treatment decisions have yet to be determined.

Reproducibility of diffusion-weighted imaging

One of the main challenges of all functional techniques that need to be assessed is the reproducibility of results within, and mainly between MRI systems. In **Chapter 3.1** we performed sequential imaging in seven healthy volunteers on five MRI systems, at three time points in two institutes in order to assess reproducibility. A secondary aim was to find a tissue that may serve as a reference tissue. Apparent diffusion coefficient values of five different tissues were obtained: submandibular gland, sternocleidomastoid muscle, spinal cord, subdigastric lymph node and tonsil. In the tonsils it was possible to place a region of interest (ROI) in 87% of images due to artifacts caused by the presence of an air-tissue boundary and (involuntary) patient movement (e.g. caused by breathing and swallowing). In the other regions a ROI could be drawn on 97-98% of images.

To deal with inter- and intra-subject variation we used a linear mixed model. Of the included sequences turbo spin-echo (TSE)-DWI had the highest standard error of measurement (SEM) ($284.5 \cdot 10^{-6}$ mm^2/s) compared to $216 \cdot 10^{-6}$ mm^2/s and $190.3 \cdot 10^{-6}$ mm^2/s for echo-planar imaging (EPI) with 2 and 6 b-values (EPI-DWI-2b and EPI-DWI-6b), respectively. The spinal cord and tonsil were the tissues with the lowest SEM $151.2 \cdot 10^{-6}$ mm/s^2 and $190.1 \cdot 10^{-6}$ mm/s^2). Variance caused by times was very limited; the mean difference in ADC values

was only $5{\cdot}10^{-6}$ mm/s for imaging performed one month later. We did find significant differences in ADC values between the different MRI systems used in this study (P <0.001). Therefore, we concluded that the smallest range of ADC values can be obtained by imaging a subject on the same MRI system with an EPI-DWI with 6 b-values. Before quantitative DWI can be used in multicenter trials this issue needs to be solved. Of the investigated tissues, the spinal cord shows the least variance and is always visible on MRI of the neck. The spinal cord is therefore a candidate to serve as reference tissue in the head and neck region.

Diagnostic capacity of diffusion-weighted imaging
Other studies already focused on the difference in ADC values between HNSCC and other malignant and benign tissues (11, 50-52). Therefore, we focused on other diagnostic challenges in HNSCC: detecting distant metastases (**Chapter 4.1**), detection of a primary tumor in patients presenting with cervical metastasis and a clinically unknown primary tumor (**Chapter 4.2**) and detecting residual HNSCC after (chemo)radiotherapy (**Chapter 4.3** and **4.4**).

When distant metastases are present, HNSCC is considered incurable and only palliative treatment options remain. To avoid futile treatment, it is essential to detect distant metastases in patients at high risk of developing them (53). In **Chapter 4.1** we assessed the feasibility of whole-body MRI including whole-body DWI background body signal suppression (DWIBS) and compared the results to ^{18}F-FDG-PET-CT and chest-CT in patients with a high risk of developing distant metastases. In two patients a second primary tumor (one renal cell carcinoma and one adrenal malignancy) was found and in one patient a distant metastasis was found. All these lesions were detected with DWIBS at the cost of seven clinically indeterminate lesions that did not progress at follow-up. With ^{18}F-FDG-PET-CT the renal cell carcinoma was missed, however only three lesions were clinically indeterminate. On chest-CT eight lesions were clinically indeterminate and the adrenal malignancy was missed. We concluded that our WB-MRI protocol including DWIBS was feasible. The addition of DWIBS to the imaging protocol allows for fast image interpretation because since malignant tissue can be detected “at-a-glance”. The higher soft tissue detail of MRI compared to ^{18}F-FDG-PET-CT and chest-CT may result in more incidental findings. To deal with this higher detail, some experience in WB-MRI is necessary. Then lesions can not only be detected, but also characterized based on imaging characteristics. This may reduce the need to take biopsies.

Another patient population at risk of over-treatment are patients presenting with a cervical lymph node metastasis containing squamous cell carcinoma without an apparent primary tumor. The current consensus is that a primary tumor must be present. Therefore, a patient with an unknown primary is staged as Tx instead of T0 in the eight edition of the American Joint Committee on cancer staging (AJCC) manual (54). If the primary tumor cannot be found despite extensive diagnostic testing, then the whole mucosal lining of the upper aerodigestive tract, where the primary tumor is expected to be located, is treated with radiotherapy (55). In **Chapter 4.2** we attempted to increase the accuracy of diagnostic imaging. Therefore, we assessed and compared the diagnostic value of DWI and ^{18}F-FDG-

PET-CT for detecting unknown primary tumors in patients presenting with nodal metastasis without an apparent primary tumor. We included 31 patients presenting with an occult primary tumor after a complete head and neck examination including flexible endoscopy at the outpatient clinic. The final diagnosis as defined by the multi-disciplinary team with access to all available diagnostic results, including the results of biopsies performed during an examination under anesthesia, was used as the reference standard. If no primary tumor was found it was defined as Tx (54). This reference standard has some limitations: 1) sampling errors with biopsies may result in missing primary tumors; 2) because patients are treated with extensive field radiotherapy these missed primary tumors will not become clinically evident during follow-up. An alternative approach is to use a detection rate as proposed by Rusthoven et al. (56). The authors assessed the additional value of ^{18}F-FDG-PET-CT to a conventional workup including CT and/or MRI. The detection rate of ^{18}F-FDG-PET-CT was defined as the rate of lesions that were detected with ^{18}F-FDG-PET-CT, but were not detected on CT or MRI. In our study we compared DWI with ^{18}F-FDG-PET-CT. Even though ^{18}F-FDG-PET-CT is generally considered to be the most valuable imaging technique for detecting occult HNSCC (56-61), there is still lack of consensus on what the conventional workup is. Therefore, we choose not to report detection rates. In this study our main focus was high image sensitivity, because we consider that the reduction of extensive field radiotherapy outweighs the risk of an extra biopsy. With qualitative image analysis we found a non-significant trend for higher sensitivity of ^{18}F-FDG-PET-CT compared to DWI (93.8% vs 81.3%) with equal specificity (73.3%). Combining both modalities did not further improve diagnostic accuracy, probably because ^{18}F-FDG-PET-CT already had a very high sensitivity. Quantitative analysis was performed with histogram analysis of ADC on DWI and with maximum standardized uptake value (SUV_{max}) on ^{18}F-FDG-PET-CT. We found that SUV_{max} was also very accurate with a sensitivity and specificity of 81.3% and 93.3%, respectively. Volume on DWI was the only MRI parameter that was significantly different between malignant and benign lesions. Because qualitative analysis of DWI did have high diagnostic accuracy compared to quantitative analysis we hypothesize that parameters unrelated to ADC value contributed to the value of qualitative analysis. The tonsils are the most frequent location of an initially unknown primary HNSCC (61). Benign tonsillar tissue consists of relatively densely packed small lymphocytes which result in some degree of diffusion restriction. In some studies benign tonsils were even found to have lower ADC values than malignant tissue (51, 58). We could not confirm these results. We found significantly lower ADC_{mean} values for malignant tonsillar tissue compared to benign tonsils (P=0.05). To conclude we found a high sensitivity for diagnostic imaging, however even with state of the art diagnostic imaging the primary tumor location remained obscure in 15 of the included patients.

In **Chapter 4.3** we focused on the added value of DWI in patients with residual ^{18}F-FDG uptake three months after (chemo)radiotherapy. This selected patient population of 24 patients was chosen because this is the population where current follow-up strategies require further improvement. With negative findings on ^{18}F-FDG-PET-CT three months after (chemo)radiotherapy, the presence of residual malignancy can be ruled out with high confidence. This is reflected by reported negative predictive values of 92-99% (62-66). The positive predictive value of ^{18}F-FDG-PET-CT is suboptimal due to inflammatory post-irradiation effects which also result in ^{18}F-FDG-uptake (62-66). This results in unnecessary

biopsies and neck dissections with associated morbidity and risk of complications (67). By assessing the value of DWI in this population we aimed to fulfill a clinical need for more reliable characterization of residual ^{18}F-FDG avid lesions after (chemo)radiotherapy. When only assessing ^{18}F-FDG-PET-CT we found a sensitivity and specificity of 100% and 47%, respectively. When the results of DWI were added to ^{18}F-FDG-PET-CT in these patients with residual ^{18}F-FDG uptake and only a positive read on both ^{18}F-FDG-PET-CT and DWI was considered to be overall positive, then sensitivity remained 80%, and specificity increased to 88%. If the combined findings of DWI and ^{18}F-FDG-PET-CT would have been used to select patients at high risk of having residual disease for an examination under anesthesia (EUA), then in only 27% of patients an EUA would have been performed. With this strategy 33% of these EUAs would have been unnecessary, however at the expense of missing one patient with tumor residue. We therefore concluded that the addition of DWI to ^{18}F-FDG-PET-CT has the potential to substantially increase the specificity of response evaluation by imaging with limited decrease of sensitivity.

In **Chapter 4.4** we compared the diagnostic and prognostic value of routinely performed DWI and ^{18}F-FDG-PET-CT performed 3-6 months after (chemo)radiotherapy in order to assess if DWI has additional value to ^{18}F-FDG-PET-CT or may even outperform ^{18}F-FDG-PET-CT. It should be noted that the patients included in **Chapter 4.3** and **4.4** are fundamentally different. In **Chapter 4.3** all had residual ^{18}F-FDG uptake, whereas routinely performed imaging was included in **Chapter 4.4**. In **Chapter 4.4** we included 82 patients. DWI was analyzed quantitatively and qualitatively. For ^{18}F-FDG-PET-CT we used the Hopkins criteria (Table 1). Primary tumors and lymph nodes were analyzed separately. For primary tumor analysis with DWI sensitivity and specificity of 57.1% and 91.9%, respectively, were found. For ^{18}F-FDG-PET-CT primary tumor sensitivity and specificity were 85.7% and 86.5%, respectively. When combining both modalities, best results were achieved with a sequential approach only including the second modality in positive reads of the first modality. It did not matter which modality was assessed first. This resulted in a sensitivity and specificity of 57.1% and 97.3%, respectively. For lymph node analysis with DWI sensitivity and specificity were 100% and 72.1%, respectively. With ^{18}F-FDG-PET-CT sensitivity and specificity were 83.3% and 92.6%, respectively. Specificity of nodal assessment with ^{18}F-FDG-PET-CT was significantly higher than DWI. When combining modalities again best results were achieved with a sequential approach only including the second modality in positive reads of the first modality resulting in a sensitivity and specificity of 83.3% and 95.6%, respectively. It did not matter which modality was assessed first.

For quantitative DWI analysis we found the results to be inferior to qualitative analysis. This is in line with literature findings (68). This implies that the decision to consider a lesion as suspicious for containing malignancy is dependent of more factors than solely an ADC value. Also factors as asymmetry, findings on other sequences and the patient's history (e.g. swallowing complaints or weight loss) which are not captured by quantitative analysis may contribute to the value of qualitative image analysis. Unfortunately, qualitative image analysis carries a certain degree of subjectivity, which was expressed by inferior interobserver agreement compared to quantitative DWI analysis.

We concluded that a sequential approach including both qualitative analysis of DWI and ^{18}F-FDG-PET-CT resulted in the best diagnostic accuracy for follow-up after (chemo) radiotherapy.

Prognostic capacity of diffusion-weighted imaging

The main application of functional imaging in the future will probably be determining patient prognosis. Therefore, we assessed the prognostic value of DWI. It is generally accepted to only assess the functional parameters in vital tumor tissue (i.e. by excluding necrosis). In **Chapter 5.1** we determined the role of contrast-enhanced imaging in determining ADC values only of the vital tumor part. In this study DWI was performed using a TSE-based sequence. At that moment we were experiencing difficulties with image distortion on EPI-based sequences, therefore we choose to use TSE-based DWI. Turbo spin-echo sequences are associated with a lower signal-to-noise ratio. Therefore we included two sets of b-values: 0-750 s/mm^2 and 0-1000 s/mm^2, because the generally accepted high b-value of 1000 s/mm^2 could have resulted in a too low image signal for reliable image analysis (69). Image quality of b750 images was rated higher than b1000 images. However, we found primary tumor volume and lymph node ADC_{1000} to be significant and independent predictors of disease-free survival. The use of contrast-enhanced T1-weighted imaging to select only the solid part of the tumor did result in larger ROIs and higher ADC values. However, the prognostic value of DWI without and with the use of T1-weighted imaging was comparable. We concluded that pretreatment DWI may be an additional tool to determine patient prognosis.

To have the most impact on treatment choice it is preferred to reliably determine prognosis before starting any treatment. Therefore, we assessed the prognostic value of pretreatment DWI and ^{18}F-FDG-PET-CT in **Chapter 5.2** in patients treated with (chemo)radiotherapy. In this study we used histogram analysis for a more thorough analysis of malignant lesions. We found that both primary tumor SUV_{max} and ADC_{max} showed additional value for treatment response prediction, compared with single parameter assessment. Furthermore, a high primary tumor total lesion glycolysis (TLG) was prognostic for locoregional recurrence. A large primary tumor volume of ADC, high primary tumor ADC_{SD} and a large primary tumor metabolically active tumor volume (MATV) were prognostic for overall survival. High SUV_{max} and TLG are hallmarks of an aggressive tumor, associated with a larger tumor size, areas of necrosis and a negative HPV status, all resulting in an adverse prognosis (70-72). A high ADC_{max} and ADC_{SD} may be indicative of a tumor with areas of a relatively low cellularity containing large areas of necrosis or fibrosis which are radiotherapy-resistant (8, 11, 73). When combining DWI and ^{18}F-FDG-PET-CT both ADC_{max} and SUV_{max} remained independent predictors of treatment failure, which indicates that pretreatment response prediction may be an application of combined PET-MRI.

Future perspectives

With head and neck cancer treatment becoming less invasive, the role for non-invasive diagnostic and prognostic biomarkers will increase in the future. Functional imaging has great potential in providing both diagnostic and prognostic information.

In MRI, current research has already shown the promising value of functional MRI techniques. With the use of DWI especially the early change in parameters after the start of treatment offers an opportunity for treatment monitoring (27, 28, 30, 74, 75). Currently both qualitative and quantitative image analysis is performed. Qualitative image assessment is dependent on the experience of the observer and always carries a degree of subjectivity. The head and neck area is an anatomically complex area characterized by many air-tissue boundaries and subject to both voluntary and involuntary movement. Moreover, some degree of diffusion restriction is physiological in salivary glands, thyroid, tonsils and benign lymph nodes (58, 76). Some studies even suggest that normal tonsils display more diffusion restriction than tonsillar carcinoma (51, 58), however we could not confirm this in **Chapter 4.2**. This makes quantification of functional parameters challenging. Nonetheless progress has been made, e.g. by using histogram analysis to capture various aspects of quantitative data of the whole lesion instead of solely the mean value.

In contrast-enhanced perfusion MRI, DCE appears to be the technique with the most promising results. Especially the use of quantitative analysis should result in reliable and reproducible results with minimal interobserver variability. However, before DCE can be implemented in clinical practice more standardization of both imaging protocols and postprocessing is essential. In two studies performed by Heye et al. (77, 78) on patients with uterine fibroids large varieties in DCE parameters were found. The black-box nature of the currently available software resulted in up to a 100-fold difference in parameters with the same name which were obtained with different software packages. With DWI comparable problems are encountered as we demonstrated in **Chapter 3.1.** Because we found the least variation in ADC values if the same patient received imaging on the same MRI system, serial DWI may be suitable for early follow-up as long as the patient receives imaging on the same MRI system.

As with MRI, ^{18}F-FDG-PET-CT can also be assessed quantitatively and qualitatively. In contrast to functional MRI techniques, for ^{18}F-FDG-PET-CT imaging standards have been developed that result in repeatable and reproducible results within and between PET-CT systems (79). With these guidelines from the European Association of Nuclear Medicine (EANM) an attempt is made to create uniformity in performing, interpreting and reporting the results from ^{18}F-FDG-PET-CT. Guidelines are given for patient preparation, with special interest to patients with diabetes, renal failure, and pregnant or breastfeeding patients. Recommendations are given for FDG and contrast agent dose and administration. Minimal requirements for the PET imaging protocol are discussed, and how to combine CT protocols with the ^{18}F-FDG-PET-CT study. Further recommendations are given for PET- and CT-reconstruction, interpretation criteria and guidelines for uniform imaging documentation and reporting.

In image analysis of ^{18}F-FDG-PET-CT efforts have been made to create uniformity in qualitative image analysis. One of the first widely accepted qualitative imaging criteria were the Deauville criteria which were developed for the assessment of interim ^{18}F-FDG-PET-CT in lymphoma (80, 81). In the Deauville criteria the ^{18}F-FDG uptake of a lesion is compared to the mediastinal blood pool uptake and the liver uptake. This results in a 5-point Likert

scale (Table 1). A Deauville score 1-3 is generally considered with a complete metabolic response, whereas a score of 4-5 is a sign a stable or progressive disease. Even though these criteria were developed for the assessment of lymphoma, Sjövall et al. assessed the value of the Deauville criteria for response assessment of HNSCC after radiotherapy (82). The authors found that the Deauville criteria outperformed SUV_{max} in the assessment of regional tumor control (area under the curve (AUC)=0.82 vs 0.67).

The Hopkins criteria were developed for the assessment of ^{18}F-FDG-PET-CT after (chemo) radiotherapy in HNSCC (83). The Hopkins criteria are comparable to the Deauville criteria, except that the internal jugular vein is used as a reference tissue instead of the mediastinal blood pool. Again, a score of 1-3 is associated with benign conditions (e.g. postradiation inflammation), whereas a score of 4-5 is a sign of residual malignancy. The authors found substantial to very good interobserver agreement with κ ranging from 0.69 to 0.89 (83, 84). Agreement was comparable for primary tumor, left neck, right neck and overall response. It has been shown that the use of the Hopkins criteria for post-treatment response assessment may result in improved prediction of survival outcomes (85). Especially the negative predictive value of the Hopkins criteria is high, with reported negative predictive values (NPV) of over 90% when performed 12 weeks after (chemo)radiotherapy (83, 86). The positive predictive value is slightly lower with reported values of 62-71% (83, 86).

Table 1 *Deauville and Hopkins criteria*

	Deauville criteria (80, 81)	Hopkins criteria (83)	Conclusion
1	No ^{18}F-FDG uptake	^{18}F-FDG uptake at the primary site and nodes less than the internal jugular vein	Complete metabolic response
2	Slight ^{18}F-FDG uptake, but less than the uptake in the mediastinal blood pool	Focal ^{18}F-FDG uptake at the primary site and nodes greater than the internal jugular vein but less than liver.	Likely complete metabolic response
3	^{18}F-FDG uptake above mediastinal bloodpool, but below or equal to ^{18}F-FDG uptake in the liver	Diffuse ^{18}F-FDG uptake at the primary site or nodes is greater than the internal jugular vein or liver.	Likely postradiation inflammation
4	^{18}F-FDG uptake slightly to moderately higher than liver	Focal ^{18}F-FDG uptake at the primary site or nodes greater than liver.	Likely residual tumor
5	Markedly increased ^{18}F-FDG uptake or any new lesion (on response evaluation)	Focal and intense ^{18}F-FDG uptake at the primary site or nodes.	Residual tumor

The advent of PET-MRI offers opportunities for combining functional parameters. We compared functional MRI and PET-CT in **Chapter 2.2, 2.3, 4.1, 4.2, 4.3, 4.4 and 5.2.** In **Chapter 2.2** we found that DCE and PET provide complementary information. In **Chapter 2.3** we found that most data on intratreatment analysis was available for DWI and ^{18}F-FDG-PET-CT. Both modalities had the most predictive value when performed within the first 3 weeks of treatment. Combined PET-MRI may result in a 'one stop shop' interim analysis. In **Chapter 4.1** we found MRI including DWIBS to be more sensitive than ^{18}F-FDG-PET-CT in the detection and characterization of distant lesions at the expense of more clinically

indeterminate lesions. Combined PET-MRI may result in even better lesion characterization with less indeterminate lesions. In **Chapter 4.2** we did not find synergy between DWI and ^{18}F-FDG-PET-CT for the detection of unknown primary HNSCC. This was probably due to a high sensitivity of ^{18}F-FDG-PET-CT, leaving little room for improvement. In **Chapter 4.3** we found that DWI analysis provided complementary information in patients with residual ^{18}F-FDG avid lesions after (chemo)radiotherapy improving specificity from 47% to 88% with a limited decrease in sensitivity. In **Chapter 4.4** we found a sequential analysis of DWI and ^{18}F-FDG-PET-CT to result in the highest diagnostic accuracy for detecting residual disease after (chemo)radiotherapy. In **Chapter 5.2** we found pretreatment primary tumor ADC_{max} and SUV_{max} to be independent predictors of treatment outcome. We therefore conclude that there are various potential applications for PET-MRI in head and neck cancer.

With improved image registration lesions can be characterized on a pixel-by-pixel basis. This can result in boosting of radiotherapy in regions which are radiation-resistant, or selecting patients who are unlikely to respond to chemoradiotherapy and are therefore more favorable candidates for primary surgical treatment or by selecting patients who have a favorable prognosis and are candidates for de-escalation of treatment. Currently de-escalation of treatment is being assessed in studies on HPV-positive patients (87). With imaging even better treatment selection may be possible. Especially the use of serial imaging has shown great promise. However, so far imaging research has been focusing on response prediction without actually using results for clinical decision making. The next step should be the advent of clinical trials incorporating imaging results in treatment decisions.

The advent of more and more imaging biomarkers will potentially result in better lesion characterization. An important challenge will be to find the most optimal combination of parameters for various applications (e.g. diagnosis, response prediction and detection of recurrent or residual disease). With the use of 'radiomics' it is attempted to find this most optimal combination. With radiomics large amounts of imaging features are extracted and mined using artificial intelligence (AI) (88-92). It is hypothesized that these imaging features will harbor information on the tumor phenotype and gene expression patterns and will result in advanced lesion characterization (92). Machine learning is often used in order to quickly identify and combine the most promising parameters derived from radiomics. With machine learning and deep learning a computer can learn concepts by itself and gain experience without being explicitly programmed. In general, a training cohort is used to find the most optimal combination of parameters which is then applied in a validation cohort. It is hypothesized that radiomics and machine learning may take over tasks of radiologist because these algorithms can train themselves. It should be noted that with machine learning predictors can be identified (91). There is however an essential difference between a correlation and a causal interference (90, 93). Another challenge will be to provide enough data for both the training cohort and the validation cohort to ensure that machine learning will provide more robust results (91). Another shortcoming of machine learning is that because the model trains itself, there is a risk of creating a so-called black-box where nobody knows how the results were obtained and any errors in a model will go unnoticed (93).

In the future the use of radiomics and machine learning may be used to extract more data from imaging. This should result in more accurate quantitative characterization of lesions. To analyze this data more sophisticated higher order statistical analyses are necessary which makes data more difficult to interpret. Before clinical application it is necessary that data obtained with radiomics can be used to make decisions for an individual patient. We expect that the radiologist will still be necessary to interpret the data provided by radiomics and to translate the information to the individual patient (90, 93). We therefore believe that the radiologist who can use AI will replace the radiologist who cannot.

REFERENCES

1. Bisdas S, Seitz O, Middendorp M, et al. An exploratory pilot study into the association between microcirculatory parameters derived by MRI-based pharmacokinetic analysis and glucose utilization estimated by PET-CT imaging in head and neck cancer. Eur Radiol. 2010;20(10):2358-66.
2. Jansen JF, Carlson DL, Lu Y, et al. Correlation of a priori DCE-MRI and (1)H-MRS data with molecular markers in neck nodal metastases: Initial analysis. Oral Oncol. 2012;48(8):717-22.
3. Jansen JF, Schoder H, Lee NY, et al. Noninvasive assessment of tumor microenvironment using dynamic contrast-enhanced magnetic resonance imaging and 18F-fluoromisonidazole positron emission tomography imaging in neck nodal metastases. Int J Radiat Oncol Biol Phys. 2010;77(5):1403-10.
4. Jansen JFA, Schöder H, Lee NY, et al. Tumor metabolism and perfusion in head and neck squamous cell carcinoma: Pretreatment multimodality imaging with 1H magnetic resonance spectroscopy, dynamic contrast-enhanced MRI, and [18F]FDG-PET. International Journal of Radiation Oncology Biology Physics. 2012;82(1):299-307.
5. Peinado H, Cano A. A hypoxic twist in metastasis. Nat Cell Biol. 2008;10(3):253-4.
6. Shukla-Dave A, Lee NY, Jansen JF, et al. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases. Int J Radiat Oncol Biol Phys. 2012;82(5):1837-44.
7. Kim S, Loevner LA, Quon H, et al. Prediction of response to chemoradiation therapy in squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. AJNR Am J Neuroradiol. 2010;31(2):262-8.
8. Chawla S, Kim S, Dougherty L, 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;200(1):35-43.
9. Ng SH, Lin CY, Chan SC, et al. Dynamic contrast-enhanced MR imaging predicts local control in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiotherapy. PLoS One. 2013;8(8):e72230.
10. Vandecaveye V, Dirix P, De Keyzer F, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. Eur Radiol. 2010;20(7):1703-14.
11. Srinivasan A, Dvorak R, Perni K, Rohrer S, Mukherji SK. Differentiation of benign and malignant pathology in the head and neck using 3T apparent diffusion coefficient values: early experience. AJNR Am J Neuroradiol. 2008;29(1):40-4.
12. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
13. Padhani AR, Koh DM, Collins DJ. Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology. 2011;261(3):700-18.
14. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46.
15. Sakamoto J, Imaizumi A, Sasaki Y, et al. Comparison of accuracy of intravoxel incoherent motion and apparent diffusion coefficient techniques for predicting malignancy of head and neck tumors using half-Fourier single-shot turbo spin-echo diffusion-weighted imaging. Magn Reson Imaging. 2014;32(7):860-6.
16. Sasaki M, Sumi M, Eida S, Katayama I, Hotokezaka Y, Nakamura T. Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors. PLoS One. 2014;9(11):e112866.
17. Sumi M, Nakamura T. Head and neck tumors: assessment of perfusion-related parameters and diffusion coefficients based on the intravoxel incoherent motion model. AJNR Am J Neuroradiol. 2013;34(2):410-6.
18. Sumi M, Van Cauteren M, Sumi T, Obara M, Ichikawa Y, Nakamura T. Salivary gland tumors: use of intravoxel

incoherent motion MR imaging for assessment of diffusion and perfusion for the differentiation of benign from malignant tumors. Radiology. 2012;263(3):770-7.

19. Yu XP, J. Hou, et al. Intravoxel Incoherent Motion Diffusion Weighted Magnetic Resonance Imaging for Differentiation Between Nasopharyngeal Carcinoma and Lymphoma at the Primary Site. J Comput Assist Tomogr. 2016.

20. Xiao Y, Pan J, Chen Y, Chen Y, He Z, Zheng X. Intravoxel Incoherent Motion-Magnetic Resonance Imaging as an Early Predictor of Treatment Response to Neoadjuvant Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma. Medicine (Baltimore). 2015;94(24):e973.

21. Xiao-ping Y, Jing H, Fei-ping L, et al. Intravoxel incoherent motion MRI for predicting early response to induction chemotherapy and chemoradiotherapy in patients with nasopharyngeal carcinoma. J Magn Reson Imaging. 2016;43(5):1179-90.

22. Guo W, Luo D, Lin M, et al. Pretreatment Intra-Voxel Incoherent Motion Diffusion-Weighted Imaging (IVIM-DWI) in Predicting Induction Chemotherapy Response in Locally Advanced Hypopharyngeal Carcinoma. Medicine (Baltimore). 2016;95(10):e3039.

23. Hauser T, Essig M, Jensen A, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. Neuroradiology. 2013;55(5):527-36.

24. Hauser T, Essig M, Jensen A, et al. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: Evaluation of lymph node metastasis. Eur J Radiol. 2014;83(5):783-7.

25. de Oliveira MV, Pereira Gomes EP, Pereira CS, et al. Prognostic value of microvessel density and p53 expression on the locoregional metastasis and survival of the patients with head and neck squamous cell carcinoma. Appl Immunohistochem Mol Morphol. 2013;21(5):444-51.

26. Gasparini G, Weidner N, Maluta S, et al. Intratumoral microvessel density and L53 protein: Correlation with metastasis in head-and-neck squamous-cell carcinoma. Int J Cancer. 1993;55(5):739-44.

27. Galban CJ, Mukherji SK, Chenevert TL, et al. A feasibility study of parametric response map analysis of diffusion-weighted magnetic resonance imaging scans of head and neck cancer patients for providing early detection of therapeutic efficacy. Transl Oncol. 2009;2(3):184-90.

28. Paudyal R, Oh JH, Riaz N, 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.

29. Tyagi N, Riaz N, Hunt M, et al. Weekly response assessment of involved lymph nodes to radiotherapy using diffusion-weighted MRI in oropharynx squamous cell carcinoma. Med Phys. 2016;43(1):137.

30. Martins EB, Chojniak R, Kowalski LP, Nicolau UR, Lima EN, Bitencourt AG. Diffusion-Weighted MRI in the Assessment of Early Treatment Response in Patients with Squamous-Cell Carcinoma of the Head and Neck: Comparison with Morphological and PET/CT Findings. PLoS One. 2015;10(11):e0140009.

31. King AD, Mo FKF, Yu KH, et al. Squamous cell carcinoma of the head and neck: Diffusion-weighted MR imaging for prediction and monitoring of treatment response. European radiology. 2010;20(9):2213-20.

32. Brun E, Kjellen E, Tennvall J, et al. FDG PET studies during treatment: prediction of therapy outcome in head and neck squamous cell carcinoma. Head Neck. 2002;24(2):127-35.

33. Brun E, Ohlsson T, Erlandsson K, et al. Early prediction of treatment outcome in head and neck cancer with 2- 18FDG PET. Acta Oncol. 1997;36(7):741-7.

34. Sheikhbahaei S, Taghipour M, Ahmad R, et al. Diagnostic Accuracy of Follow-Up FDG PET or PET/CT in Patients With Head and Neck Cancer After Definitive Treatment: A Systematic Review and Meta-Analysis. AJR Am J Roentgenol. 2015;205(3):629-39.

35. Castaldi P, Rufini V, Bussu F, et al. Can “early” and “late”18F-FDG PET-CT be used as prognostic factors for the

clinical outcome of patients with locally advanced head and neck cancer treated with radio-chemotherapy? Radiother Oncol. 2012;103(1):63-8.

36. Truong MT, Saito N, Ozonoff A, et al. Prediction of locoregional control in head and neck squamous cell carcinoma with serial CT perfusion during radiotherapy. AJNR Am J Neuroradiol. 2011;32(7):1195-201.

37. Abramyuk A, Hietschold V, Appold S, von Kummer R, Abolmaali N. Radiochemotherapy-induced changes of tumour vascularity and blood supply estimated by dynamic contrast-enhanced CT and fractal analysis in malignant head and neck tumours. Br J Radiol. 2015;88(1045):20140412.

38. Dirix P, Vandecaveye V, De Keyzer F, Stroobants S, Hermans R, Nuyts S. 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. 2009;50(7):1020-7.

39. King AD, Chow KK, Yu KH, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology. 2013;266(2):531-8.

40. Lambrecht M, Van Calster B, Vandecaveye V, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiother Oncol. 2014;110(3):429-34.

41. Matoba M, Tuji H, Shimode Y, et al. 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. 2014;35(2):379-85.

42. Schouten CS, de Bree R, van der Putten L, 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;4(4):239-50.

43. Baer AH, Hoff BA, Srinivasan A, Galban CJ, Mukherji SK. Feasibility analysis of the parametric response map as an early predictor of treatment efficacy in head and neck cancer. AJNR Am J Neuroradiol. 2015;36(4):757-62.

44. Min M, Lin P, Lee MT, et al. Prognostic role of metabolic parameters of (18)F-FDG PET-CT scan performed during radiation therapy in locally advanced head and neck squamous cell carcinoma. Eur J Nucl Med Mol Imaging. 2015;42(13):1984-94.

45. Min M, Lin P, Lee M, et al. 18F-FDG PET-CT performed before and during radiation therapy of head and neck squamous cell carcinoma: Are they independent or complementary to each other? J Med Imaging Radiat Oncol. 2016;60(3):433-40.

46. Lin P, Min M, Lee M, et al. Nodal parameters of FDG PET/CT performed during radiotherapy for locally advanced mucosal primary head and neck squamous cell carcinoma can predict treatment outcomes: SUVmean and response rate are useful imaging biomarkers. European journal of nuclear medicine and molecular imaging. 2016;44(5):801-11.

47. Pollom EL, Song J, Durkee BY, et al. Prognostic value of midtreatment FDG-PET in oropharyngeal cancer. Head & neck. 2016;38(10):1472-8.

48. Farrag A, Ceulemans G, Voordeckers M, Everaert H, Storme G. Can 18F-FDG-PET response during radiotherapy be used as a predictive factor for the outcome of head and neck cancer patients? Nucl Med Commun. 2010;31(6):495-501.

49. Hentschel M, Appold S, Schreiber A, et al. 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. 2011;38(7):1203-11.

50. Wang J, Takashima S, Takayama F, et al. Head and neck lesions: characterization with diffusion-weighted echo-planar MR imaging. Radiology. 2001;220(3):621-30.

51. Bhatia KS, King AD, Yeung DK, et al. Can diffusion-weighted imaging distinguish between normal and squamous cell carcinoma of the palatine tonsil? Br J Radiol. 2010;83(993):753-8.

52. Abdel Razek AAK, Gaballa G, Elhawarey G, Megahed AS, Hafez M, Nada N. Characterization of pediatric head and neck masses with diffusion-weighted MR imaging. Eur Radiol. 2009;19(1):201-8.

53. de Bree R, Deurloo EE, Snow GB, Leemans CR. Screening for distant metastases in patients with head and neck cancer. Laryngoscope. 2000;110(3 Pt 1):397-401.

54. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(2):122-37.

55. Shoushtari A, Saylor D, Kerr KL, et al. Outcomes of patients with head-and-neck cancer of unknown primary origin treated with intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys. 2011;81(3):e83-91.

56. Rusthoven KE, Koshy M, Paulino AC. The role of fluorodeoxyglucose positron emission tomography in cervical lymph node metastases from an unknown primary tumor. Cancer. 2004;101(11):2641-9.

57. Lee HJ, Kim J, Woo HY, Kang WJ, Lee JH, Koh YW. 18F-FDG PET-CT as a supplement to CT/MRI for detection of nodal metastasis in hypopharyngeal SCC with palpably negative neck. Laryngoscope. 2015;125(7):1607-12.

58. Choi YJ, Lee JH, Kim HO, et al. Histogram Analysis of Apparent Diffusion Coefficients for Occult Tonsil Cancer in Patients with Cervical Nodal Metastasis from an Unknown Primary Site at Presentation. Radiology. 2016;278(1):146-55.

59. Lee JR, Kim JS, Roh JL, et al. Detection of occult primary tumors in patients with cervical metastases of unknown primary tumors: comparison of (18)F FDG PET/CT with contrast-enhanced CT or CT/MR imaging-prospective study. Radiology. 2015;274(3):764-71.

60. Regelink G, Brouwer J, de Bree R, et al. Detection of unknown primary tumours and distant metastases in patients with cervical metastases: value of FDG-PET versus conventional modalities. Eur J Nucl Med Mol Imaging. 2002;29(8):1024-30.

61. Johansen J, Buus S, Loft A, et al. Prospective study of 18FDG-PET in the detection and management of patients with lymph node metastases to the neck from an unknown primary tumor. Results from the DAHANCA-13 study. Head Neck. 2008;30(4):471-8.

62. Ryan WR, Fee WE, Jr., Le QT, Pinto HA. Positron-emission tomography for surveillance of head and neck cancer. Laryngoscope. 2005;115(4):645-50.

63. Malone JP, Gerberi MA, Vasireddy S, et al. Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography. Arch Otolaryngol Head Neck Surg. 2009;135(11):1119-25.

64. Nakajo M, Nakajo M, Kajiya Y, et al. FDG PET/CT and Diffusion-Weighted Imaging of Head and Neck Squamous Cell Carcinoma Comparison of Prognostic Significance Between Primary Tumor Standardized Uptake Value and Apparent Diffusion Coefficient. Clin Nucl Med. 2012;37(5):475-80.

65. Ho AS, Tsao GJ, Chen FW, et al. Impact of positron emission tomography/computed tomography surveillance at 12 and 24 months for detecting head and neck cancer recurrence. Cancer. 2013;119(7):1349-56.

66. van der Putten L, Hoekstra OS, de Bree R, et al. 2-Deoxy-2[F-18]FDG-PET for detection of recurrent laryngeal carcinoma after radiotherapy: interobserver variability in reporting. Mol Imaging Biol. 2008;10(5):294-303.

67. de Bree R, Putten Lvd, Brouwer J, Castelijns JA, Hoekstra OS, René Leemans C. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. Oral Oncol. 2009;45(4-5):386-93.

68. 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;23(2):562-9.

69. Pipe JG, Farthing VG, Forbes KP. Multishot diffusion-weighted FSE using PROPELLER MRI. Magn Reson Med. 2002;47(1):42-52.

70. Paidpally V, Chirindel A, Lam S, Agrawal N, Quon H, Subramaniam RM. FDG-PET/CT imaging biomarkers in head and neck squamous cell carcinoma. Imaging Med. 2012;4(6):633-47.

71. Jung YS, Najy AJ, Huang W, et al. HPV-associated differential regulation of tumor metabolism in oropharyngeal head and neck cancer. Oncotarget. 2017;8(31):51530-41.

72. Chen SW, Hsieh TC, Yen KY, Liang JA, Kao CH. Pretreatment (18)F-FDG PET/CT in whole-body total lesion glycolysis to predict survival in patients with pharyngeal cancer treated with definitive radiotherapy. Clin Nucl Med. 2014;39(5):e296-300.

73. Foo SS, Abbott DF, Lawrentschuk N, Scott AM. Functional imaging of intratumoral hypoxia. Mol Imaging Biol. 2004;6(5):291-305.

74. Kim S, Loevner L, Quon H, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clinical Cancer Research. 2009;15(3):986-94.

75. King AD, Mo FK, Yu KH, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. European radiology. 2010;20(9):2213-20.

76. Vandecaveye V, De Keyzer F, Dirix P, Lambrecht M, Nuyts S, Hermans R. Applications of diffusion-weighted magnetic resonance imaging in head and neck squamous cell carcinoma. Neuroradiology. 2010;52(9):773-84.

77. Heye T, Davenport MS, Horvath JJ, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part I. Perfusion characteristics in the female pelvis by using multiple computer-aided diagnosis perfusion analysis solutions. Radiology. 2013;266(3):801-11.

78. Heye T, Merkle EM, Reiner CS, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology. 2013;266(3):812-21.

79. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-54.

80. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. Leuk Lymphoma. 2009;50(8):1257-60.

81. Meignan M, Gallamini A, Haioun C, Polliack A. Report on the Second International Workshop on interim positron emission tomography in lymphoma held in Menton, France, 8-9 April 2010. Leuk Lymphoma. 2010;51(12):2171-80.

82. Sjovall J, Bitzen U, Kjellen E, Nilsson P, Wahlberg P, Brun E. Qualitative interpretation of PET scans using a Likert scale to assess neck node response to radiotherapy in head and neck cancer. Eur J Nucl Med Mol Imaging. 2015.

83. Marcus C, Ciarallo A, Tahari AK, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes. J Nucl Med. 2014;55(9):1411-6.

84. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

85. Wray R, Sheikhbahaei S, Marcus C, et al. Therapy Response Assessment and Patient Outcomes in Head and Neck Squamous Cell Carcinoma: FDG PET Hopkins Criteria Versus Residual Neck Node Size and Morphologic Features. AJR Am J Roentgenol. 2016;207(3):641-7.

86. Van den Wyngaert T, Helsen N, Carp L, et al. Fluorodeoxyglucose-Positron Emission Tomography/Computed Tomography After Concurrent Chemoradiotherapy in Locally Advanced Head-and-Neck Squamous Cell Cancer: The ECLYPS Study. J Clin Oncol. 2017;35(30):3458-64.

87. Determination of Cetuximab Versus Cisplatin Early and Late Toxicity Events in HPV+ OPSCC. https://ClinicalTrials.gov/show/NCT01874171.

88. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441-6.

89. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. Front Oncol. 2015;5:272.

90. Quer G, Muse ED, Nikzad N, Topol EJ, Steinhubl SR. Augmenting diagnostic vision with AI. The Lancet. 2017;390(10091):221.

91. Obermeyer Z, Emanuel EJ. Predicting the Future- Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-9.

92. Caudell JJ, Torres-Roca JF, Gillies RJ, et al. The future of personalised radiotherapy for head and neck cancer. The Lancet Oncology. 2017;18(5):e266-e73.

93. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318(6):517-8.

6

CHAPTER 7

SUMMARY IN DUTCH/ NEDERLANDSE SAMENVATTING

SUMMARY IN DUTCH/ NEDERLANDSE SAMENVATTING

Hoofd-halskanker
Het plaveiselcelcarcinoom van het hoofd-halsgebied is het meest voorkomende kankertype in het hoofd-halsgebied. Het hoofd-halsgebied omvat het gebied tussen schedelbasis en de sleutelbeenderen. In het hoofd-halsgebied bevinden zich onder andere de neus met bijholten, neuskeelholte (nasopharynx), de mondholte, de mondkeelholte (orofarynx), de onderste keelholte (hypofarynx), het strottenhoofd (larynx) en de speekselklieren. Ongeveer 3% van alle patiënten met kanker heeft een tumor in het hoofd-halsgebied. Van deze tumoren bestaat 90% uit plaveiselcelcarcinoom. In Nederland worden elk jaar ongeveer 3000 nieuwe patiënten met hoofd-halskanker gediagnosticeerd.

De ziekte komt ongeveer drie keer vaker voor bij mannen dan bij vrouwen. De gemiddelde leeftijd bij diagnose is 60 jaar. De belangrijkste risicofactoren voor het ontwikkelen van een plaveiselcelcarcinoom van het hoofd-halsgebied zijn roken en alcoholgebruik. Daarnaast kunnen virussen een rol spelen bij het ontstaan van hoofd-halskanker.

Tumoren van de neuskeelholte zijn sterk geassocieerd met het Epstein-Barr-virus, vooral bekend als verwekker van de ziekte van Pfeiffer. Deze nasopharynxtumoren komen relatief vaker voor in Azië.

Het humaan papillomavirus (HPV), en dan met name het subtype HPV-16, is geassocieerd met plaveiselcelcarcinomen van de keelamandelen (tonsillen) en de tongbasis. Het HPV-virus is verder bekend als veroorzaker van wratten en baarmoederhalskanker. Patiënten met een HPV-positieve tumor zijn over het algemeen jonger en hebben meestal geen andere risicofactoren voor hoofd-halskanker zoals alcoholgebruik en roken. Er zijn aanwijzingen dat patiënten met een HPV-positieve tumor beter reageren op chemotherapie en radiotherapie en daarmee ook een betere prognose hebben.

Behandeling
De behandeling van hoofd-halskanker hangt af van de lokalisatie van de tumor en de uitgebreidheid van de tumor: hoe groot de tumor zelf is, de ingroei in omliggende structuren, en of er al uitzaaiingen zijn naar lymfklieren in de omgeving. Als de ziekte is uitgezaaid buiten het hoofd-halsgebied, meestal naar de longen, lever en skelet, dan kan de ziekte niet meer genezen worden.

Voor de minst uitgebreide tumoren, die zich in een vroeg stadium bevinden, zal de behandeling meestal bestaan uit een operatie of bestraling (radiotherapie). Als de tumor niet in een keer verwijderd kan worden, of als deze om belangrijke structuren zoals de halsslagader zit. Of als opereren zou lijden tot een verlies van functies zoals eten, spreken en slikken. Dan bestaat de behandeling meestal uit bestraling. Bij grote of uitgebreide tumoren wordt hier nog chemotherapie aan toegevoegd. Bij grote of uitgebreide tumoren, die zich in een vergevorderd stadium bevinden, wordt aan chirurgie nog radiotherapie met of zonder chemotherapie toegevoegd, of wordt aan de primaire radiotherapie nog chemotherapie toegevoegd en wordt chirurgie achter de hand gehouden voor residu of recidief tumor.

Afhankelijk van de tumorlokalisatie en -stadium kan met primaire chemoradiatie een vergelijkbare kans op overleving en een beter behoud van functies (eten, spreken en slikken) verkregen worden dan wanneer primair chirurgisch ingegrepen wordt.

Dit proefschrift beschrijft de rol van beeldvormend onderzoek bij hoofd-halskanker. Wij hebben onderzocht of het met nieuwe functionele beeldvormende technieken beter mogelijk is om vast te stellen of iemand hoofd-halskanker heeft. Hierbij hebben we met name functionele magnetische resonantie imaging (MRI) en positron emissie tomografie (PET) onderzocht. In het bijzonder door te kijken of er sprake is van uitzaaiingen op afstand en of we met deze technieken beter een onbekende primaire tumor kunnen detecteren. Verder hebben we gekeken of op basis van beeldvorming voorspeld kan worden welke patiënten goed zullen reageren op de therapie. Daarnaast hebben we onderzocht of de resultaten van deze nieuwe MRI technieken reproduceerbaar zijn.

Beeldvormend onderzoek
Beeldvormend onderzoek wordt bij hoofd-halskanker met name gebruikt voor het bepalen van de uitgebreidheid van de ziekte, de aanwezigheid van uitzaaiingen en voor het vervolgen van het oorspronkelijke tumorgebied na behandeling. Hierbij wordt onder andere gebruikt gemaakt van echografie, computer tomografie (CT), MRI en PET. Echografie en MRI hebben als voordeel dat er geen ioniserende straling nodig is om het onderzoek te kunnen verrichten. Een ander voordeel van MRI ten opzichte van CT is het hogere weke-delencontrast. Daarnaast is voor een groot deel van de MRI technieken geen intraveneus contrast nodig. Een nadeel van MRI is de relatief lange onderzoektijd waarin de patiënt helemaal stil moet liggen. Voor alle beeldvormende technieken geldt dat de bevindingen bij voorkeur worden bevestigd door histopathologisch onderzoek van een te nemen biopt.

Bovengenoemde conventionele modaliteiten (met uitzondering van PET) zijn vooral gericht op bepalen waar de tumor zit en in welke omliggende structuren de tumor groeit. Met functionele beeldvorming wordt gekeken naar hoe weefsel zich gedraagt. Bij PET-CT wordt hiervoor een radioactieve tracer gebruikt, meestal is dit een radioactieve suiker analoog: 18F-Fludeoxyglucose (^{18}F-FDG).

De functionele imaging technieken die in dit proefschrift aan de orde komen zijn: diffusie-gewogen MR imaging (DWI) en het hiervan afgeleide intravoxel incoherent motion (IVIM), dynamic contrast enhanced imaging (DCE) en PET.

Diffusie-gewogen MRI (DWI) is gebaseerd op de diffusiemogelijkheden van water in weefsel. In weefsel met veel cellen die dicht tegen elkaar aanliggen, zoals tumorweefsel, zal de diffusiemogelijkheid van water beperkt zijn. Echter in weefsel met veel vocht tussen de cellen, zoals bij ontstoken weefsel, zal water juist vrij kunnen diffunderen. Om een diffusie-gewogen beeld te verkrijgen is een bepaalde diffusie-gradiënt nodig. Deze sterkte wordt uitgedrukt in een b-waarde. Bij lage b-waarden wordt zowel een diffusie- als een perfusie-effect gemeten, terwijl bij hoge waarden de pure diffusie wordt gemeten. Middels de ADC kan het verval van diffusie over b-waarden in een getal worden uitgedrukt. Meestal worden hiervoor b-waarden van 0 s/mm^2 en van 1000 s/mm^2 gebruikt. De IVIM techniek is

ontwikkeld om deze twee eigenschappen van het diffusiesignaal beide te kunnen meten. Op deze manier wordt gepoogd om tumoren nog beter te kunnen karakteriseren. In **Hoofdstuk 2.2** hebben we middels een systematische review het potentieel van IVIM in kaart gebracht. We vonden sterke aanwijzingen dat het combineren van IVIM parameters zeer geschikt is voor betrouwbaar onderscheid tussen verschillende soorten tumoren zoals plaveiselcelcarcinomen, lymfomen en speekselkliertumoren. Daarnaast lijken IVIM parameters voorspellend te zijn voor de respons op behandeling. Wel dient te worden opgemerkt dat er nog veel verschillen zitten in de imaging protocollen en wijzen van analyseren. Dit zorgt ervoor dat IVIM op dit moment niet toepasbaar is in de kliniek. Hiervoor moeten eerst universele standaarden ontwikkeld worden.

In **Hoofdstuk 4.1** hebben we de mogelijkheden van DWI bij het detecteren van uitzaaiingen op afstand onderzocht. Genezing is niet meer mogelijk als uitzaaiingen worden gevonden. Dan zal de behandeling meer op comfort (palliatie) gericht zijn. Hiervoor hebben we 'whole body' MRI met DWIBS (diffusion-weighted imaging with background body signal suppression), een whole-body DWI techniek, vergeleken met PET-CT en een CT van de thorax bij patiënten met een hoog risico op uitzaaiingen op afstand. Hierbij vonden we dat MRI met DWI een gevoeligere techniek was voor het ontdekken van uitzaaiingen op afstand, maar dat dit wel ten koste ging van een hoger aantal fout-positieve bevindingen. Doordat we meerdere sequenties gebruikten in het protocol was het over het algemeen goed mogelijk om een laesie niet alleen te detecteren, maar ook direct te karakteriseren zonder erin te prikken. Een voordeel van de DWIBS techniek is dat laesies direct 'oplichten' waardoor de beelden relatief snel te beoordelen zijn.

In **Hoofdstuk 4.2** hebben we gekeken naar een andere patiëntengroep die risico loopt op overbehandeling, namelijk patiënten die zich presenteren met een onbekende primaire tumor. Deze patiënten presenteren zich over het algemeen met een zwelling in de hals op basis van een uitzaaiing in een lymfklier. Als de primaire tumor niet wordt gevonden, dan wordt het hele gebied bestraald waar de tumor zich zou kunnen bevinden. Dit kan met veel en ernstige bijwerkingen gepaard gaan. Het vinden van een primaire tumor maakt dus een meer gerichte therapie mogelijk. We hebben gekeken of het gebruik van DWI meerwaarde heeft ten opzichte PET-CT. We vonden een niet-significante trend voor een hogere sensitiviteit van PET-CT ten opzichte van DWI (93.8% vs 81.3%) met een gelijke specificiteit (73.3%). Het combineren van beide technieken gaf geen beter resultaat dan analyse van één van beide technieken apart. We concludeerden daarom dat zowel DWI als PET-CT zeer sensitieve technieken zijn. Aangezien met PET-CT meer ervaring is, is het aan te bevelen om deze modaliteit voor deze indicatie te gebruiken. Diffusie-gewogen MRI is echter wel een geschikt alternatief als PET-CT niet beschikbaar is. Bovendien is DWI goedkoper dan PET-CT. Het dient overigens te worden opgemerkt dat bij 15 van de 31 geïncludeerde patiënten uiteindelijk geen primaire tumor werd gevonden, ook niet na een onderzoek onder narcose met het nemen van biopten.

In **Hoofdstuk 4.3** en **4.4** hebben we onderzocht of DWI van aanvullende waarde is bij het opsporen van recidief tumor na behandeling met chemoradiatie. In **Hoofdstuk 4.3** hebben we gekeken naar de aanvullende waarde van DWI bij patiënten met FDG-opname op de

PET-CT drie maanden na chemoradiatie. We hebben voor deze patiëntenpopulatie gekozen omdat met name bij deze patiënten de PET-CT relatief vaak fout-positief is. Dit resulteert dan in onnodige biopten. Na chemoradiatie is de kans op complicaties zoals bloedingen en infecties na een biopt verhoogd. Bij patiënten zonder verhoogde FDG-opname is de negatief voorspellende waarde van PET-CT erg hoog. Met PET-CT vonden we een sensitiviteit van 100%, echter de specificiteit was slechts 47%. Door DWI en PET-CT te combineren steeg de specificiteit naar 88%, maar dit ging gepaard met een daling van de sensitiviteit naar 80%. Wanneer imaging met DWI en PET-CT gebruikt was om patiënten te selecteren voor een onderzoek onder narcose, dan had slechts 27% van de totale patiëntenpopulatie een onderzoek onder narcose hoeven te ondergaan. Dit was dan wel ten koste gegaan van het missen van één tumor residu. We concludeerden daarom dat het toevoegen van DWI leidde tot een substantiële verbetering van de specificiteit ten opzichte van alleen PET-CT, echter ten koste van een beperkte daling van sensitiviteit.

In **Hoofdstuk 4.4** hebben we gekeken naar de waarde van routinematig uitgevoerde DWI en PET-CT na (chemo)radiatie. We hebben primaire tumoren en klieruitzaaiingen apart geanalyseerd. De bevindingen waren vergelijkbaar. Bij zowel primaire tumoren als klieruitzaaiingen werden de beste resultaten behaald door eerst één techniek te analyseren en, indien deze een positief resultaat gaf (verdacht was voor restziekte), vervolgens de andere techniek te analyseren. Het maakte hierbij niet uit welke techniek als eerste werd uitgevoerd. Alleen indien beide positief waren werd de uiteindelijke verdenking op restziekte gesteld.

Een andere veelbelovende functionele MRI techniek is perfusie-MRI, hiervoor wordt het meest gebruik gemaakt van DCE. Bij DCE wordt intraveneus contrast toegediend. Na toediening wordt in korte tijd een groot aantal opnamen gemaakt waarmee informatie wordt verkregen over de doorbloeding of perfusie van weefsels. Hierbij kan zowel het uittreden van contrast van de bloedvaten naar de weefsels als andersom getalsmatig worden uitgedrukt aan de hand van wiskundige modellen. Kwaadaardige tumoren zijn in staat om zelf bloedvaten te ontwikkelen welke echter van lage kwaliteit zijn en gemakkelijk lekken. In **Hoofdstuk 2.1** hebben wij gepoogd een overzicht te creëren van de mogelijkheden van DCE bij hoofd-halskanker. In een systematische review hebben wij 33 artikelen geanalyseerd. Een belangrijke toepassing van DCE lijkt het opsporen van hypoxische tumoren welke slecht zullen reageren op chemoradiatie. In de toekomst kan het ertoe lijden dat aan deze patiënten een andere therapie kan worden aangeboden. Op dit moment geldt echter ook voor DCE dat er nog grote verschillen zitten in de imaging protocollen en gebruikte analytische modellen. Hierdoor is ook deze techniek momenteel nog niet geschikt voor routinematige toepassing in de kliniek.

Een belangrijk onderzoeksonderwerp van functionele imaging is de mogelijkheid om de respons op behandeling te voorspellen. Met name bij een behandeling met chemoradiatie is veel behoefte aan mogelijkheden om in een vroeg stadium te voorspellen welke patiënten niet reageren op de behandeling. Met name omdat deze therapie zeven weken duurt. Wanneer dit mogelijk is kan bij deze patiënten worden voorkomen dat zij onnodig bijwerkingen ervaren van de chemoradiatie en kan een meer optimale behandeling voor

de patiënt worden gekozen. In **Hoofdstuk 5** hebben wij gekeken naar de voorspellende waarde van DWI voor de start van behandeling met chemoradiatie bij hoofd-halskanker. In **Hoofdstuk 5.1** hebben gekeken wat de aanvullende waarde is van contrast-gewogen MRI op de interpretatie van DWI. De consensus is dat DWI de meeste waarde heeft, wanneer de ADC waarde wordt bepaald van het vitale deel van de tumor. Necrose kan een foutief verhoogd signaal geven op DWI-beelden. Met contrast-gewogen MRI kan necrose worden opgespoord: vitaal tumorweefsel kleurt aan, terwijl necrose dit niet doet. In dit onderzoek werden de DWI beelden eerst geanalyseerd zonder de contrast-gewogen MRI beelden. In een tweede sessie enkele weken later werden de DWI beelden geanalyseerd met de contrast-gewogen MRI beelden erbij. Met het gebruik van contrast-gewogen MRI werden laesies groter ingeschat en was de ADC waarde iets hoger. Dit zou erop kunnen duiden dat er gebieden van necrose werden meegenomen door het bepalen van ADC waarden met hulp van contrast-gewogen MRI. Echter de voorspellende waarde van DWI met of zonder contrast-gewogen MRI was vergelijkbaar. Derhalve concludeerden wij dat contrast-gewogen MRI een beperkte meerwaarde heeft in het verbeteren van de prognostische waarde van DWI.

In **Hoofdstuk 5.2** hebben we de voorspellende waarde van DWI en PET-CT voor de start van behandeling onderzocht. Hierbij hebben we gebruik gemaakt van een meer geavanceerde kwantitatieve analyse, de zogeheten histogramanalyse. Hiermee kan een laesie beter kwantitatief gekarakteriseerd worden dan wanneer slechts een gemiddelde kwantitatieve waarde wordt bepaald. Respons op behandeling kon het best voorspeld worden met de maximale ADC waarde en de maximale FDG-opname (uitgedrukt in een SUV_{max}) van de primaire tumor. Een hoge ADC_{max} is een teken dat een tumor gebieden bevat met veel bindweefsel en necrose die relatief ongevoelig zijn voor (chemo)radiatie. Een hoge SUV_{max} is juist een teken van een metabool actieve tumor die snel kan groeien en snel uit kan zaaien. Omdat zowel DWI als PET parameters onafhankelijke voorspellers waren van respons op behandeling zou dit een potentiële toepassing kunnen zijn van PET-MRI.

In **Hoofdstuk 2.3** geven wij een overzicht van alle beschikbare literatuur van functionele imaging tijdens de eerste vier weken van behandeling met chemoradiatie. We hebben voor de eerste vier weken gekozen omdat het dan nog mogelijk is om de therapie aan te passen. Chemoradiatie duurt immers zeven weken. Hierbij hebben we onder andere gekeken naar functionele CT, functionele MRI en PET-CT. De meeste data waren beschikbaar voor DWI en PET-CT. We hebben eerst naar de korte termijn gekeken, dat wil zeggen het voorspellen van de respons op behandeling na 3-6 maanden. Daarnaast hebben we ook gekeken naar de lange termijn, of de resultaten ook voorspellend zijn voor het voorspellen van de kans op locoregionale controle. Voor DWI werden de beste resultaten behaald door DWI zowel voor behandeling als 2-3 weken na de start van behandeling te verrichten om vervolgens de verandering in ADC waarde te berekenen. Een stijging van ADC waarde is suggestief voor een gunstige respons op behandeling. Na drie weken lijkt de ADC waarde een plateau te bereiken. Voor PET-CT is de timing van imaging wat kritischer. Initieel zal de FDG-opname dalen door verval van tumorcellen die veel suiker verbruiken. Om al deze cellen op te ruimen zal een ontstekingsreactie ontstaan waarbij de ontstekingscellen eveneens veel suiker verbruiken. Derhalve is PET-CT na behandeling pas drie maanden

na behandeling betrouwbaar voor het bepalen van de respons op behandeling. Uit onze review bleek echter dat daling van FDG-opname in de eerste 2-3 weken van de behandeling ook voorspellend is voor de respons op behandeling. We concludeerden daarom dat functionele imaging technieken zeker van waarde kunnen zijn bij het voorspellen van de respons op behandeling tijdens chemoradiatie. In de toekomst kan dit ertoe lijden dat een interim analyse kan gaan bepalen hoe het tweede deel van de therapie eruit ziet. Echter is er meer standaardisatie met universele afkapwaarden nodig voordat toepassing in de klinische praktijk mogelijk is.

Toekomstperspectief
Alvorens functionele imaging op grote schaal in de kliniek toe te kunnen passen is het nodig om de validiteit en reproduceerbaarheid te bepalen. In **Hoofdstuk 3.1** hebben wij de reproduceerbaarheid van DWI bij zeven gezonde vrijwilligers bepaald door ze op verschillende apparaten en op verschillende tijdstippen te scannen. Hierbij vonden we een significant verschil tussen beelden gemaakt op verschillende apparaten, terwijl het verschil in ADC waarden tussen verschillende tijdstippen, maar op hetzelfde apparaat verkregen, beperkt was. We concludeerden derhalve dat patiënten die een MRI met DWI krijgen bij voorkeur elke keer op hetzelfde apparaat gescand worden. In het ruggenmerg bleken de diffusie waarden het meest stabiel te zijn. Daarom lijkt het ruggenmerg het meest ideale weefsel om te gebruiken als referentieweefsel.

Met de komst van de PET-MRI scanner is het sinds kort mogelijk om beide technieken te combineren. Op basis van dit proefschrift zien wij zeker meerwaarde voor de combinatie van PET en MRI boven de nu op grote schaal beschikbare PET-CT. In dit proefschrift zijn PET en MRI samen onderzocht in **Hoofdstuk 2.2, 2.3, 4.1, 4.2, 4.3, 4.4** en **5.2**. In **Hoofdstuk 2.2** vonden we dat PET en DCE onafhankelijke voorspellende informatie opleverden. In **Hoofdstuk 4.1** vonden we dat whole-body MRI gevoeliger was, maar ook meer niet te determineren laesies opleverde dan met PET-CT. Met de combinatie van PET en MRI wordt het wellicht beter om laesies middels een onderzoek te vinden én karakteriseren. In **Hoofdstuk 4.2** vonden we geen meerwaarde voor het toevoegen van DWI aan PET-CT bij de detectie van een onbekende primaire tumor. Waarschijnlijk doordat de resultaten voor PET-CT al dusdanig goed waren dat er weinig ruimte was voor verbetering. In **Hoofdstuk 4.3** vonden we dat DWI aanvullende waarde had bij patiënten met residuale FDG-opname na (chemo)radiatie. De specificiteit verbeterde van 47% naar 88% met een beperkte daling van sensitiviteit. In **Hoofdstuk 4.4** vonden we de grootste diagnostische waarde bij een sequentiële aanpak van DWI en PET-CT voor het detecteren van residu na (chemo)radiatie. Het maakte hierbij niet uit welke van beide modaliteiten als eerste werd verricht. In **Hoofdstuk 5.2** vonden we dat ADC_{max} en SUV_{max} voor start van behandeling onafhankelijke predictoren waren voor behandeluitkomst. Derhalve concluderen wij dat er meerdere potentiële toepassingen zijn voor PET-MRI bij patiënten met hoofd-halskanker.

Ook voor de andere functionele technieken is meer onderzoek nodig naar de betrouwbaarheid van de bevindingen. Daarnaast moet voor elke techniek afzonderlijk het meest optimale tijdstip worden vastgesteld om ingezet te worden. Zo zullen DWI en DCE parameters eerder beginnen te veranderen tijdens behandeling dan andere parameters

als tumorgrootte en activiteit op de PET-scan. Verder moet de meest optimale combinatie van functionele imaging worden bepaald in verder onderzoek. Het combineren van een grote hoeveelheid imaging parameters om hier vervolgens de meest optimale combinatie uit te destilleren wordt 'radiomics' genoemd, een term die is afgeleid van de genetica. Om deze meest optimale combinatie te vinden wordt in toenemende mate gebruik gemaakt van kunstmatige intelligentie (KI), het zogeheten 'machine learning' of 'deep learning'. Hierbij kan een computer zichzelf algoritmes aanleren voor het interpreteren van beelden zonder dat deze hier expliciet voor geprogrammeerd hoeft te worden.

Er wordt gedacht dat de radioloog (gedeeltelijk) overbodig zal worden door KI. Wij verwachten dat dit niet op korte termijn zal gebeuren. De computer kan immers alleen predictoren opsporen, maar is niet in staat tot het leggen van causale verbanden of het interpreteren van de context. Daarnaast zijn modellen gebaseerd op informatie van een grote database en wordt een behandeling toegepast bij een individuele patiënt. Wij verwachten dat de radioloog nodig zal blijven om causale verbanden te leggen en de vertaalslag te maken naar de individuele patiënt. Wel verwachten wij dat de radioloog die met KI om kan gaan de radioloog gaat vervangen die dat niet kan.

7

CHAPTER 8

DANKWOORD

Dat er één naam op de cover staat wekt ten onrechte de indruk dat promoveren een soloprestatie zou zijn. Het is een teamprestatie van formaat. Daarom wil ik graag de nodige mensen bedanken die direct of indirect hebben bijgedragen aan dit proefschrift. Om te beginnen ben ik dank verschuldigd aan alle medewerkers van de afdelingen Radiologie, Nucleaire Geneeskunde , KNO/hoofd-hals chirurgie en Radiotherapie.

Beste promotoren en co-promotor, professor J.A. Castelijns, professor R. de Bree en dr. P. de Graaf. Dankzij jullie is dit proefschrift van de grond gekomen en geworden tot wat het is.

Promotoren
Jonas Castelijns, dankzij uw verhaal over diffusie-MRI met de in selecte kringen beroemde groene, gele en rode smileys ben ik uiteindelijk via het Honours Programma binnengekomen op de afdeling Radiologie. Al vrij snel sprak u het vertrouwen in mij uit de te vragen wat mijn lange termijn plannen waren en of ik hier niet op wilde promoveren. Met het schrijven van dit proefschrift heb ik aangetoond dat dit vertrouwen niet voor niets was. U was altijd erg betrokken en stond altijd klaar als er iets geregeld moest worden, groot of klein. Dankzij uw inspanningen kon ik mijn verlengde wetenschappelijke stage nog verder verlengen en nog tegen betaling ook. Ik kan u zeggen dat dit heeft geleid tot wat scheve gezichten bij medestudenten. Er was ook ruimte voor overleg en ik voelde mij vrijwel direct een volwaardig lid van het team. Dit heeft geleid tot constructieve discussies met creatieve oplossingen. Toch heb ik nog wel een klein aandachtspuntje: Dat voortijds weggaan bij Ajax-PSV, dat kon echt niet.

Remco de Bree, heel erg bedankt voor uw immer kritische en scherpe blik. Uw inbreng werd door mij altijd bijzonder gewaardeerd omdat u altijd naar het hele plaatje keek. Daarbij verdient ook de snelheid waarmee u altijd weet te reageren een compliment. Soms kwam uw reactie nog eerder dan een ‘out of office’ melding als u niet in het ziekenhuis aanwezig was. Uw overstap naar het UMC Utrecht veranderde hier gelukkig niets aan.

Pim de Graaf, het was soms even zoeken naar een goede tijd om af te spreken, maar als we elkaar eenmaal gevonden hadden kon er snel zaken gedaan worden. Vooral je mamadagen waren hiervoor erg geschikt. Juist doordat je je breed inzet op zowel klinisch als wetenschappelijk gebied kwam je altijd met zeer bruikbare feedback.

Co-auteurs
Marcus de Jong, de MC, je hebt zoveel voor mijn wetenschappelijk carrière gedaan dat het moeilijk was om te bedenken waar ik je in dit dankwoord moest zetten. Met jouw hulp heeft mijn wetenschappelijke carrière een reuzensprong kunnen maken. Van het opzetten van artikelen, statistische analyse, snel en overzichtelijk databases aanleggen tot een straffe bak koffiezetten, muizen bestrijden en een plant in leven houden. Dankzij deze reviewkoning van de VU ben ik in korte tijd flink gegroeid als onderzoeker. Ik ben er dan ook supertrots op om jou als paranimf een officiële rol te kunnen geven bij mijn promotie. Al had je ook niet misstaan als opponent, copromotor of rector magnificus.

Roland Martens, mijn opvolger en de man die iedereen kent. Als we door het ziekenhuis liepen om patiënten van A naar B te brengen dat maakte het niet uit welke route we liepen, jij kwam overal wel een bekende tegen. Dus als we iemand nodig hadden om ons te helpen, dan wist jij die persoon te vinden. Zo zaten we opeens op het 3D lab om uiteindelijk te vinden wat we zochten aan het eind van de gang. Ik wens je heel veel succes met het voortzetten van onze zoektocht naar de grenzen van de mogelijkheden van functionele imaging.

Thomas Koopman, zo kom je elkaar nog eens tegen. Van de Rome-reis op de middelbare school komen we elkaar daarna opeens weer tegen in het ziekenhuis. Met jouw programmeer skills heb jij mijn onderzoek in een stroomversnelling kunnen brengen. 'Daar moet een app voor zijn' heb jij vertaald naar 'Daar moet een applicatie in matlab voor te maken zijn'. Het tempo waarmee jij die applicaties ontwikkelt en vervolgens toepast zorgde ervoor dat we veel meer konden doen met onze data dan wij oorspronkelijk voor mogelijk hielden.

Otto Hoekstra, heel erg bedankt voor de begeleiding van het nucleaire gedeelte van dit proefschrift. Uw tips en adviezen werden door mij zeer gewaardeerd. Met name omdat u altijd kritische het hele artikel doorkeek en u niet beperkte tot de nucleaire geneeskunde. Uw volledigheid heeft alle artikelen waar uw naam op te vinden is een kwaliteitsinjectie gegeven.

Redina Ljumanovic, bedankt voor jouw introductie op de afdeling Radiologie en veel plezier en geluk toegewenst op het altijd zonnige Curaçao!

Petra Pouwels, bedankt voor het delen van jouw fysische kennis van MRI en het zoeken naar oplossingen voor alle beren die onderweg opdoken.

Tim Marcus en Christina Lavini, ik wil jullie graag bedanken met jullie werk naar de toekomst van functionele imaging. Waar ik begonnen ben met 'eenvoudige' diffusie-MRI, heb ik dankzij jullie alvast een kijkje kunnen in de toekomst van imaging met IVIM en DCE imaging.

Vincent Vandecaveye en Frederik de Keyzer, bedankt dat we een keer bij jullie in de Leuvense keuken mee mochten kijken om te zien hoe het er bij jullie aan toe gaat.

Dirk Knol, bedankt voor uw statistische adviezen.

Indra Pieters, bedankt voor jouw inzet in het whole-body project waardoor ik uiteindelijk de kans heb gekregen om mijn onderzoek te presenteren op de ECR in Wenen.

Els, Emile, Daniela, Rinze, Viresh, Fieke, Patricia, Ben, Ronald, René, Meedi en Lieven bedankt voor jullie input in mijn artikelen.

Kamergenoten
Iris!Iris!Iris!Iris!, mijn eerste onderzoeksbuurvrouw, ster met FLAIR, sprintkanon, bedankt voor al jouw hulp om van mij een goede onderzoeker en een goede dokter te maken. Op een paar incidentjes met PEC na hadden we het altijd erg gezellig op tha west side.

Laurien, weer zo'n student, nee PhD student (!), die gelukkig bleef plakken om te promoveren. Limburgse gezelligheid in het westen. Als er iets door jou georganiseerd wordt, dan denk je werkelijk aan alles. Even hop-hop, dan bla-bla en het is af. Dankzij jou weet ik waar ik moet zijn als ik ergens lekker wil eten of borrelen. Doordat je even op en neer ging naar New York om je onderzoek net dat beetje extra te geven, weet ik nu ook alle hippe plekken in the big Apple.

Robin, die jongen die komt er wel. Een echte prijzenpakker. Bedankt voor de gezellige sfeer, het mogen lenen van je database skills en voor de culturele opvoeding met uitstapjes naar het concertgebouw. Follow the rules, and you will have success. Eigenlijk bedoel ik 'keep having success', want met artikelen in Radiology en Oncotarget ben je al een gearriveerde naam in de wereld van radiomics.

Alette, Hester, Sjirk, Firazia en Fenna, in wisselende samenstelling hebben we samen PK-1X100 bewoond, bedankt voor de koffiemomentjes, foute uren en de gezelligheid!

Bedankt voor de gezelligheid op de gang Martijn (ook bedankt voor de snoeppot), Gijs, Veronica, Quinten, Svenja, Tineke en Mike.

Medewerkers Radiologie en KNO
Professor Van Kuijk: Hoewel de zaken in mijn geval voornamelijk via Jonas liepen wil ik u bedanken voor de goede organisatie van de afdeling.

Kees, Hesdy en Dennis, bedankt voor jullie geduld als ik weer eens, echt per ongeluk (echt waar!), een wachtrij had opgeblazen door even een MRI scan of 100 tegelijk binnen te halen.

Sandra, Greet, Letty, Angelique, Regina en Amanda, bedankt voor jullie ondersteuning vanuit het secretariaat en het verwerken en regelen van alle aanvragen en declaratieverzoeken.

Charlotte, bedankt voor het samen scoren en de klinische input bij al die imaging.

Interne collega's
Peter en Jeanette, bedankt voor de mogelijkheid om klinische ervaring op te doen als ANIOS Hematologie in de Daniel den Hoed kliniek. Ik heb de werksfeer als bijzonder prettig ervaren. De flexibele opstelling in het mogelijk maken om congressen te bezoeken en om mijn proefschrift af te kunnen worden door mij bijzonder op prijs gesteld.

Yvonne en Silvia, bedankt voor het vertrouwen dat ik van jullie kreeg om tijdens een sollicitatiegesprek al een contract te mogen tekenen als ANIOS Interne Geneeskunde. Ik

hoop dat ik het vertrouwen heb waargemaakt. Hopelijk kan ik met de ervaringen die ik heb opgedaan in het Franciscus Gasthuis nu de stap maken naar AIOS Interne Geneeskunde.

Nick, Christian, Aileen en Annemiek, STAMCELLEN!!! What else? Ik heb het genot gehad om met jullie te mogen werken in de Daniel den Hoed. Gezellig muziekje erbij, en als de agenda's het toelieten gezellig samen een vorkje prikken. Hopelijk kom ik jullie, als vertegenwoordigers van diverse specialismen of directe collega, nog vaak tegen.

Franny, Karlijn, Rachel, Vinnie, Rozanne, Gijs, Adem, Anja, Tim, Marijke, Nick, Roos en alle andere assistenten van het Franciscus Gasthuis waar ik mee heb mogen samen werken. Bedankt voor de fijne sfeer, de lekkere koffie en jullie steun. In jullie omgeving heb ik mij verder kunnen verbreden en ben ik een meer complete arts geworden.

Vrienden
Diallo, wieler- (en daardoor wijn-)kenner, beklimmer van hors categorie bergen, laminaatlegger, Vitesse fan. In het ziekenhuis moet ik het zelf uitzoeken, maar daarbuiten sta je altijd voor me klaar waarvoor bijzonder veel dank.

Floris, Chantal, bedankt voor de vele leuke momenten in Amsterdam en Egmond. Gezellig hardlopen en fietsen in Egmond, Schoorl en co-cycling. Bedankt dat jullie deur altijd open staat en ik altijd welkom ben.

Edwin, Casper. Van dokters is spe, naar echte artsen. Bedankt voor de vele gezellige momenten, etentjes en spelletjes avonden.

Papa en mama, heel erg bedankt voor alles wat jullie voor mij gedaan hebben en nog steeds doen. Dankzij jullie sta ik waar ik nu sta. Jullie hebben me altijd gesteund en ervoor gezorgd dat ik eruit heb kunnen halen wat erin zit. Zonder schulden afgestudeerd en altijd welkom voor een (soms zeer kort van tevoren aangekondigd) bezoek. Als ik alle specifieke situaties af zou gaan waar ik jullie voor moet bedanken, dan zou de helft van dit proefschrift uit het dankwoord bestaan. Samen met jullie en Kim zijn we team Noij en staan we er voor elkaar.

Kim, mijn grote zus die er altijd voor me is in voor- en tegenspoed. Als een echte verkenner liep jij een jaar voor mij uit. Dankzij jou wist ik telkens wat ik kon verwachten vanaf de basisschool tot en met de universiteit (ook al wist ik eerder dat ik Geneeskunde wilde studeren). Inmiddels ben je uitgewaaid naar Boston, waar ik zeker nog een keer die marathon kom lopen. Dat broer en zus prima samen kunnen werken hebben we laten zien in onze gezamenlijke publicatie. Daarom ben ik er ook supertrots op dat je mijn paranimf wilt zijn.

CHAPTER 9

BIBLIOGRAPHY

BIBLIOGRAPHY

Noij DP, Martens RM, Koopman T, Hoekstra OS, Comans EF, Zwezerijnen B, de Bree R, de Graaf P, Castelijns JA. Use of diffusion-weighted imaging and ^{18}F-FDG-PET/CT in the response assessment for (chemo)radiotherapy in head and neck squamous cell carcinoma. *Accepted for publiction in Clinical Oncology*

Martens RM, **Noij DP**, Koopman T, Zwezerijnen B, Heymans M, Hoekstra OS, Vergeer MR, de Bree R, Leemans CR, de Graaf P, Boellaard R, Castelijns JA. Predictive and prognostic value of quantitative ^{18}F-FDG-PET and diffusion-weighted imaging in head and neck squamous cell cancer. *Submitted*

Noij DP, Martens RM, de Graaf P, Zwezerijnen B, Koopman T, de Bree R, Hoekstra OS, Castelijns JA. Diagnostic value of diffusion-weighted imaging and ^{18}F-FDG-PET/CT for the detection of unknown primary head and neck cancer in patients presenting with cervical metastasis. Eur J Radiol. 2018;107:20-25

Martens RM, **Noij DP**, Ali M, Koopman T, Marcus JT, Vergeer MR, de Vet H, de Jong MC, Leemans CR, Hoekstra OS, de Bree R, de Graaf P, Boellaard R, Castelijns JA. Functional imaging early during (chemo)radiotherapy for response prediction in head and neck squamous cell carcinoma: a systematic review and meta-analysis. *Submitted*

Noij DP, Jagesar VA, de Graaf P, de Jong MC, Hoekstra OS, de Bree R, Castelijns JA. 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;124:296-305.

Noij DP, Martens RM, Marcus JT, de Bree R, Leemans CR, Castelijns JA, de Jong MC, de Graaf P. Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value. Oral Oncol. 2017;68:81-91.

Noij DP, Van der Linden PW. Spontaneous regression of hepatocellular carcinoma in a Caucasian male: a case report and review of the literature. Mol Clin Oncol. 2017;6:225-228.

Noij KS, **Noij DP**, Borgstein J. Healing of the tympanic membrane after surgical intervention for atelectasis of the middle ear in sixty-two adults. Clin Otolaryngol. 2017;42:450-454.

De Jong MC, van der Meer FJ, Göricke SL, Brisse HJ, Galluzzi P, Maeder P, Sirin S, De Francesco S, Sastre-Garau X, Metz KA, Cerase A, **Noij DP**, van der Valk P, Moll AC, Castelijns JA, de Graaf P; European Retinoblastoma Imaging Collaboration. Diagnostic Accuracy of Intraocular Tumor Size Measured with MR Imaging in the Prediction of Postlaminar Optic Nerve Invasion and Massive Choroidal Invasion of Retinoblastoma. Radiology. 2016;279:817-26.

Noij DP, de Jong MC, Mulders LG, Marcus JT, de Bree R, Lavini C, de Graaf P, Castelijns JA. Contrast-enhanced perfusion magnetic resonance imaging for head and neck squamous cell carcinoma: a systematic review. Oral oncol. 2014. 2015;51:124-138.

Noij DP, Pouwels PJ, Ljumanovic R, Knol DL, Doornaert P, de Bree R, Castelijns JA, de Graaf P. Predictive value of diffusion-weighted imaging without and with including contrast-enhanced magnetic resonance imaging in image analysis of head and neck squamous cell carcinoma. Eur J Radiol. 2015;84:108-16.

Kolff-Gart AS, Pouwels PJ, **Noij DP**, Ljumanovic R, Vandecaveye V, de Keyzer F, de Graaf P, Knol DL, Castelijns JA. 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;36:384-90. *Nominated for 2014 Lucien Levy Best Research Article Award*

Schouten CS, de Bree R, van der Putten L, **Noij DP**, Hoekstra OS, Comans EF, Witte BI, Doornaert PA, Leemans CR, Castelijns JA. Diffusion-weighted EPI- and HASTE-MRI and ^{18}F-FDG-PET-CT early during chemoradiotherapy in advanced head and neck cancer. Quant Imaging Med Surg. 2014;4:239-50.

Noij DP, Boerhout EJ, Pieters-van den Bos IC, Comans EF, Oprea-Lager D, Reinhard R, Hoekstra OS, de Bree R, de Graaf P, Castelijns JA. 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. 2014;83:1144-51.

de Jong MC, de Graaf P, **Noij DP**, Göricke S, Maeder P, Galluzzi P, Brisse HJ, Moll AC, Castelijns JA; European Retinoblastoma Imaging Collaboration (ERIC). Diagnostic performance of magnetic resonance imaging and computed tomography for advanced retinoblastoma: a systematic review and meta-analysis. Ophthalmology. 2014;121:1109-18.

CHAPTER 10

CURRICULUM VITAE

CURRICULUM VITAE

Daniel Peter Noij was born on April 8, 1991 in Hoofddorp, the Netherlands. After primary school he attended the Katholieke Scholengemeenschap Hoofddorp, where he obtained his VWO diploma, cum laude, in 2009. The next year Daniel went to the VU University Medical Center in Amsterdam to start his medical training. With a mother who had worked as a radiographer he already had a special interest in Radiology. During his bachelor in Medicine, which was obtained cum laude, he participated in the Honours Programme and started his research at the Radiology Department, which resulted in this thesis. In 2013 Daniel decided to participate in Alpe d'HuZes together with his father to do something back as team PayBackTime to the society who gave him the opportunity to develop himself in this way. In 2015 Daniel obtained his Master degree in Medicine cum laude which resulted in the 'third star'. After a full-time period of 8 months as a PhD student at the Radiology Department in the VU University Medical Center, Daniel went to the Daniel den Hoed Cancer Center in Rotterdam to work as a doctor in the Hematology Department and thereby obtaining clinical experience. Since 2018 he is working in the Franciscus Gasthuis in Rotterdam as a resident in the Internal Medicine department to further broaden his knowledge.

Daniel Peter Noij werd geboren in Hoofddorp, Nederland, op 8 april 1991. Na de lagere school ging hij naar de Katholieke Scholengemeenschap Hoofddorp waar hij cum laude zijn VWO diploma behaalde in 2009. Vervolgens ging Daniel naar het VU medisch centrum te Amsterdam om Geneeskunde te studeren. Met een moeder die röntgenlaborant was geweest had Daniel al langere tijd interesse in de Radiologie. Tijdens zijn bachelor Geneeskunde, welke uiteindelijk cum laude werd behaald, kwam hij via het Honours Programma bij de afdeling Radiologie terecht waar hij begon met wetenschappelijk onderzoek wat uiteindelijk heeft geleid tot dit promotieonderzoek. Daniel besloot om in 2013 samen met zijn vader deel te nemen aan Alpe d'HuZes om als team PayBackTime iets terug te doen voor de maatschappij die hem de kans heeft geboden om zich op deze manier te kunnen ontplooien. In 2015 rondde Daniel zijn Master Geneeskunde cum laude af waarmee de 'derde ster' binnen was. Na een full-time periode van 8 maanden als promovendus bij de afdeling Radiologie in het VU medisch centrum ging Daniel in 2016 naar de Daniel den Hoed kliniek in Rotterdam om als ANIOS Hematologie klinische ervaring op te doen. Per 2018 is hij werkzaam als ANIOS Interne Geneeskunde in het Franciscus Gasthuis in Rotterdam om zich verder te verbreden.

CHAPTER 11

ABBREVIATION LIST

ABBREVIATION LIST

^{18}F-FDG-PET-CT	= ^{18}F-Fluorodeoxyglucose positron emission tomography combined with computed tomography
^{18}F-MISO	= ^{18}F-Fluoromisonidazole
5-FU	= 5-Fluoruracil
A	= signal intensity
ADC	= apparent diffusion coefficient
AH	= amplitude scaling constant.
AIF	= arterial input function
AJCC	= American Joint Committee on Cancer
AR	= ascending rate
AUC/AUGC	= area under the (gadolinium concentration) curve
BF	= blood flow
BV	= blood volume
CE	= contrast-enhanced
CHESS	= chemical shift selective
CI	= contrast-Index
CT	= computed tomography
D	= diffusion coefficient
D*	= pseudodiffusion coefficient
DCE	= dynamic contrast enhanced
DFS	= disease-free survival
DSC	= dynamic susceptibility contrast
DWI/ DW-MRI	= diffusion-weighted imaging
DWIBS	= diffusion-weighted whole-body imaging with background body signal suppression
E	= post-radiotherapy enhancement value
EANM	= European association of nuclear medicine
EBV	= Epstein-Barr virus
EGFR	= Epidermal growth factor receptor
ENS	= extranodal spread

EPI	= echo-planar imaging
EUA	= examination under anesthesia
f	= perfusion fraction
FDG	= ^{18}Fluoro-2-deoxyglucose
FHX4	= 3-[^{18}F]fluoro-2-(4-((2-nitro-1H-imidazol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propan-1-ol
FLT	= 3'-Deoxy-3'-18F-fluorothymidine
FWHM	= full width at half maximum
FN	= false negative
FP	= false positive
GRE	= gradient-echo
GTV	= gross tumor volume
HASTE	= half-fourier acquisition single-shot turbo spin-echo
HNC	= head and neck cancer
HNSCC	= head and neck squamous cell carcinoma
HPV	= human papillomavirus
ICC	= intraclass correlation coefficient
IQR	= interquartile range
IVIM	= intravoxel incoherent motion
K21	= exchange rate constant
k_{ep}	= rate constant between interstitial space and plasma
K^{trans}	= volume transfer constant between plasma and interstitial space
LD-CT	= low dose computed tomography
LoA	= limits of agreement
LRC	= locoregional control
LRF	= locoregional failure
MP	= maximum probability model
MRI	= magnetic resonance imaging
MRgl	= metabolic response in FDG-uptake
MRS	= magnetic resonance spectroscopy
MSI/MSIR	= maximum slope of increase (ratio)

MVD = microvessel density

NA = negative agreement

NAC = neo-adjuvant chemotherapy

NPC = nasopharyngeal carcinoma

NPV = negative predictive value

NR = non-linear model

NS = not statistically significant

OA = overall survival

PA = positive agreement

PSA = proportion specific agreement

PCNA = proliferating cell nuclear antigen labeling index

PET = positron emission tomography

PFS = progression-free survival

PP = perfusion-related parameter

PPV = positive predictive value

PRISMA = preferred reporting items for systematic reviews and meta-analyses

PROPELLER = periodically rotated overlapping parallel lines with enhanced reconstruction

PS = permeability surface

QUADAS-2 = quality assessment of studies of diagnostic accuracy included in systematic reviews

QUIPS = quality in prognostic studies

R = regression coefficient

rTBV = tumor blood volume ratio

RFS = recurrence-free survival

ROC = receiver operating characteristic

ROI = region of interest

RT = radiotherapy

S1DE = relative slope of the signal intensity corrected to TE = 0

S2DE = relative change of the spin-spin relaxation rate

SCC = squamous cell carcinoma

SCM = sternocleidomastoid muscle

SD = standard deviation

SDG LN = subdigastric lymph node

SE/N = signal enhancement to noise ratio

SEM = standard error of measurement

SG = salivary gland

SMG = submandibular gland

SOM = suspicion of malignancy

SPIR = spectral presaturation with inversion recovery

SPT = second primary tumor

SRR-P = SUV reduction ratio in primary tumor

SS-SE-EPI = single-shot spin-echo echo planar imaging

STARD = standards for reporting of diagnostic accuracy

STIR = short tau inversion recovery

SUV = standardized uptake value

T0 = lag time

T1 = intravascular mean transit time

T1-WI = T1-weighted imaging

TA = time of arrival

TIC = times-signal intensity curve

TIM = total imaging matrix

TN = true negative

TOF = time of flight

TP = true positive

TSE = turbo spin-echo

TTP = time to peak

UP = unknown primary

v_e = interstitial fractional volume ($v_e = K^{trans}/ k_{ep}$)

VEGF = vascular endothelial growth factor

VIBE = volumetric interpolated breath-hold

WB-MRI = whole-body magnetic resonance imaging

YI = youden index

CHAPTER 12

APPENDICES

APPENDICES CHAPTER 2.1

Appendix A *Medline and Embase search*

Index test MRI (1)
("Magnetic Resonance Imaging"[Mesh] OR (magnetic[tw] AND resonance[tw]) OR NMR[tw] OR MRI[tw] OR MR[tw])

Technique (2)
(DCE*[tw] OR dynamic*[tw] OR pharmacokinetic[tw] OR DSC[tw] OR "Perfusion Imaging"[Mesh] OR perfusion[tw] OR bolustracking[tw] OR permeability[tw])

Target condition (location specific) (3)
"head and neck neoplasms"[Mesh] OR "Nasopharyngeal Neoplasms"[Mesh]

Target condition (location aspecific) (4)
("neoplasms, squamous cell"[Mesh] OR squamous[tw] OR "carcinoma, squamous cell"[Mesh] OR carcinoma[tw] OR cancer[tw] OR malignan*[tw] OR tumor[tw] OR tumour[tw] OR neoplasm*[tw])

Location (5)
("Neck"[Mesh] OR neck[tw] OR "Head"[Mesh] OR head[tw] OR "Pharynx"[Mesh] OR pharynx[tw] OR pharyngeal[tw] OR oropharynx[tw] OR oropharyngeal[tw] OR hypopharynx[tw] OR hypopharyngeal[tw] OR "Larynx"[Mesh] OR larynx[tw] OR laryngeal[tw] OR laryngopharyngeal[tw] OR nasopharyn* OR "Tongue"[Mesh] OR tongue[tw] OR "Palatine Tonsil"[Mesh] OR "Palate"[Mesh] OR palat*[tw] OR tonsil*[tw] OR "Vocal Cords"[Mesh] OR vocal cords[tw] OR "Lip"[Mesh] OR lip[tw] OR "Mouth"[Mesh] OR mouth[tw])

Limits (6)
English[lang] OR German[lang] OR Dutch[lang]

Final Medline search
1 AND 2 AND (3 OR (4 AND 5)) AND 6

Final Embase search
1 AND 2 AND 4 AND 5 AND 6

The Embase search was the same as the Medline search, but with only 'text words' and excluding Medline studies. However '3' was not included, because the Mesh-term is covered by '4' and 5'

Appendix B *Patient characteristics*

Author	Eligible patients (n)	Patients included	Mean age (SD; range)	% male	Stage	Nodes	Treatment	Reference standard
Sumi (41), 2014	131	79	59 (14-29)	53%	…	…	…	Histopathology
Ai (40), 2013	79	46	35 (18.3;7-69)	54%	…	…	…	Histopathology
Chawla (34), 2013	32	24	57.8 (11.1)	81%	...	...	CRx	Clinical assessment
Ng (39), 2013	78	58	48.5[a] (34-78)	93%	II (n=3); III (n=3); IVa (n=38); IVb (n=14)	N0 (n=10); N1 (n=4); N2 (n=33); N3 (n=11)	CRx	Histopathology and follow-up
Agrawal (13), 2012	...	21	43.9 (4.7)	71%	II (n=1); III (n=7); IV (n=13)	...	CRx	Clinical assessment
Chikui (33), 2012	42	29	60.2 (16.4)	66%	II (n=15); III (n=4); IV (n=6)	...	iCRx + Sx	Histopathology and follow-up
Jansen (29), 2012	46	12	57.0 (13; 40-79)	67%	III (n=1); IV (n=11)	...	Sx	Immunohistochemistry
Jansen (30), 2012	76	16	58.0 (7.0)	81%	III (n=1); IV (n=15)	...	CRx	PET
Lee (16), 2012	...	63	...	...	...	...	...	Histopathology
Shukla-Dave (32), 2012	...	74	... (38-83)	85%	III (n=12); IV (n=63)	...	CRx (n=61). Sx (n=13)	Clinical follow-up
Wang (37), 2012	14	13	56.9 (39-83)	86%	III (n=1); IV (n=13)	...	CRx	Clinical follow-up
Wendl (19), 2012	...	10	62.0 (51-83)	50%	...	...	Sx	Histopathology
Chawla (38), 2011	66	57	59.7 (10.4)[b]	82%b	III (n=1);. IV (n=65)	...	CRx	Clinical follow-up
Chikui (18), 2011	27	23	62.4 (14.4)	65%	...	...	iCRx + Sx	Histopathology and follow-up
Sumi (20), 2011	...	43	62.0 (37-82)	86%	...	N0 (n=17); N1 (n=4); N2 (n=8)	...	Histopathology
Bisdas (17), 2010	...	27	68[a] (49-81)	56%	T2 (n=3);. T3 (n=18); T4 (n=6)	...	...	PET-CT
Jansen (31), 2010	...	13	58.2 (9.3; 47-80)	100%	III (n=2); IV (n=11)	...	CRx	18F-FMISO/FDG PET
Kim (35), 2010	...	33	60.8 (10.8; 31-79)	85%	IV (n=33)	...	iCx (n=17); CRx (n=32); Sx (n=21)	Histopathology and follow-up
Unetsubo (21), 2009	150	28	65.9 (16.8;	50%	T2 (n=24);	N0 (n=2); N2 (n=9); N3 (n=3)	Sx	Immunohistochemistry

Appendix B *continued*

Author	Eligible patients (n)	Patients included	Mean age (SD; range)	% male	Stage	Nodes	Treatment	Reference standard
			23-90)		T3 (n=4)			
Van Cann (14), 2008	33	25	54.0 (48-76)	60%	CM (n=12); C n=9); O (n=5)	Benign (n=5); malignant (n=5)	Sx	Histopathology
Ariyoshi (22), 2006	...	20	56.1 (26-73)	65%	T2 (n=8); T3 (n=6); T4 (n=6)	...	Sx, Cx and CRx	Histopathology
Tomura (23), 2005	...	27	62.7 (8.6; 48-77)	89%	T2 (n=10); T3 (n=4); T4 (n=12); Tx (n=1)	N1 (n=3); N2 (n=56); N3 (n=7)	Rx + Sx	Histopathology
Hietschold (36), 2004	20	18	55.8 (43.8-67.5)	...	...	N0 (n=14); N1 (n=3); N2 (n=6)	Rx	PET and Eppendorf pO2
Fischbein (24), 2003	27	21	62.9 (12.1; 44-85)	62%	I (n=3); II (n=3); III (n=5); IV (n=10)	Cervical Lymphadenopathy	Sx +/- CRx	Histopathology
Helbig (15), 2003	...	13	51.8 (38-67)	85%	IV (n=13)	...	Rx or CRx	Clinical follow-up
Konouchi (25), 2003	50	30	69.3 (47-93)	53%	...	ENS+ (n=26); ENS- (n=28)	Sx	Immunohistochemistry
Tomura (26), 2002	28	10	59.6 (45-77)	76%	T1 (n=4); T2 (n=11); T3 (n=3); T4 (n=3)	1 (n=9); 2 (n=3); 3 (n=1)	Sx	Histopathology
Hoskin (27), 1999	...	13	...	...	III (n=4);IV (n=9)	N1 (n=2;. N2 (n=29); N3 (n=2)	Rx	Clinical follow-up
Baba (28), 1997	...	76	... (20-82)	76%	...	...	CRx (n=76) and Sx (n=35)	Histopathology
Razek (44), 2011	81	78	43.0 (18-71)	64%	...	...	Sx	Histopathology
Abdel Razek (45), 2011	47	45	45.0 (17-72)	...	...	N2 (n=23)	...	Histopathology
Bisdas (43), 2009	...	23	55.8 (8.9)	61%	T4 (n=12); T3 (n=9); T2 (n=2)	...	...	(DCE) CT
Wu (42), 2004	...	18	54.0 (15-82)	78%	...	...	...	Histopathology

[a] Median age
[b] n=66

Abbreviations: CR = chemoradiotherapy; CT = computed tomography; DCE = dynamic contrast-enhanced; ENS = extranodal spread; iCx = induction chemotherapy; PET = positron emission tomography; Rx = radiotherapy; Sx = surgery; SD = standard deviation; Tx = unknown primary tumor

Appendix C *List of abbreviations and relevant definitions*

A	= signal intensity
AH	= amplitude scaling constant. AH is a parameter affected by several factors (e.g. the dose of contrast agent and the ratio of the extracellular extravascular space).
AIF	= arterial input function
AUC/AUGC	= area under the (Gadolinium concentration) curve
AR	= ascending rate
BF	= blood flow
BV	= blood volume
CE	= contrast-enhanced
CI	= contrast-index
CRT/CRx	= chemoradiotherapy
DCE	= dynamic contrast enhanced
DSC	= dynamic susceptibility contrast
E	= post-radiotherapy enhancement value
ENS	= extranodal spread
FDG	= ^{18}Fluoro-2-deoxyglucose
^{18}F-MISO	= ^{18}F-Fluoromisonidazole
GTV	= gross tumor volume
HNSCC	= head and neck squamous cell carcinoma
K21	= exchange rate constant
k_{ep}	= rate constant between interstitial space and plasma
K^{trans}	= volume transfer constant between plasma and interstitial space
MSI/MSIR	= maximum slope of increase (ratio)
MVD	= microvessel density
PCNA	= proliferating cell nuclear antigen labeling index

PET	= positron emission tomography
PFS	= progression-free survival
PRISMA	= Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	= permeability surface
QUADAS-2	= QUality Assessment of studies of Diagnostic Accuracy included in Systematic reviews
QUIPS	= QUality In Prognosis Studies tool
R	= regression coefficient
ROC analysis	= receiver operating characteristic analysis
ROI	= region of interest
S1DE	= relative slope of the signal intensity corrected to TE = 0
S2DE	= relative change of the spin-spin relaxation rate
SE/N	= signal enhancement to noise ratio
STARD	= checklist to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study (internal validity) and to evaluate its generalizability (external validity).
SUV	= standardized uptake value
T0	= lag time
T1	= intravascular mean transit time
TA	= time of arrival
TTP	= time to peak
TIC	= times-signal intensity curve
v_e	= interstitial fractional volume ($v_e = K^{trans}/ k_{ep}$)
VEGF	= vascular Endothelial Growth Factor
YI	= youden's Index rates the performance of a diagnostic test: (J=Sensitivity + Specificity -1)

Appendix D *Results of diagnostic studies*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Sumi (41), 2014					79
TIC pattern	Histopathological diagnosis	…	…	SCC >< benign: Sens: 76% Spec: 57%[a] SCC >< lymphoma: Sens: 94% Spec: 50%[b]	
Ai (40), 2013					46
TIC pattern	Histopathological diagnosis[c]	…	..	Sens: 79% Spec: 91%	
Jansen (29), 2012					12
Ktrans median	VEGF	Spearman rank correlation	NS	---	
Ktrans median	Ki-67	Spearman rank correlation	NS	+	
Ktrans median	CA-IX	Spearman rank correlation	NS	+	
Ktrans median	HIF1a	Spearman rank correlation	NS	++	
Ktrans median	HE (necrosis)	Spearman rank correlation	NS	--	
Ktrans SD	VEGF	Spearman rank correlation	NS	+++	
Ktrans SD	Ki-67	Spearman rank correlation	<0.01	---	
Ktrans SD	CA-IX	Spearman rank correlation	NS	+	
Ktrans SD	HIF1a	Spearman rank correlation	NS	++	
Ktrans SD	HE (necrosis)	Spearman rank correlation	NS	--	
Kep median	VEGF	Spearman rank correlation	NS	++	
Kep median	Ki-67	Spearman rank correlation	NS	+/-	
Kep median	CA-IX	Spearman rank correlation	NS	++	
Kep median	HIF1a	Spearman rank correlation	NS	+/-	
Kep median	HE (necrosis)	Spearman rank correlation	NS	+/-	
Kep SD	VEGF[d]	Spearman rank correlation	<0.01	R = +++ Sens: 73% Spec: 100% AUC: 0.85 (CI95%: 0.64-1.00) YI: 0.73	
Kep SD	Ki-67	Spearman rank correlation	NS	--	
Kep SD	CA-IX	Spearman rank correlation	NS	+	
Kep SD	HIF1a	Spearman rank correlation	NS	-	
Kep SD	HE (necrosis)	Spearman rank correlation	NS	+/-	
Ve median	VEGF	Spearman rank correlation	NS	---	
Ve median	Ki-67	Spearman rank correlation	NS	+/-	
Ve median	CA-IX	Spearman rank correlation	NS	-	
Ve median	HIF1a	Spearman rank correlation	NS	+/-	
Ve median	HE (necrosis)	Spearman rank correlation	NS	-	
Ve SD	VEGF	Spearman rank correlation	NS	+++	
Ve SD	Ki-67	Spearman rank correlation	<0.01	---	
Ve SD	CA-IX	Spearman rank correlation	NS	+/-	
Ve SD	HIF1a	Spearman rank correlation	NS	--	
Ve SD	HE (necrosis)	Spearman rank correlation	NS	-	
Jansen (30), 2012					16
Ktrans mean	SUVmax	Spearman rank correlation	NS	+	
Ktrans mean	SUVmax	Spearman rank correlation	NS	+	
Ktrans mean	SUVmean	Spearman rank correlation	NS	++	
Ktrans mean	TLG[e]	Spearman rank correlation	NS	++	
Ktrans SD	SUVmax	Spearman rank correlation	NS	++	
Ktrans SD	SUVmean	Spearman rank correlation	NS	+	
Ktrans SD	TLG[e]	Spearman rank correlation	NS	--	
Kep mean	SUVmax	Spearman rank correlation	NS	--	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Kep mean	SUVmean	Spearman rank correlation	NS	-	
Kep mean	TLG[e]	Spearman rank correlation	NS	-	
Kep SD	SUVmax	Spearman rank correlation	NS	-	
Kep SD	SUVmean	Spearman rank correlation	NS	--	
Kep SD	TLG[e]	Spearman rank correlation	NS	++	
Ve mean	SUVmax	Spearman rank correlation	NS	++	
Ve mean	SUVmean	Spearman rank correlation	NS	+	
Ve mean	TLGee	Spearman rank correlation	NS	+	
Ve SD	SUVmax	Spearman rank correlation	NS	+	
Ve SD	SUVmean	Spearman rank correlation	NS	-	
Ve SD	TLG[e]	Logistic regression	NS	AUC: 0.50	
Ktrans SD	Short term response	Logistic regression	NS	AUC: 0.50	
AUC90 mean	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 88% Spec: 68%	
AUC60 25%	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 100% Spec: 36%	
Wendl (19), 2012					10
TTP	Histopathological diagnosis	ROC analysis	NS	Sens: 100% Spec: 40% AUC: 0.68 (CI95%: 0.33-1.00) YI: 0.40	
Maximal signal rise	Histopathological diagnosis	ROC analysis	<0.01	Sens: 100% Spec: 100% AUC: 1.00 (CI95%: 1.00-1.00) YI: 1.00	
Sumi (20), 2011					43
Type 2 TIC	Presence of ENS	ROC analysis	…	Sens: 77% Spec: 100%	
Type 1 TIC	Presence of ENS	Spearman rank correlation	NS	-	
Type 2 TIC	Presence of ENS	Spearman rank correlation	<0.001	+++	
Type 3 TIC	Presence of ENS	Spearman rank correlation	NS	+/-	
Type 4 TIC	Presence of ENS	Spearman rank correlation	<0.001	---	
Bisdas (17), 2010					27
Ktrans	SUVmean	Spearman rank correlation	NS	+/-	
Ktrans	SUVmax	Spearman rank correlation	NS	+	
Kep	SUVmean	Spearman rank correlation	NS	-	
Kep	SUVmax	Spearman rank correlation	NS	+/-	
Ve	SUVmean	Spearman rank correlation	<0.05	++	
Ve	SUVmax	Spearman rank correlation	NS	++	
iAUC	SUVmean	Spearman rank correlation	<0.001	+++	
iAUC	SUVmax	Spearman rank correlation	<0.001	+++	
Jansen (31), 2010					13
Ktrans median	FMISO SUV	Spearman rank correlation	NS	--	
Ktrans SD	FMISO SUV	Spearman rank correlation	NS	+/-	
Ktransskewness	FMISO SUV	Spearman rank correlation	NS	++	
Kep median	FMISO SUV	Spearman rank correlation	<0.05	---	
Kep SD	FMISO SUV	Spearman rank correlation	NS	+	
Kepskewness	FMISO SUV	Spearman rank correlation	NS	++	
Ve median	FMISO SUV	Spearman rank correlation	NS	+	
Ve SD	FMISO SUV	Spearman rank correlation	NS	+	
Veskewness	FMISO SUV	Spearman rank correlation	NS	--	
Unetsubo (21), 2009					28
Maximum CI	PCNA labelling index	Spearman rank correlation	NS	+	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Maximum CI	PCNA labelling index[f]	ROC analysis	NS	Sens: 100% Spec: 31% AUC: 0.52 (CI95%: 0.28-0.75) YI: 0.31	
Maximum CI	MVD	Spearman rank correlation	NS	+	
Maximum CI	MVD[g]	ROC analysis	NS	Sens: 44% Spec: 95% AUC: 0.70 (CI95%: 0.49-0.90) YI: 0.39	
Maximum CI gain	PCNA labelling index	Spearman rank correlation	<0.05	++	
Maximum CI gain	PCNA labelling index[f]	ROC analysis	NS	Sens: 100% Spec: 54% AUC: 0.63 (CI95%: 0.40-0.87) YI: 0.54	
Maximum CI gain	MVD	Spearman rank correlation	<0.01	++	
Maximum CI gain	MVD[g]	ROC analysis	<0.01	Sens: 100% Spec: 58% AUC: 0.83 (CI95%: 0.67-0.99) YI: 0.58	
CI-gain/CI-max ratio	PCNA labelling index	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	PCNA labelling index[f]	ROC analysis	<0.05	Sens: 100% Spec: 46% AUC: 0.75 (CI95%: 0.57-0.94) YI: 0.46	
CI-gain/CI-max ratio	MVD	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	MVD[g]	ROC analysis	<0.01	Sens: 89% Spec: 79% AUC: 0.83 (CI95%: 0.68-0.98) YI: 0.68	
Van Cann (14), 2008					25
Ktrans	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 92% Spec: 85% AUC: 0.94 (CI95%: 0.86-1.00) YI: 0.76	
Kep	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 100% Spec: 92% AUC: 0.99 (CI95%: 0.97-1.00) YI: 0.92	
Ve	Medullary mandibular invasion	ROC analysis	NS	Sens: 67% Spec: 54% AUC: 0.49 (CI95%: 0.26-0.73) YI: 0.21	
Ariyoshi (22), 2006					20
Max SE/N	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
Max SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	
%wash out	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
%wash out	Tumor invasion central area	Mann-Whitney U test	NS	…	
T-max SE/N	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
T-max SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	
AR SE/N	Tumor invasion peripheral area	Mann-Whitney U test	<0.05	…	
AR SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Tomura (23), 2005					27
MSIR[h]	Histologic grading	Mann-Whitney U test	<0.05	Sens: 77% Spec: 100%	
Hietschold (36), 2004					18
S1DE rel slope	SUV mean initial + 50Gy	Linear regression	≤0.05	0.19	
S1DE rel slope	SUV mean	Linear regression	≤0.05	0.21	
S2 flow	pO2 <5mmHg initial + 50 Gy	Linear regression	≤0.05	0.21	
S2DE flow	SUVmean	Linear regression	≤0.05	0.15	
S2DE flow	volume PET initial	Linear regression	≤0.05	0.29	
Fischbein (24), 2003					21
Peak time	Presence of tumor	T-test	<0.001	…	
Peak enhancement	Presence of tumor	T-test	<0.05	…	
Maximum slope	Presence of tumor	T-test	<0.01	…	
Washout slope	Presence of tumor	T-test	<0.05	…	
Konouchi (25), 2003					30
Maximum CI	PCNA[f]	Student's t-test → ROC analysis	<0.001	R= +++ Sens: 94% Spec: 92% AUC: 0.96 (CI95%: 0.90-1.00) YI: 0.86	
Maximum CI gain	PCNA[f]	Student's t-test → ROC analysis	<0.01	R= +++ Sens: 56% Spec: 92% AUC: 0.79 (CI95%: 0.62-0.95) YI: 0.47	
Tomura (26), 2002					10
Peak time	Presence of tumor	…	…	Sens: 80%	
Razek (44), 2011					78
DSC%	Histopathological diagnosis	ANOVA	NS	…	
Abdel Razek (45), 2011					45
DSC%[i]	SCC >< NHL	ROC analysis	…	Sens: 95% Spec: 91% AUC: 0.97	
Bisdas (43), 2009					23
Blood flow (MR)	Blood flow (CT)	Parametric comparison test	<0.01	…	
T1 (MR)	T1 (CT)	Parametric comparison test	≤0.001	…	
T0 (MR)	T0 (CT)	Parametric comparison test	≤0.001	…	
PS (MR)	PS (CT)	Parametric comparison test	<0.01	…	
E (MR)	E (CT)	Parametric comparison test	NS	…	
Ve (MR)	Ve (CT)	Parametric comparison test	NS	…	
Vp (MR)	Vp (CT)	Parametric comparison test	≤0.001	…	
Wu (42), 2004					18
rBV	Histopathological diagnosis	…	…	Sens: 100% Spec: 100%	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Baba (28), 1997					76
Enhancement (early vs late)	Tumor regression	…	…	Sens: 94% Spec: 93%	

[a] Cut-off at TIC 2/3
[b] Cut-off at TIC 3/4
[c] Cut-off at TIC A/B
[d] Cut-off at -/+
[e] n=14
[f] Cut-off at mean PCNA labeling index
[g] Cut-off at mean MVD
[h] Cut-off at MSIR <2.5
[i] Cut-off at DSC% of 43.5%

Correlations are reported as follows: -1.0 to -0.5 = - - -; - 0.5 to -0.3 = - -; -0.3 to -0.1 = -; -0.1 to 0.1 = +/-; 0.1 to 0.3 = +; 0.3 to 0.5 = + +; 0.5 to 1.0 = + + +

Abbreviations: ANOVA = analysis of variance; AR = ascending rate; AUC = area under the curve; CA-IX = carbonic anhydrase; CI = contrast index; DSC = dynamic susceptibility contrast; E = extraction ratio; ENS = extranodal spread; FMISO = 18F-fluoromisonidazol; HE = haematoxylin-eosin; HIF1α = Hypoxia-inducible factor 1-alpha; iAUC = initial area under the curve; Kep = rate constant between extracellular extravascular space and blood plasma; Ki-67 = marker of cellular proliferation; Ktrans = volume transfer constant between plasma and extracellular extravascular space; MSIR = maximum slope of increase ratio; MVD = microvessel density; NHL = non-Hodgkin lymphoma; NS = not significant; PCNA = proliferating cell nuclear antigen; PET = positron emission tomography; PS = permeability surface; R = regression coefficient; rBV = relative blood volume; ROC = receiver operating characteristic; S1DE = relative slope of the signal intensity corrected to TE = 0; S2DE = relative change of the spin-spin relaxation rate; SCC = squamous cell carcinoma; SD = standard deviation; SE/N = signal enhancement to noise ratio; SUV = standardized Uptake Value; T0= lag time; T1= intravascular blood volume; TIC = time-signal intensity curve; TLG = total lesion glycolysis; TTP = time to peak; Undiff = undifferentiated carcinoma; Ve= volume of extravascular extracellular space per unit volume of tissue; VEGF = vascular endothelial growth factor; YI = Youden Index

Appendix E *Results of prognostic studies*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
Chawla (34), 2013						24
Median 23.72 months (range 2.37-49.9months)	Ktrans primary tumor	Clinical response	Student t test	NS	…	
	Ktrans node	Clinical response	Student t test	<0.01	…	
	Ktrans node	Clinical response	Univariate analysis	NS	Sens: 44% Spec: 88% AUC: 0.56	
	Ve primary tumor	Clinical response	Student t test	NS	…	
	Ve node	Clinical response	Student t test	NS	…	
	Vp primary tumor	Clinical response	Student t test	NS	…	
	Vp node	Clinical response	Student t test	NS	…	
Ng (39), 2013						58
Median 19.2 months (range 9-32.3 months)	Ktrans primary tumor	2-year control	Log rank test	<0.05	Hazard ratio: 0.34	
	Ktrans primary tumor	Local control	Univariable logistic regression	0.001		
	Ktrans primary tumor	Local control	Multivariable logistic regression → ROC analysis	<0.05	Sens: 71% Spec: 83%	
	Ktrans primary tumor	Local control	Student t test	0.01	Odds ratio: 0.06	
	Ve primary tumor	Local control	Univariable logistic regression	<0.05	…	
	Ve primary tumor	2-year control	Log rank test	NS	…	
	Ve primary tumor	Local control	Student t test	NS	…	
	Vp primary tumor	2-year control	Log rank test	NS	…	
	Vp primary tumor	Local control	Student t test	NS	…	
	Kep primary tumor	2-year control	Log rank test	NS	…	
	Kep primary tumor	Local control	Student t test	NS	…	
Agrawal (13), 2012						21
6 weeks	BF	Clinical response	Independent samples t-test	0.001	…	
	BF	T-stage	Independent samples t-test	<0.05	…	
	BF	Control at primary	Independent samples t-test	NS	…	
	BF	Control at node	Independent samples t-test	NS	…	
	BV	Clinical response	Independent samples t-test	0.001	…	
	BV	T-stage	Independent samples t-test	0.05	…	
	BV	Control at primary	Independent samples t-test	0.05	…	
	BV	Control at node	Independent samples t-test	NS	…	
Chikui (33), 2012						29
…	Ktrans pre CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ktrans post CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ktrans ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	Kep pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	Kep post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ve pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ve post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	Ve ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	Vp pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Vp post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Vp ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AUGC pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AUGC post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
	AUGC ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
Shukla-Dave (32), 2012[a]						74
Median 40 months (range 13-64 months)	Ktransskewness	PFS	Cox regression analyses	<0.05	…	
	Ktrans median	PFS	Cox regression analyses	NS	…	
	Ktransskewness	OS	Cox regression analyses	<0.05	…	
	Ktrans median	OS	Cox regression analyses	NS	…	
	Ktrans SD	OS	Cox regression analyses	NS	…	
	Stage IV only ktransskewness	PFS	Cox regression analyses	<0.001	…	
	Stage IV only ktransskewness	OS	Cox regression analyses	<0.001	…	
	Ve SD	PFS	Cox regression analyses	<0.05	…	
	Ve median	OS	Cox regression analyses	<0.01	…	
Wang (37), 2012						13
Median 19.6 months (range 14.1-36.4 months)	Subvolume with low BF pre-Rx	Local control	Mann–Whitney U test	NS	…	
	Subvolume with low BF 2wk	Local control	Mann–Whitney U test	0.05	…	
	Mean ΔBF	Local control	Mann–Whitney U test	NS	…	
	Subvolume with low BV pre-Rx	Local control	Mann–Whitney U test	<0.05	…	
	Subvolume with low BV 2wk	Local control	Mann–Whitney U test	0.01	…	
	Mean ΔBV	Local control	Mann–Whitney U test	<0.05	…	
	GTV w/ low BV pretreatment	Local control	ROC analysis	…	Sens: 75% Spec: 100% AUC (SD): 0.92 (0.09) 95%CI: 0.74-1.00 YI: 0.75	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	GTV w/ low BV 2wk	Local control	ROC analysis	…	Sens: 100% Spec: 77.8% AUC (SD): 0.94 (0.07) 95%CI: 0.81-1.00) YI: 0.78	
	Change BV of GTV 2wk	Local control	ROC analysis	…	Sens: 85% Spec: 83% AUC (SD): 0.90 (0.08)	
	Change BV of pre GTV	Local control	ROC analysis	…	Sens: 85% Spec: 76%	
Chawla (38), 2011						57
Median 30 months (range 13-48 months)	Ktrans	Disease-free survival	Hazard ratio	<0.05	3.81 (95%CI: 1.04-13.9)	
Chikui (18), 2011						23
54.4±5.9 days	Kep pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AH pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AH ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH ΔCRT	Ohboshi and Shimosato criteria	Tukey's HSD	<0.05	Sens: 83% Spec: 71%	
	Kel pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kel post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kel ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
Kim (35), 2010						33
Disease status at the end of CRT	Ktrans	Clinical response (CR/PR)	2-tailed t test with unequal variance	0.001	…	
	Ktrans	Clinical response (CR/PR)	ROC analysis	<0.01	Sens: 89% Spec: 63% AUC: 0.83 (95%CI: 0.68-0.98) YI: 0.51	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	Ve	Clinical response (CR/PR)	2-tailed t test with unequal variance	NS	…	
	Ve	Clinical response (CR/PR)	ROC analysis	NS	Sens: 33% Spec: 88% AUC: 0.54 (95%CI: 0.31-0.78) YI: 0.21	
	τi	Clinical response (CR/PR)	2-tailed t test with unequal variance	NS	…	
	τi	Clinical response (CR/PR)	ROC analysis	NS	Sens: 56% Spec: 63% AUC: 0.51 (95%CI: 0.28-0.73) YI: 0.18	
Helbig (15), 2003						13
>9 months	A primary tumor	Clinical response (CR/PR)	…	…	CR: increase > SD/PR: increase	
	A node	Clinical response (CR/PR)	…	…	CR: increase SD/PR: decrease	
	K21 primary	Clinical response (CR/PR)	…	…	CR: decrease PR: increase	
	K21 node	Clinical response (CR/PR)	…	…	CR: decrease PR: increase	
Hoskin (27), 1999						13
11.5 months (range 6-20 months)	ΔTTP	Locoregional control	…	…	Sens: 86% Spec: 67%	
	Peak enhancement post-Rx	Locoregional control	Student t test	<0.01	…	
	Peak enhancement post-Rx	Locoregional control	ROC analysis	<0.01	Sens: 100% Spec: 80% AUC: 0.95 (95%CI: 0.83-1.00) YI: 0.80	
	ΔPeakenhancment	Locoregional control	…	…	Sens: 75% Spec: 60%	
	ΔMaximum slope	Locoregional control	…	…	Sens: 67% Spec: 33%	

[a] Not mentioned by the authors are Ktrans vs PFS; Kep median vs PFS; Kep skewness vs PFS; Kep SD vs PFS; Kep median vs OS; Kep skewness vs OS; Kep SD vs OS; Ve SD vs OS; Ve median vs PFS; Ve skewness vs PFS; Ve skewness vs OS

Abbreviations: A = signal intensity; AH = amplitude scaling constant; AUGC = area under the gadolinium concentration curve; AUC = area under the curve; BF = blood flow; BV = blood volume; CR = complete response; CRT = chemoradiotherapy; GTV = gross tumor volume; K21 = exchange rate constant; Ktrans = volume transfer constant between plasma and extracellular extravascular space; Kel = elimination of contrast medium from the central compartment; Kep = rate constant between extracellular extravascular space and blood plasma; NS = not significant; OS = overall survival; PFS = progression-free survival; PR = partial response; ROC = receiver operating curve; Rx = radiation therapy; SD = standard deviation; Sens = sensitivity; Spec = specificity; TA = time of arrival; Ti = intracellular water lifetime; TTP = time to peak; Ve = extravascular volume fraction; Vp = plasma volume fraction; YI = Youden index; 95%CI = 95% confidence interval;

Appendix F *QUADAS-2: Summary of the risk of bias and applicability concerns: review of authors' judgment about each domain of all included studies: '✓' indicates a low risk of bias; '?' an unclear risk and 'x' indicates a high risk of bias*

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Agrawal (13), 2012	✓	?	?	✓	✓	?	?
Ai (40), 2013	?	✓	✓	✓	✓	✓	✓
Ariyoshi (22), 2006	?	?	?	?	✓	✓	✓
Baba (28), 1997	✓	?	✓	✓	?	?	?
Bisdas (17), 2010	?	✓	?	✓	✓	✓	x
Bisdas (43), 2009	?	?	x	✓	✓	?	x
Chawla (34), 2013	?	x	x	x	✓	✓	?
Chawla (38), 2011	?	?	?	x	✓	✓	?
Chikui (18), 2011	?	✓	?	?	✓	✓	✓
Chikui (33), 2012	x	?	?	✓	✓	x	✓
Fischbein (24), 2003	?	✓	?	?	✓	✓	✓
Helbig (15), 2003	?	?	?	?	✓	✓	✓
Hietschold (36), 2004	?	?	?	?	✓	✓	✓
Hoskin (27), 1999	x	?	?	?	✓	✓	✓
Jansen (29), 2012	?	?	x	?	✓	✓	✓
Jansen (30), 2012	?	?	?	x	✓	✓	x
Jansen (31), 2010	?	?	?	✓	✓	✓	✓
Kim (35), 2010	?	?	?	?	✓	✓	✓
Konouchi (25), 2003	?	?	?	?	✓	✓	✓
Lee (16), 2012	x	?	?	?	✓	?	✓
Ng (39), 2013	?	x	?	?	✓	✓	✓
Razek (44), 2011	✓	✓	?	?	✓	✓	✓
Abdel Razek (45), 2011	?	?	?	?	✓	✓	✓
ShuklaxDave (32), 2012	?	✓	?	?	✓	✓	✓
Sumi (20), 2011	✓	?	?	?	✓	✓	✓
Sumi (41), 2014	x	?	?	✓	✓	✓	✓
Tomura (23), 2005	?	?	?	✓	✓	✓	✓
Tomura (26), 2002	?	?	?	?	✓	✓	✓
Unetsubo (21), 2009	?	?	✓	?	✓	✓	✓
Van Cann (14), 2008	?	x	?	?	✓	?	✓
Wang (37), 2012	?	?	?	?	✓	x	✓
Wendl (19), 2012	x	?	?	?	✓	✓	✓
Wu (42), 2004	x	?	?	?	?	✓	✓

Appendix G *QUIPS: Summary of the risk of bias on six domains: review of authors' judgment about each domain of the included prognostic studies: '✓' indicates a low risk of bias; '?' a moderate risk and 'x' indicates a high risk of bias*

	1. Study participation	2. Study attrition	3. Prognostic factor measurement	4. Outcome measurement	5. Study confounding	6. Statistical analysis and reporting
Agrawal (13), 2012	?	?	✓	x	x	?
Chawla (34), 2013	?	x	?	x	?	?
Chawla (38), 2011	✓	?	x	x	?	✓
Chikui (18), 2011	?	?	✓	✓	?	?
Chikui (33), 2012	?	x	?	✓	?	?
Helbig (15), 2003	?	✓	✓	?	?	?
Hoskin (27), 1999	x	✓	x	?	x	x
Kim (35), 2010	?	✓	?	x	?	✓
Ng (39), 2013	✓	?	?	x	x	✓
Shukla-Dave (32), 2012	?	✓	✓	?	✓	✓
Wang (37), 2012	x	✓	?	?	?	✓

CHAPTER 12

APPENDICES

APPENDICES CHAPTER 2.1

Appendix A *Medline and Embase search*

Index test MRI (1)
("Magnetic Resonance Imaging"[Mesh] OR (magnetic[tw] AND resonance[tw]) OR NMR[tw] OR MRI[tw] OR MR[tw])

Technique (2)
(DCE*[tw] OR dynamic*[tw] OR pharmacokinetic[tw] OR DSC[tw] OR "Perfusion Imaging"[Mesh] OR perfusion[tw] OR bolustracking[tw] OR permeability[tw])

Target condition (location specific) (3)
"head and neck neoplasms"[Mesh] OR "Nasopharyngeal Neoplasms"[Mesh]

Target condition (location aspecific) (4)
("neoplasms, squamous cell"[Mesh] OR squamous[tw] OR "carcinoma, squamous cell"[Mesh] OR carcinoma[tw] OR cancer[tw] OR malignan*[tw] OR tumor[tw] OR tumour[tw] OR neoplasm*[tw])

Location (5)
("Neck"[Mesh] OR neck[tw] OR "Head"[Mesh] OR head[tw] OR "Pharynx"[Mesh] OR pharynx[tw] OR pharyngeal[tw] OR oropharynx[tw] OR oropharyngeal[tw] OR hypopharynx[tw] OR hypopharyngeal[tw] OR "Larynx"[Mesh] OR larynx[tw] OR laryngeal[tw] OR laryngopharyngeal[tw] OR nasopharyn* OR "Tongue"[Mesh] OR tongue[tw] OR "Palatine Tonsil"[Mesh] OR "Palate"[Mesh] OR palat*[tw] OR tonsil*[tw] OR "Vocal Cords"[Mesh] OR vocal cords[tw] OR "Lip"[Mesh] OR lip[tw] OR "Mouth"[Mesh] OR mouth[tw])

Limits (6)
English[lang] OR German[lang] OR Dutch[lang]

Final Medline search
1 AND 2 AND (3 OR (4 AND 5)) AND 6

Final Embase search
1 AND 2 AND 4 AND 5 AND 6

The Embase search was the same as the Medline search, but with only 'text words' and excluding Medline studies. However '3' was not included, because the Mesh-term is covered by '4' and 5'

Appendix B *Patient characteristics*

Author	Eligible patients (n)	Patients included	Mean age (SD; range)	% male	Stage	Nodes	Treatment	Reference standard
Sumi (41), 2014	131	79	59 (14-29)	53%	…	…	…	Histopathology
Ai (40), 2013	79	46	35 (18.3;7-69)	54%	…	…	…	Histopathology
Chawla (34), 2013	32	24	57.8 (11.1)	81%	...	...	CRx	Clinical assessment
Ng (39), 2013	78	58	48.5[a] (34-78)	93%	II (n=3); III (n=3); IVa (n=38); IVb (n=14)	N0 (n=10); N1 (n=4); N2 (n=33); N3 (n=11)	CRx	Histopathology and follow-up
Agrawal (13), 2012	...	21	43.9 (4.7)	71%	II (n=1); III (n=7); IV (n=13)	...	CRx	Clinical assessment
Chikui (33), 2012	42	29	60.2 (16.4)	66%	II (n=15); III (n=4); IV (n=6)	...	iCRx + Sx	Histopathology and follow-up
Jansen (29), 2012	46	12	57.0 (13; 40-79)	67%	III (n=1); IV (n=11)	...	Sx	Immunohistochemistry
Jansen (30), 2012	76	16	58.0 (7.0)	81%	III (n=1); IV (n=15)	...	CRx	PET
Lee (16), 2012	...	63	...	...	...	...	...	Histopathology
Shukla-Dave (32), 2012	...	74	... (38-83)	85%	III (n=12); IV (n=63)	...	CRx (n=61). Sx (n=13)	Clinical follow-up
Wang (37), 2012	14	13	56.9 (39-83)	86%	III (n=1); IV (n=13)	...	CRx	Clinical follow-up
Wendl (19), 2012	...	10	62.0 (51-83)	50%	...	...	Sx	Histopathology
Chawla (38), 2011	66	57	59.7 (10.4)[b]	82%b	III (n=1);. IV (n=65)	...	CRx	Clinical follow-up
Chikui (18), 2011	27	23	62.4 (14.4)	65%	...	...	iCRx + Sx	Histopathology and follow-up
Sumi (20), 2011	...	43	62.0 (37-82)	86%	...	N0 (n=17); N1 (n=4); N2 (n=8)	...	Histopathology
Bisdas (17), 2010	...	27	68[a] (49-81)	56%	T2 (n=3);. T3 (n=18); T4 (n=6)	...	...	PET-CT
Jansen (31), 2010	...	13	58.2 (9.3; 47-80)	100%	III (n=2); IV (n=11)	...	CRx	18F-FMISO/FDG PET
Kim (35), 2010	...	33	60.8 (10.8; 31-79)	85%	IV (n=33)	...	iCx (n=17); CRx (n=32); Sx (n=21)	Histopathology and follow-up
Unetsubo (21), 2009	150	28	65.9 (16.8;	50%	T2 (n=24);	N0 (n=2); N2 (n=9); N3 (n=3)	Sx	Immunohistochemistry

Appendix B *continued*

Author	Eligible patients (n)	Patients included	Mean age (SD; range)	% male	Stage	Nodes	Treatment	Reference standard
			23-90)		T3 (n=4)			
Van Cann (14), 2008	33	25	54.0 (48-76)	60%	CM (n=12); C n=9); O (n=5)	Benign (n=5); malignant (n=5)	Sx	Histopathology
Ariyoshi (22), 2006	...	20	56.1 (26-73)	65%	T2 (n=8); T3 (n=6); T4 (n=6)	...	Sx, Cx and CRx	Histopathology
Tomura (23), 2005	...	27	62.7 (8.6; 48-77)	89%	T2 (n=10); T3 (n=4); T4 (n=12); Tx (n=1)	N1 (n=3); N2 (n=56); N3 (n=7)	Rx + Sx	Histopathology
Hietschold (36), 2004	20	18	55.8 (43.8-67.5)	...	...	N0 (n=14); N1 (n=3); N2 (n=6)	Rx	PET and Eppendorf pO2
Fischbein (24), 2003	27	21	62.9 (12.1; 44-85)	62%	I (n=3); II (n=3); III (n=5); IV (n=10)	Cervical Lymphadenopathy	Sx +/- CRx	Histopathology
Helbig (15), 2003	...	13	51.8 (38-67)	85%	IV (n=13)	...	Rx or CRx	Clinical follow-up
Konouchi (25), 2003	50	30	69.3 (47-93)	53%	...	ENS+ (n=26); ENS- (n=28)	Sx	Immunohistochemistry
Tomura (26), 2002	28	10	59.6 (45-77)	76%	T1 (n=4); T2 (n=11); T3 (n=3); T4 (n=3)	1 (n=9); 2 (n=3); 3 (n=1)	Sx	Histopathology
Hoskin (27), 1999	...	13	...	...	III (n=4);IV (n=9)	N1 (n=2;. N2 (n=29); N3 (n=2)	Rx	Clinical follow-up
Baba (28), 1997	...	76	... (20-82)	76%	...	...	CRx (n=76) and Sx (n=35)	Histopathology
Razek (44), 2011	81	78	43.0 (18-71)	64%	...	...	Sx	Histopathology
Abdel Razek (45), 2011	47	45	45.0 (17-72)	...	...	N2 (n=23)	...	Histopathology
Bisdas (43), 2009	...	23	55.8 (8.9)	61%	T4 (n=12); T3 (n=9); T2 (n=2)	...	...	(DCE) CT
Wu (42), 2004	...	18	54.0 (15-82)	78%	...	...	...	Histopathology

[a] Median age
[b] n=66

Abbreviations: CR = chemoradiotherapy; CT = computed tomography; DCE = dynamic contrast-enhanced; ENS = extranodal spread; iCx = induction chemotherapy; PET = positron emission tomography; Rx = radiotherapy; Sx = surgery; SD = standard deviation; Tx = unknown primary tumor

Appendix C *List of abbreviations and relevant definitions*

A	= signal intensity
AH	= amplitude scaling constant. AH is a parameter affected by several factors (e.g. the dose of contrast agent and the ratio of the extracellular extravascular space).
AIF	= arterial input function
AUC/AUGC	= area under the (Gadolinium concentration) curve
AR	= ascending rate
BF	= blood flow
BV	= blood volume
CE	= contrast-enhanced
CI	= contrast-index
CRT/CRx	= chemoradiotherapy
DCE	= dynamic contrast enhanced
DSC	= dynamic susceptibility contrast
E	= post-radiotherapy enhancement value
ENS	= extranodal spread
FDG	= ^{18}Fluoro-2-deoxyglucose
^{18}F-MISO	= ^{18}F-Fluoromisonidazole
GTV	= gross tumor volume
HNSCC	= head and neck squamous cell carcinoma
K21	= exchange rate constant
k_{ep}	= rate constant between interstitial space and plasma
K^{trans}	= volume transfer constant between plasma and interstitial space
MSI/MSIR	= maximum slope of increase (ratio)
MVD	= microvessel density
PCNA	= proliferating cell nuclear antigen labeling index

PET	= positron emission tomography
PFS	= progression-free survival
PRISMA	= Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	= permeability surface
QUADAS-2	= QUality Assessment of studies of Diagnostic Accuracy included in Systematic reviews
QUIPS	= QUality In Prognosis Studies tool
R	= regression coefficient
ROC analysis	= receiver operating characteristic analysis
ROI	= region of interest
S1DE	= relative slope of the signal intensity corrected to TE = 0
S2DE	= relative change of the spin-spin relaxation rate
SE/N	= signal enhancement to noise ratio
STARD	= checklist to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study (internal validity) and to evaluate its generalizability (external validity).
SUV	= standardized uptake value
T0	= lag time
T1	= intravascular mean transit time
TA	= time of arrival
TTP	= time to peak
TIC	= times-signal intensity curve
v_e	= interstitial fractional volume ($v_e = K^{trans}/ k_{ep}$)
VEGF	= vascular Endothelial Growth Factor
YI	= youden's Index rates the performance of a diagnostic test: (J=Sensitivity + Specificity -1)

Appendix D *Results of diagnostic studies*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Sumi (41), 2014					79
TIC pattern	Histopathological diagnosis	…	…	SCC >< benign: Sens: 76% Spec: 57%[a] SCC >< lymphoma: Sens: 94% Spec: 50%[b]	
Ai (40), 2013					46
TIC pattern	Histopathological diagnosis[c]	…	..	Sens: 79% Spec: 91%	
Jansen (29), 2012					12
Ktrans median	VEGF	Spearman rank correlation	NS	---	
Ktrans median	Ki-67	Spearman rank correlation	NS	+	
Ktrans median	CA-IX	Spearman rank correlation	NS	+	
Ktrans median	HIF1a	Spearman rank correlation	NS	++	
Ktrans median	HE (necrosis)	Spearman rank correlation	NS	--	
Ktrans SD	VEGF	Spearman rank correlation	NS	+++	
Ktrans SD	Ki-67	Spearman rank correlation	<0.01	---	
Ktrans SD	CA-IX	Spearman rank correlation	NS	+	
Ktrans SD	HIF1a	Spearman rank correlation	NS	++	
Ktrans SD	HE (necrosis)	Spearman rank correlation	NS	--	
Kep median	VEGF	Spearman rank correlation	NS	++	
Kep median	Ki-67	Spearman rank correlation	NS	+/-	
Kep median	CA-IX	Spearman rank correlation	NS	++	
Kep median	HIF1a	Spearman rank correlation	NS	+/-	
Kep median	HE (necrosis)	Spearman rank correlation	NS	+/-	
Kep SD	VEGF[d]	Spearman rank correlation	<0.01	R = +++ Sens: 73% Spec: 100% AUC: 0.85 (CI95%: 0.64-1.00) YI: 0.73	
Kep SD	Ki-67	Spearman rank correlation	NS	--	
Kep SD	CA-IX	Spearman rank correlation	NS	+	
Kep SD	HIF1a	Spearman rank correlation	NS	-	
Kep SD	HE (necrosis)	Spearman rank correlation	NS	+/-	
Ve median	VEGF	Spearman rank correlation	NS	---	
Ve median	Ki-67	Spearman rank correlation	NS	+/-	
Ve median	CA-IX	Spearman rank correlation	NS	-	
Ve median	HIF1a	Spearman rank correlation	NS	+/-	
Ve median	HE (necrosis)	Spearman rank correlation	NS	-	
Ve SD	VEGF	Spearman rank correlation	NS	+++	
Ve SD	Ki-67	Spearman rank correlation	<0.01	---	
Ve SD	CA-IX	Spearman rank correlation	NS	+/-	
Ve SD	HIF1a	Spearman rank correlation	NS	--	
Ve SD	HE (necrosis)	Spearman rank correlation	NS	-	
Jansen (30), 2012					16
Ktrans mean	SUVmax	Spearman rank correlation	NS	+	
Ktrans mean	SUVmax	Spearman rank correlation	NS	+	
Ktrans mean	SUVmean	Spearman rank correlation	NS	++	
Ktrans mean	TLG[e]	Spearman rank correlation	NS	++	
Ktrans SD	SUVmax	Spearman rank correlation	NS	++	
Ktrans SD	SUVmean	Spearman rank correlation	NS	+	
Ktrans SD	TLG[e]	Spearman rank correlation	NS	--	
Kep mean	SUVmax	Spearman rank correlation	NS	--	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Kep mean	SUVmean	Spearman rank correlation	NS	-	
Kep mean	TLG[e]	Spearman rank correlation	NS	-	
Kep SD	SUVmax	Spearman rank correlation	NS	-	
Kep SD	SUVmean	Spearman rank correlation	NS	--	
Kep SD	TLG[e]	Spearman rank correlation	NS	++	
Ve mean	SUVmax	Spearman rank correlation	NS	++	
Ve mean	SUVmean	Spearman rank correlation	NS	+	
Ve mean	TLGee	Spearman rank correlation	NS	+	
Ve SD	SUVmax	Spearman rank correlation	NS	+	
Ve SD	SUVmean	Spearman rank correlation	NS	-	
Ve SD	TLG[e]	Logistic regression	NS	AUC: 0.50	
Ktrans SD	Short term response	Logistic regression	NS	AUC: 0.50	
AUC90 mean	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 88% Spec: 68%	
AUC60 25%	SCC vs Undiff	Kruskal–Wallis test → ROC analysis	<0.01	Sens: 100% Spec: 36%	
Wendl (19), 2012					10
TTP	Histopathological diagnosis	ROC analysis	NS	Sens: 100% Spec: 40% AUC: 0.68 (CI95%: 0.33-1.00) YI: 0.40	
Maximal signal rise	Histopathological diagnosis	ROC analysis	<0.01	Sens: 100% Spec: 100% AUC: 1.00 (CI95%: 1.00-1.00) YI: 1.00	
Sumi (20), 2011					43
Type 2 TIC	Presence of ENS	ROC analysis	…	Sens: 77% Spec: 100%	
Type 1 TIC	Presence of ENS	Spearman rank correlation	NS	-	
Type 2 TIC	Presence of ENS	Spearman rank correlation	<0.001	+++	
Type 3 TIC	Presence of ENS	Spearman rank correlation	NS	+/-	
Type 4 TIC	Presence of ENS	Spearman rank correlation	<0.001	---	
Bisdas (17), 2010					27
Ktrans	SUVmean	Spearman rank correlation	NS	+/-	
Ktrans	SUVmax	Spearman rank correlation	NS	+	
Kep	SUVmean	Spearman rank correlation	NS	-	
Kep	SUVmax	Spearman rank correlation	NS	+/-	
Ve	SUVmean	Spearman rank correlation	<0.05	++	
Ve	SUVmax	Spearman rank correlation	NS	++	
iAUC	SUVmean	Spearman rank correlation	<0.001	+++	
iAUC	SUVmax	Spearman rank correlation	<0.001	+++	
Jansen (31), 2010					13
Ktrans median	FMISO SUV	Spearman rank correlation	NS	--	
Ktrans SD	FMISO SUV	Spearman rank correlation	NS	+/-	
Ktransskewness	FMISO SUV	Spearman rank correlation	NS	++	
Kep median	FMISO SUV	Spearman rank correlation	<0.05	---	
Kep SD	FMISO SUV	Spearman rank correlation	NS	+	
Kepskewness	FMISO SUV	Spearman rank correlation	NS	++	
Ve median	FMISO SUV	Spearman rank correlation	NS	+	
Ve SD	FMISO SUV	Spearman rank correlation	NS	+	
Veskewness	FMISO SUV	Spearman rank correlation	NS	--	
Unetsubo (21), 2009					28
Maximum CI	PCNA labelling index	Spearman rank correlation	NS	+	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Maximum CI	PCNA labelling index[f]	ROC analysis	NS	Sens: 100% Spec: 31% AUC: 0.52 (CI95%: 0.28-0.75) YI: 0.31	
Maximum CI	MVD	Spearman rank correlation	NS	+	
Maximum CI	MVD[g]	ROC analysis	NS	Sens: 44% Spec: 95% AUC: 0.70 (CI95%: 0.49-0.90) YI: 0.39	
Maximum CI gain	PCNA labelling index	Spearman rank correlation	<0.05	++	
Maximum CI gain	PCNA labelling index[f]	ROC analysis	NS	Sens: 100% Spec: 54% AUC: 0.63 (CI95%: 0.40-0.87) YI: 0.54	
Maximum CI gain	MVD	Spearman rank correlation	<0.01	++	
Maximum CI gain	MVD[g]	ROC analysis	<0.01	Sens: 100% Spec: 58% AUC: 0.83 (CI95%: 0.67-0.99) YI: 0.58	
CI-gain/CI-max ratio	PCNA labelling index	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	PCNA labelling index[f]	ROC analysis	<0.05	Sens: 100% Spec: 46% AUC: 0.75 (CI95%: 0.57-0.94) YI: 0.46	
CI-gain/CI-max ratio	MVD	Spearman rank correlation	0.001	+++	
CI-gain/CI-max ratio	MVD[g]	ROC analysis	<0.01	Sens: 89% Spec: 79% AUC: 0.83 (CI95%: 0.68-0.98) YI: 0.68	
Van Cann (14), 2008					25
Ktrans	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 92% Spec: 85% AUC: 0.94 (CI95%: 0.86-1.00) YI: 0.76	
Kep	Medullary mandibular invasion	ROC analysis	<0.001	Sens: 100% Spec: 92% AUC: 0.99 (CI95%: 0.97-1.00) YI: 0.92	
Ve	Medullary mandibular invasion	ROC analysis	NS	Sens: 67% Spec: 54% AUC: 0.49 (CI95%: 0.26-0.73) YI: 0.21	
Ariyoshi (22), 2006					20
Max SE/N	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
Max SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	
%wash out	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
%wash out	Tumor invasion central area	Mann-Whitney U test	NS	…	
T-max SE/N	Tumor invasion peripheral area	Mann-Whitney U test	NS	…	
T-max SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	
AR SE/N	Tumor invasion peripheral area	Mann-Whitney U test	<0.05	…	
AR SE/N	Tumor invasion central area	Mann-Whitney U test	NS	…	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Tomura (23), 2005					27
MSIR[h]	Histologic grading	Mann-Whitney U test	<0.05	Sens: 77% Spec: 100%	
Hietschold (36), 2004					18
S1DE rel slope	SUV mean initial + 50Gy	Linear regression	≤0.05	0.19	
S1DE rel slope	SUV mean	Linear regression	≤0.05	0.21	
S2 flow	pO2 <5mmHg initial + 50 Gy	Linear regression	≤0.05	0.21	
S2DE flow	SUVmean	Linear regression	≤0.05	0.15	
S2DE flow	volume PET initial	Linear regression	≤0.05	0.29	
Fischbein (24), 2003					21
Peak time	Presence of tumor	T-test	<0.001	…	
Peak enhancement	Presence of tumor	T-test	<0.05	…	
Maximum slope	Presence of tumor	T-test	<0.01	…	
Washout slope	Presence of tumor	T-test	<0.05	…	
Konouchi (25), 2003					30
Maximum CI	PCNA[f]	Student's t-test → ROC analysis	<0.001	R= +++ Sens: 94% Spec: 92% AUC: 0.96 (CI95%: 0.90-1.00) YI: 0.86	
Maximum CI gain	PCNA[f]	Student's t-test → ROC analysis	<0.01	R= +++ Sens: 56% Spec: 92% AUC: 0.79 (CI95%: 0.62-0.95) YI: 0.47	
Tomura (26), 2002					10
Peak time	Presence of tumor	…	…	Sens: 80%	
Razek (44), 2011					78
DSC%	Histopathological diagnosis	ANOVA	NS	…	
Abdel Razek (45), 2011					45
DSC%[i]	SCC >< NHL	ROC analysis	…	Sens: 95% Spec: 91% AUC: 0.97	
Bisdas (43), 2009					23
Blood flow (MR)	Blood flow (CT)	Parametric comparison test	<0.01	…	
T1 (MR)	T1 (CT)	Parametric comparison test	≤0.001	…	
T0 (MR)	T0 (CT)	Parametric comparison test	≤0.001	…	
PS (MR)	PS (CT)	Parametric comparison test	<0.01	…	
E (MR)	E (CT)	Parametric comparison test	NS	…	
Ve (MR)	Ve (CT)	Parametric comparison test	NS	…	
Vp (MR)	Vp (CT)	Parametric comparison test	≤0.001	…	
Wu (42), 2004					18
rBV	Histopathological diagnosis	…	…	Sens: 100% Spec: 100%	

Appendix D *continued*

MRI parameter	Outcome measurement	Test of significance	*P* value	Test outcome	n
Baba (28), 1997					76
Enhancement (early vs late)	Tumor regression	…	…	Sens: 94% Spec: 93%	

[a] Cut-off at TIC 2/3
[b] Cut-off at TIC 3/4
[c] Cut-off at TIC A/B
[d] Cut-off at -/+
[e] n=14
[f] Cut-off at mean PCNA labeling index
[g] Cut-off at mean MVD
[h] Cut-off at MSIR <2.5
[i] Cut-off at DSC% of 43.5%

Correlations are reported as follows: -1.0 to -0.5 = - - -; - 0.5 to -0.3 = - -; -0.3 to -0.1 = -; -0.1 to 0.1 = +/-; 0.1 to 0.3 = +; 0.3 to 0.5 = + +; 0.5 to 1.0 = + + +

Abbreviations: ANOVA = analysis of variance; AR = ascending rate; AUC = area under the curve; CA-IX = carbonic anhydrase; CI = contrast index; DSC = dynamic susceptibility contrast; E = extraction ratio; ENS = extranodal spread; FMISO = 18F-fluoromisonidazol; HE = haematoxylin-eosin; HIF1α = Hypoxia-inducible factor 1-alpha; iAUC = initial area under the curve; Kep = rate constant between extracellular extravascular space and blood plasma; Ki-67 = marker of cellular proliferation; Ktrans = volume transfer constant between plasma and extracellular extravascular space; MSIR = maximum slope of increase ratio; MVD = microvessel density; NHL = non-Hodgkin lymphoma; NS = not significant; PCNA = proliferating cell nuclear antigen; PET = positron emission tomography; PS = permeability surface; R = regression coefficient; rBV = relative blood volume; ROC = receiver operating characteristic; S1DE = relative slope of the signal intensity corrected to TE = 0; S2DE = relative change of the spin-spin relaxation rate; SCC = squamous cell carcinoma; SD = standard deviation; SE/N = signal enhancement to noise ratio; SUV = standardized Uptake Value; T0= lag time; T1= intravascular blood volume; TIC = time-signal intensity curve; TLG = total lesion glycolysis; TTP = time to peak; Undiff = undifferentiated carcinoma; Ve= volume of extravascular extracellular space per unit volume of tissue; VEGF = vascular endothelial growth factor; YI = Youden Index

Appendix E *Results of prognostic studies*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
Chawla (34), 2013						24
Median 23.72 months (range 2.37-49.9months)	Ktrans primary tumor	Clinical response	Student t test	NS	…	
	Ktrans node	Clinical response	Student t test	<0.01	…	
	Ktrans node	Clinical response	Univariate analysis	NS	Sens: 44% Spec: 88% AUC: 0.56	
	Ve primary tumor	Clinical response	Student t test	NS	…	
	Ve node	Clinical response	Student t test	NS	…	
	Vp primary tumor	Clinical response	Student t test	NS	…	
	Vp node	Clinical response	Student t test	NS	…	
Ng (39), 2013						58
Median 19.2 months (range 9-32.3 months)	Ktrans primary tumor	2-year control	Log rank test	<0.05	Hazard ratio: 0.34	
	Ktrans primary tumor	Local control	Univariable logistic regression	0.001		
	Ktrans primary tumor	Local control	Multivariable logistic regression → ROC analysis	<0.05	Sens: 71% Spec: 83%	
	Ktrans primary tumor	Local control	Student t test	0.01	Odds ratio: 0.06	
	Ve primary tumor	Local control	Univariable logistic regression	<0.05	…	
	Ve primary tumor	2-year control	Log rank test	NS	…	
	Ve primary tumor	Local control	Student t test	NS	…	
	Vp primary tumor	2-year control	Log rank test	NS	…	
	Vp primary tumor	Local control	Student t test	NS	…	
	Kep primary tumor	2-year control	Log rank test	NS	…	
	Kep primary tumor	Local control	Student t test	NS	…	
Agrawal (13), 2012						21
6 weeks	BF	Clinical response	Independent samples t-test	0.001	…	
	BF	T-stage	Independent samples t-test	<0.05	…	
	BF	Control at primary	Independent samples t-test	NS	…	
	BF	Control at node	Independent samples t-test	NS	…	
	BV	Clinical response	Independent samples t-test	0.001	…	
	BV	T-stage	Independent samples t-test	0.05	…	
	BV	Control at primary	Independent samples t-test	0.05	…	
	BV	Control at node	Independent samples t-test	NS	…	
Chikui (33), 2012						29
…	Ktrans pre CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ktrans post CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ktrans ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	Kep pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	Kep post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ve pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Ve post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	Ve ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	0.001	…	
	Vp pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Vp post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Vp ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AUGC pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AUGC post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
	AUGC ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.01	…	
Shukla-Dave (32), 2012[a]						74
Median 40 months (range 13-64 months)	Ktransskewness	PFS	Cox regression analyses	<0.05	…	
	Ktrans median	PFS	Cox regression analyses	NS	…	
	Ktransskewness	OS	Cox regression analyses	<0.05	…	
	Ktrans median	OS	Cox regression analyses	NS	…	
	Ktrans SD	OS	Cox regression analyses	NS	…	
	Stage IV only ktransskewness	PFS	Cox regression analyses	<0.001	…	
	Stage IV only ktransskewness	OS	Cox regression analyses	<0.001	…	
	Ve SD	PFS	Cox regression analyses	<0.05	…	
	Ve median	OS	Cox regression analyses	<0.01	…	
Wang (37), 2012						13
Median 19.6 months (range 14.1-36.4 months)	Subvolume with low BF pre-Rx	Local control	Mann–Whitney U test	NS	…	
	Subvolume with low BF 2wk	Local control	Mann–Whitney U test	0.05	…	
	Mean ΔBF	Local control	Mann–Whitney U test	NS	…	
	Subvolume with low BV pre-Rx	Local control	Mann–Whitney U test	<0.05	…	
	Subvolume with low BV 2wk	Local control	Mann–Whitney U test	0.01	…	
	Mean ΔBV	Local control	Mann–Whitney U test	<0.05	…	
	GTV w/ low BV pretreatment	Local control	ROC analysis	…	Sens: 75% Spec: 100% AUC (SD): 0.92 (0.09) 95%CI: 0.74-1.00 YI: 0.75	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	GTV w/ low BV 2wk	Local control	ROC analysis	…	Sens: 100% Spec: 77.8% AUC (SD): 0.94 (0.07) 95%CI: 0.81-1.00) YI: 0.78	
	Change BV of GTV 2wk	Local control	ROC analysis	…	Sens: 85% Spec: 83% AUC (SD): 0.90 (0.08)	
	Change BV of pre GTV	Local control	ROC analysis	…	Sens: 85% Spec: 76%	
Chawla (38), 2011						57
Median 30 months (range 13-48 months)	Ktrans	Disease-free survival	Hazard ratio	<0.05	3.81 (95%CI: 1.04-13.9)	
Chikui (18), 2011						23
54.4±5.9 days	Kep pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kep ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	TA ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AH pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	AH ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	<0.05	…	
	AH ΔCRT	Ohboshi and Shimosato criteria	Tukey's HSD	<0.05	Sens: 83% Spec: 71%	
	Kel pre-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kel post-CRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
	Kel ΔCRT	Clinical response (IUCC criteria)	Wilcoxon signed-rank test	NS	…	
Kim (35), 2010						33
Disease status at the end of CRT	Ktrans	Clinical response (CR/PR)	2-tailed t test with unequal variance	0.001	…	
	Ktrans	Clinical response (CR/PR)	ROC analysis	<0.01	Sens: 89% Spec: 63% AUC: 0.83 (95%CI: 0.68-0.98) YI: 0.51	

Appendix E *continued*

Follow-up	Predictor	Outcome	Test of significance	*P* value	Test outcome	n
	Ve	Clinical response (CR/PR)	2-tailed t test with unequal variance	NS	…	
	Ve	Clinical response (CR/PR)	ROC analysis	NS	Sens: 33% Spec: 88% AUC: 0.54 (95%CI: 0.31-0.78) YI: 0.21	
	τi	Clinical response (CR/PR)	2-tailed t test with unequal variance	NS	…	
	τi	Clinical response (CR/PR)	ROC analysis	NS	Sens: 56% Spec: 63% AUC: 0.51 (95%CI: 0.28-0.73) YI: 0.18	
Helbig (15), 2003						13
>9 months	A primary tumor	Clinical response (CR/PR)	…	…	CR: increase > SD/PR: increase	
	A node	Clinical response (CR/PR)	…	…	CR: increase SD/PR: decrease	
	K21 primary	Clinical response (CR/PR)	…	…	CR: decrease PR: increase	
	K21 node	Clinical response (CR/PR)	…	…	CR: decrease PR: increase	
Hoskin (27), 1999						13
11.5 months (range 6-20 months)	ΔTTP	Locoregional control	…	…	Sens: 86% Spec: 67%	
	Peak enhancement post-Rx	Locoregional control	Student t test	<0.01	…	
	Peak enhancement post-Rx	Locoregional control	ROC analysis	<0.01	Sens: 100% Spec: 80% AUC: 0.95 (95%CI: 0.83-1.00) YI: 0.80	
	ΔPeakenhancment	Locoregional control	…	…	Sens: 75% Spec: 60%	
	ΔMaximum slope	Locoregional control	…	…	Sens: 67% Spec: 33%	

[a] Not mentioned by the authors are Ktrans vs PFS; Kep median vs PFS; Kep skewness vs PFS; Kep SD vs PFS; Kep median vs OS; Kep skewness vs OS; Kep SD vs OS; Ve SD vs OS; Ve median vs PFS; Ve skewness vs PFS; Ve skewness vs OS

Abbreviations: A = signal intensity; AH = amplitude scaling constant; AUGC = area under the gadolinium concentration curve; AUC = area under the curve; BF = blood flow; BV = blood volume; CR = complete response; CRT = chemoradiotherapy; GTV = gross tumor volume; K21 = exchange rate constant; Ktrans = volume transfer constant between plasma and extracellular extravascular space; Kel = elimination of contrast medium from the central compartment; Kep = rate constant between extracellular extravascular space and blood plasma; NS = not significant; OS = overall survival; PFS = progression-free survival; PR = partial response; ROC = receiver operating curve; Rx = radiation therapy; SD = standard deviation; Sens = sensitivity; Spec = specificity; TA = time of arrival; Ti = intracellular water lifetime; TTP = time to peak; Ve = extravascular volume fraction; Vp = plasma volume fraction; YI = Youden index; 95%CI = 95% confidence interval;

Appendix F *QUADAS-2: Summary of the risk of bias and applicability concerns: review of authors' judgment about each domain of all included studies: '✓' indicates a low risk of bias; '?' an unclear risk and 'x' indicates a high risk of bias*

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Agrawal (13), 2012	✓	?	?	✓	✓	?	?
Ai (40), 2013	?	✓	✓	✓	✓	✓	✓
Ariyoshi (22), 2006	?	?	?	?	✓	✓	✓
Baba (28), 1997	✓	?	✓	✓	?	?	?
Bisdas (17), 2010	?	✓	?	✓	✓	✓	x
Bisdas (43), 2009	?	?	x	✓	✓	?	x
Chawla (34), 2013	?	x	x	x	✓	✓	?
Chawla (38), 2011	?	?	?	x	✓	✓	?
Chikui (18), 2011	?	✓	?	?	✓	✓	✓
Chikui (33), 2012	x	?	?	✓	✓	x	✓
Fischbein (24), 2003	?	✓	?	?	✓	✓	✓
Helbig (15), 2003	?	?	?	?	✓	✓	✓
Hietschold (36), 2004	?	?	?	?	✓	✓	✓
Hoskin (27), 1999	x	?	?	?	✓	✓	✓
Jansen (29), 2012	?	?	x	?	✓	✓	✓
Jansen (30), 2012	?	?	?	x	✓	✓	x
Jansen (31), 2010	?	?	?	✓	✓	✓	✓
Kim (35), 2010	?	?	?	?	✓	✓	✓
Konouchi (25), 2003	?	?	?	?	✓	✓	✓
Lee (16), 2012	x	?	?	?	✓	?	✓
Ng (39), 2013	?	x	?	?	✓	✓	✓
Razek (44), 2011	✓	✓	?	?	✓	✓	✓
Abdel Razek (45), 2011	?	?	?	?	✓	✓	✓
ShuklaxDave (32), 2012	?	✓	?	?	✓	✓	✓
Sumi (20), 2011	✓	?	?	?	✓	✓	✓
Sumi (41), 2014	x	?	?	✓	✓	✓	✓
Tomura (23), 2005	?	?	?	✓	✓	✓	✓
Tomura (26), 2002	?	?	?	?	✓	✓	✓
Unetsubo (21), 2009	?	?	✓	?	✓	✓	✓
Van Cann (14), 2008	?	x	?	?	✓	?	✓
Wang (37), 2012	?	?	?	?	✓	x	✓
Wendl (19), 2012	x	?	?	?	✓	✓	✓
Wu (42), 2004	x	?	?	?	?	✓	✓

Appendix G *QUIPS: Summary of the risk of bias on six domains: review of authors' judgment about each domain of the included prognostic studies: '✓' indicates a low risk of bias; '?' a moderate risk and 'x' indicates a high risk of bias*

	1. Study participation	2. Study attrition	3. Prognostic factor measurement	4. Outcome measurement	5. Study confounding	6. Statistical analysis and reporting
Agrawal (13), 2012	?	?	✓	x	x	?
Chawla (34), 2013	?	x	?	x	?	?
Chawla (38), 2011	✓	?	x	x	?	✓
Chikui (18), 2011	?	?	✓	✓	?	?
Chikui (33), 2012	?	x	?	✓	?	?
Helbig (15), 2003	?	✓	✓	?	?	?
Hoskin (27), 1999	x	✓	x	?	x	x
Kim (35), 2010	?	✓	?	x	?	✓
Ng (39), 2013	✓	?	?	x	x	✓
Shukla-Dave (32), 2012	?	✓	✓	?	✓	✓
Wang (37), 2012	x	✓	?	?	?	✓

APPENDICES CHAPTER 2.3

APPENDIX A *Medline and Embase search*

Index test (1)
("Magnetic Resonance Imaging"[Mesh] OR MRI[tiab] OR MRIs[tiab] OR NMR[tiab] OR NMRs[tiab] OR fMRI[tiab] OR "magnetic resonance"[tiab] OR MR[tiab] OR "nuclear magnetic"[tiab] OR (tomogra*[tiab] AND (MR[tiab] OR "proton spin"[tiab])) OR ("chemical shift"[tiab] OR "chemical shifts"[tiab]) OR "Magnetization Transfer"[tiab] OR "Magnetisation Transfer"[tiab] OR "resonance magnetic"[tiab] OR MR imag*[tiab] OR "functional magnetic resonance imaging"[tiab] OR DCE*[tiab] OR "dynamic contrast"[tiab] OR DSC[tiab] OR "Perfusion Imaging"[Mesh:NoExp] OR Perfusion Imag*[tiab] OR "Image Enhancement/methods"[MeSH Terms] OR diffusion-weighted imag*[tiab] OR "MR-Diffusion"[tiab] OR DWI [tiab] OR "Diffusion Magnetic Resonance Imaging"[Mesh] OR "Magnetic Resonance Spectroscopy" [Mesh] OR MRS [tiab] OR "Choline/metabolism"[MeSH Terms] OR "Contrast Media"[MeSH Terms] OR "Radiography"[MeSH Terms]) OR ((deoxyglucose[mh] OR deoxyglucose[tiab] OR desoxyglucose[tiab] OR deoxy-glucose[tiab] OR desoxy-glucose[tiab] OR deoxy-d-glucose[tiab] OR desoxy-d-glucose[tiab] OR 2deoxyglucose[tiab] OR 2deoxy-d-glucose[tiab] OR fluorodeoxyglucose[tiab] OR fluorodesoxyglucose[tiab] OR fludeoxyglucose[tiab] OR fluordeoxyglucose[tiab] OR fluordesoxyglucose[tiab] OR 18fluorodeoxyglucose[tiab] OR 18fluorodesoxyglucose[tiab] OR 18fluordeoxyglucose[tiab] OR fdg*[tiab] OR 18fdg*[tiab] OR 18f-dg*[tiab] OR 18f-fdg*[tiab] OR ((fluor[tiab] OR 2fluor*[tiab] OR fluoro[tiab] OR fluorodeoxy[tiab] OR fludeoxy[tiab] OR fluorine[tiab] OR 18f[tiab] OR 18flu*[tiab]) AND glucose[tiab])) AND (pet[tiab] OR pet/*[tiab] OR petscan*[tiab] OR tomography, emission-computed[mh] OR (emission[tiab] AND (tomograph [tiab] OR tomographs [tiab] OR tomographic*[tiab] OR tomography[tiab] OR tomographies[tiab] OR scan[tiab])))) OR ("Tomography, X-Ray Computed"[Mesh] OR (compute*[tiab] AND tomograph*[tiab]) OR ((CAT OR CT) AND (scan OR scans OR scanning OR X-ray OR X-rays)))) OR ("tomography, emission-computed, single-photon"[MeSH Terms] OR "single-photon emission-computed tomography"[tiab] OR "spect"[tiab])

Target location specified condition (2)
(("Head and Neck Neoplasms"[Mesh:NoExp] OR "Mouth Neoplasms"[Mesh] OR "Otorhinolaryngologic Neoplasms"[Mesh] OR "Tracheal Neoplasms"[Mesh] OR "Neoplasms, Squamous Cell"[Mesh] OR squamous[tiab] OR "carcinoma, squamous cell"[Mesh] OR HNSCC*[tiab] OR carcinom*[tiab] OR cancer*[tiab] OR malignan*[tiab] OR tumor*[tiab] OR tumour*[tiab] OR neoplas*[tiab] OR oncolog*[tiab] OR metasta*[tiab] OR carcinogen*[tiab] OR oncogen*[tiab] OR SCC[tiab]) AND ("Oropharyngeal"[Mesh] OR "Neck"[Mesh] OR neck[tiab] OR "Head"[Mesh] OR head[tiab] OR "Pharynx"[Mesh] OR pharynx[tiab] OR Otorhinolaryngological*[tiab] OR Otorhinolaryngeal*[tiab] OR "laryngopharynx" [tiab] OR pharyn*[tiab] OR oropharyn*[tiab] OR hypopharyn*[tiab] OR "Larynx"[Mesh] OR laryn*[tiab] OR laryngopharynx*[tiab] OR nasopharyn*[tiab] OR "Tongue"[Mesh] OR tongue[tiab] OR "Palatine Tonsil"[Mesh] OR "Palate"[Mesh] OR palat*[tiab] OR tonsil*[tiab] OR "Vocal Cords"[Mesh] OR vocal cord*[tiab] OR "Lip"[Mesh] OR lip[tiab] OR lips[tiab] OR "Mouth"[Mesh] OR mouth[tiab]))

Prognostic filter (3)
("sensitivity and specificity"[MeSH Terms] OR sensitiv*[tiab] OR specifici*[tiab] OR diagnos*[tiab] OR diagnos*[tiab] OR pre-test[tiab] OR pretest[tiab] OR post-test[tiab] OR posttest[tiab] OR "Probability"[Mesh] OR probabilit*[tiab] OR predict*[tiab] OR value[tiab] OR "likelihood ratio"[tiab] OR prognos*[tiab]) OR misdiagnos*[tiab] OR "diagnostic Errors"[Mesh] OR correlation [tiab] OR detection[tiab] OR "predictive value of tests"[Mesh] OR "ROC Curve"[Mesh] OR sroc[tiab] OR receiver operating characteristic*[tiab] OR receiver operator characteristic*[tiab] OR (diagnos*[tiab] AND (accuracy[tiab] OR error*[tiab] OR efficacy[tiab] OR "Follow-Up Studies"[Mesh] OR "Prospective Studies"[MeSH Terms] OR "Kaplan-Meier Estimate"[MeSH Terms] OR response[tiab] OR responsive[tiab] OR responding[tiab] OR (("Follow-Up"[tiab] OR "Disease-Free Survival"[Mesh] OR "Survival Analysis"[MeSH Terms] OR "Treatment Failure"[MeSH Terms] OR "Treatment Outcome"[MeSH Terms] OR "Survival Rate"[Mesh] OR survival*[tiab] OR "disease free"[tiab] OR "Recurrence"[Mesh] OR "Neoplasm Recurrence, Local"[Mesh] OR recurren*[tiab] OR relapse*[tiab] OR residual*[tiab] OR detection[tiab] OR intratherap*[tiab] OR intra-therap*[tiab] OR during therap*[tiab] OR post-therap*[tiab] OR posttherap*[tiab] OR "work-up"[tiab] OR "Early Detection of Cancer"[Mesh] OR "Outcome Assessment"[MeSH] OR "Neoplasm, Residual"[Mesh] OR "Neoplasm Recurrence, Local"[Mesh]) AND (preliminary[tiab] OR early[tiab] OR "short term"[tiab] OR week*[tiab] OR day* [tiab] OR month*[tiab]))

Final Medline search

1 AND 2 AND 3

Final Embase search

1 AND 2 AND 3

The Embase search was the same as the Medline search. but with only 'text words' and excluding Medline studies. '3' was not included however. because the Mesh-term is covered by '4' and 5'

Appendix B1 *Baseline characteristics and technical specifics CT studies*

Author, year	Imaging technique	Design	Diag Prog	Eligibl patients	Incl pts	% Male	Median Age	HPV +/-	NAC (pt)	RT (pt)	Total dose (Gy)	CRT (pt)	Chemotherapy	Mean follow-up Mo (SD)	CR	LRC
Truong, 2011 (36)	CTP	PS	P	15	15	80	57	-	0	15	70	0	cisplatin 7wks 30mg/m2 / 1 wk cetuximab 250mg/m2	28	-	13(87%)
Abramyuk, 2015 (37)	CTP	PS	P	15	15	-	51	-	0	15	72	1	Cisplatin 3wks 100mg/m2 / Mitomycin	24	-	14(93%)
Rana, 2015 (93)	CTP	PS	D	24	24	95.8	57.9 (mean)	-	0	-	66	24	Cisplatin 3wks 100mg/m2	-	12(50%)	-
Ursino, 2016 (38)	CTP	PS	D	27	25	-	-	-	0	25	66	25	Cisplatin 3wks 100mg/m2	3	16(64%)	-

Abbreviations: 5-FU = 5-Fluorouracil; D = Diagnostic study, evaluating treatment response; EGFR = epidermal growth factor receptor-inhibitor immunotherapy (Cetuximab); P = Prognostic, evaluating long-term outcome; Plat = platinum based chemotherapy (Cisplatin etc.); PS = Prospective study; RS = Retrospective study; RT = radiotherapy; Tax = Taxus based chemotherapy (Paclitaxel etc.); VEGF = vascular endothelial growth factor inhibitor receptor (Bevacizumab)

Appendix B2 *Baseline characteristics MRI studies*

Author, year	Imaging	Study design	Prog/ Diag	Eligible patients	Incl. patients	% Male	HPV/ P16	Mean/ median age (range)	Mean follow-up (range)	Radio-therapy (Gy)	Treatment regimen	Reference standard	CR	LRC	OS
Bhatia, 2010(39)	T1/T2	RS	P	89	69	91	...	59 (47-75)	41 (...)	72	Plat	HP	-	46 (67%)	-
King, 2010(40)	MRS	PS	P/D	60	47	92	...	57 (43-74)	39 (13-64)	...	...	HP	-	32 (68%)	27 (57%)
Baer, 2015(41)	DCE	PS	P	24	10	80	...	59 (43-83)	64.9 (33-75)	70	Plat+5-FU/EGFR	HP	-	-	5 (50%)
Cao, 2008(16)	DCE	PS	P	14	14	86	...	58 (39-83)	9.7 (5-27)	70	Plat/Tax+EGFR	HP	-	9 (64%)	10 (71%)
Dirix, 2009 (42)	DCE/DWI	PS	P/D	15	15	87	...	57 (46-61)	31 (6-56)	72	Plat	Physical exam/ CT	-	8 (53%)	13 (87%)
Jansen, 2016 (43)	DCE	RS	P	19	19	84	...	57 (41-74)	32 (15-76)	70	Plat/EGFR/VEGF	HP	-	17 (89%)	-
Wang, 2012(44)	DCE	PS	P/D	14	14	86	...	53 (...)	20(...)	70	Plat/EGFR/VEGF	HP	-	9 (64%)	10 (71%)
Yoo, 2012 (45)	DCE	PS	P	29	28	79	15/27	53 (39-76)	46(25-63)	70	Plat	HP	27 (96%)	23 (82%)	24 (86%)
Wong 2017 (59)	DWI	PS	P	40	35	100		61 (34-69)	14 (5-33)	65	Plat	HP	-	27 (77%)	31 (89%)
Galban, 2009 (46)	DWI	PS	P	23	15	0	...	...	6 (...)	70	Plat/Tax/5-FU	HP	-	12 (80%)	13 (87%)
Hoang, 2014 (47)	DWI	PS	P	18	16	100	...	53 (...)	31 (27-36)	70	Plat	HP	-	14 (88%)	-
Kim, 2009 (48)	DWI	PS	P/D	40	33	79	...	61 (50-72)	12 (...)	70	Plat+EGFR	HP/clinical exam	26 (79%)	-	-
King, 2010 (84)	DWI	PS	P/D	64	50	90	...	58 (45-73)	27 (5-60)	72	Plat	HP	39 (78%)	30 (60%)	-
King, 2013 (50)	DWI	RS	P/D	56	37	87	...	57 (45-71)	44 (24-78)	...	...	HP	-	21 (57%)	-
Lambrecht, 2014 (86)	DWI	PS	P	20	20	95	...	55 (39-78)	57 (7-76)	70	RT	HP	-	11 (55%)	-
Martins, 2015(52)	DWI	PS	P	23	23	91	...	56 (46-69)	...	...	...	HP	2 (11%)	-	-
Matoba, 2014(87)	DWI	RS	P/D	40	35	86	...	66 (33-79)	31 (7-56)	60-70	Plat+5-FU	Clinical exam / MRI	-	21 (60%)	32 (91%)
Scalco, 2016 (53)	D WI	PS	P/D	30	30	90	...	58 (28-82)	27 (7-50)	70	Plat	HP	-	19 (63%)	-
Schouten, 2014(54)	DWI	PS	P	8	8	75	...	61 (51-68)	38 (17-60)	70	Plat+EGFR	DWI/ FDG-PET/ clinical exam	-	5 (29%)	5 (29%)
Tyagi, 2016 (55)	DWI	PS	P	10	10	80	...	58 (51-61)	...	70	Plat/Plat+Tax	HP	-	-	-
Vandecaveye, 2010(56)	DWI	PS	P/D	31	30	94	...	53 (38-66)	24 (...)	72	...	HP	-	15 (50%)	21 (70%)
Ding, 2015 (57)	IVIM	PS	P/D	40	31	94	31/31	57 (44-78)	...	70	Plat/EGFR/Tax	HP	15 (52%)	-	-
Paudyal, 2016(58)	IVIM	PS	P	40	34	91	30/34	57 (30-82)	3-6 (...)	70	Plat	HP	40 (88%)	-	-
Marzi 2017 (92)	IVIM	PS	R	51	34	88	33	55 (28-82)	27 (12-50)	70	Plat	HP	-	24 (71%)	-

Appendix B3 *Baseline characteristics PET*

Author. year	Imaging technique	Design PS / RS	Diag Prog	Eligible pt	Incl pts	% Male	Age Median	HPV +/-	NAC (pt)	RT only (pt)	Total dose (Gy)	CRT (pt)	Chemotherapy	Mean follow-up in months (SD)	Ref stand	CR	LRC	OS
Brun 1997 (85)	FDG	PS	P / D	19	17	100	52 (21-73)	NS	0	12	66	5	NS	NS	FNAC	13(76%)	3(76%)	12(71%)
Brun 2002 (7)	FDG	PS	P	48	47	85.1	55 (22-85)	NS	10	0	66-70	NS	NS	40 (14-83)	HP	36(80%)	5(78%)	30(65%)
Dirix 2009 (42)	FDG	PS	P	15	15	86.7	57 (46-61)	NS	0	0	72	0	Plat 3 wk 100	31 (6-56)	NS	13(87%)	0(67%)	-
Farrag 2010 (62)	FDG	PS	P	43	43	86.0	56 (34-80)	NS	0	27	66-71	16	Plat 3 wk 100	13 (3-35)	HP	-	22(52%)	28 (66%)
Ceulemans 2011 (64)	FDG	PS	P	40	40	82.5	56 (34-80)	NS	0	24	71	16	Plat 3 wk 100	26 (7-50)	HP	-	24(80%)	30(75%)
Hentschel 2011 (65)	FDG	PS	P	43	37	89.1	55 (37-70)	NS	0	0	72	37	Plat 1 wk 30	26 (8-50)	HP	19(51%)	20(54%)	20(54%)
Castaldi 2012 (66)	FDG	PS	P	26	26	88.5	63 (41-79)	NS	22c	0	70-78	26	Plat 1wk 30 / 3wk 100	29 (3-56)	HP	-	17(65%)	19(70%)
Hoshikawa 2013 (68)	FDG	PS	P	30	30	96.7	69 (40-83)	NS	0	6	60-70	24	Plat 3wk 70 + 5-FU 1000	32 (12-75)	HP	-	27(90%)	25(83%)
Schouten 2014 (54)	FDG	PS	P	8	8	75.0	61* (51-68)	NS	0	0	70	8	NS	38 (17-60)	HP	-	6(75%)	7(88%)
Chen 2014 (73)	FDG	PS	P	51	51	94.1	52 (29-69)	NS	0	7	68-74	41	Plat 80-100	23 (7-53)	HP	-	27(53%)	29(57%)
Min 2015 (69)	FDG	RS	P	72	72	84.7	60 (39-75)	NS	15c	0	60-70	57	Plat (42). Carb (12).Cet (15)	25 (6-70)	HP	-	51(71%)	56(78%)
Min 2016 (88)	FDG	RS	P	100	100	84.0	61 (37-81)	NS	17c	15	60-70	68	Plat (53). Carb (12). Cet (17)	20 (4-70)	HP	-	75(75%)	80(80%)
Min 2016 (71)	FDG	RS	P	72	69	88.4	65 (39-81)	NS	15	15	60-70	46	Plat (33). Carb (10. Cet (13)		HP	29(57%)	48(70%)	51(74%)
Lin 2016 (74)	FDG	RS	P	75	75	89.3	89 (39-80)	NS	17	0	60-70	75	Plat (50). Carb (14). Cet (11)	28 (9-70)	HP	54(72%)	54(72%)	55(73%)
Pollom 2016 (75)	FDG	RS	P	74	74	NS	60 (27-83)	NS	11	0	74	74	Plat (34). Carb (9). Cet (31)	18 (3-67)	HP	-	65(88%)	72(97%)
Thorwarth 2007 (76)	FMISO	NS	P	15	10	86.7	57(46-69)	NS	0	10	70	0	NA	NS	HP	-	8(53%)	-
Zips 2012 (77)	FMISO	PS	P	28	25	88.0	59 (45-76)	NS	0	0	69-72	25	5-FU 5 days 1000	12 (2-58)	HP	-	19(57%)	9(32%)
Wiedenmann 2015 (30)	FMISO	PS	P	16	16	NS	NS	NS	0	0	70	16	Plat 3 wk 100	44 (1-51)	HP	15(94%)	11(69%)	8(50%)
Zschaeck 2015 (78)	FMISO	PS	P	25	12	88.0	54 (45-64)	NS	0	0	72	25	Plat (21) wk 30	39	HP	-	5(42%)	-
Lee 2016 (9)	FMISO	PS	P	33	33	NS	58 (26-79)	33	0	0	70	33	Plat (28) Carb+5-FU (2)+pac(2)	32	HP	-	32(97%)	33(100%)
Grkovski 2017 (80)	FMISO	RS	P	128	72	92	60 (49-71)	47	0	0	70	72	Plat (55) Carb+5FU(3) Cet(7)	26 (5-73)	HP	-	66(92%)	68(94%)

Appendix C1 *Technical details CT*

Author. year	Imaging technique	Design PS/RS	Scanner	Voltage (kV)	Ampere (mA)	Thickness (mm)
Truong, 2011(36)	CTP	PS	-	120	50	-
Abramyuk, 2015(37)	CTP	PS	-	120	150	1.5
Rana, 2015(93)	CTP	PS	-	-	-	-
Ursino, 2016 (38)	CTP	PS	Discovery CT750	100	20	2.5

Appendix C2 *Technical details MRI*

Author, year	Imaging	Tesla	Coils	Imaging system	Sequence	TR/TE(ms)	Matrix	FOV (mm)	Slices	Gap (mm)	Slice thickness (mm)	Pharmacokinetic model	b-values	IVIM fit	IVIM parameter
Bhatia, 2010	T1/T2	1.5	HN	Intera NT, Philips	...	2500/100	512*512	...	...	0	4	...	...	...	...
King, 2010	MRS	1.5	Circular	Intera NT, Philips	...	...	...	...	...	...	...	...	...	...	...
Baer, 2015	DCE	3.0	...	Achieva, Philips	3D-GRE	5,1/1,1	...	...	...	...	2	Mod Tofts	...	...	...
Cao, 2008	DCE	...	...	NS	...	...	...	...	...	...	...	Tofts	...	...	...
Dirix, 2009	DCE	1.5	2-Ch pa	Biograph, Siemens	3D fGRE	DCE 4,3/1,6	...	200*250	...	...	...	...	...	...	...
	DWI				(...)-EPI	DWI	...	...	...	...	...		0, 50, 100, 500, 750, 1000	...	...
Jansen, 2016	DCE	1.5	8-Ch NV	General Electric	3D fGRE	7400/84	...	180-200	3-8	0	5-6	Tofts	...	...	...
Wang, 2012	DCE	3.0	...	Phllips	3D GRE	7,8/1,9	...	...	...	...	2	Tofts	...	...	...
Yoo, 2012	DCE	1.5	...	Signa Excite, GE	3D fGRE	5,1/1,1	238*238	240*240	48		2	Tofts	...	...	...
Wong 2017	DCE	1.5	Flex&spline	Aera, Siemens	3D fGRE	5,4/2,0	...	...	...	...	...	Ext Kety	...	...	...
	DWI	1.5	Flex&spline		SE-EPI	7.2/2.4	96*96	...	...	...	...	...	50,400,800	...	...
Galban, 2009	DWI	3.0	8-Ch HN pa	Achieva, Philips	SS-EPI	13400/61	205*205	270*270	24	...	4	...	0, 800	...	...
Hoang, 2014	DWI	1.5	4-Ch flex pa	Signa Excite, GE	(SE)-EPI	1.000/64	256*256	240*240	...	...	5	...	250, 500, 750, 1000	...	...
Kim, 2009	DWI	1.5/ 3.0	NV		SS-EPI	4000/89	...	260*260	...	...	5	...	0, 500, 1000	...	...
King, 2010	DWI	1.5	...	Intera NT, Philips	SS-EPI	2000/75	230*250	230*250	...	...	4	...	0, 100, 200, 300, 400, 500	...	...
King, 2013	DWI	1.5	...	Intera NT, Philips	SS-EPI	2000/75	112*112	...	22	...	4	...	0, 100, 200, 300, 400, 500	...	...
Lambrecht, 2014	DWI	1.5	2-Ch HNS pa	Sonata, Siemens	SS-EPI	7100/84	104*128	200*250	44	0.4	4	...	0, 50, 100, 500, 750, 1000	...	...
Martins, 2015	DWI	1.5	Cervical	Signa Excite, GE	SS-EPI	2000/75	256*256	230*230	...	...	4	...	0, 1000	...	...
Matoba, 2014	DWI	1.5	NV	Avanto, Siemens	SS-EPI	4000/68/180	512*256	250*250	...	3	6	...	0, 90, 800	...	...
Scalco, 2016	DWI	1.5	16Ch HNS	Optima, GE	SS-(SE)-EPI	4500/77	128*128	260-280	...	5	5	...	0, 500, 800	...	...
Schouten, 2014	DWI	1.5	3-Ch HNS pa	Sonata, Siemens	HASTE	5000/105	...	...	22	0.5	...	...	0, 500, 1000	...	...
					EPI	900/110						...	0, 800	...	...
Tyagi, 2016	DWI	3.0	16 ante 44 post	Ingenia, Philips	SS-EPI	5000/65	128*128	...	300	...	5	...	0, 50, 100, 500, 750, 1000	...	...
Vandecaveye, 2010	DWI	1.5	2-Ch pa	Sonata, Siemens	SS-EPI	7100/84	104*128	200*250	44	...	4	...	0, 20, 40, 60, 80, 100, 120,	...	...
Ding, 2015	IVIM	3.0	6-element pa	Discovery, GE	SS-(SE)-EPI	3600/80	128*128	...	30	...	...	...	150, 200, 400, 600, 800	NLLS	D*, D, f
Paudyal, 2016	IVIM	3.0	NV	Ingenia, Philips	SS-EPI	4000/80-100	128*128	200-240		...	5	...	0, 20, 50, 80, 200, 300, 500, 800, 1500, 2000		
Marzi 2017	IVIM	1.5	HNS	Optima MR 450	SS-EPI	4500/77	128*128	260-280	8-10	5	4	...	0,25,50,75,100,150,300,500,800	NLLS	D*, D, f

Appendix C3 *Technical details PET*

Author. year	Imaging system	PET-tracer	Fasted (hr)	Glucosis mean (range)	MBq injected	Time per bed position (min)	Time scan after injection (min)	Matrix	FO\	Voxel size	CT kV	mAs
Brun 1997	PC384-7B	FDG	6	7-38	37-74		52-66	64*64 128*128		3.4x3.4 2.2x2.2		
Lowe 1997	ECAT	FDG	0		370	10	50					
Brun 2002	Scanditronix	FDG	4		79 (51-97)			129x129	26.(	2.2x2.2		
Dirix 2009	Biograph	FDG					120-240					
Farrag 2010	ECAT	FDG	6									
Ceulemans 2011	ECAT	FDG	6	(6.7-14.0)	281-581	3	60					
Hentschel 2011	Biograph Duo	FDG			259-341	3	50-70				120	100
Castaldi 2012	Gemini TF	FDG										
Hoshikawa 2013	ECAT	FLT/FDG	1		3.5	2-3	60		15.!	4.6	120	210
Schouten 2014	ECAT / Gemini TF	FDG	6	6.5(4.3-11.2)	186-367	5 / 2	45-75	128x128		5x5	120	50
Chen 2014	Discoverty	FDG	4		370		60					
Min 2015	Gemini TF	FDG	4	4.5-6.9	4.1/5.18	1.5-2.4	59	144*144 256*255			120- 139	30- 39
Min 2016	Gemini TF	FDG	4	4.5-6.9	4.1/5.18	1.5-2.4	59	144*144 256*255			120- 139	30- 39
Min 2016	Gemini TF	FDG	4	10	4.1/5.18	1.5-2.4	60	144*144 256*255			120- 139	30- 39
Lin 2016	Gemini TF / Discoverty	FDG	4		4.1/5.18							
Pollom 2016	NS	FDG	8		10	3-5	45-60			1.3x1.3x1.3		
Wong 2017	Gemini TF	FDG	6	<10		3	60					50
Thorwarth 2007	Advance	MISO	6		400		120-240					
Zips 2012	Biograph Duo	MISO			250-300	15	120-240					
Wiedermann 2015	NS	MISO	0		400		150			4.3*4.3*3.4		
Zschaik 2015	Biograph Duo	MISO	4		250-300		120 / 240				120	100
Lee 2016	NS	MISO			269-370		90-180					
Grkovski 2017	NS	MISO			376-404		87-174					
Lock 2017	Biograph Duo	MISO			250-300	12						
Troost 2010	Biograph Duo	FLT			250	7	60				130	80
Kishino 2012	ECAT	FLT										
Hoeben 2013	Biograph Duo	FLT	5		250	7	60			5.31x5.31x3.38		
Nyfflot 2015	Discoverty	FLT	4				120-180			5		
Zegers 2014	Biograph Duo	FHX4 PET			294-462		90/180/240					

Appendix D1 *Baseline TNM stage of included CT-studies*

Author, year	T1 (Tx)	T2	T3	T4	N0 (Nx)	N1	N2a (N2x)	N2b	N2c	N3	M1	AJCC
Truong, 2011	1	7	3	4	5	1	1	3	4	1	0	3 / 4
Abramyuk, 2015	-	-	-	-	-	-	-	-	-	-	-	-
Rana, 2015	-	-	-	-	-	-	-	-	-	-	-	3 (19) / 4 (5)
Ursino, 2016	-	-	-	-	-	-	-	-	-	-	-	-

Appendix D2 *Baseline TNM stage of included MRI-studies*

Author, year	MRI Sequence	Tx	T1	T2	T3	T4	T4a	T4b	Nx	N0	N1	N2x	N2a	N2b	N2c	N3	M	AJCC
Bhatia, 2010	T1/T2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/IVa
King, 2010	MRS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/IVa
Baer, 2015	DCE	0	1	5	1	3	0	0	0	2	0	0	1	4	0	2	0	-
Cao, 2008	DCE	1	2	6	1	4	0	0	0	2	0	0	1	5	3	3	0	-
Dirix, 2009	DCE/DWI	0	2	3	3	7	0	0	0	1	4	0	0	4	6	0	0	-
Jansen, 2016	DCE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/IV
Wang, 2012	DCE	1	2	6	1	4	0	0	0	2	0	0	1	5	3	3	0	-
Yoo, 2012	DCE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/IV
Wong 2017	DCE/DWI	-	20	-	15	-	-	-	-	12	-	23	-	-	-	-	0	-
Galban, 2009	DWI	1	2	7	2	3	0	0	0	1	0	2	2	5	1	4	0	III/IV
Hoang, 2014	DWI	0	6	8	1	1	0	0	0	1	1	0	1	10	3	0	0	III/IV
Kim, 2009	DWI	7	1	9	5	6	3	1	0	0	2	3	5	14	7	2	0	-
King, 2010	DWI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/IV
King, 2013	DWI	0	9	0	8	20	0	0	0	0	0	0	0	0	0	0	0	III/IV
Lambrecht, 2014	DWI	0	0	0	10	10	0	0	0	0	1	0	1	8	10	0	0	IV
Martins, 2015	DWI	1	1	3	13	5	0	0	0	0	7	3	0	5	3	5	0	III/IV
Matoba, 2014	DWI	0	1	14	8	12	0	0	0	3	7	20	0	0	0	5	0	-
Scalco, 2016	DWI	1	6	11	5	7	0	0	0	0	6	0	4	6	8	6	0	-
Schouten, 2014	DWI	0	0	3	3	2	0	0	0	0	1	0	1	2	3	1	0	III/IVa
Tyagi, 2016	DWI	3	0	4	3	0	0	0	0	0	0	5	0	0	4	1	0	III/IVa
Vandecaveye, 2010	DWI	0	2	5	12	12	0	0	0	3	8	0	1	8	9	1	0	-
Ding, 2015	IVIM	0	1	14	8	8	0	0	0	1	3	0	2	13	12	0	0	III/IVa
Paudyal, 2016	IVIM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	III/Iva
Marzi 2017	IVIM	-	4	13	6	7	-	-	-	-	6	23	-	-	-	5	-	-

Appendix D3 *Baseline TNM stage of included PET-studies*

Author, year	T1 (Tx)	T2	T3	T4	N0 (Nx)	N1	N2a(x)	N2b	N2c	N3	M1
Brun, 1997	4	6	3	4	2	7	0(5)	0	0	3	0
Lowe, 1997	3(2)	5	5	12	4(1)	2	4(1)	3	4	7	1
Brun, 2002	27	0	20	0	0(24)	0	0(23)	0	0	0	0
Dirix, 2009	2	3	3	7	1	4	0	4	6	0	0
Farrag, 2010	9	15	13	6	17	2	24	0	0	0	0
Ceulemans, 2011	-	-	-	-	-	-	-	-	-	-	-
Hentschel, 2011	3(2)	11	11	10	7(1)	3	0(24)	0	0	2	0
Castaldi, 2012	3	9	6	8	3	7	2(3)	3	8	0	0
Hoshikawa, 2013	2	19	7	2	10	9	2	7	2	0	0
Schouten, 2014	0	3	3	2	0	1	1	2	3	1	0
Chen, 2014	-	-	-	-	-	-	-	-	-	-	-
Min, 2015	6	25	31	10	9	11	0(47)	0	0	5	0
Min, 2016	8	40	38	14	21	14	60	0	0	5	0
Min, 2016	4	28	27	10	16	10	0(39)	0	0	4	0
Lin, 2016	6	31	27	11	0	11	38	12	9	5	0
Pollom, 2016	10	25	18	21	2	3	5	38	19	7	0
Wong, 2017	20	-	15	-	12	-	23	-	-	-	0
Thorwarth, 2007	-	-	-	-	-	-	-	-	-	-	-
Zips, 2012	0	0	9	16	5	2	0	3	12	2	0
Wiedermann, 2015	-	-	-	-	-	-	-	-	-	-	-
Zschaik, 2015	0	0	4	8	1	1	0(1)	2	7	0	0
Lee, 2016	6	17	7	3	0	1	30	-	-	2	0
Grkovski, 2017	16	36	10	8	0	7	7	31	25	2	0
Lock, 2017	0	0	9	16	5	2	0	3	13	2	0
Troost, 2010	1	7	2	0	5	2	0	3	0	0	0
Kishino, 2012	1	17	7	3	7	9	2	8	2	0	0
Hoeben, 2013	1	25	17	7	22	6	0	11	11	0	0
Nyfflot, 2015	0	4	2	4	0	0	0	5	5	0	0
Zegers, 2014	1	6	11	2	9	3	1	7	0	0	0

Appendix E1 *Location primary tumor in CT studies*

Author, year	Oropharynx	Hypofarynx	Larynx	Oral Cavity	Tonsils	Nasopharynx
Truong, 2011	2	1	3	1	3	3
Abramyuk, 2015	-	-	-	-	-	-
Rana, 2015	-	-	-	-	-	-
Ursino, 2016	7	2	5	5	-	6

Appendix E2 *Primary tumor location in MRI studies*

Author, year	Imaging	Nasopharynx	Oral cavity	Oropharynx	Hypopharynx	Larynx	Piriform sinus	Tonsil	Unknown	Nasal cavity	Paranasal sinus	Esophagus	Nodes
Bhatia, 2010	T1/T2	3	24	0	24	18	0	0	0	0	0	0	0
King, 2010	MRS	0	22	0	0	0	0	0	0	3	0	2	25
Baer, 2015	DCE	1	1	2	1	1	0	4	0	0	0	0	0
Cao, 2008	DCE	1	2	0	1	1	1	7	1	0	0	0	0
Dirix, 2009	DCE/DWI	0	1	6	3	5	0	0	0	0	0	0	0
Jansen, 2016	DCE	-	-	19	-	-	-	-	-	-	-	-	-
Wang, 2012	DCE	1	0	3	1	1	1	6	1	0	0	0	0
Yoo, 2012	DCE	1	2	12	-	1	0	11	2	0	0	0	0
Wong 2017	DCE/DWI	0	0	29	6	0	0	0	0	0	0	0	0
Galban, 2009	DWI	1	1	3	-	0	1	8	1	0	0	0	0
Hoang, 2014	DWI	1	1	4	-	0	0	10	0	0	0	0	0
Kim, 2009	DWI	-	-	12	-	6	0	10	5	0	0	0	0
King, 2010	DWI	0	9	0	13	4	0	0	0	2	1	1	21
King, 2013	DWI	0	14	0	20	0	0	0	0	2	1	0	0
Lambrecht, 2014	DWI	0	0	10	7	3	0	0	0	0	0	0	0
Martins, 2015	DWI	0	0	16	2	4	0	0	1	0	0	0	52
Matoba, 2014	DWI	0	4	9	9	13	0	0	0	0	0	0	0
Scalco, 2016	DWI	11	0	12	5	1	0	0	1	0	0	0	0
Schouten, 2014	DWI	0	0	2	0	0	1	5	0	0	0	0	0
Tyagi, 2016	DWI	0	4	0	0	0	0	3	3	0	0	0	15
Vandecaveye, 2010	DWI	0	0	6	1	9	5	5	0	0	0	0	0
Ding, 2015	IVIM	0	0	15	0	0	0	16	0	0	0	0	0
Paudyal, 2016	IVIM	0	0	32	0	0	0	0	2	0	0	0	0
Marzi, 2017	IVIM	13	0	14	6	0	0	0	1	0	0	0	0

Appendix E3 *Location primary tumor in PET studies*

Author, year	Oropharynx	Hypofarynx	Larynx	Oral Cavity	Tonsils	Nasoparynx	Other
Brun, 1997	2	1	2	5	6	1	-
Lowe, 1997	13	6	3	1	-	2	2
Brun, 2002	21	4	6	9	-	6	1
Dirix, 2009	6	3	5	1	-	-	-
Farrag, 2010	12	15	3	9	-	4	-
Ceulemans, 2011	12	13	3	12	-	4	-
Hentschel, 2011	17	13	3	4	-	-	-
Kikuchi, 2011	6	7	-	3	-	-	-
Castaldi, 2012	12	2	9	-	-	3	-
Hoshikawa, 2013	10	10	10	-	-	-	-
Kikuchi, 2013	27	8	14	1	-	-	-
Schouten, 2014	2	1	-	-	5	-	-
Chen, 2014	20	21	-	-	-	10	-
Min, 2015	47	6	16	3	-	-	-
Gavid, 2015	6	15	-	-	-	-	-
Min, 2016	64	6	26	4	-	-	-
Min, 2016	41	5	20	3	-	-	-
Lin, 2016	56	5	11	3	-	-	-
Pollom, 2016	74	-	-	-	-	-	-
Wong, 2017	29	6	-	-	-	-	-
Thorwarth, 2007	-	-	-	-	-	-	-
Zips, 2012	9	7	3	6	-	-	-
Wiedermann, 2015	16	16	16	16	-	-	-
Zschaik, 2015	3	3	-	6	-	-	-
Lee, 2016	16	-	-	-	17	-	-
Grkovski, 2017	39	1	2	-	28	-	-
Lock, 2017	-	-	-	-	-	-	-
Troost, 2010	2	-	-	1	7	-	-
Kishino, 2012	9	9	8	1	-	28	-
Hoeben, 2013	27	8	14	1	-	-	-
Nyfflot, 2015	6	-	-	-	4	-	-
Zegers, 2014	7	5	8	-	-	-	-

Appendix F1 *Early response results in CT studies*

Author, year	Time Imaging	Parameter	Statistical Test	Outcome	Sen (%)	Spec (%)	PPV	NPV	HR	CI	P-value
Rana, 2015	4	BV40Gy	Wilson Score	CR							0.461
	4	BV40Gy	Wilson Score	CR							0.351
	4	BV40Gy	Wilson Score	CR							0.860
	4	BV40Gy	Wilson Score	CR							0.456
	0	BF	Wilson Score	CR							0.006
	0	BV	Wilson Score	CR							0.128
	0	MTT	Wilson Score	CR							0.523
	0	PS	Wilson Score	CR							0.386
	6	BF	Wilson Score	CR							0.974
	6	BV	Wilson Score	CR							0.356
	6	MTT	Wilson Score	CR							0.034
	6	PS	Wilson Score	CR							0.272
	0-6	Delta BF	Wilson Score	CR							0.002
	0-6	Delta BV	Wilson Score	CR							0.005
	4	Delta MTT	Wilson Score	CR							0.024
	0-4	Delta PS	Wilson Score	CR							0.233
	0-6	Delta BF	Wilson Score	CR							0.050
	0-6	Delta PS	Wilson Score	CR							0.032
	0-6	BF	Wilson Score	CR				1.38-18.2	5.0	-	0.004
	0-6	BV	Wilson Score	CR				0.73-3.76	1.65	-	0.413
	0-6	MTT	Wilson Score	CR				0.24–1.44	0.59	-	0.413
	0-6	PS	Wilson Score	CR				0.31–1.62	0.71	-	0.683
	0	High BF	Wilson Score	CR	83.3 (55.2-95.3)	83.3 (55.2-95.3)	83.3 (55.2-95.3)	83.3 (55.2-95.3)	5.0	1.8-13.9	
	0	High BV	Wilson Score	CR	58.3 (32.0-80.7)	62.5 (39.1-86.2)	63.6 (35.4-84.8)	61.5 (35.5-82.3)	1.8	0.9–3.5	
	0	Low MTT	Wilson Score	CR	66.7 (39.1-86.2)	62.5 (32.0-80.7)	61.5 (35.5-82.3)	63.6 (35.4-84.8)	1.6	1.0–2.7	
	0	Low PS	Wilson Score	CR	58.3 (32.0-80.7)	58.3 (32.0-80.7)	58.3 (32.0-80.7)	58.3 (32.0-80.7)	1.4	0.8–2.5	
	0	High BF & high BV	Wilson Score	CR	58.3 (32.0-80.7)	70.8 (55.2-95.3)	77.8 (45.3-93.7)	66.7 (41.7-84.8)	3.5	1.1–11.4	
	0	High BF & low MTT	Wilson Score	CR	58.3 (32.0-80.7)	75.0 (64.6-98.5)	87.5 (52.9-97.8)	68.8 (44.4-85.8)	7	0.8–60.7	
	0	High BF & low PS	Wilson Score	CR	50.0 (25.4-74.6)	75.0 (75.8-100)	100 (61.0-100.0)	66.7 (43.8-83.7)	-	-	

Appendix F1 *continued*

Author, year	Time Imaging	Parameter	Statistical Test	Outcome	Sen (%)	Spec (%)	PPV	NPV	HR	CI	P-value
Ursino, 2016	0-3	deltaBV		CR							0.001
	3-12	deltaBV		CR							0.066
	0-3	deltaBF		CR							0.002
	3-12	deltaBF		CR							0.061
	0-3	deltaMTT		CR							0.722
	3-12	deltaMTT		CR							0.036
	0-3	deltaPS		CR							0.004
	3-12	deltaPS		CR							0.089
	0-3	deltaBV	Univariate	CR							0.030
	0-3	deltaBF	Univariate	CR							0.042
	0-3	deltaMTT	Univariate	CR							0.998
	0-3	deltaPS	Univariate	CR							0.014
	3-12	deltaBV	Univariate	CR							0.038
	3-12	deltaBF	Univariate	CR							0.032
	3-12	deltaMTT	Univariate	CR							0.692
	3-12	deltaPS	Univariate	CR							0.005
	0-3	deltaBV	Multivariate	CR							0.923
	0-3	deltaBF	Multivariate	CR							0.800
	0-3	deltaMTT	Multivariate	CR							-
	0-3	deltaPS	Multivariate	CR							0.099
	3-12	deltaBV	Multivariate	CR							0.446
	3-12	deltaBF	Multivariate	CR							0.472
	3-12	deltaMTT	Multivariate	CR							-
	3-12	deltaPS	Multivariate	CR							0.037

Appendix F2 *Early response results in MRI studies*

Author, year	Time of Image	Follow-up	RT until imaging	Early response Parameter	Group 1 vs Group 2		Statistical Test	Out-come	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P-value / ROC*	AUC	95% confidence Interval
Yoo, 2012	1	30d post-therapy	CRT (erlotinib/ bevacizumab)	Ktrans	Complete response	Non-complete response	Simon's two stage optimal design	CR	...	...	...	...	...	...	...	
Galban, 2009	3	6 mo post-therapy	Cisplatin/ Carbotaxol/ 5-FU + 30 Gy	V (primary)	Baseline	3 weeks	Unpaired 2-tailed student's T-test	CR/ PR	...	...	...	...	...	P=0.005		
				V (nodal)	Baseline	3 weeks								P<0.0001		
				V (primary)	Complete resp	Non-complete								P<0.05		
				V (nodal)	Complete resp	Non-complete								P<0.05		
				ADC(nodal)	Baseline	3 weeks								P<0.0001		
				ADC(primary)	Compete resp	Non-complete								P=NS		
				ADC (complete resp)	Baseline	3 weeks	ROC analysis							P<0.0001		
				ADC (partial resp)	Baseline	3 weeks								P=NS		
				ΔVolumetry (pre v intra)	Baseline	3 weeks									0.758	
				(PRM)ADC		3 weeks									0.825	
King, 2010	2	6 mo post-therapy	Cisplatin + 18 Gy	ΔADC	Correlation outcome at 6 months		ROC analysis	LRF	100	92	88	100	95		...	
				ADC	Correlation outcome at 6 months		Fisher's exact T							P=0.98		
				ΔADC (resolve group)	Correlation outcome at 6 months									P<0.0001		
				ΔADC (residual group)	Correlation outcome at 6 months									P=0.0001		
Martins, 2015	2	14d intra-therapy	CRT(ns)	ADC (complete resp)	Baseline	2 weeks	Mann-Whitney U	CR/ PR/ SD/ PD	...	...	...	...	...	P=0.04	...	
				ADC (incomplete resp)	Baseline	2 weeks								P=NS		
				ΔADC	Stable disease	Partial response								P=NS		
Tyagi, 2016	1-4	3 mo post-therapy	Cisplatin/ Carboplatin/ Paclitaxel +10-40 Gy	ΔADC	Correlation with ΔGTV		Spearman Rho correlation coefficient	CR	...	...	...	...	...	Rho = 0.15	...	
Ding, 2015	3-4	2-3 mo post-therapy	Cetuximab/ Cisplatin/ Carboplatin/ Docetaxel +37 Gy	ΔD+ΔADC	Correlation with CR		ROC analysis	CR/ PR	63	85	...	...	78		0.87	0.79-0.96
				ADC	CR	Non-CR	Mann-Whitney U							P<0.0003		
				ΔADC	CR	Non-CR								P<0.003		
				ΔD	CR	Non-CR								P<0.003		
				ΔADC	Baseline	3/4 weeks								P<0.0001		
				ΔD	Baseline	3/4 weeks								P<0.0001		
				Δf	Baseline	3/4 weeks								P<0.0001		
				D*	Baseline	3/4 weeks								P=NS		
				stratified ADC (node)	Complete resp	Non-Complete								P<0,0001		
				Stratified ADC (tumor)	Complete resp	Non-Complete								P<0,0005		
				delta D (poststrat)	Complete resp	Non-Complete								P=0.01		
				delta f (poststrat)	Complete resp	Non-Complete								P=0.01		
				delta D* (poststrat)	Complete resp	Non-Complete								P=NS		
				norm ΔADC (poststrat)	Complete resp	Non-Complete								p<0,0001		
				norm ΔD (poststrat)	Complete resp	Non-Complete								p<0,0001		
				norm Δf (poststrat)	Complete resp	Non-Complete								P=NS		
				norm ΔD* (poststrat)	Complete resp	Non-Complete								P=NS		
				Norm ΔADC	Complete resp	Non-Complete								P<0.00027		
				Norm ΔD	Complete resp	Non-complete								P<0.00024		
				Norm Δf	Complete resp	Non-complete	ROC analysis							P<0.024		

Appendix F2 *continued*

Author, year	Time of Image	Follow-up	RT until imaging	Early response Parameter	Group 1 vs Group 2		Statistical Test	Out-come	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P-value / ROC*	AUC	95% confidence Interval
Paudyal, 2016	1-3	3-6 mo post-therapy	Cisplatin/ Cetuximab/ Carboplatin/ 5-FU +10-30 Gy	MTV (ΔV CR)	Baseline	3	Wilcoxon rank sum	CR/ PR/ SD/ PD	...	...	...	...	...	P<0,003	...	
				MTV (ΔV non-CR)	Baseline	3								P>0,05		
	1			MTV (ΔV CR)	LRC	LRF								P=0,016		
	2			MTV (ΔV CR)	LRC	LRF								P>0,05		
	3			MTV (ΔV CR)	LRC	LRF								P>0,05		
				ADC	LRC	LRF								P>0,003		
				D	LRC	LRF								P>0,003		
				f	LRC	LRF								P>0,003		
				D*	LRC	LRF								P>0,003		
				ADC (ΔADC CR)	Baseline	1								P<0,003		
				ADC (ΔADC CR)	Baseline	2								P<0,003		
				ADC (ΔADC CR)	Baseline	3								P<0,003		
				D (ΔD CR)	Baseline	1								P<0,003		
				D (ΔD CR)	Baseline	2								P<0,003		
				D (ΔD CR)	Baseline	3								P<0,003		
				f (Δf CR)	Baseline	2								P>0,003		
				f (Δf CR)	Baseline	3								P>0,003		
				D* (ΔD* CR)	Baseline	1								P>0,003		
				D* (ΔD* CR)	Baseline	2								P>0,003		
				D* (ΔD* CR)	Baseline	3								P>0,003		
				ADC (ΔADC non-CR)	Baseline	1								P>0,003		
				ADC (ΔADC non-CR)	Baseline	2								P>0,003		
				ADC (ΔADC non-CR)	Baseline	3								P>0,003		
				D (ΔD non-CR)	Baseline	1								P>0,003		
				D (ΔD non-CR	Baseline	2								P>0,003		
				D (ΔD non-CR)	Baseline	3								P>0,003		
				f (Δf non-CR)	Baseline	1/2/3								P=NS		
				D* (ΔD* non-CR)	CR	1/2/3								P=NS		
				D	CR	Non-CR								P=0,007		
				ADC	CR	Non-CR								P>0,05		
				D (Δwk3-wk0)	CR	Non-CR								P=0,017		
				ADC (Δwk3-wk0)	CR	Non-CR								P=0,06		
				f (Δwk 3-wk0)	CR	Non-CR								P>0,003		

Appendix F3 *Early response results in PET studies*

Author, year	Time imaging	RT until imaging (Gy)	Early response PET parameter	Group 1 vs Group 2		Statistics	Outcome	Sen(%)	Spec (%)	PPV	NPV	Accuracy	P-value
Brun, 1997	1		Metabolic rate + initial	<20 µmol	>20µmol	NS	CR						NS
Brun, 2002	1-3		PET1	<26	>26	MW	CR						0.14
			Metabolic rate metastasis	<18	>18	MW	CR						1.0
			SUV tumor	<9	>9	MW	CR						0.01
			SUV metastasis	<6	>6	MW	CR						1.0
			PET2 Metabolic rate PT	<16	>16	MW	CR						0.007
			PET2 Metabolic rate LN	<14	>14	MW	CR						0.25
			PET2 SUV PT	<5	>5	MW	CR						0.07
			PET2 SUV LN	<4.5	>4.5	MW	CR						0.45
			MR FDG PET unadjusted	>716 (median)	<716 (median)	Fisher	CR						-
			TNM	T1-2	T3-4	Fisher	CR						0.46
			TNM	N0-1	N2-3	Fisher	CR						0.14
			UICC	2-3	4	Fisher	CR						0.12
			Tumor site	Oropharynx	Other	Fisher	CR						0.48
			Histology	1-2	3	Fisher	CR						0.26
			Hb g/L	<135	>135	Fisher	CR						0.26
			Age	<55	>55	Fisher	CR						0.46
Ceulemans, 2011	0.4	47	Qualitative SUV	CMR	Non CMR	-	CR	28.6	81.8	31.0	80.0	42.5	0.50

Appendix G1 *Long-term outcome results in CT studies*

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	P-value
Truong, 2011	28 (6-44)	2.4	34, 70	mean GTV PT	LRC	LRF	Indep-samples t		0.003
	28 (6-44)	0	0	mean GTV PT	>39.46	<39.46	Indep-samples t		0.85
	28 (6-44)	0	0	mean cutoff GTV PT	>39.46	<39.46	Indep-samples t		0.64
	28 (6-44)	0	0	mean cutoff GTV PT	>39.46	<39.46	Indep-samples t		0.71
	28 (6-44)	0	0	mean cutoff GTV PT	>39.46	<39.46	Indep-samples t		0.22
	28 (6-44)	0	0	BF (mL/100g/min)	LRC	LRF	Indep-samples t		0.004
	28 (6-44)	0	0	BV (mL/100g)	LRC	LRF	Indep-samples t		0.546
	28 (6-44)	0	0	MTT (s)	LRC	LRF	Indep-samples t		0.461
	28 (6-44)	0	0	CP	LRC	LRF	Indep-samples t		0.02
	28 (6-44)	2	34	BF	LRF	LRC	Indep-samples t		0.0008
	28 (6-44)	4	70	BF	LRF	LRC	Indep-samples t		0.598
	28 (6-44)	2	34	BV	LRF	LRC	Indep-samples t		0.081
	28 (6-44)	4	70	BV	LRF	LRC	Indep-samples t		0.721
	28 (6-44)	2	34	MTT	LRF	LRC	Indep-samples t		0.379
	28 (6-44)	4	70	MTT	LRF	LRC	Indep-samples t		0.438
	28 (6-44)	2	34	CP	LRF	LRC	Indep-samples t		0.098
	28 (6-44)	4	70	CP	LRF	LRC	Indep-samples t		0.694
	28 (6-44)	0-2	0-34	Delta BV in %	Baseline	2	Indep-samples t		0.053
	28 (6-44)	0-2	0-34	Delta BF in %	Baseline	2	Indep-samples t		0.046
	28 (6-44)	0-2	0-34	Delta CP in %	Baseline	2	Indep-samples t		0.106
Abramyuk, 2015	24	2-5		Delta Ktrans	LRC	LRF	Welch mod. T-		<0.05
	24	2-5		Delta rTBV	LRC	LRF	Welch mod. T		<0.05
	24	2-5		KtransSlope L & box	Slopelog(lacune)	log(box size)	Welch mod. T		<0.05
	24	2-5		rTBV Slope L & box	Slopelog(lacune)	log(box size)	Welch mod. T		<0.05
	24			Ktrans	% difference	lacun Ktrans	Correlation coeff	0.78	
	24			rTBV	% difference	lacun Ktrans	Correlation coeff	0.48	

Appendix G2 *Long-term outcome results in MRI studies*

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	Hazard ratio	95% CI HR	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P value	AUC
Bhatia, 2010	1 y	2	Cisplatin	Volumetry (primary tumor)	<10.6 cm3	>10.6 cm3	ROC analysis	LC/ LF	...	...	76	74	62	85	75		0.721
			+18 Gy	Volumetry (hypopharynx)	<10.6 cm3	>10.6 cm3	ROC analysis				78	100	100	80	88		0.873
				ΔV (pre vs intra)	<9.4%	>9.4%	ROC analysis				41	87	64	73	71		0.689
				ΔV (pre vs intra), hypopharynx	<47.2%	>47.2%	ROC analysis				78	63	70	71	71		0.635
				V (2 wk)	LC	LF	Mann Whitney U									0.009	
				ΔV	LC	LF	Mann Whitney U									0.021	
				V (2 wk), hypopharynx	LC	LF	Mann Whitney U									0.011	
				ΔV, hypopharynx	LC	LF	Mann Whitney U									0.321	
				ΔV (rise)	LC	LF	Fisher's exact									0.095	
King, 2010	1 y	2	+18 Gy	Cho: water (rise/fall)	LRC vs LRF		Log regression	LRC/LRF/DM	...	...	...	...	...	....	...	0.2-0.84	...
				Cho: cr (rise/fall)	LRC vs LRF		Log regression	/OS								0.2-0.84	
				Intra-treatment cho (persistence/ resolution)	LRC vs LRF		Chi-squared test	/CRD								0.4-1.0	
Baer, 2015	37 mo	2	Carboplatin/	(PRM)Ktrans	Baseline	2 weeks	Log rank test	OS	...	...	...	...	...	...	...	0.002	...
			Cisplatin/	%ΔKtrans	Baseline	2 weeks										0.041	
			Paclitaxel/	PRM(NAUC60)	Baseline	2 weeks										0.022	
			5-FU/	%ΔPRM(NAUC60)	Baseline	2 weeks										0.024	
			Cetuximab/	PRMVp	Baseline	2 weeks										0.308	
			Erlotinib	%ΔPRMVp	Baseline	2 weeks										0.068	
			+20 Gy														
Cao, 2008	9.7 mo	2	Carboplatin/	GTV (primary)	Local control	Local failure	Mann Whitney U	LRC/LRF/LF	...	...	...	...	...	...	...	NS	...
			Paclitaxel/	GTV (nodal)	Local control	Local failure										NA	
			Cisplatin/	%ΔGTV (primary)	Local control	Local failure										NS	
			Cetuximab	%ΔGTV (nodal)	Local control	Local failure										NA	
			+20 Gy	%ΔTBV (primary)	Local control	Local failure										0.03	
				%ΔTBV (nodal)	Local control Local	Local failure										NA	
				%ΔTBF (primary)	control	Local failure										NS	
				%ΔTBF (nodal)	Local control	Local failure										NA	
Dirix, 2009	30.7 mo	(2,3)4	Cisplatin	ADC (primary+nodal)	LRC	LRF	2-tailed students T	DFS	...	...	...	...	...	...	...	0.01	...
			+10 Gy	DCE (primary+nodal)	Baseline	Week 4	2-tailed students T	/								0.03	
				GTV (DW)	DWI	T1	Paired students T	LRR/								<0.01	
				GTV (DW)	DWI	T2	Paired students T	LRC								<0.01	
Jansen, 2016	32 mo	2	Cetuximab/	Volume (cm3)	Baseline	Intratreatment	Paired students T	LRC/LRF	...	...	...	...	...	...	...	0.02	...
			Bevacizumab	Ktrans (mean)	Baseline	Intratreatment	Paired students T									0.36	
			/Cisplatin	Ktrans (SD)	Baseline	Intratreatment	Paired students T									0.99	
			+20-28 Gy	Ktrans (energy)	Baseline	Intratreatment	Paired students T									0.18	
				Ktrans (homogeneity)	Baseline	Intratreatment	Paired students T									0.61	
				Ve (mean)	Baseline	Intratreatment	Paired students T									0.09	
				Ve (SD)	Baseline	Intratreatment	Paired students T									0.26	
				Ve (energy)	Baseline	Intratreatment	Paired students T									0.04	
				Ve (homogeneity)	Baseline	Intratreatment	Paired students T									0.78	
				Volume (cm3)	LRC	LRF	Mann Whitney U									0.54	
				ΔVolume	LRC	LRF	Mann Whitney U									0.64	
				Volume (mean)	LRC	LRF	Mann Whitney U									0.42	
				ΔVolume (mean)	LRC	LRF	Mann Whitney U									0.42	
				Ktrans (SD)	LRC	LRF	Mann Whitney U									0.74	
				Ktrans (SD)	LRC	LRF	Mann Whitney U									0.49	
				Ktrans (E)	LRC	LRF	Mann Whitney U									1	

Appendix G2 *continued*

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	Hazard ratio	95% CI HR	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P value	AUC
				Ktrans (E)	LRC	LRF	Mann Whitney U									0.57	
				Ktrans (H)	LRC	LRF	Mann Whitney U									0.57	
				Ktrans (H)	LRC	LRF	Mann Whitney U									0.35	
				Ve (mean)	LRC	LRF	Mann Whitney U									0.42	
				Ve (mean)	LRC	LRF	Mann Whitney U									0.35	
				Ve (SD)	LRC	LRF	Mann Whitney U									0.14	
				Ve (SD)	LRC	LRF	Mann Whitney U									0.10	
				Ve (E)	LRC	LRF	Mann Whitney U									0.1	
				Ve (E)	LRC	LRF	Mann Whitney U									0.07	
				Ve (H)	LRC	LRF	Mann Whitney U									0.18	
				Ve (H)	LRC	LRF	Mann Whitney U									0.14	
Wong 2017	0	14 mo	Cisplatin	ADC primary	LRC	LRF	Mann Whitney U									0.009	0.829
	0-1		+11/20Gy	Delta ADC primary	LRC	LRF	Mann Whitney U									0.274	
	0-2			Delta ADC primary	LRC	LRF	Mann Whitney U									0.000	0.937
	0			R2 primary	LRC	LRF	Mann Whitney U									0.685	
	0-1			Delta R2 primary	LRC	LRF	Mann Whitney U									0.555	
	0-2			Delta R2 primary	LRC	LRF	Mann Whitney U									0.791	
	0			Ktrans primary	LRC	LRF	Mann Whitney U									0.735	
	0-1			Delta Ktrans primary	LRC	LRF	Mann Whitney U									0.331	
	0-2			Delta Ktrans primary	LRC	LRF	Mann Whitney U									0.012	0.813
	0			Ve primary	LRC	LRF	Mann Whitney U									0.003	0.864
	0-1			Delta Ve primary	LRC	LRF	Mann Whitney U									0.836	
	0-2			Delta Ve primary	LRC	LRF	Mann Whitney U									0.047	0.774
	0			Vp primary	LRC	LRF	Mann Whitney U									0.072	
	0-1			Delta Vp primary	LRC	LRF	Mann Whitney U									0.219	
	0-2			Delta Vp primary	LRC	LRF	Mann Whitney U									0.288	
	0			ADC lymphnode	LRC	LRF	Mann Whitney U									0.456	
	0-1			ADC lymphnode	LRC	LRF	Mann Whitney U									0.239	
	0-2			ADC lymphnode	LRC	LRF	Mann Whitney U									0.900	
	0			R2 lymphnode	LRC	LRF	Mann Whitney U									0.611	
	0-1			R2 lymphnode	LRC	LRF	Mann Whitney U									0.677	
	0-2			R2 lymphnode	LRC	LRF	Mann Whitney U									0.455	
	0			Ktrans lymphnode	LRC	LRF	Mann Whitney U									0.184	
	0-1			Ktrans lymphnode	LRC	LRF	Mann Whitney U									1.00	
	0-2			Ktrans lymphnode	LRC	LRF	Mann Whitney U									0.102	
	0			Ve lymphnode	LRC	LRF	Mann Whitney U									0.020	0.783
	0-1			Ve lymphnode	LRC	LRF	Mann Whitney U									0.244	
	0-2			Ve lymphnode	LRC	LRF	Mann Whitney U									0.694	
	0			Vp lymphnode	LRC	LRF	Mann Whitney U									0.229	
	0-1			Vp lymphnode	LRC	LRF	Mann Whitney U									0.944	
	0-2			Vp lymphnode	LRC	LRF	Mann Whitney U									0.813	
Wang, 2012	19.6 mo	2	Carboplatin/	ΔGTV (volumetry)			ROC analysis	LRC/DM/LC/	...	...	85	91	...	...	...		0.723
			Paclitaxel/	ΔBV			ROC analysis	LRF/RF			85	83.0					0.903
			Cisplatin/	ΔBV (post correction)			ROC analysis										0.947
			Cetuximab/	BV (2 weeks)	Local control	Local failure	Mann-Whitney U									0.01	
			+20 Gy	BF (2 weeks)	Local control	Local failure	Mann-Whitney U									0.05	
				BF+BV (2 weeks)	Local control	Local failure	Mann-Whitney U									0.01	
				%ΔGTV (2 weeks)	Local control	Local failure	Mann-Whitney U									0.33	
				ΔBV	Local control	Local failure	Mann-Whitney U									0.03	
				ΔBF	Local control	Local failure	Mann-Whitney U									0.71	
Hoang, 2014	2 y	2	Cisplatin	%ΔADC(primary)	Baseline	Change 2 weeks	Fisher's exact	RES/REC	...	...	...	...	...	...	...	0.12	...
			+12 Gy	%ΔADC(nodal)	Baseline	Change 2 weeks	Fisher's exact									0.01	
Kim, 2009	1 y	1	Cisplatin/	ΔADC	<1.11*E-3	>1.11E-3	ROC analysis	CR/ PR	...	...	86	83	...	...	...		0.88

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	Hazard ratio	95% CI HR	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P value	AUC
			Cetuximab	ADC (median)	CR	PR	Mann Whitney U									<0.01	
			+8 Gy	ADC (normalized)	CR	PR	Mann Whitney U									<0.05	
				ΔADC	CR	PR	Mann Whitney U									<0.01	
				ΔADC (complete resp)	Baseline	1 week	Mann Whitney U									<0.01	
				ΔADC (partial resp)	Baseline	1 week	Mann Whitney U									NS	
				Volumetry (median)	Baseline	1 week	Mann Whitney U									<0.01	
King, 2010	1 y	2	Cisplatin	ΔADC	Correlation: outcome (12mo)		ROC analysis	LRF/ LC	...	...	88	92	88	92	90		...
			+18 Gy	ΔADC	Correlation outcome (follow-up)		ROC analysis				80	100	100	83	90		
				ADC	Correlation: outcome (12mo)		Logistic regression									0.89	
				ADC	Correlation: final outcome		Logistic regression									0.5	
				ΔADC (resolve group)	Correlation: outcome (12 mo)		Logistic regression									<0.0001	
				ΔADC (resolve group)	Correlation: final outcome		Logistic regression									<0.0001	
				ΔADC (residual group)	Correlation: outcome (12 mo)		Logistic regression									0.0008	
				ΔADC (residual group)	Correlation: final outcome		Logistic regression									0.0007	
King, 2013	2 y	2	CRT (NS)	Kurtosis (ADC)	<3.9	>3.9	ROC analysis	LC/ LF	4.4	...	77	65	63	79	70		...
				Skewness (ADC)	<0.4	>0.4	ROC analysis		1.5		77	77	71	86	77		
				%ΔADC	<15.5%	>15.5%	ROC analysis		0.97		77	69	77	69	73		
				Mean ADC	LC	LF	Univariate analysis									0.112	
				Kurtosis ADC	LC	LF	Univariate analysis									0.04	
				Skewness ADC	LC	LF	Univariate analysis									0.015	
				Δ mean ADC	LC	LF	Univariate analysis									0.016	
				Δ kurtosis (%)	LC	LF	Univariate analysis									0.238	
				Δ skewness (%)	LC	LF	Univariate analysis									0.316	
				Intra Kurtosis ADC	<3.9	>3.9	Univariate analysis									0.033	
				Intra Skewness ADC	<0.4	>0.4	Univariate analysis									0.009	
				Skewness ADC	Correlation LRF		Univariate analysis		12.3	1.3-116.5						0.029	
				Kurtosis ADC	Correlation LRF		Univariate analysis		2.1	1.03-4.3						0.041	
				Δ mean ADC	Correlation LRF		Univariate analysis		0.96	0.92-0.99						0.03	
				Skewness ADC	Correlation TTR		Univariate analysis		4.4	1.3-14.7						0.015	
				Kurtosis ADC	Correlation TTR		Univariate analysis		1.5	1.1-2.1						0.013	
				Δ mean ADC	Correlation TTR		Univariate analysis		0.97	0.94-0.99						0.024	
				Volumetry	Correlation LF		Univariate analysis		1.20	1.05-1.36						0.008	
				%Δ mean V	Correlation LF		Univariate analysis		1.04	1.01-1.07						0.016	
				Volumetry	Correlation: TTR		Univariate analysis		1.06	1.02-1.09						0.001	
				%Δ mean V	Correlation: TTR		Univariate analysis		1.02	1.01-1.04						0.002	
Lambrecht, 2014	57.3 mo	2	20-23 Gy	%ΔADC	Poor response	Good response	Mann-Whitney U	LRC/ LRR	...	...	...	...	...	...	...	<0.001	...
				ΔADC	No TPS warp	With TPS warp	Paired T-Test									<0.001	
				Markers vs reg error	Poor response	Good response	Linear regression									<0.001	
				ADC (IQR)	Poor response	Good response	Mann-Whitney U									0.7	
				ΔADC (registr)	Poor response	Good response	Mann-Whitney U									0.3	
				ΔADC (auto registr)	Poor response	Good response	Mann-Whitney U									0.03	
				ΔADC (marker registr)	Poor response	Good response	Mann-Whitney U									0.001	
Matoba, 2014	30.8 mo	3	S-1/Cisplatin	ΔADC (primary tumor)	<0.24	>0.24	ROC analysis	PFS	...	...	100	78.7	76.7	100	84.8		0.9
			+30 Gy	ΔADC (primary tumor)	<0.24	>0.24	Log rank test	PFS								<0.05	
				ADC (primary)	LRC	LRF	Mann Whitney U									NS	
				ADC (nodal)	LRC	LRF	Mann Whitney U									NS	
				ΔADC (primary)	LRC	LRF	Mann Whitney U									0.0003	
				ΔADC (nodal)	LRC	LRF	Mann Whitney U									0.01	
				V (primary)	LRC	LRF	Mann Whitney U									0.03	
				V (nodal)	LRC	LRF	Mann Whitney U									0.01	
				ΔV (primary)	LRC	LRF	Mann Whitney U									NS	
				ΔV (nodal)	LRC	LRF	Mann Whitney U									0.02	
				ADC (primary)	LRC		Nominal logistic									NS	

Appendix G2 *continued*

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	Hazard ratio	95% CI HR	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P value	AUC
				ADC (nodal)	LRC		Nominal logistic									NS	
				ΔADC (primary)	LRC		Nominal logistic									0.004	
				ΔADC (nodal)	LRC		Nominal logistic									0.03	
				V (primary)	LRC		Nominal logistic									NS	
				V (nodal)	LRC		Nominal logistic									NS	
				ΔV (primary)	LRC		Nominal logistic									NS	
				ΔV (nodal)	LRC		Nominal logistic									0.002	
				ΔADC (primary)	LRC		Nominal logistic									0.04	
				ΔADC (nodal)	LRC		Nominal logistic									NS	
				ΔV (nodal)	LRC		Nominal logistic									NS	
Scalco, 2016	>6 mo	3	Cisplatin +34-36 Gy	$\Delta\sigma^2$			LDA analysis	LRC/LRF	...	...	45.5	85.7	...	...	68.0		...
				ΔFD			LDA analysis				36.4	90.9			63.5		
				ΔADC			LDA analysis				36.4	85.7			64.0		
				$\Delta\sigma^2$-ΔFD			LDA analysis				63.6	72.7			68.2		
				$\Delta\sigma^2$-ΔADC			LDA analysis				45.5	83.3			65.2		
				ΔADC-ΔFD			LDA analysis				45.5	88.9			5.0		
				GLCM_cont_T2wMR-ΔFD			LDA analysis				63.6	100			81.8		
				$\Delta\sigma^2$-FD1_T2wMRI			LDA analysis				63.6	85.7			76.0		
				ΔADC-E_{cc1}			LDA analysis				54.5	92.9			76.0		
				ADC	Control	Failure	Unpaired T test									NS	
				ΔADC	Control	Failure	Unpaired T test									NS	
Schouten, 2014	38 mo	2	Cisplatin/ Cetuximab +20 Gy	ΔADC (tumor) EPI	Baseline	During	Wilcoxon signed rank	LRC	...	...	...	...	...	...	...	0.05	...
				ADC (tumor) EPI	LRC	Recurrent										0.89	
				ADC (tumor) HASTE	LRC	Recurrent										0.74	
				ΔADC low (tumor) EPI	LRC	Recurrent										0.18	
				ΔADC low (tumor) HASTE	LRC	Recurrent										0.18	
				ADC (tumor) EPI	ADCEPI	ΔSUVmean										0.8	
				ADC (tumor) EPI	ADCEPI	ΔSUVmax										0.8	
				ADC (nodal) HASTE	ADCHASTE	SUVmean										0.6	
				ADC (nodal) HASTE	ADCHASTE	SUVmax										0.6	
				ADC (nodal) EPI	ADCEPI	SUVmean										0.19	
				ADC (nodal) HASTE	ADCHASTE	SUVmax										0.6	
				ADC multiparametric	ΔADCHASTE	ΔSUVmax										0.04	
				ADC multiparametric	ΔADCHASTE	ΔSUVmean										0.01	
				ΔVolumetry (tumor)	LRC	LRF										0.74	
Tyagi, 2016	1 y	1-4	Cisplatin/ Carboplatin/ Paclitaxel +10-40 Gy	ΔADC	Correlation with ΔGTV		Spearman Rho correlation coefficient	LRC	...	...	...	...	...	...	...	Rho=0.15	...
Vandecaveye, 2010	2 y	2	CRT (NS)	ΔADC (primary tumor)	<14%	>14%	ROC analysis	LRC	...	...	88	91	78	96	90		0.94
		4		ΔADC (nodal)	<14.61%	>14.61%	ROC analysis				80	89	62	95	87		0.83
		2		ΔADC (primary tumor)	<25%	>25%	ROC analysis				100	91	80	100	94		0.97
		4		ΔADC (nodal)	<19%	>19%	ROC analysis				80	96	80	96	93		0.90
		2		ΔV (primary tumor)	<20%	>20%	ROC analysis				88	57	41	93	65		0.78
		4		ΔV (nodal)	<33%	>33%	ROC analysis				90	48	28	96	56		0.68
		2		ΔV (primary tumor)	<65%	>65%	ROC analysis				75	57	38	87	61		0.64
		4		ΔV (nodal)	<50%	>50%	ROC analysis				70	61	29	90	63		0.68
		2		ΔV	Correlation LRC		Multivariate									>0.05	
		4		ΔV	Correlation LRC		Multivariate									>0.05	
		2		V (primary)	Recurrence	Responder	Mann-Whitney U									0.03	
		4		V (primary)	Recurrence	Responder	Mann-Whitney U									NS	
		2		ΔADC (primary+nodal)	Recurrence	Responder	Mann-Whitney U									<0.001	
		4		ΔADC (primary+nodal)	Recurrence	Responder	Mann-Whitney U									<0.0001	

Author, year	Follow-up	Time of Imaging	C+RT at imaging	Prognostic parameter	Group 1 vs Group 2		Statistical Test	Outcome	Hazard ratio	95% CI HR	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)	P value	AUC
		2		ΔADC (primary+nodal)	Correlation LRC		Multivariate									0.003	
		4		ΔADC (primary+nodal)	Correlation LRC		Multivariate									<0.0001	
Marzi 2017	27.6	0	Cisplatin	D pretreatment	RC	RF	Mann-Whitney U	LRC								0.038	
		3	100mg	D midtreatment	RC	RF	Mann-Whitney U	LRC								0.025	
		3	+ 34 Gy	D Change at midRT	RC	RF	Mann-Whitney U	LRC								0.387	
		3		D Change % at midRT	RC	RF	Mann-Whitney U	LRC								0.760	
		0		D*	RC	RF	Mann-Whitney U	LRC								0.774	
		3		D*	RC	RF	Mann-Whitney U	LRC								0.127	
		3		D*	RC	RF	Mann-Whitney U	LRC								0.263	
		3		D*	RC	RF	Mann-Whitney U	LRC								0.134	
		0		F	RC	RF	Mann-Whitney U	LRC								0.205	
		3		F	RC	RF	Mann-Whitney U	LRC								0.252	
		3		F	RC	RF	Mann-Whitney U	LRC								0.015	
		3		F	RC	RF	Mann-Whitney U	LRC								0.008	
		0		D* x f	RC	RF	Mann-Whitney U	LRC								0.381	
		3		D* x f	RC	RF	Mann-Whitney U	LRC								0.037	
		3		D* x f	RC	RF	Mann-Whitney U	LRC								0.003	
		3		D* x f	RC	RF	Mann-Whitney U	LRC								<0.001	
		0		ADC	RC	RF	Mann-Whitney U	LRC								0.041	
		3		ADC	RC	RF	Mann-Whitney U	LRC								0.286	
		3		ADC	RC	RF	Mann-Whitney U	LRC								0.286	
		3		ADC	RC	RF	Mann-Whitney U	LRC								0.186	
		0		D pretreatment	>0.97	<0.97	ROC	LRF			70	75	54	86			0.73
		0		ADC pretreatment	>1.20	<1.20	ROC	LRF			70	79.2	58	86			0.73
		3		D midtreatment	>1.12	<1.12	ROC	LRF			100	60	50	10091			0.78
		3		D* midtreatment	≤15.6	>15.6	ROC	LRF			88	50	42	89			0.69
		3		D* x f midtreatment	≤92.9	>92.9	ROC	LRF			75	80	61	95			0.76
		3		Deltaf	≤ -19.1%	> -19.1%	ROC	LRF			88	90	78	95			0.83
		3		Delta(D* x f)	≤ -48.3	> -48.3	ROC	LRF			88	90	79	88			0.88
		3		deltaADC	≤15.5%	>15.5%	ROC	LRF			75	70	50				0.66

Appendix G3 *Long-term outcome results in PET studies*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
Brun, 2002	1-3	Metabolic rate tumor	CR	LCR		OS	Cox's pr. HR	0.001	4.5	1.5-13
	1-3	Metabolic rate meta	CR	LCR		OS	Cox's pr. HR	0.008	2.8	0.90-8.6
	1-3	SUV tumor	CR	LCR		OS	Cox's pr. HR	0.07	2.8	9.8-8.1
	1-3	SUV meta	CR	LCR		OS	Cox's pr. HR	0.28	2.0	0.66-6.3
		PET2 Metabolic rate PT	<16	>16		LRC	MW	0.002	2.6	0.95-7.3
		PET2 Metabolic rate LN	<14	>14		LRC	MW	0.07	1.4	0.54-3.9
		PET2 SUV PT	<5	>5		LRC	MW	0.031	2.6	0.74-9.2
		PET2 SUV LN	<4.5	>4.5		LRC	MW	0.15	2.2	0.76-6.4
		TNM	T1-2	T3-4		LRC	Fisher	0.07	0.83	0.31-2.2
		TNM	N0-1	N2-3		LRC	Fisher	0.035	0.4	0.14-1.1
		UICC	2-3	4		LRC	Fisher	0.06	1.4	0.54-3.9
		Tumor site	Oropharynx	Other		LRC	Fisher	0.73		
		Histology	1-2	3		LRC	Fisher	0.73		
		Hb g/L	<135	>135		LRC	Fisher	1		
		Age	<55	>55		LRC	Fisher	1		
Dirix, 2009	0.2.3.4	Gross tumor volume	GTVDWI	GTVT1	30.7		Pair. stud. t-test	<0.01		
	0.2.3.4	Gross tumor volume	GTVDWI	GTVT2	30.7		Pair. stud. t-test	<0.01		
	4	Gross tumor volume	LRC	LRF			Unp. 2-tail. stud t test	0.01		
	0.4	T/Bmax FMISO	pretreatment	Wk 4	30.7	DFS	Unp. 2-tail. stud t test	0.02		
	4	DWI: ADC	LRC	LRF		LRC		0.01		
	4	DCE: Initial slope height	pretreatment	Wk 4		LRC		0.03		
Farrag, 2010	4	SUVmax (median = 4.03)	Low <median	High >median	24	OS	Log rank test	0.026		
	4	SUVmax (median = 4.03)	Low <median	High >median	24	Caus spec surv.	Log rank test	0.12		
	4	SUVmax (median = 4.03)	Low <median	High >median	24	DFS	Log rank test	0.32		
	4	SUVmax (median = 4.03)	Low <median	High >median	24	LRRFS	Log rank test	0.27		
	4	SUVmax (median = 4.03)	Low <median	High >median	24	MFS	Log rank test	0.29		
	4	TNMstage + SUVmax	Low <median	High >median	24	OS	Multivariate analysis	0.044		
	4	Visual evaluation (expert opinion)	CMR	Non-CMR	24	OS	Log rank test	0.5		
	4	Visual evaluation (expert opinion)	CMR	Non-CMR	24	CSS	Log rank test	0.42		
	4	Visual evaluation (expert opinion)	CMR	Non-CMR	24	DFS	Log rank test	0.86		
	4	Visual evaluation (expert opinion)	CMR	Non-CMR	24	LRRFS	Log rank test	0.97		
	4	Visual evaluation (expert opinion)	CMR	Non-CMR	24	MFS	Log rank test	0.96		
	0	TNM	T1-2	T3-4	24	OS	Log-rank test	0.49		
	0	TNM	T1-2	T3-4	24	DFS	Log-rank test	0.90		
	0	TNM	T1-2	T3-4	24	LRRFS	Log-rank test	0.60		
	0	TNM	N0	N+	24	OS	Log-rank test	0.96		
	0	TNM	N0	N+	24	DFS	Log-rank test	0.56		
	0	TNM	N0	N+	24	LRRFS	Log-rank test	0.46		
	0	AJCC	1.2	3.4	24	OS	Log-rank test	0.72		
	0	AJCC	1.2	3.4	24	DFS	Log-rank test	0.18		
	0	AJCC	1.2	3.4	24	LRRFS	Log-rank test	0.68		
	0	Age	<60	>60	24	OS	Log-rank test	0.19		
	0	Age	<60	>60	24	DFS	Log-rank test	0.84		
	0	Age	<60	>60	24	LRRFS	Log-rank test	0.61		
	0	Sex	Male	Female	24	OS	Log-rank test	0.38		
	0	Sex	Male	Female	24	DFS	Log-rank test	0.96		
	0	Sex	Male	Female	24	LRRFS	Log-rank test	0.72		
	0	Chemotherapy	Yes	No	24	OS	Log-rank test	0.25		
	0	Chemotherapy	Yes	No	24	DFS	Log-rank test	0.58		
	0	Chemotherapy	Yes	No	24	LRRFS	Log-rank test	0.34		
Ceulemans, 2011	4	Rate of survival	CR	NCR	24	OS	Log rank test	0.50		

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	4	Complete response	CRT	RT	24	OS	Log rank test	0.51		
	4	Non-Complete response	CRT	RT	24	OS	Log rank test	0.35		
Hentschel, 2011	0 / 1-2	Delta SUVmax 0Gy vs 10/20gy	≥50%	<50%	24	OS	Log RT + Cox-m	0.02		
		Delta SUVmax 0Gy vs 10/20gy	≥50%	<50%	24	DFS	Log RT + Cox-m.	0.1		
		Delta SUVmax 0Gy vs 10/20gy	≥50%	<50%	24	LRC	Log RT + Cox-m	0.06		
		Delta SUVmean 0Gy vs 10/20gy	≥40%	<40%	24	OS	Log RT + Cox-m	0.15		
		Delta SUVmean 0Gy vs 10/20gy	≥40%	<40%	24	DFS	Log RT + Cox-m	0.12		
		Delta SUVmean 0Gy vs 10/20gy	≥40%	<40%	24	LRC	Log RT + Cox-m	0.05		
		SUVmax	30gy/40gy	50gy/60gy	24	OS / DFS/ LRC	Stratified an.	NS		
		SUVmean	30gy/40gy	50gy/60gy	24	OS / DFS/ LRC	Stratified an.	NS		
Castaldi, 2012	0→2	DeltaSUVmax reduction	Primary	Lymphnodes	29.2	Linear relation	Cox regression	0.004		
	2	SUVmax	CMR	Non CMR	29.2	DSS	Log rank test	0.2		
	2	SUVmax	CMR	Non CMR	29.2	RFS	Log rank test	0.6		
Hoshikawa, 2013	0	FDG SUV	Mod diff	Well diff	32	Diff	Mann W	0.27		
	0	FDG SUV	poorly diff	Well diff	32	Diff	Mann W	0.16		
	0	FLT SUV	Mod diff	Well diff	32	Diff	Mann W	0.17		
	3-5	FLT SUV	Mod diff	Well diff	32	Diff	Mann W	0.17		
	3 → 4w p	FDG FLT SUV decr.	3wk	4wk post treat	32	Diff	Wilcox	<0.05		
	3 → 6w p	FDG FLT SUV decr.	3wk	6wk post treat	32	Diff	Wilcox	<0.05		
Schouten, 2014	0→2	Delta ADC PT	Pre	2wk	38		Spearman	0.05		
	2	ADC EPI PT	LRC	LRF	38	LRC	Mann W U	0.89		
	2	ADC HASTE PT	LRC	LRF	38	LRC	Mann W U	0.74		
	0→2	Delta ADC low EPI PT	LRC	LRF	38	LRC	Spearman	0.18		
	0→2	Delta ADC low HASTE PT	LRC	LRF	38	LRC	Spearman	0.18		
	2	ADC EPI PT	ADC	deltaSUVmean	38		Mann W U	0.8		
	2	ADC EPI PT	ADC	delta SUVmax	38		Mann W U	0.8		
	2	ADC HASTE LN	ADC	SUVmean	38		Mann W U	0.6		
	2	ADC HASTE LN	ADC	SUVmax	38		Mann W U	0.6		
	2	ADC EPI LN	ADC	SUVmean	38		Mann W U	0.19		
	2	ADC EPI LN	ADC	SUVmax	38		Mann W U	0.6		
	2	ADC multiparametric	deltaADC	deltaSUVmax	38		Mann W U	0.04		
	2	ADC multiparametric	deltaADC	deltaSUVmean	38		Mann W U	0.01		
	2	SUVmean	LRC	LRF	38		Mann W U	0.08		
	2	MRI volumetry PT	LRC	LRF	38		Mann W U	0.74		
Chen, 2014	0	Tstage	T12	T34	23	OS	Cox reg	0.35	1.14	0.57-2.29
	0	Tstage	T12	T34	23	DFS	Cox reg	0.42	1.32	0.69-2.52
	0	Tstage	T12	T34	23	PRFS	Cox reg	0.71	1.41	0.56-3.50
	4	SUVmean	=<3.9	>3.9	23	OS	Cox reg	0.63	1.04	0.32-3.36
	4	SUVmean	=<3.9	>3.9	23	DFS	Cox reg	0.84	1.14	0.37-3.50
	4	SUVmean	=<3.9	>3.9	23	PRFS	Cox reg	0.27	1.32	0.23-4.38
	4	SUV reduction ratio	=>0.64	<0.64	23	OS	Cox reg	0.035	2.64	1.08-6.49
	4	SUV reduction ratio	=>0.64	<0.64	23	DFS	Cox reg	0.045	2.33	1.02-5.35
	4	SUV reduction ratio	=>0.64	<0.64	23	PRFS	Cox reg	0.05	2.87	0.99-8.26
	4	SUVmean	=<10.5	>10.5	23	OS	Cox reg	0.84	1.45	0.15-13.70
	4	SUVmean	=<10.5	>10.5	23	DFS	Cox reg	0.6	5.95	0.52-46.67
	4	SUVmean	=<10.5	>10.5	23	PRFS	Cox reg	0.29	2.01	0.22-17.80
	4	Volume reduction ratio	≥0.33	<0.33	23	OS	Cox reg	0.51	1.22	0.41-3.68
	4	Volume reduction ratio	≥0.33	<0.33	23	DFS	Cox reg	0.86	1.14	0.42-2.81
	4	Volume reduction ratio	≥0.33	<0.33	23	PRFS	Cox reg	0.58	1.92	0.47-7.70
	4	SUVmean oropharynx	LRC	LRF	23	LRC	Mann-whitney U	0.07		
	4	SUVmean hypopharynx	LRC	LRF	23	LRC	Mann-whitney U	0.015		
	4	SUVmean hypopharynx	Low SSR	High SRR	23	LRC	Mann-whitney U	0.04		
	4	SUVmean	Small SUVm	Large SUVm	23	NRFS	Log rank t	0.043		

Appendix G3 *continued*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	4	VRR LNN	<0.37	0.37	23	NRFS	Log rank t	0.009	5.5	1.52-20.4
	4	SUVmean reduction ratio PT	>0.64	<0.64	23	OS	Cox reg	0.028		
	4	SUVmean reduction ratio PT	>0.64	<0.64	23	DFS	Cox reg	0.039		
	4	SUVmean reduction ratio PT	>0.64	<0.64	23	PRFS	Cox reg	0.041		
	4	Reduction ratio PT	SUVmean	Tumor Volume	23	LRC	Pearsons R	0.16		
	4	Reduction ratio N	SUVmean	Tumor Volume	23	LRC	Pearsons R	0.008		
	4	SUVmean PT	LRC	LRF	23	LRC	Mann-whitney U	0.029		
	4	SUV reduction ratio	LRC	LRF	23	LRC	Mann-whitney U	0.038		
	4	Volume in nl	LRC	LRF	23	LRC	Mann-whitney U	0.1		
	4	Volume reduction ratio	LRC	LRF	23	LRC	Mann-whitney U	0.75		
	4	SUVmean LN	LRC	LRF	23	LRC	Mann-whitney U	0.012		
	4	Suv reduction ratio	LRC	LRF	23	LRC	Mann-whitney U	0.44		
	4	Volume in nl	LRC	LRF	23	LRC	Mann-whitney U	0.039		
	4	Volume reduction ratio	LRC	LRF	23	LRC	Mann-whitney U	0.002		
Min 2015	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Log-rank test		0.025	
	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Univariate an		0.034	2.975
	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Multivariate an		0.03	4.182
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Log-rank test		0.03	
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Univariate an		0.045	3.666
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Multivariate an		0.074	3.875
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Log-rank test		0.52	
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Univariate an		0.525	1.490
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Multivariate an		0.706	1.353
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Log-rank test		0.345	
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Univariate an		0.353	1.616
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Multivariate an		0.136	2.549
	3	SUVmax RT + Cetuxi			24	DFS			0.025	
	3	SUVmax RT + Cetuxi			24	LRFS			0.03	
	3	SUVmax RT + Cetuxi			24	OS			0.345	
	3	SUVmax RT - Cetuxi			24	DFS			0.064	
	3	SUVmax RT - Cetuxi			24	LRFS			0.126	
	3	SUVmax RT - Cetuxi			24	OS			0.308	
	3	SUVmax LN+ & low metab			24	DFS			0.005	
	3	SUVmax LN+ &low metab			24	LRDS			0.0004	
	3	SUVmax LN + & low metab			24	MFFS			0.073	
	3	SUVmax LN + &low metab			24	OS			0.077	
	3	MTV	≤3.3	>3.3	24	DFS	Log-rank test		0.018	
	3	MTV	≤3.3	>3.3	24	DFS	Univariate an		0.026	3.143
	3	MTV	≤3.3	>3.3	24	DFS	Multivariate an		0.02	4.929
	3	MTV	≤3.3	>3.3	24	LRFS	Log-rank test		0.022	
	3	MTV	≤3.3	>3.3	24	LRFS	Univariate an		0.035	3.913
	3	MTV	≤3.3	>3.3	24	LRFS	Multivariate an		0.057	4.725
	3	MTV	≤3.3	>3.3	24	MFFS	Log-rank test		0.436	
	3	MTV	≤3.3	>3.3	24	MFFS	Univariate an		0.443	1.618
	3	MTV	≤3.3	>3.3	24	MFFS	Multivariate an		0.613	1.502
	3	MTV	≤3.3	>3.3	24	OS	Log-rank test		0.512	
	3	MTV	≤3.3	>3.3	24	OS	Univariate an		0.516	1.389
	3	MTV	≤3.3	>3.3	24	OS	Multivariate an		0.33	1.868
	3	MTV RT + Cetuxi			24	DFS			0.018	
	3	MTV RT + Cetuxi			24	LRFS			0.022	
	3	MTV RT + Cetuxi			24	OS			0.512	
	3	MTV RT - Cetuxi			24	DFS			0.072	
	3	MTV RT - Cetuxi			24	LRFS			0.132	

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	3	MTV RT - Cetuxi			24	OS			0.68	
	3	MTV Positive LN + low metab			24	DFS			0.024	
	3	MTV Positive LN + low metab			24	LRDS			0.0004	
	3	MTV Positive LN + low metab			24	MFFS			0.042	
	3	MTV Positive LN + low metab			24	OS			0.049	
	3	TLG	≤9.4	>9.4	24	DFS			0.005	
	3	TLG	≤9.4	>9.4	24	DFS			0.011	4.14
	3	TLG	≤9.4	>9.4	24	DFS			0.005	7.756
	3	TLG	≤9.4	>9.4	24	LRFS			0.005	
	3	TLG	≤9.4	>9.4	24	LRFS			0.015	6.312
	3	TLG	≤9.4	>9.4	24	LRFS			0.016	8.305
	3	TLG	≤9.4	>9.4	24	MFFS			0.488	1.537
	3	TLG	≤9.4	>9.4	24	MFFS			0.494	1.502
	3	TLG	≤9.4	>9.4	24	MFFS			0.613	
	3	TLG	≤9.4	>9.4	24	OS			0.279	1.736
	3	TLG	≤9.4	>9.4	24	OS			0.288	2.488
	3	TLG			24	OS			0.154	
	3	TLG RT + Cetuxi			24	DFS			0.005	
	3	TLG RT + Cetuxi			24	LRFS			0.005	
	3	TLG RT + Cetuxi			24	OS			0.279	
	3	TLG RT - Cetuxi			24	DFS			0.022	
	3	TLG RT - Cetuxi			24	LRFS			0.038	
	3	TLG RT - Cetuxi			24	OS			0.37	
	3	TLG Positive LN + low metab			24	DFS			0.017	
	3	TLG Positive LN + low metab			24	LRDS			0.001	
	3	TLG Positive LN + low metab			24	MFFS			0.034	
	3	TLG Positive LN + low metab			24	OS			0.045	
Min 2016	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Log-rank test		0.002	
	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Univariate an		0.021	3.174
	3	SUVmax	≤ 4.25	> 4.25	24	DFS	Multivariate an		0.058	2.853
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Log-rank test		0.029	
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Univariate an		0.042	3.657
	3	SUVmax	≤ 4.25	> 4.25	24	LRFS	Multivariate an		0.211	2.365
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Log-rank test		0.284	
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Univariate an		0.294	1.863
	3	SUVmax	≤ 4.25	> 4.25	24	MFFS	Multivariate an		0.455	1.650
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Log-rank test		0.143	
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Univariate an		0.153	2.011
	3	SUVmax	≤ 4.25	> 4.25	24	OS	Multivariate an		0.255	1.855
	3	MTV	≤ 3.3	> 3.3	24	DFS	Log-rank test		0.013	
	3	MTV	≤ 3.3	> 3.3	24	DFS	Univariate an		0.019	3.001
	3	MTV	≤ 3.3	> 3.3	24	DFS	Multivariate an		0.045	2.903
	3	MTV	≤ 3.3	> 3.3	24	LRFS	Log-rank test		0.04	
	3	MTV	≤ 3.3	> 3.3	24	LRFS	Univariate an		0.051	3.057
	3	MTV	≤ 3.3	> 3.3	24	LRFS	Multivariate an		0.236	2.148
	3	MTV	≤ 3.3	> 3.3	24	MFFS	Log-rank test		0.358	
	3	MTV	≤ 3.3	> 3.3	24	MFFS	Univariate an		0.366	1.657
	3	MTV	≤ 3.3	> 3.3	24	MFFS	Multivariate an		0.414	1.716
	3	MTV	≤ 3.3	> 3.3	24	OS	Log-rank test		0.267	
	3	MTV	≤ 3.3	> 3.3	24	OS	Univariate an		0.274	1.655
	3	MTV	≤ 3.3	> 3.3	24	OS	Multivariate an		0.391	1.607
	3	TLG	≤ 9.4	> 9.4	24	DFS	Log-rank test		0.002	
	3	TLG	≤ 9.4	> 9.4	24	DFS	Univariate an		0.004	4.734

Appendix G3 *continued*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	3	TLG	≤ 9.4	> 9.4	24	DFS	Multivariate an		0.012	4.638
	3	TLG	≤ 9.4	> 9.4	24	LRFS	Log-rank test		0.003	
	3	TLG	≤ 9.4	> 9.4	24	LRFS	Univariate an		0.011	6.767
	3	TLG	≤ 9.4	> 9.4	24	LRFS	Multivariate an		0.062	4.636
	3	TLG	≤ 9.4	> 9.4	24	MFFS	Log-rank test		0.212	
	3	TLG	≤ 9.4	> 9.4	24	MFFS	Univariate an		0.225	2.055
	3	TLG	≤ 9.4	> 9.4	24	MFFS	Multivariate an		0.402	1.76
	3	TLG	≤ 9.4	> 9.4	24	OS	Log-rank test		0.068	
	3	TLG	≤ 9.4	> 9.4	24	OS	Univariate an		0.078	2.388
	3	TLG	≤ 9.4	> 9.4	24	OS	Multivariate an		0.173	2.182
	3	OC pre- & midtreatment	≤ OC	> OC	24	DFS	Log-rank test		0.001	
	3	OC pre- & midtreatmen	≤ OC	> OC	24	DFS	Univariate an		0.004	4.734
	3	OC pre- & midtreatmen	≤ OC	> OC	24	DFS	Multivariate an		0.011	3.388
	3	OC pre- & midtreatmen	≤ OC	> OC	24	LRFS	Log-rank test		0.007	
	3	OC pre- & midtreatmen	≤ OC	> OC	24	LRFS	Univariate an		0.011	6.767
	3	OC pre- & midtreatmen	≤ OC	> OC	24	LRFS	Multivariate an		0.103	2.494
	3	OC pre- & midtreatmen	≤ OC	> OC	24	MFFS	Log-rank test		0.059	
	3	OC pre- & midtreatmen	≤ OC	> OC	24	MFFS	Univariate an		0.225	2.055
	3	OC pre- & midtreatmen	≤ OC	> OC	24	MFFS	Multivariate an		0.114	2.903
	3	OC pre- & midtreatmen	≤ OC	> OC	24	OS	Log-rank test		0.042	
	3	OC pre- & midtreatmen	≤ OC	> OC	24	OS	Univariate an		0.078	2.388
	3	OC pre- & midtreatmen	≤ OC	> OC	24	OS	Multivariate an		0.092	2.553
Min 2016	3	FDG-PET	CMR	Non-CMR	24	LRFS	Log rank		0.062	
	3	FDG-PET	CMR	Non-CMR	24	LRFS	Univariate an		0.085	3.684
	3	FDG-PET	CMR	Non-CMR	24	LRFS	Multivar		0.175	3.297
	3	FDG-PET	CMR	Non-CMR	24	DFS	Log rank		0.132	
	3	FDG-PET	CMR	Non-CMR	24	DFS	Univariate an		0.148	2.234
	3	FDG-PET	CMR	Non-CMR	24	DFS	Multivar an		0.216	2.333
	3	FDG-PET	CMR	Non-CMR	24	OS	Log rank		0.328	
	3	FDG-PET	CMR	Non-CMR	24	OS	Univariate an		0.337	1.732
	3	FDG-PET	CMR	Non-CMR	24	OS	Multivariate an		0.240	2.369
	3	FDG-PET – Primary and nodal	CMR	Non-CMR	24	LRFS			0.038	
	3	FDG-PET – Primary and nodal	CMR	Non-CMR	24	DFS			0.208	
	3	FDG-PET – Primary and nodal	CMR	Non-CMR	24	OS			0.370	
	3	FDG-PET – Nodal	CMR	Non-CMR	24	LRFS			0.033	
	3	FDG-PET – Visual assessment PT FDG-PET	CMR	Non-CMR	24	LRFS			0.062	
	3	FDG-PET	CMR	Non-CM	24	LRFS - grade 1		88	13	23
	3	FDG-PET	CMR	Non-CM	24	LRFS - grade 2		88	21	25
	3	FDG-PET	CMR	Non-CM	24	LRFS - grade 3		88	38	30
	3	FDG-PET	CMR	Non-CM	24	DFS - grade 1		86	13	30
	3	FDG-PET	CMR	Non-CM	24	DFS - grade 2		81	19	30
	3	FDG-PET	CMR	Non-CM	24	DFS - grade 3		81	38	36
Pollom 2016	0-4	FDG – MTV red 50%	≤2			PFS	Cox(univariate)		<0.001	1.24
	0-4	FDG – MTV red 50%	≤2			OS	Cox (univariate)		<0.001	1.21
	0-4	FDG – MTV red 50%	≤2			PFS	Cox (univariate)		<0.001	1.26
	0-4	FDG – MTV red 50%	≤2			OS	Cox (univariate)		<0.001	1.22
	0-4	FDG – MTV red 50%	≤2			PFS	Cox (univariate)		0.015	
	0-4	FDG – MTV red 50%	≤2			PFS	Cox (univariate)		0.16	
	0-4	FDG – MTV red 50%	≤2			PFS	Cox (univariate)		0.007	1.2
	4	TLG - primary	≤2			PFS	Cox (univariate)		0.3	
	4	TLG – nodal	≤2			PFS	Cox (univariate)		0.004	
	4	TLG - combined	≤2			PFS	Cox (univariate)		0.007	
	4	TLG - primary	≤2			OS	Cox (univariate)		0.9	

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	4	TLG - nodal	≤2			OS	Cox (univariate)		0.06	
	4	TLG - combined	≤2			OS	Cox (univariate)		0.09	
	4	TLG in HPV - primary	≤2			PFS	Cox (univariate)		0.3	
	4	TLG in HPV - nodal	≤2			PFS	Cox (univariate)		0.003	
	4	TLG in HPV – combined	≤2			PFS	Cox (univariate)		0.005	
	4	TLG in HPV - primary	≤2			OS	Cox (univariate)		0.6	
	4	TLG in HPV - nodal	≤2			OS	Cox (univariate)		0.07	
	4	TLG in HPV - combined	≤2			OS	Cox (univariate)		0.07	
	4	TLG – Velocity >5% decrease	≤2			PFS	Cox (univariate)		0.04	0.37
	4	TLG HPV - Velocity >5% decrease	≤2			PFS	Cox (univariate)		0.005	0.24
	4	Volume + TLG	Volume delta	TLG delta					<0.001	0.57
	4	Volume nodal velocity				PFS	Cox(univariate)		0.5	
	4	Volume nodal velocity				OS	Cox (univariate)		0.8	
	4	Volume nodal velocity >1%decrease				PFS	Cox (univariate)		0.5	
	4	Volume nodal velocity >1%decrease				OS	Cox (univariate)		0.5	
	4	Baseline smoking >10yr				PFS	Cox (univariate)		0.49	0.71
	4	Baseline smoking >10yr				OS	Cox (univariate)		0.93	1.06
	4	Baseline ECOG >1				PFS	Cox (univariate)		0.32	2.12
	4	Baseline ECOG >1				OS	Cox (univariate)		0.09	3.91
	4	Baseline HPV pos	<T4	<T4		PFS	Cox (univariate)		0.66	0.76
	4	Baseline HPV pos	<T4	<T4		OS	Cox (univariate)		0.52	0.6
	4	Baseline T stadium	<N3	>N3		PFS	Cox (univariate)		0.02	2.9
	4	Baseline T stadium	<N3	>N3		OS	Cox (univariate)		0.21	2.33
	4	Baseline N stadium	Platinum	Cetuximab		PFS	Cox (univariate)		0.46	1.41
	4	Baseline N stadium	Platinum	Cetuximab		OS	Cox (univariate)		0.32	1.92
	4	Baseline Chemo				PFS	Cox (univariate)		0.22	0.57
	4	Baseline Chemo				OS	Cox (univariate)		0.15	0.36
	4	Baseline IC				PFS	Cox (univariate)		0.04	2.84
	4	Baseline IC				OS	Cox (univariate)		0.07	3.22
	4	MTV2.0				PFS	Cox (univariate)		<0.001	1.24
	4	MTV2.0				OS	bivariate analysis		0.009	1.21
	4	SUVmax per 10cc				PFS	bivariate analysis		0.96	0.98
	4	SUVmax per 10cc				OS	bivariate analysis		0.18	0.21
	4	MTV2.0 (10 cc) + ICC				PFS	bivariate analysis		<0.001	1.25
	4	MTV2.0 (10 cc) + ICC				OS	bivariate analysis		0.003	1.29
	4	MTV2.0 + IC				PFS	bivariate analysis		0.02	3.37
	4	MTV2.0 + IC				OS	bivariate analysis		0.01	6.18
	4	nodal TLG + IC				PFS	bivariate analysis		0.01	1.04
	4	nodal TLG + IC				OS	bivariate analysis		0.07	1.04
	4	nodal TLG + IC				PFS	bivariate analysis		0.03	3
	4	nodal TLG + IC				OS	bivariate analysis		0.03	3.91
	4	MTV2.0 + Tstage	T4	<T4		PFS	bivariate analysis		0.009	1.2
	4	MTV2.0 + Tstage	T4	<T4		OS	bivariate analysis		0.03	1.22
	4	MTV2.0 + Tstage	T4	<T4		PFS	bivariate analysis		0.4	1.53
	4	MTV2.0 + Tstage	T4	<T4		OS	bivariate analysis		0.8	0.82
	4	nodal TLG + Tstage	T4	<T4		PFS	bivariate analysis		0.04	1.04
	4	nodal TLG + Tstage	T4	<T4		OS	bivariate analysis		0.2	1.03
	4	nodal TLG + Tstage	T4	<T4		PFS	bivariate analysis		0.3	1.79
	4	nodal TLG + Tstage	T4	<T4		OS	bivariate analysis		0.6	1.48
Wong, 2017	0	TLG40%	LRC	LRF			Mann–Whitney U test		0.192	
	1	Delta TLG40%	LRC	LRF			Mann–Whitney U test		0.007	0.825
	2	Delta TLG40%	LRC	LRF			Mann–Whitney U test		0.131	
	0	SUVmax	LRC	LRF			Mann–Whitney U test		0.269	

Appendix G3 *continued*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	1	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.034	0.764
	2	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.158	
	0	TLG	LRC	LRF			Mann–Whitney U test		0.104	
	1	Delta TLG	LRC	LRF			Mann–Whitney U test		0.069	
	2	Delta TLG	LRC	LRF			Mann–Whitney U test		0.310	
	0	SUVmax	LRC	LRF			Mann–Whitney U test		0.498	
	1	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.050	
	2	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.224	
	0	TLG	LRC	LRF			Mann–Whitney U test		0.532	
	1	Delta TLG	LRC	LRF			Mann–Whitney U test		0.085	
	2	Delta TLG	LRC	LRF			Mann–Whitney U test		0.790	
	0	SUVmax	LRC	LRF			Mann–Whitney U test		0.859	
	1	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.391	
	2	Delta SUVmax	LRC	LRF			Mann–Whitney U test		0.533	
Zips, 2012	1	FDG – volumetry delineated				LRC	Univariate		0.34	
	1	CT-delineated volume				LRC	Multivariate HR		1.0	
	1	HPV status				LRC	Univariate		0.24	
	1	FDG – hypoxic volume	≤1.4	>1.4			Univariate HR Cox		0.047	1.02
	1	FDG – hypoxic volume	≤1.6	>1.6			Univariate HR		0.047	1.02
	2	FDG – hypoxic volume	≤1.8	>1.8			Univariate HR		0.040	1.03
	2	FDG – hypoxic volume	≤2.0	>2.0			Univariate HR		0.024	1.06
	2	FDG – hypoxic volume	≤1.4	>1.4			Univariate HR		0.035	1.02
	2	FDG – hypoxic volume	≤1.6	>1.6			Univariate HR		0.021	1.03
	1	FDG – hypoxic volume	≤1.8	>1.8			Univariate HR		0.011	1.07
	1	FDG – hypoxic volume	≤2.0	>2.0			Univariate HR		0.013	1.14
	1	FDG – hypoxic volume	≤1.4	>1.4			Multivariate HR		0.068	1.02
	2	FDG – hypoxic volume	≤1.6	>1.6			Multivariate HR		0.033	1.04
	2	FDG – hypoxic volume	≤1.8	>1.8			Multivariate HR		0.023	1.07
	2	FDG – hypoxic volume	≤2.0	>2.0			Multivariate HR		0.020	1.12
	2	FDG – hypoxic volume	≤1.4	>1.4			Multivariate HR		0.042	1.05
	1	FDG – hypoxic volume	≤1.6	>1.6			Multivariate HR		0.045	1.08
	1	FDG – hypoxic volume	≤1.8	>1.8			Multivariate HR		0.042	1.12
	2	FDG – hypoxic volume	≤2.0	>2.0			Multivariate HR		0.047	1.62
	1	FMISO (8-10Gy) SUVmax					Univariate HR		0.024	2.72
	2	FMISO (18-20Gy) SUVmax					Univariate HR		0.015	3.85
	1	FMISO (8-10Gy) SUVmax					Multivariate HR		0.06	3.79
	1	FMISO (18-20Gy) SUVmax					Multivariate HR		0.169	3.72
	2	FMISO TBRmax					Univariate HR		0.025	2.76
	1	FMISO TBRmax					Univariate HR		0.008	4.25
	2	FMISO TBRmax					Multivariate HR		0.04	6.04
	2	FMISO TBRmax					Multivariate HR		0.061	6.19
	1	FMISO >= 2.16	≤2.16	>2.16		LPFS	Log rank		0.236	
	2	FMISO ->=1.93	≤1.93	>1.93		LPFS	Log rank		0.001	
	0	FMISO TBRmax ≥2.23	≤2.23	>2.23		LPFS	Log rank		0.937	
Wiedenmann, 2015	0-2	FMISOPET mean TBRmax	Pretreatment	Wk2			Paired Wilcoxon		0.035	
	0-2	FMISOPET mean TBRmax	Pretreatment	Wk2			Paired Wilcoxon		0.061	
	0-5	FMISOPET mean TBRmax	Pretreatment	Wk5			Paired Wilcoxon		0.003	
		FMISOPET mean TBRmax	Recurrence	Without recur.			Mann-whitney U		0.002	
	0	FMISOPET mean TBRmax basel	LRF	LRC			Log rank		0.031	
	2	FMISOPET mean TBRmax 2wk	LRF	LRC			Log rank		0.016	
	0	FDG TBRmax vs FMISOhyp basel	FDG TBRmax	FMISO TBR			Pearson correlation		0.714	
		FDT PT survival					Log rank		0.568	
		TBRmax LN survival		TBRmax			Log rank		0.935	

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
		TBRmax PT volume survival		TBRmax			Log rank		0.697	
	0	FMISOPET mean TBRmax	LRC	TBRmax			Log rank		0.031	
	2	FMISOPET mean TBRmax	LRC				Log rank		0.0164	
	5	FMISOPET mean TBRmax	LRC				Log rank		0.0719	
Zschaeck, 2015	1.2	FMISO PET volume variation	Pre	Wk 1 or 2					NS	
Lee, 2016	0-1	FMISO PET	LRC							100%
Grkovski, 2017	0-2	FMISO PET TBR	Pretreatment	Wk2			Unp. Stud.t		<0.01	
	0-2	FMISO PET K3	Pretreatment	Wk2			Unp. Stud.t		<0.01	
	0-2	FMISO PET distribution volume	Pretreatment	Wk2			Unp. Stud.t		<0.01	
	0-2	FMISO PET K1	Pretreatment	Wk2			Unp. Stud.t		0.20	
	0-2	FMISO PET v_B	Pretreatment	Wk2			Unp. Stud.t		0.28	
	0-2	FMISO PET deltak3 – deltaTBR	deltak3	deltaTBR			Pearson		<0.01	r=0.70
	0-2	FMISO PET deltaDV	deltak3	deltaDV			Pearson		<0.01	r=-0.56
	0-2	FMISO PET deltaTBR	wk 0	wk2			Pearson		<0.01	r=0.58
	0-2	FMISO PET deltak3	wk 0	wk2			Pearson		<0.01	r=0.53
	0-2	FMISO PET deltak1	wk 0	wk2			Pearson		<0.01	r=0.41
	0-2	FMISO PET deltav$_B$	wk 0	wk2			Pearson		<0.01	r=0.39
	0-2	FMISO PET TBR - deltaTBR	wk 0	deltaTBR			Pearson		<0.01	r=-0.9
	0-2	FMISO PET k3 – deltak3	wk 0	deltak3			Pearson		<0.01	r=-0.75
	0-2	FMISO PET DV - deltaDV	wk 0	deltaDV			Pearson		<0.01	r=-0.53
	0-2	FMISO PET K1 – deltaK1	wk 0	deltaK1			Pearson		<0.01	r=0.43
	0-2	FMISO PET v_B - deltav$_B$	wk 0	deltav$_B$			Pearson		<0.01	r=-0.48
Lock, 2017	0	FMISO PET $HV_{1.6}$	Cohort 1	Cohort 2			Log rank		0.008	
	0	FMISO PET TBR_{peak}	Cohort 1	Cohort 2			Log rank		0.093	
	0-1	FMISO PET $HV_{1.6}$	wk 0	wk 1			Log rank		0.01	
	0-1	FMISO PET TBR_{peak}	wk 0	wk 1			Log rank		0.14	
	0-2	FMISO PET $HV_{1.6}$	wk 0	wk 2			Log rank		0.008	
	0-2	FMISO PET TBR_{peak}	wk 0	wk 2			Log rank		0.017	
	1	FMISO PET $HV_{1.6}$	$HV_{1.6}$				Univ. cox		0.021	
	1	FMISO PET TBR_{peak}	TBR_{peak}				Univ. cox		0.019	
	2	FMISO PET $HV_{1.6}$	$HV_{1.6}$				Univ. cox		0.032	
	2	FMISO PET TBR_{peak}	TBR_{peak}				Univ. cox		0.012	
Troost. 2010	0-2	FLT SUVmax PT					Friedman t. Wilcox		<0.0001	
	0-4	FLT SUVmax PT					Friedman t. Wilcox		<0.0001	
	0-2	FLT SUVmax LN					Friedman t. Wilcox		<0.0001	
	0-4	FLT SUVmax LN					Friedman t. Wilcox		<0.001	
	0-2	FLT SUVmax PT					Friedman t. Wilcox		<0.0001	
	0-4	FLT SUVmean PT					Friedman t. Wilcox		<0.0001	
	0-2	FLT SUVmean LN					Friedman t. Wilcox		<0.001	
	0-4	FLT SUVmean LN					Friedman t. Wilcox		<0.001	
	0-2	FLT deltaSUVmax PT					Friedman t. Wilcox	-		55%
	2-4	FLT deltaSUVmax PT					Friedman t. Wilcox	-		34%
	0-2	FLT deltaSUVmax LN					Friedman t. Wilcox	-		44%
	2-4	FLT deltaSUVmax LN					Friedman t. Wilcox	-		47%
Kishino. 2012	4	FDG - SUVmax – all lesions	LRC	LRF	39		Accuracy	30		
	4	FDG - SUVmax – PT	LRC	LRF	39		Accuracy	19		
	4	FDG – SUVmax – LN	LRC	LRF	39		Accuracy	41		
	4	FLT – SUVmax – all lesions	LRC	LRF	39		Accuracy	74		
	4	FLT – SUVmax – PT	LRC	LRF	39		Accuracy	65		
	4	FLT – SUVmax - LN	LRC	LRF	39		Accuracy	71		
	0-4	SUV FLT during T			39				<0.001	
	4	SUV FDG during T			39				<0.01	
	4	PT SUV disappearance			39				<0.001	

Appendix G3 *continued*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
Hoeben. 2013	0	TNM-stage	T3	T4	36	LRC			0.004	
	0	TNM-stage	N0-N1	N2c	36	LRC			0.028	
	0	Age	≤60y	>60	36	LRC			0.041	
	0	TNM-stage and GTV on CT	T-stage	segm GTV CT	36	LRC			0.003	
	0	TNM-stage and GTV on PET	T-stage	segm GTV CT	36	LRC			0.007	
	0	FLT – SUVmax in all pts	≥6.1	<6.1	36	LRC			0.17	
	0	FLT – SUVmax in all pts	≥6.1	<6.1	36	DFS			0.12	
	0	FLT – SUVmax in all pts	≥6.1	<6.1	36	OS			0.52	
	0-2	FLT - Decrease SUVmax in all pts	<6.5 cm3	36	36	LRC			0.066	
	0-2	FLT - Decrease SUVmax in all pts	<6.5 cm3	36	36	DFS			0.035	
	0-2	FLT - Decrease SUVmax in all pts	<6.5 cm3	36	36	OS			0.23	
	0-2	FLT – SUVmax in RT	≥6.6	<6.6	36	LRC			0.075	
	0-2	FLT - SUVmax in RT	≥6.6	<6.6	36	DFS			0.044	
	0-2	FLT - SUVmax in RT	≥6.6	<6.6	36	OS			0.590	
	0-2	FLT - SUVmax in CRT	≥ 45%	<45%	36	LRC			0.044	
	0-2	FLT - SUVmax in CRT	≥ 45%	<45%	36	DFS			0.044	
	0-2	FLT - SUVmax in CRT	≥ 45%	<45%	36	OS			0.080	
	0-2	FLT - Δ decrease SUVmax - CRT	45% decrease	2e scan	36	LRC			0.081	
	0-2	FLT - Δ decrease SUVmax - CRT	45% decrease	2e scan	36	DFS			0.081	
	0-2	FLT - Δ decrease SUVmax - CRT	45% decrease	2e scan	36	OS			0.1	
	0-2	FLT - Δ GTV visual delineated	≥31%	<31%	36	LRC			0.63	
	0-2	FLT - Δ GTV visual delineated	≥31%	<31%	36	DFS			0.63	
	0-2	FLT - Δ GTV visual delineated	≥31%	<31%	36	OS			0.66	
	0	GTV - CT - CRT	<6.5 cm3	≥6.5	36	LRC			0.039	
	0	GTV - CT - CRT	<6.5 cm3	≥6.5	36	DFS			0.039	
	0	GTV - CT - CRT	<6.5 cm3	≥6.5	36	OS			0.079	
	0-2	GTV visual delineated - CRT	<6.5 cm3	≥6.5	36	LRC			0.63	
	0-2	GTV visual delineated - CRT	<6.5 cm3	≥6.5	36	DFS			0.63	
	0-2	GTV visual delineated - CRT	<6.5 cm3	≥6.5	36	OS			0.66	
	0-2	Δ GTV visual delineated - CRT	≥31%	<31%	36	LRC			0.039	
	0-2	Δ GTV visual delineated - CRT	≥31%	<31%	36	DFS			0.039	
	0-2	Δ GTV visual delineated - CRT	≥31%	<31%	36	OS			0.079	
	0-4	Decrease Δ GTVvis - RT			36	LRC			0.0001	
Nyflot. 2015	0-3	FLT Δ SUVmean							0.008	
	0-3	FLT Δ SUVpeak							0.05	
	0-3	FLT Δ K_{FLT}							<0.001	
	0-5	Decline FLT Δ SUVmean							<0.001	
	0-5	Decline FLT Δ SUVpeak							<0.001	
	0-5	Decline FLT Δ KFLT							0.90	
	0-3	Cu-ATSM Δ SUVmean							0.046	
	0-3	Cu-ATSM Δ SUVpeak							0.24	
	0-5	Cu-ATSM Δ SUVmean							0.048	
	0-5	Cu-ATSM Δ SUVpeak							0.11	
	3	Enhancement HU							0.02	
	3	Perfusion bloodflow							0.09	
	3	perfusion bloodvolume							0.10	
	3	perfusion meantransit							0.93	
	3	perfusion permeability							0.95	
Lin. 2017	3	FDG - SUVmax –Index LN	≤4.5	>4.5	28	LRFFS	Log-rank test		0.449	
	3	FDG - SUVmax–Index LN	≤4.5	>4.5	28	RFFS	Log-rank test		0.450	
	3	FDG - SUVmax–Index LN	≤4.5	>4.5	28	DMFFS	Log-rank test		0.985	
	3	FDG - SUVmax–Index LN	≤4.5	>4.5	28	DFS	Log-rank test		0.206	
	3	FDG - SUVmax–Index LN	≤4.5	>4.5	28	OS	Log-rank test		0.138	

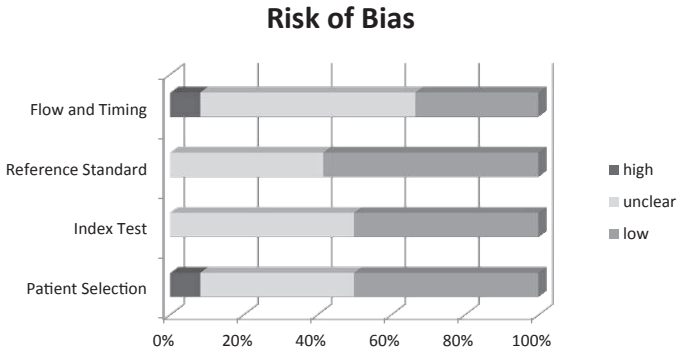
Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	3	FDG - SUVmean–Index LN	≤2.95	>2.95	28	LRFFS	Log-rank test		0.178	
	3	FDG - SUVmean–Index LN	≤2.95	>2.95	28	RFFS	Log-rank test		0.343	
	3	FDG - SUVmean–Index LN	≤2.95	>2.95	28	DMFFS	Log-rank test		0.454	
	3	FDG - SUVmean–Index LN	≤2.95	>2.95	28	DFS	Log-rank test		0.041	
	3	FDG - SUVmean–Index LN	≤2.95	>2.95	28	OS	Log-rank test		0.032	
	3	FDG - MTV–Index LN	≤2.85	>2.85	28	LRFFS	Log-rank test		0.228	
	3	FDG - MTV–Index LN	≤2.85	>2.85	28	RFFS	Log-rank test		0.167	
	3	FDG - MTV–Index LN	≤2.85	>2.85	28	DMFFS	Log-rank test		0.561	
	3	FDG - MTV–Index LN	≤2.85	>2.85	28	DFS	Log-rank test		0.107	
	3	FDG - MTV–Index LN	≤2.85	>2.85	28	OS	Log-rank test		0.021	
	3	FDG - TLG–Index LN	≤9.0	>9.0	28	LRFFS	Log-rank test		0.777	
	3	FDG - TLG–Index LN	≤9.0	>9.0	28	RFFS	Log-rank test		0.359	
	3	FDG - TLG–Index LN	≤9.0	>9.0	28	DMFFS	Log-rank test		0.837	
	3	FDG - TLG–Index LN	≤9.0	>9.0	28	DFS	Log-rank test		0.349	
	3	FDG - TLG–Index LN	≤9.0	>9.0	28	OS	Log-rank test		0.122	
	3	FDG - SUVmax – total LNs	≤5.05	>5.05	28	LRFFS	Log-rank test		0.031	
	3	FDG - SUVmax– total LNs	≤5.05	>5.05	28	RFFS	Log-rank test		0.019	
	3	FDG - SUVmax– total LNs	≤5.05	>5.05	28	DMFFS	Log-rank test		0.138	
	3	FDG - SUVmax– total LNs	≤5.05	>5.05	28	DFS	Log-rank test		0.011	
	3	FDG - SUVmax– total LNs	≤5.05	>5.05	28	OS	Log-rank test		0.007	
	3	FDG - SUVmean– total LNs	≤3.25	>3.25	28	LRFFS	Log-rank test		0.003	
	3	FDG - SUVmean– total LNs	≤3.25	>3.25	28	RFFS	Log-rank test		0.012	
	3	FDG - SUVmean– total LNs	≤3.25	>3.25	28	DMFFS	Log-rank test		0.328	
	3	FDG - SUVmean– total LNs	≤3.25	>3.25	28	DFS	Log-rank test		0.005	
	3	FDG - SUVmean– total LNs	≤3.25	>3.25	28	OS	Log-rank test		0.002	
	3	FDG - MTV(cm3) – total LNs	≤5.23	>5.23	28	LRFFS	Log-rank test		0.368	
	3	FDG - MTV(cm3) – total LNs	≤5.23	>5.23	28	RFFS	Log-rank test		0.099	
	3	FDG - MTV(cm3) – total LNs	≤5.23	>5.23	28	DMFFS	Log-rank test		0.837	
	3	FDG - MTV(cm3) – total LNs	≤5.23	>5.23	28	DFS	Log-rank test		0.147	
	3	FDG - MTV(cm3) – total LNs	≤5.23	>5.23	28	OS	Log-rank test		0.112	
	3	FDG - TLG– total LNs	≤14.0	>14.0	28	LRFFS	Log-rank test		0.24	
	3	FDG - TLG– total LNs	≤14.0	>14.0	28	RFFS	Log-rank test		0.038	
	3	FDG - TLG– total LNs	≤14.0	>14.0	28	DMFFS	Log-rank test		0.561	
	3	FDG - TLG– total LNs	≤14.0	>14.0	28	DFS	Log-rank test		0.116	
	3	FDG - TLG– total LNs	≤14.0	>14.0	28	OS	Log-rank test		0.077	
	3	FDG - SUVmax reduct- Index LN	≥50%	<50%	28	LRFFS	Log-rank test		0.235	
	3	FDG - SUVmax reduct- Index LN	≥50%	<50%	28	RFFS	Log-rank test		0.226	
	3	FDG - SUVmax reduct- Index LN	≥50%	<50%	28	DMFFS	Log-rank test		0.867	
	3	FDG - SUVmax reduct- Index LN	≥50%	<50%	28	DFS	Log-rank test		0.283	
	3	FDG - SUVmax reduct- Index LN	≥50%	<50%	28	OS	Log-rank test		0.158	
	3	FDG – SUVmean reduc–Index LN	≥50%	<50%	28	LRFFS	Log-rank test		0.503	
	3	FDG – SUVmean reduct –Index LN	≥50%	<50%	28	RFFS	Log-rank test		0.551	
	3	FDG - SUVmean reduct –Index LN	≥50%	<50%	28	DMFFS	Log-rank test		0.61	
	3	FDG - SUVmean reduct –Index LN	≥50%	<50%	28	DFS	Log-rank test		0.336	
	3	FDG - SUVmean reduct –Index LN	≥50%	<50%	28	OS	Log-rank test		0.279	
	3	FDG - MTV reduct –Index LN	≥50%	<50%	28	LRFFS	Log-rank test		0.0407	
	3	FDG - MTV reduct –Index LN	≥50%	<50%	28	RFFS	Log-rank test		0.214	
	3	FDG - MTV reduct –Index LN	≥50%	<50%	28	DMFFS	Log-rank test		0.45	
	3	FDG - MTV reduct –Index LN	≥50%	<50%	28	DFS	Log-rank test		0.123	
	3	FDG - MTV reduct –Index LN	≥50%	<50%	28	OS	Log-rank test		0.036	
	3	FDG - TLG reduct –Index LN	≥50%	<50%	28	LRFFS	Log-rank test		0.024	
	3	FDG - TLG reduct –Index LN	≥50%	<50%	28	RFFS	Log-rank test		0.044	
	3	FDG - TLG reduct –Index LN	≥50%	<50%	28	DMFFS	Log-rank test		0.125	

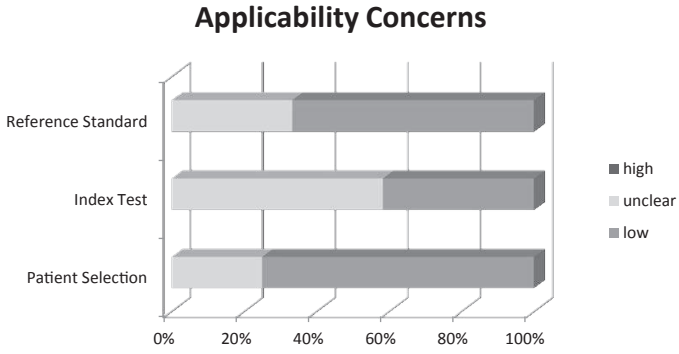
Appendix G3 *continued*

Author, year	Time Imaging	Prognostic parameter	Group 1	Group 2	Follow-up	Outcome	Statistical Test	P value	Hazard ratio	CI 95%
	3	FDG - TLG reduct –Index LN	≥50%	<50%	28	DFS	Log-rank test		0.389	
	3	FDG - TLG reduct –Index LN	≥50%	<50%	28	OS	Log-rank test		0.157	
	3	FDG – SUVmax reduct – total LNs	≥50%	<50%	28	LRFFS	Log-rank test		0.143	
	3	FDG – SUVmax reduct – total LNs	≥50%	<50%	28	RFFS	Log-rank test		0.099	
	3	FDG – SUVmax reduct – total LNs	≥50%	<50%	28	DMFFS	Log-rank test		0.835	
	3	FDG – SUVmax reduct – total LNs	≥50%	<50%	28	DFS	Log-rank test		0.228	
	3	FDG – SUVmax reduct – total LNs	≥50%	<50%	28	OS	Log-rank test		0.106	
	3	FDG – SUVmean reduct – total LNs	≥50%	<50%	28	LRFFS	Log-rank test		0.321	
	3	FDG – SUVmean reduct – total LNs	≥50%	<50%	28	RFFS	Log-rank test		0.272	
	3	FDG – SUVmean reduct – total LNs	≥50%	<50%	28	DMFFS	Log-rank test		0.907	
	3	FDG – SUVmean reduct – total LNs	≥50%	<50%	28	DFS	Log-rank test		0.258	
	3	FDG – SUVmean reduct – total LNs	≥50%	<50%	28	OS	Log-rank test		0.185	
	3	FDG – MTV reduct – total LNs	≥50%	<50%	28	LRFFS	Log-rank test		0.050	
	3	FDG – MTV reduct – total LNs	≥50%	<50%	28	RFFS	Log-rank test		0.220	
	3	FDG – MTV reduct – total LNs	≥50%	<50%	28	DMFFS	Log-rank test		0.629	
	3	FDG – MTV reduct – total LNs	≥50%	<50%	28	DFS	Log-rank test		0.011	
	3	FDG – MTV reduct – total LNs	≥50%	<50%	28	OS	Log-rank test		0.009	
	3	FDG – TLG reduct – total LNs	≥50%	<50%	28	LRFFS	Log-rank test		0.002	
	3	FDG – TLG reduct – total LNs	≥50%	<50%	28	RFFS	Log-rank test		0.032	
	3	FDG – TLG reduct – total LNs	≥50%	<50%	28	DMFFS	Log-rank test		0.94	
	3	FDG – TLG reduct – total LNs	≥50%	<50%	28	DFS	Log-rank test		0.005	
	3	FDG – TLG reduct – total LNs	≥50%	<50%	28	OS	Log-rank test		0.002	
	3	FDG - SUVmean–Index LN	≥50%	<50%	28	LRFFS	Multivariate analysis		0.113	
	3	FDG - SUVmean–Index LN	≥50%	<50%	28	RFFS	Multivariate analysis		0.277	
	3	FDG - SUVmean–Index LN	≥50%	<50%	28	DFS	multivariate analysis		0.033	
	3	FDG - SUVmean–Index LN	≥50%	<50%	28	OS	multivariate analysis		0.003	
	3	FDG - MTV reduction –Index LN	≥50%	<50%	28	LRFFS	multivariate analysis		0.008	
	3	FDG - MTV reduction –Index LN	≥50%	<50%	28	RFFS	multivariate analysis		0.078	
	3	FDG - MTV reduction –Index LN	≥50%	<50%	28	DFS	multivariate analysis		0.027	
	3	FDG - MTV reduction –Index LN	≥50%	<50%	28	OS	multivariate analysis		0.038	
	3	FDG - TLG reduction –Index LN	≤5.05	>5.05	28	LRFFS	multivariate analysis		0.011	
	3	FDG - TLG reduction –Index LN	≤5.05	>5.05	28	RFFS	multivariate analysis		0.029	
	3	FDG - TLG reduction –Index LN	≤5.05	>5.05	28	DFS	multivariate analysis		0.204	
	3	FDG - TLG reduction –Index LN	≤5.05	>5.05	28	OS	multivariate analysis		0.198	
	3	FDG - SUVmax – total LNs	≤5.05	>5.05	28	LRFFS	multivariate analysis		0.131	
	3	FDG - SUVmax– total LNs	≤5.05	>3.25	28	RFFS	multivariate analysis		0.084	
	3	FDG - SUVmax– total LNs	≤3.25	>3.25	28	DFS	multivariate analysis		0.042	
	3	FDG - SUVmax– total LNs	≤3.25	>3.25	28	OS	multivariate analysis		0.026	
	3	FDG - SUVmean– total LNs	≥50%	<50%	28	LRFFS	multivariate analysis		0.003	
	3	FDG - SUVmean– total LNs	≥50%	<50%	28	RFFS	multivariate analysis		0.075	
	3	FDG - SUVmean– total LNs	≥50%	<50%	28	DFS	multivariate analysis		0.025	
	3	FDG - SUVmean– total LNs	≥50%	<50%	28	OS	multivariate analysis		0.014	
	3	FDG – MTV reduction – total LNs	≥50%	<50%	28	LRFFS	multivariate analysis		0.026	
	3	FDG - MTV reduction – total LNs	≥50%	<50%	28	RFFS	multivariate analysis		0.147	
	3	FDG - MTV reduction – total LNs	≥50%	<50%	28	DFS	multivariate analysis		0.003	
	3	FDG - MTV reduction – total LNs	≥50%	<50%	28	OS	multivariate analysis		0.014	
	3	FDG – TLG reduction– total LNs	≥50%	<50%	28	LRFFS	multivariate analysis		0.001	
	3	FDG - TLG reduction – total LNs	≥50%	<50%	28	RFFS	multivariate analysis		0.016	
	3	FDG - TLG reduction – total LNs	≥50%	<50%	28	DFS	multivariate analysis		0.001	
	3	FDG - TLG reduction – total LNs	≥50%	<50%	28	OS	multivariate analysis		0.004	

Appendix H *Quality assessment of diagnostic accuracy studies (QUADAS-2) of included treatment response studies. The risk of bias and applicability concerns are shown, which were scored as high (x), moderate (?) and low (✓).*

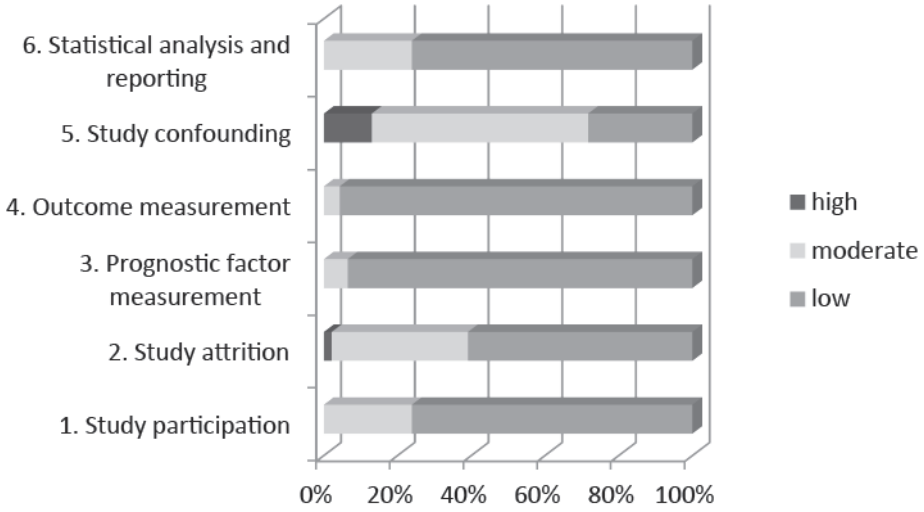
				Risk of Bias				Applicability Concerns		
				Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
CT	CTp	Rana	2015	x	?	?	✓	?	?	?
		Ursino	2016	?	✓	✓	?	?	✓	✓
MRI	DCE	Yoo	2012	✓	?	?	?	✓	?	✓
	DWI	Kim	2009	✓	✓	✓	?	✓	?	✓
		Galban	2009	?	✓	✓	✓	?	✓	✓
		King	2010	✓	✓	✓	X	✓	✓	✓
		Martins	2015	✓	?	?	?	✓	?	?
		Ding	2015	✓	?	?	?	✓	?	?
		Tyagi	2016	?	✓	✓	✓	✓	✓	✓
	IVIM	Paudyal	2016	✓	✓	?	✓	✓	✓	?
PET	FDG	Brun	1997	?	?	✓	?	✓	?	✓
		Brun	2002	?	?	✓	?	✓	?	✓





Appendix I *Quality in Prognostic Studies (QUIPS) of included long-term outcome studies The risk of bias and applicability concerns are shown, which were scored as high (x), moderate (?) and low (✓).*

				1. Study participation	2. Study attrition	3. Prognostic factor measurement	4. Outcome measurement	5. Study confounding	6. Statistical analysis and reporting
CT	CTp	Truong	2011	?	?	✓	✓	?	✓
		Abramyuk	2015	?	?	?	✓	?	✓
		Rana	2015	?	?	✓	?	?	✓
		Ursino	2016	?	✓	✓	✓	✓	✓
MRI	DCE	Cao	2008	?	✓	✓	✓	✓	?
		Yoo	2012	✓	✓	✓	✓	?	?
		Wang	2012	?	✓	✓	✓	?	✓
		Baer	2015	✓	?	✓	✓	?	✓
		Jansen	2016	✓	x	✓	✓	x	?
		Wong	2017	✓	✓	✓	✓	✓	✓
	DWI	Dirix	2009	?	?	✓	✓	?	?
		Kim	2009	✓	✓	✓	✓	?	✓
		Bhatia	2010	✓	?	✓	✓	?	?
		King	2010	✓	?	✓	✓	?	?
		Vandecaveye	2010	?	✓	✓	✓	?	✓
		King	2010	?	?	✓	✓	x	✓
		King	2013	✓	?	✓	✓	?	✓
		Hoang	2014	✓	✓	✓	?	✓	?
		Lambrecht	2014	✓	✓	✓	✓	?	✓
		Matoba	2014	✓	✓	✓	✓	✓	✓
		Schouden	2014	?	?	✓	✓	✓	?
		Tyagi	2016	?	✓	✓	✓	✓	?
	IVIM	Scalco	2016	✓	✓	?	✓	?	✓
		Marzi	2017	✓	?	✓	✓	?	✓
PET	FDG	Brun	2002	✓	?	✓	✓	x	✓
		Farrag	2010	✓	✓	✓	✓	?	✓
		Ceulemans	2011	✓	✓	✓	✓	?	✓
		Hentschel	2011	✓	?	✓	✓	?	✓
		Castaldi	2012	✓	✓	✓	✓	✓	✓
		Zips	2012	✓	?	✓	✓	?	✓
		Kishino	2012	✓	✓	✓	✓	?	✓
		Hoshikawa	2013	✓	✓	✓	✓	✓	✓
		Chen	2014	✓	✓	✓	✓	?	✓
		Min	2015	✓	✓	✓	✓	✓	✓
		Pollom	2016	✓	✓	✓	✓	?	✓
		Min	2016	✓	✓	✓	✓	x	✓
		Min	2016	✓	✓	✓	✓	✓	✓
		Lin	2017	✓	✓	✓	✓	?	✓
		Wong	2017	✓	✓	✓	✓	✓	✓
	FMISO	Wiedenmann	2015	✓	✓	?	✓	?	✓
		Grkovski	2017	✓	?	✓	✓	?	✓
		Zschaeck	2015	✓	?	✓	✓	✓	✓
		Lock	2017	✓	✓	✓	✓	?	?
	FLT	Hoeben	2013	✓	?	✓	✓	x	✓
		Byflot	2015	✓	✓	✓	✓	x	?
		Troost	2010	✓	✓	✓	✓	?	✓


6. Statistical analysis and reporting
5. Study confounding
4. Outcome measurement
3. Prognostic factor measurement
2. Study attrition
1. Study participation
0%
20%
40%
60%
80%
100%
high
moderate
low

APPENDIX CHAPTER 4.2

Appendix A *Quantitative image results of DWI and ^{18}F-FDG-PET-CT in the unknown primary group. Only parameters which were significantly different, with the Mann-Whitney U test, (P<0.05) between benign and malignant tissue were selected for ROC analysis. First sensitivity and specificity with the highest Youden Index are presented. In the last columns diagnostic accuracy is presented with a sensitivity of at least 90%*

Parameter	n	AUC (95%CI)	Cut-off	Sensitivity (%, 95%CI, ratio)	Specificity (%, 95%CI, ratio)	Youden Index	Cut-off, with sensitivity >90%	Sensitivity >90% (%, 95%CI, ratio)	Specificity, with sensitivity >90% (%, 95%CI, ratio)
SUV_{max} patient	31	0.900 (0.788-1.000)	4.56	81.3 (53.6-95.0, 13/16)	93.3 (66.0-99.7) 14/15	0.746	3.38	93.8 (67.8-99.7, 15/16)	53.3 (27.4-77.8, 8/15)
SUV_{max}/BP_{max} patient	31	0.906 (0.803-1.0000)	2.58	87.5 (60.4-97.8, 14/16)	86.7 (58.4-97.7, 13/15)	0.742	2.05	93.8 (67.8-99.7, 15/16)	66.7 (38.7-87.0, 10/15)
SUV_{max} lesion	128	0.811 (0.668-0.954)	4.58	68.8 (41.5-87.8, 11/16)	91.1 (83.8-95.4, 102/112)	0.599	1.96	93.8 (67.7-99.7, 15/16)	31.3 (18.0-32.7, 35/112)[a]
SUV_{max}/BP_{max} lesion	128	0.814 (0.678-0.950)	2.58	68.8 (41.5-87.8, 11/16)	89.3 (81.7-94.1, 100/112)	0.581	1.14	93.8 (67.7-99.7, 15/16)	35.7 (27.0-45.4, 40/112)
ADC_{volume} lesion	104	0.812 (0.683-0.942)	1.39 cm^2	66.7 (38.7-87.0, 10/15)[b]	90.0 (81.2-95.0, 80/89)	0.567	0.44 cm^2	93.3 (66.0-99.7, 14/15)	40.4 (30.3-51.4, 36/89)
ADC_{mean} tonsil	50	0.720 (0.553-0.887)	1.08 $\cdot 10^{-3}$ mm^2/sec	87.5 (46.7-99.3, 7/8)	59.5 (44.3-74.0, 25/42)[c]	0.470	1.21 $\cdot 10^{-3}$ mm^2/sec	100.0 (59.8-100.0, 8/8)	33.3 (20.0-49.6, 14/42)[d]
ADC_{min} tonsil	50	0.807 (0.669-0.944)	0.59 $\cdot 10^{-3}$ mm^2/sec	100.0 (59.8-100.0, 8/8)	54.8 (38.8-69.8, 23/42)[c]	0.548	0.59 $\cdot 10^{-3}$ mm^2/sec	100.0 (59.8-100.0, 8/8)	54.8 (38.8-69.8, 23/42)[d]
ADC_{median} tonsil	50	0.732 (0.581-0.883)	1.06 $\cdot 10^{-3}$ mm^2/sec	87.5 (46.7-99.3, 7/8)	59.5 (44.3-74.0, 25/42)[c]	0.470	1.18 $\cdot 10^{-3}$ mm^2/sec	100 (59.8-100, 8/8)	45.2 (30.2-61.1, 19/42)[d]
ADC_{volume} tonsil	50	0.845 (0.726-0.964)	0.73 cm^3	100.0 (59.8-100.0, 8/8)	64.3 (48.0-78.0, 27/42)	0.643	0.73 cm^3	100.0 (59.8-100.0, 8/8)	64.3 (48.0-78.0, 27/42)
SUV_{max} tonsil	60	0.945 (0.886-1.000)	4.58	100.0 (59.8-100.0, 7/8)	82.7 (69.2-91.3, 43/52)	0.827	0.945 (0.886-1.000)	100.0 (59.8-100.0, 8/8)	82.7 (69.2-91.3, 43/52)
SUV_{max}/BP_{max} tonsil	60	0.923 (0.852-0.994)	2.58	100.0 (59.8-100.0, 7/8)	80.8 (67.0-89.9, 42/52)	0.808	0.923 (0.852-0.994)	100.0 (59.8-100.0, 8/8)	80.8 (67.0-89.9, 42/52)

[a] Significantly lower than ADC_{volume} lesion (P=0.008)
[b] In one patient with a hypopharyngeal tumor we could not assess this subsite due to local image artifacts. We therefore excluded this lesion from the per-lesion analysis, because no ADC value could be obtained.
[c] Significantly lower than SUV_{max} tonsil: ADC_{mean} tonsil (P=0.039); ADC_{min} tonsil (P=0.021); ADC_{50} tonsil (P=0.039); ADC_{60} tonsil (P=0.022)
[d] Significantly lower than SUV_{max} tonsil: ADC_{mean} tonsil (P<0.001); ADC_{min} tonsil (P=0.021); ADC_{10} tonsil (P<0.001); ADC_{40} tonsil (P<0.001); ADC_{50} tonsil (P<0.001); ADC_{60} tonsil (P<0.001); ADC_{70} tonsil (P<0.001)

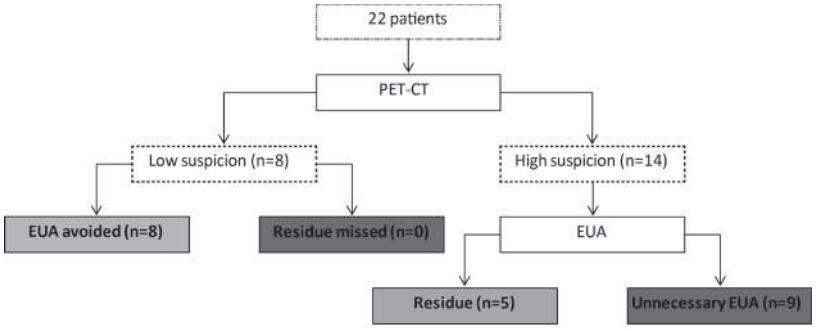
Abbreviations: AUC = area under the curve; BP = blood pool; YI = Youden Index; 95%CI = 95% confidence interval

APPENDICES CHAPTER 4.3

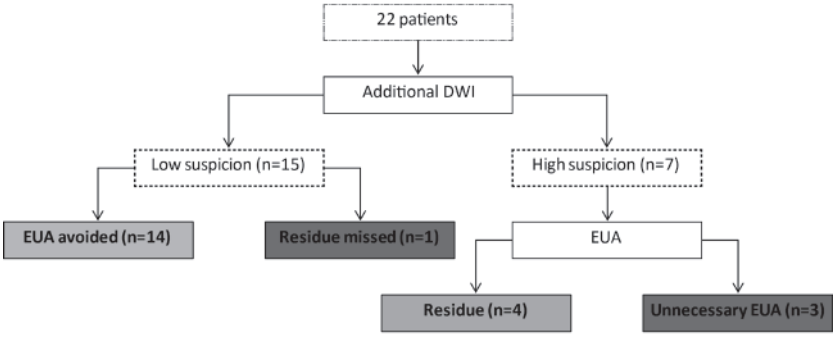
Appendix A *Comparison of DWI findings of both radiologists.*

	SOM score radiologist 2				
SOM score radiologist 1	**1**	**2**	**3**	**4**	**Total**
1	11	0	0	0	11
2	5	0	0	0	5
3	1	0	0	3	4
4	0	0	0	2	2
Total	17	0	0	5	22

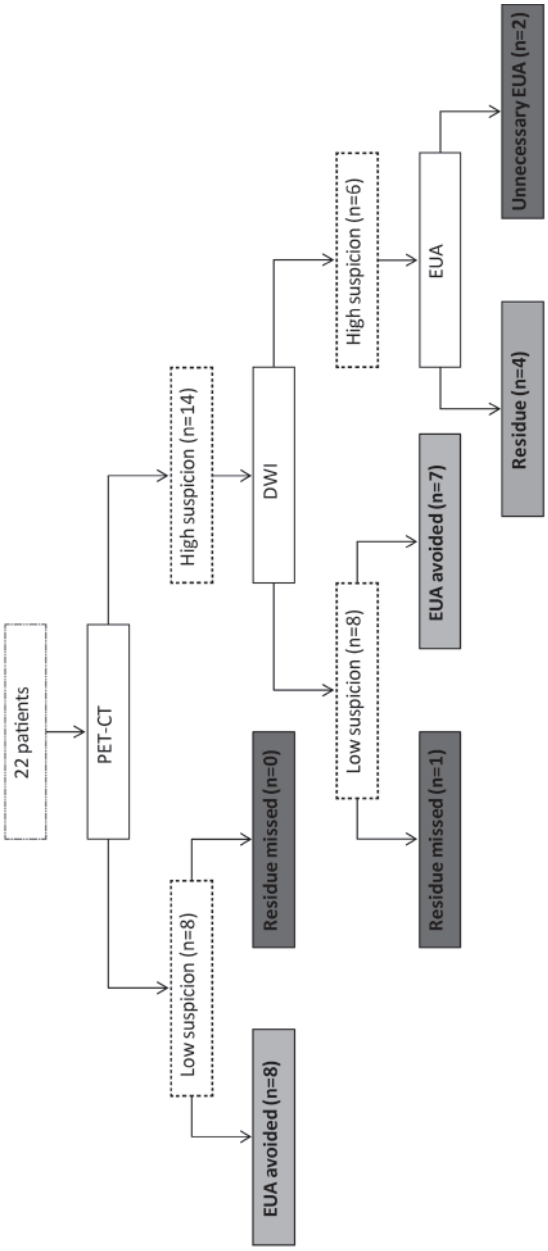
Appendix B1 *Imaging strategy when only PET-CT is used in the decision to perform EUA*



Appendix B2 *Imaging strategy when only DWI is used in the decision to perform EUA*



Appendix B3 *Imaging strategy when PET-CT and DWI are combined in the decision to perform EUA*



Appendix A Imaging algorithms

Different algorithms for primary tumor assessment (n=81)

Parameter	Sensitivity (%, 95%CI; ratio)	Specificity (%, 95%CI; ratio)	Positive predictive value (%, 95%CI; ratio)	Negative predictive value (%, 95%CI; ratio)
DWI	57.1 (20.2-88.2; 4/7)	91.9 (82.7-96.7; 68/74)	40.0 (13.7-72.6; 4/10)	95.8 (87.3-98.9; 68/71)
^{18}F-FDG-PET-CT	85.7 (42.0-99.2; 6/7)	86.5 (76.1-93.0; 64/74)	37.5 (16.3-64.1; 6/16)	98.5 (90.6-99.9; 64/65)
DWI AND ^{18}F-FDG-PET-CT conservative[a]	42.9 (11.8-79.8; 3/7)	97.3 (89.6-99.5; 72/74)	60.0 (17.0-92.7; 3/5)	94.7 (86.4-98.3; 72/76)
DWI AND ^{18}F-FDG-PET-CT sensitive[b]	100 (56.1-100.0; 7/7)	79.7(68.5-87.8; 59/74)	31.8 (14.7-54.9; 7/22)	100 (92.3-100; 59/59)
First ^{18}F-FDG-PET-CT, then DWI	57.1 (20.2-88.2; 4/7)	97.3 (89.7-99.5; 72/75)	66.7 (24.1-94.0; 4/6)	96.0 (87.8-99.0; 72/74)
First DWI, then ^{18}F-FDG-PET-CT	57.1 (20.2-88.2; 4/7)	97.3 (89.7-99.5; 72/75)	66.7 (24.1-94.0; 4/6)	96.0 (87.8-99.0; 72/74)

Different algorithms for lymph node assessment (n=74)

Parameter	Sensitivity (%, 95%CI; ratio)	Specificity (%, 95%CI; ratio)	Positive predictive value (%, 95%CI; ratio)	Negative predictive value (%, 95%CI; ratio)
DWI	100 (51.7-100; 6/6)	72.1 (59.7-81.9; 49/68)	24.0 (10.2-45.5; 6/25)	100 (90.9-100; 49/49)
^{18}F-FDG-PET-CT	83.3 (36.5-99.1; 5/6)	92.6 (83.0-97.3; 63/68)	50.0 (20.1-79.9; 5/10)	98.4 (90.5-99.9; 63/64)
DWI AND ^{18}F-FDG-PET-CT conservative[c]	33.3 (6.0-75.9; 2/6)	97.1 (8.8-99.5; 66/68)	50.0 (9.2-90.8; 2/4)	94.3 (85.3-98.2; 66/70)
DWI AND ^{18}F-FDG-PET-CT sensitive[d]	83.3 (36.5-99.2; 5/6)	83.8 (72.5-91.3; 57/68)	31.3 (12.1-58.5; 5/16)	98.3 (89.5-99.9; 57/58)
First ^{18}F-FDG-PET-CT, then DWI	83.3 (36.5-99.1; 5/6)	95.6 (86.8-98.9; 65/68)	62.5 (25.9-89.8; 5/8)	98.5 (90.7-99.9; 65/66)
First DWI, then ^{18}F-FDG-PET-CT	83.3 (36.5-99.1; 5/6)	95.6 (86.8-98.9; 65/68)	62.5 (25.9-89.8; 5/8)	98.5 (90.7-99.9; 65/66)

[a] A patient was considered positive if both modalities are positive
[b] A patient was considered positive if at least one modality is positive
[c] A patient was considered positive if both modalities are positive
[d] A patient was considered positive if at least one modality is positive

Appendix B *Interobserver agreement*

DWI primary tumor

Original data weighted κ=0.516 (95%CI=0.280-0.751)
Dichotomous κ=0.584 (95%CI=0.257-0.911)
Positive agreement=0.56
Negative agreement=0.92
ICC_{ADC}=0.916 (95%CI=0.849-0.954)

		Observer 1		
		1	2	3
Observer 2	1	63	6	0
	2	3	0	1
	3	2	2	4

DWI lymph node

Original data weighted κ=0.618 (95%CI=0.462-0.775)
Dichotomous κ=5.575 (95%CI=0.315-0.836)
Positive agreement=0.77
Negative agreement=0.88
ICC_{ADC}=0.996 (95%CI=0.993-0.998)

		Observer 1		
		1	2	3
Observer 2	1	42	9	1
	2	1	7	1
	3	1	5	7

^{18}F-FDG-PET-CT primary tumor

Original data weighted κ=0.406 (95%CI=0.242-0.569)
Three point weighted κ=0.411 (95%CI=0.240-0.581)
Dichotomous κ=0.544 (95%CI=0.312-0.776)
Positive agreement= 0.62
Negative agreement=0.92

		Observer 1				
		1	2	3	4	5
Observer 2	1	33	0	6	2	0
	2	0	0	0	0	0
	3	15	2	5	4	3
	4	2	0	0	3	2
	5	0	0	0	1	3

^{18}F-FDG-PET-CT lymph node

Original data weighted κ=0.533 (95%CI=0.357-0.708)
Three point weighted κ=0.512 (95%CI=0.312-0.713)
Dichotomous κ=0.485 (95%CI=0.230-0.739)
Positive agreement=0.55
Negative agreement=0.92

		Observer 1				
		1	2	3	4	5
Observer 2	1	48	1	7	0	0
	2	0	0	0	0	0
	3	0	0	2	0	0
	4	2	2	1	0	0
	5	2	0	3	2	4

APPENDICES CHAPTER 5.2

Appendix A *Differences of functional parameters in HPV-status and correlation with functional parameters in primary tumours and nodal metastasis. Between the HPV positive (+) and negative (-) patients the values of PT FDG-PET and DW-MRI (except skewness and kurtosis) were significantly higher in the HPV negative patients. Although there were significant correlations between HPV-status and functional imaging parameters, no correlation coefficient was higher than (-).5, which implies no robust parameter could predict the HPV-status.*

			HPV status			**Correlation**	
			HPV - Mean	HPV + Mean	P value	Pearson r	P value
Primary Tumour	FDG-PET	T1gtv (cm3)	14.067	8.148	0.057	-0.200	0.043
		MATV (cm3)	12.21	6.54	0.003	-0.250	0.011
		SUV_{max}	9.76	6.74	<0.001	-0.320	<0.01
		SUV_{peak}	7.68	5.24	0.001	-0.326	<0.01
		SUV_{mean}	6.13	4.33	0.001	-0.325	<0.01
		TLG	86.11	46.57	0.05	-0.190	0.055
	DW-MRI	ADC_{GTV} (cm3)	9.052	5.359	0.126	-0.173	0.103
		ADCmean ($x10^{-3}mm^2/s$)	1.319	1.080	<0.001	-0.452	<0.01
		ADC_{SD}	263.1	221.6	0.015	-0.290	0.006
		ADC_{min} ($x10^{-3}mm^2/s$)	0.598	0.479	0.016	-0.232	0.029
		ADC_{max} ($x10^{-3}mm^2/s$)	2.193	1.919	<0.001	-0.319	0.002
		$ADC_{skewness}$	0.356	0.429	.773	0.074	0.493
		$ADC_{kurtosis}$	3.48	3.89	.182	0.183	0.086
		ADC_{p10} to ADC_{p90PT} ($x10^{-3}mm^2/s$)	1001.8	817.8	<0.001	-0.4 to -0.48	<0.01
Lymph node metastasis	FDG-PET	MATV	2.80	3.10	0.228	0.107	0.491
		SUVmax	7.41	6.13	0.222	-0.119	0.254
		SUVpeak	5.60	4.63	0.481	-0.111	0.288
		SUVmean	4.74	3.77	0.236	-0.139	0.182
		TLG	36.89	38.30	0.315	-0.099	0.345
	DW-MRI	ADCmean ($x10^{-3}mm^2/s$)	1.206	1.206	0.738	0	0.998
		ADCsd	256.59	298.0	0.156	0.170	0.149
		ADCmin ($x10^{-3}mm^2/s$)	0.535	0.507	0.332	-0.065	0.580
		ADCmax ($x10^{-3}mm^2/s$)	2.015	2.222	0.084	0.193	0.100
		ADCskewness	3.539	3.488	0.935	-0.022	0.850
		ADCkurtosis	0.3478	0.4678	0.316	0.134	0.256
		ADCp10 to p90 ($x10^{-3}mm^2/s$)	0.903-1.550	0.867-1.624	0.341-0.536	-0.107 to -0.133	-0.26 to 0.95

Appendix B *Univariate parameters for predicting treatment response, recurrence-free survival and overall survival for FDG-PET and DWI separately. The odds ratio with the standard error for treatment failure and the hazard ratios with 95%CI for a recurrence and death is shown.*

			Treatment response			Recurrence-free survival			Overall survival		
			P value	OR	SE	P value	HR	95%CI	P value	HR	95%CI
Baseline characteristics		T-stage	0.309			0.098			0.004		
										3.413[a]	0.432-26.949
										6.083[b]	0.790-46.852
										11.579[c]	1.551-86.436
		N-stage	0.651			0.064			0.632		
		HPV-status	0.047	0.213	0.47-0.977	0.007	4.299	1.494-12.370	<0.001	0.151	0.054-0.423
		Location PT	0.312			0.119			0.299		
		Age	0.584			0.291			0.254		
Primary tumor	FDG-PET	MATV	0.028	1.051	1.005-1.098	0.001	1.051	1.019-1.084	<0.001[d]	1.061	1.037-1.085
		SUV_{max}	0.012	1.231	1.047-1.447	0.330			0.069		
		SUV_{peak}	0.011	1.284	1.058-1.557	0.200			0.025	1.119	1.014-1.235
		SUV_{mean}	0.016	1.348	1.061-1.764	0.245			0.039	1.143	1.007-1.298
		TLG	0.039	1.005	1.000-1.009	0.003	1.006	1.002-1.011	0.001[d]	1.004	1.002-1.006
	DWI	ADC_{GTV}	0.766			0.953			<0.001	1.000	1.000-1.000
		ADC_{mean}	0.237			0.055			0.197	1.001	1.000-1.002
		ADC_{SD}	0.023	2.542	1.140-5.669	0.094			0.031	1.718	1.049-2.812
		ADC_{min}	0.883			0.283			0.207		
		ADC_{max}	0.033	4.905	1.138-21.132	0.064			0.007	3.311	1.397-7.845
		$ADC_{Skewness}$	0.450			0.949			0.582		
		$ADC_{Kurtosis}$	0.600			0.361			0.838		
		ADC_{p10}	0.514			0.109			0.434		
		ADC_{p20}	0.484			0.102			0.361		
		ADC_{p30}	0.394			0.357			0.308		
		ADC_{p40}	0.357			0.339			0.257		
		ADC_{p50}	0.317			0.060			0.217		
		ADC_{p60}	0.263			0.055			0.204		
		ADC_{p70}	0.199			0.049	3.939	1.004-15.448	0.177		
		ADC_{p80}	0.143			0.047	3.779	1.019-14.017	0.142		
		ADC_{p90}	0.082			0.195			0.091		
Nodal metastasis	FDG-PET	MATV	0.892			0.828			0.620		
		SUV_{max}	0.485			0.975			0.450		
		SUV_{peak}	0.393			0.744			0.305		
		SUV_{mean}	0.288			0.648			0.211		
		TLG	0.025	1.008	1.000-1.018	0.096			0.077		
	DWI	ADC_{GTV}	0.597			0.955			0.972		
		ADC_{mean}	0.076			0.261			0.695		
		ADC_{SD}	0.441			0.985			0.920		
		ADC_{min}	0.883			0.191			0.503		
		ADC_{max}	0.158			0.712			0.704		
		$ADC_{Skewness}$	0.657			0.758			0.482		
		$ADC_{Kurtosis}$	0.212			0.347			0.173		
		ADC_{p10}	0.062			0.233			0.558		
		ADC_{p20}	0.036	38.825	1.269-1188	0.092			0.666		
		ADC_{p30}	0.048	21.99	1.028-470.3	0.151			0.500		
		ADC_{p40}	0.062			0.182			0.530		
		ADC_{p50}	0.079			0.244			0.567		
		ADC_{p60}	0.101			0.316			0.631		
		ADC_{p70}	0.118			0.356			0.682		
		ADC_{p80}	0.155			0.423			0.774		
		ADC_{p90}	0.200			0.574			0.978		

[a] T2 compared to T1
[b] T3 compared to T1
[c] T4 compared to T1
[d] Subgroup analysis of tumors larger than 4.2ml showed significant predictive for MATV (p<0.001) and TLG (p=0.009)

Abbreviations: OR = HR = Hazard ratio; Odds ratio; SE = standard error