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_20) 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 $K^{\text{trans}}$ (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, subdiagastric 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$^2$/s and $190.1 \cdot 10^{-6}$ mm$^2$/s). 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 eighth 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 18F-FDG-PET-CT to a conventional workup including CT and/or MRI. The detection rate of 18F-FDG-PET-CT was defined as the rate of lesions that were detected with 18F-FDG-PET-CT, but were not detected on CT or MRI. In our study we compared DWI with 18F-FDG-PET-CT. Even though 18F-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 18F-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 18F-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 18F-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 18F-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 18F-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 18F-FDG-PET-CT is suboptimal due to inflammatory post-irradiation effects which also result in 18F-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 of 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. post-radiation inflammation), whereas a score of 4-5 is a sign of residual malignancy. The authors found substantial to very good interobserver agreement with \( \kappa \) 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>
<thead>
<tr>
<th>Deauville criteria (80, 81)</th>
<th>Hopkins criteria (83)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 No {^{18}F}-FDG uptake</td>
<td>{^{18}F}-FDG uptake at the primary site and nodes less than the internal jugular vein</td>
<td>Complete metabolic response</td>
</tr>
<tr>
<td>2 Slight {^{18}F}-FDG uptake, but less than the uptake in the mediastinal blood pool</td>
<td>Focal {^{18}F}-FDG uptake at the primary site and nodes greater than the internal jugular vein but less than liver.</td>
<td>Likely complete metabolic response</td>
</tr>
<tr>
<td>3 {^{18}F}-FDG uptake above mediastinal bloodpool, but below or equal to {^{18}F}-FDG uptake in the liver</td>
<td>Diffuse {^{18}F}-FDG uptake at the primary site or nodes is greater than the internal jugular vein or liver.</td>
<td>Likely post-radiation inflammation</td>
</tr>
<tr>
<td>4 {^{18}F}-FDG uptake slightly to moderately higher than liver</td>
<td>Focal {^{18}F}-FDG uptake at the primary site or nodes greater than liver.</td>
<td>Likely residual tumor</td>
</tr>
<tr>
<td>5 Markedly increased {^{18}F}-FDG uptake or any new lesion (on response evaluation)</td>
<td>Focal and intense {^{18}F}-FDG uptake at the primary site or nodes.</td>
<td>Residual tumor</td>
</tr>
</tbody>
</table>

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 18F-FDG-PET-CT for the detection of unknown primary HNSCC. This was probably due to a high sensitivity of 18F-FDG-PET-CT, leaving little room for improvement. In Chapter 4.3 we found that DWI analysis provided complementary information in patients with residual 18F-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 18F-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


Chapter 6


35. Castaldi P, Rufini V, Bussu F, et al. Can “early” and “late”18F-FDG PET-CT be used as prognostic factors for the


87. Determination of Cetuximab Versus Cisplatin Early and Late Toxicity Events in HPV+ OPSCC. https://ClinicalTrials.gov/show/NCT01874171.


