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 18F-FDG-PET uptake followed by an intratreatment reduction of 18F-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 18F-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).
**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 \(^{18}\text{F}-\text{FDG-PET-CT}\) 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 ($\Delta$, 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

<table>
<thead>
<tr>
<th>Studies</th>
<th>Patients</th>
<th>Follow-up</th>
<th>Treatment</th>
<th>Complete response</th>
<th>Locoregional control</th>
<th>Overall survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Treatment response</td>
<td>Long-term outcome</td>
<td>Follow-up</td>
<td>Treatment outcome</td>
<td>Range (%)</td>
</tr>
<tr>
<td>CT</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>79 56 (51-58)</td>
<td>80-96</td>
<td>26 24-28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Follow-up (months)</td>
<td>RT (Gy)</td>
<td>Plat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfusion</td>
<td>23</td>
<td>7</td>
<td>18</td>
<td>588 54 (28-83)</td>
<td>75-100</td>
<td>29 5-76</td>
</tr>
<tr>
<td>MRI</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCE</td>
<td>17</td>
<td>3</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWI</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVIM</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>26</td>
<td>3</td>
<td>23</td>
<td>1003 61 (22-118)</td>
<td>75-100</td>
<td>30 1-83</td>
</tr>
<tr>
<td>FDG-PET</td>
<td>16</td>
<td>3</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMISO</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FHX4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Matoba[97] - Treatment 60-70Gy
b Follow-up was not specified in 1 CT, 3 MRI 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-Fluorouracil; DCE = dynamic contrast-enhanced; DWI = diffusion weighted imaging; EGFR = Epidermal growth factor receptor-targeting chemotherapy; FDG-PET = Fluorodeoxyglucose positron emission tomography; FHX4 = 3-[18F]fluoro-2-(4-[(2-nitro-1H-imidazol-1-yl)methyl]-1H-1,2,3-triazol-1-y]propan-1-ol; FLT = 3Deoxy-3[18F]-fluorothymidine; FMISO = [18F]-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 \(K_{\text{trans}}\) (product of blood flow, permeability and capillary surface area) combined with a \(K_{\text{trans}}\) reduction during the first week of treatment in patients with a complete response.

**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\(_{\text{mean}}\) 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\(_{\text{median}}\) at 2 weeks an increase of 24% was found in CR \((n=10)\) (55); at 3 weeks, six studies found a 36-42% ADC\(_{\text{mean}}\) 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\(_{\text{mean}}\) 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 $\Delta^{18}$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).

<table>
<thead>
<tr>
<th>Study</th>
<th>Parameter</th>
<th>Cut-off</th>
<th>Outcome</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brun, 1997</td>
<td>LN %ΔMRgl at week 1</td>
<td>10%</td>
<td>Early CR</td>
<td>1.00 (0.16; 1.00)</td>
<td>0.50 (0.21; 0.79)</td>
</tr>
<tr>
<td>Brun, 1997</td>
<td>PT %ΔMRgl at week 1</td>
<td>10%</td>
<td>Early CR</td>
<td>0.25 (0.01; 0.81)</td>
<td>0.70 (0.35; 0.93)</td>
</tr>
<tr>
<td>Brun, 2002</td>
<td>LN MRgl at week 2</td>
<td>14</td>
<td>CR</td>
<td>0.67 (0.30; 0.93)</td>
<td>0.62 (0.42; 0.79)</td>
</tr>
<tr>
<td>Brun, 2002</td>
<td>PT MRgl at week 2</td>
<td>16</td>
<td>CR</td>
<td>0.89 (0.52; 1.00)</td>
<td>0.64 (0.46; 0.79)</td>
</tr>
<tr>
<td>Brun, 2002</td>
<td>LN SUV at week 2</td>
<td>4.5</td>
<td>CR</td>
<td>0.67 (0.30; 0.93)</td>
<td>0.55 (0.36; 0.74)</td>
</tr>
<tr>
<td>Brun, 2002</td>
<td>PT SUV at week 2</td>
<td>5.0</td>
<td>CR</td>
<td>0.78 (0.40; 0.97)</td>
<td>0.58 (0.41; 0.74)</td>
</tr>
</tbody>
</table>

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

Long-term outcome prediction
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 ($\Delta TBV$) 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 $K_{\text{trans}}$ ($P=0.012$) 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 $\Delta ADC_{\text{mean}}$ was determined per study, which was most prognostic for long-term outcome (Figure 3a). Pooled analysis showed that a $\Delta ADC_{\text{mean}}$ 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.

<table>
<thead>
<tr>
<th>Study</th>
<th>Parameter</th>
<th>Tissue</th>
<th>Cut-off</th>
<th>Outcome</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambrecht, 2014</td>
<td>%ΔADC at week 2</td>
<td>PT + LN</td>
<td>10.8%</td>
<td>LRC</td>
<td>10.8%</td>
<td>130.33 (4.71; 3609.26)</td>
</tr>
<tr>
<td>Vandecaveye, 2010</td>
<td>%ΔADC at week 2</td>
<td>PT + LN</td>
<td>14.0%</td>
<td>LRC</td>
<td>14.0%</td>
<td>2.56 (0.94; 8.17)</td>
</tr>
<tr>
<td>King, 2013</td>
<td>%ΔADC at week 2</td>
<td>PT</td>
<td>15.5%</td>
<td>LRC</td>
<td>15.5%</td>
<td>2.83 (0.80; 8.95)</td>
</tr>
<tr>
<td>Marzi, 2017</td>
<td>%ΔADC at week 3</td>
<td>PT</td>
<td>15.5%</td>
<td>LRC</td>
<td>15.5%</td>
<td>2.83 (0.80; 8.95)</td>
</tr>
<tr>
<td>Matoba, 2014</td>
<td>%ΔADC at week 3</td>
<td>PT</td>
<td>24.0%</td>
<td>LRC</td>
<td>24.0%</td>
<td>2.83 (0.80; 8.95)</td>
</tr>
<tr>
<td>Vandecaveye, 2010</td>
<td>%ΔADC at week 4</td>
<td>PT + LN</td>
<td>25.0%</td>
<td>LRC</td>
<td>25.0%</td>
<td>2.83 (0.80; 8.95)</td>
</tr>
</tbody>
</table>

Week 2

Subtotal

61.3% 19.34 (6.28; 59.51)

Heterogeneity: df = 2, P = 0.48; I² = 0%

Test for overall effect: P < 0.001

Week 3

Subtotal

38.7% 19.79 (1.06; 369.52)

Heterogeneity: df = 1, P = 0.08; I² = 68%

Test for overall effect: P = 0.05

Total

100.0% 16.91 (6.19; 46.15)

Heterogeneity: df = 4, P = 0.30; I² = 19%

Test for overall effect: P = 0.001

Test for subgroup differences: df = 1, P = 0.99; I² = 0%

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
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). TBRpeak 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.
Functional imaging during (chemo)radiotherapy in HNSCC

A

<table>
<thead>
<tr>
<th>Study</th>
<th>Parameter</th>
<th>Cut-off</th>
<th>Log Hazard Ratio</th>
<th>SE</th>
<th>Weight</th>
<th>Hazard Ratio IV Random (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin, 2017</td>
<td>LN TLG at week 3</td>
<td>14.6</td>
<td>1.73</td>
<td>0.53</td>
<td>43.8%</td>
<td>5.63 (1.99; 15.93)</td>
</tr>
<tr>
<td>Min, 2015</td>
<td>PT TLG at week 3</td>
<td>9.4</td>
<td>2.05</td>
<td>0.73</td>
<td>23.0%</td>
<td>7.76 (1.85; 32.57)</td>
</tr>
<tr>
<td>Min, 2016</td>
<td>PT TLG at week 3</td>
<td>9.4</td>
<td>1.53</td>
<td>0.61</td>
<td>33.2%</td>
<td>4.64 (1.40; 15.32)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0% 5.68 (2.86; 11.31)</td>
</tr>
</tbody>
</table>

Heterogeneity: df = 2, P = 0.86; F = 0%
Test for overall effect: P < 0.001

B

<table>
<thead>
<tr>
<th>Study</th>
<th>Parameter</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
<th>Hazard Ratio IV Random (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin, 2017</td>
<td>LN TLG at week 3</td>
<td></td>
<td></td>
<td>3.91 (1.75; 8.72)</td>
</tr>
<tr>
<td>Min, 2015</td>
<td>PT TLG at week 3</td>
<td></td>
<td></td>
<td>2.49 (0.71; 8.72)</td>
</tr>
<tr>
<td>Min, 2016</td>
<td>PT TLG at week 3</td>
<td></td>
<td></td>
<td>2.18 (0.71; 6.71)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>3.04 (1.70; 5.42)</td>
</tr>
</tbody>
</table>

Heterogeneity: df = 2, P = 0.67; F = 0%
Test for overall effect: P < 0.001

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 $K^{\text{trans}}$ 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\textsubscript{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\textsubscript{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 18F-FDG avid due to glucose consumption by activated macrophages, limits defining correct tumor volumes mainly from the 3\textsuperscript{rd} 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\textsubscript{max} (reduction; i.e. a low intratreatment absolute SUV\textsubscript{max}) at week 2, 3 or 4 intratreatment was associated with LRC and OS (62, 65, 66, 69, 71). A lower absolute SUV\textsubscript{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 $\Delta$TLG (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 $K_{trans}$ 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 ($\leq$4.25 (69, 88), $\leq$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


