Discussion and future perspectives
Discussion and future perspectives

In chapter two we discussed the literature with respect to tumor perfusion measurements with PET and DCE-MRI. Antiangiogenic agents are thought to alter tumor vasculature by normalizing vessel structure, inhibition of lymphangiogenesis and reduction of tumor interstitial pressure [1]. Several of these consequences of antiangiogenic treatment can be monitored with imaging techniques. DCE-MRI and \( H_2^{15}O \) PET measure the endothelial transfer coefficient and tumor perfusion, respectively. Accuracy studies for vascular measurements with PET have not been performed in tumors, mainly because typical characteristics of tumor vasculature like heterogeneity in afferent and efferent vessels and perfusion inhomogeneity are difficult to model. However, the one-compartment model has been validated for perfusion measurements in other tissues like myocardium and brain [2-8]. Although perfusion inhomogeneity introduces bias in regional perfusion calculations, measurements of average tumor perfusion are expected to be accurate because the model is still linear in the range of observed tumor perfusion values.

For DCE-MRI, accuracy is more difficult to determine because the signal reflects both perfusion and permeability. However, data supports the theory that enhancement of the MRI signal reflects vessel permeability and tumor perfusion [9-11]. Although validation is limited for both techniques, they reflect changes of tumor vasculature and are the most sophisticated way to non-invasively study tumor vasculature until more validated methods become available.

The different nature of their signal offers interesting opportunities when both techniques are used side-by-side. \( H_2^{15}O \) PET specifically measures tumor perfusion due to the free distribution of water, while DCE-MRI measures a combination of vessel permeability, surface area and perfusion because gadolinium is not freely diffusible. Therefore, the two modalities can complement each other. In theory, both perfusion and permeability can be isolated when perfusion \( (F) \) and the endothelial transfer coefficient \( (K^{\text{trans}}) \) are known.

Validation of PET imaging biomarkers

The results obtained in chapter two in combination with the biological background of combined EGFR and VEGF treatment lead to the design of validation and qualification studies aiming to detect target modulation and to produce surrogate endpoint biomarkers. First we focussed on validation of the imaging biomarkers of interest, and in particular on repeatability.

In chapter three the repeatability of \( H_2^{15}O \), a marker of perfusion, was evaluated. The study was performed using a PET only system with a 10-15 min. transmission scan followed by a dynamic emission
Discussion and future perspectives

Volumes of interest (VOIs) were defined on subsequently derived $^{18}$F-FLT images because of superior contrast between tumor and background. When patients are followed over time, changes of more than 18% in F and more than 32% in the volume of distribution ($V_T$) are likely to represent a true change, rather than measurement error.

In chapter four, $^{18}$F-FDG, a marker of glucose metabolism, was evaluated. A meta-analysis of the published data was performed to evaluate the pooled test-retest variability. Recently, Wahl et al introduced the term PERCIST [12] as a PET counterpart of the CT biomarker RECIST which is a FDA approved surrogate endpoint biomarker. Validation of the PERCIST concept was mandatory before the qualification process could be initiated and the present meta-analysis aimed to do so. In accordance with the study of Wahl et al, percentage repeatability was found to be a function of the level of uptake. For serial PET scans, a threshold of a combination of 20% as well as 1.2 SUV$_{mean}$ units was found to be more than 95% likely to be a true change of glucose metabolism, rather than measurement error. After adjusting for uptake rate, tumor volume had only minimal effect on repeatability. Compared to SUV$_{max}$, SUV$_{mean}$ showed better repeatability.

In chapter five the repeatability of $^{18}$F-FLT, a marker of tissue proliferation, was evaluated. When patients are followed over time, changes of more than 15% in SUV$_{mean}$ and 20–25% in $K_i$ and SUV$_{max}$ are likely to represent a change in tumor proliferation, rather than measurement error. $^{18}$F-FLT is being trapped in the cell due to phosphorylation by thymidine kinase 1. Theoretically, $k_3$ is the most specific parameter to study TK1 activity and probably best linked to proliferation, because SUV and $K_i$ are also subject to perfusion and extraction. Unfortunately, $k_3$ showed poor repeatability. However, $K_i$ and SUV were found to be accurate markers of proliferation in validation studies using histological proliferation markers like Ki-67 [13-15].

QUALIFICATION OF $^{15}$O AND $^{18}$F-FDG PET DERIVED IMAGING BIOMARKERS

Next we designed a clinical phase II trial where NSCLC patients were treated with BE. Regarding PET, three tracers were available to evaluate target modulation. The short radioactive half-life of $^{15}$O (2 min) allows a scan to be followed by a second scan with a different tracer to study additional aspects of tumor biology. However, the longer half-life of $^{18}$F does not allow a sequential scan with a different $^{18}$F tracer within the same imaging session. Therefore either $^{18}$F-FDG or $^{18}$F-FLT had to be selected. Both glucose metabolism and cellular proliferation are downstream effects of the EGFR pathway, as discussed in the introduction. The detailed information of the connection between the EGFR pathway and glucose metabolism favored $^{18}$F-FDG. Recently the superiority of $^{18}$F-FDG (compared to $^{18}$F-FLT) as surrogate endpoint biomarker was shown in a comparative response monitoring study in NSCLC patients treated with an EGFR TKI [16].

Because $^{15}$O and $^{18}$F-FDG were not qualified for integral use (i.e. for clinical decision making, see Table 1 in chapter 1), the biomarkers were not used to influence trial design, but as pharmacodynamic
measurements to study target modulation and explored as surrogate endpoint biomarkers.

Chapter six and seven describe the results of a clinical phase II study where a group of chemotherapy-naive patients with stage IV non-squamous NSCLC were treated with BE until disease progression (assessed with RECIST). The value of imaging biomarkers to predict progression-free survival (PFS) was explored using predefined cutoff values.

The primary end point, a non-progression rate (NPR) of 75% after six weeks of treatment, was met. The objective response rate (ORR) of 25% was not different from that what can be obtained with cytotoxic chemotherapy. Disappointingly however, median overall survival (OS) was only 6.9 (95% CI 5.5–8.4) months. Importantly only half of the patients received second-line platinum-based chemotherapy upon disease progression. This was mainly due to patient refusal and deterioration of physical performance due to disease progression, indicating that early use of sensitive biomarkers is essential for optimal patient treatment.

In patients with longer PFS than the mean, F, $K_{trans}$, and SUV decreased by a mean of 20% (IQR -48% to +1%), 17% (IQR -61% to -10%) and 17% (IQR -39% to +2%), respectively, indicating vascular and metabolic change. In contrast, mean F and SUV did not change in patients with lower PFS than the mean; -3% (IQR -15% to -4%) and 6% (IQR -2% to +22%), respectively. Mean $K_{trans}$, however, decreased irrespective of PFS benefit (-17%, IQR -49% to +10%). This contrast between F and $K_{trans}$ might be explained by a decrease in vessel permeability and/or surface area which does not translate into a PFS benefit. However, tumor $K_{trans}$ is a difficult parameter to interpret. At baseline, the extravasation of contrast agent is not restricted by the permeability of the numerous leaky blood vessels and is primarily flow driven. After therapy, the extravasation becomes more restricted by the normalized vessels. This variable pathophysiologic character of $K_{trans}$ might explain the difference in Δ perfusion and Δ $K_{trans}$. Other explanations could be error in $K_{trans}$ measurements due to the semi-quantitative nature of the DCE-MRI technique (caused by the nonlinear relationship between the degree of signal enhancement and gadolinium concentration) or the use of a population-based AIF (due to failure of patient-based AIF measurements).

To qualify the imaging parameters as surrogate endpoint biomarkers, we evaluated their predictive value for PFS after three weeks of treatment. With RECIST, 14% of the patients had a PR at week 3 and 78% had SD, while these numbers were 26% and 63% for tumor perfusion, 18% and 68% for SUV and 43% and 43% for $K_{trans}$, respectively, using predefined cutoff values. Although a partial response (according to RECIST) at week 3 was associated with prolonged PFS (4.6 months for responders vs. 2.9 months for all other patients; HR 0.39; 95% CI 0.18–0.84; $p = 0.017$), 40% of patients that had a PR/CR somewhere during the course of treatment still had stable disease at the 3-week time point. This corroborates previous observations that size based response criteria have low discriminative value early in the course of treatment. In these patients tumor perfusion was already decreased and $K_{trans}$ hetero-
geneity showed no increase. The results indicate that the latter parameters are more discriminative than RECIST at the 3-week time point. Regardless of tumor size change, patients that had a metabolic response (SUV decrease > 20%) after three weeks had longer PFS (9.7, 95% CI 1.8–17.6, months) than those without (2.8, 95% CI 2.0–3.5, months; HR 0.38; 95% CI 0.18–0.79; \( p = 0.01 \)).

Although the change in \( K^{\text{trans}} \) indicated vascular modulation, the parameter could not be qualified as a surrogate endpoint biomarker because it was not related to PFS (\( p = 0.39 \)). Exploratory pixel-by-pixel analysis of \( K^{\text{trans}} \) showed that patients with an increase of more than 15% in the standard deviation of tumor \( K^{\text{trans}} \) (\( K^{\text{trans}} \) heterogeneity), i.e. an increase in regions with low or high \( K^{\text{trans}} \) values, after three weeks had shorter PFS (2.3 vs. 7.0 months; HR 4.4; 95% CI 1.5–13.3; \( p = 0.008 \)).

An increase in \( K^{\text{trans}} \) heterogeneity indicates a less even distribution of gadolinium contrast extravasation into the EES, reflecting a pathologic vessel bed with areas of leaky vessels and increased perfusion and areas with severe hypoperfusion, both pathological and thus absence of vascular normalization. Although promising, \( K^{\text{trans}} \) heterogeneity analysis has yet to be validated and qualified. The cutoff value of 15% to define a relevant change was most discriminative, but whether this holds for future studies remains speculative.

A greater than 20% decrease in tumor perfusion, as measured with PET, was also associated with PFS benefit; 12.5 months vs. 2.9 months for patients without a decrease (HR 0.25; 95% CI 0.09–0.70; \( p = 0.009 \)). Although both \( K^{\text{trans}} \) and \( F \) decreased during treatment, indicating a reduction in perfusion and vessel permeability, no clear correlation was found between them, possibly reflecting the different nature of the signals, or because of the already mentioned difficulty of quantitative \( K^{\text{trans}} \) measurements due to signal saturation and the use of a population based AIF. This could well explain why \( K^{\text{trans}} \) heterogeneity was more discriminative than mean \( K^{\text{trans}} \). \( \Delta \) (mean) \( K^{\text{trans}} \) requires the absolute value of the week 3 scan to be compared with that of baseline, while \( K^{\text{trans}} \) heterogeneity is a reflection of the heterogeneity within the tumor and is a measure of the distribution of values within the tumor at a single time-point and thus expected to be less dependent on error in AIF measurement and absolute signal value. The signal intensity of pixels is being compared to the signal intensity of the other pixels in the tumor within the same scan.

Although tumor perfusion was not related to glucose metabolism at baseline and after three weeks of treatment, a non-significant positive relation was observed after three weeks, possibly reflecting a combination of an overall reduction in tumor perfusion together with normalization of the remaining vessel bed.

In conclusion, this thesis shows that PET derived parameters of glucose metabolism, proliferation and perfusion are repeatable. DCE-MRI derived \( K^{\text{trans}} \) heterogeneity and PET derived measurements of tumor perfusion and glucose metabolism were qualified as surrogate endpoint biomarkers.
FUTURE PERSPECTIVES

At present, several studies with targeted agents are actively recruiting unselected cohorts of patients (www.clinicaltrials.gov). This strategy does not only cause harm by denying patients the best treatment on the individual level, but can actually negatively influence their survival. This was recently demonstrated in a trial evaluating combined treatment targeting c-MET and EGFR. Although patients with high MET expression benefited from the addition of a monoclonal antibody (MAb) against MET to an EGFR TKI, patients with low MET expression experienced decreased survival when a MET MAb was added [17]. This illustrates the importance of continuous comparison of emerging data with preclinical results. Biomarkers can aid this process by reflecting target modulation and by relating this to patient outcome. Considering individual pros and cons of the different biomarkers, the solution might be to regard them as additive or even synergistic, rather than competitive. Tumor biopsy is an excellent method to screen multiple pathways and receptor mutation and expression profiles within a single tumor sample, impossible to obtain by imaging alone. Exploratory tissue analyses have been shown to generate predictive biomarkers that can be used for patient stratification (e.g. EGFR mutation status). Recently, our group developed and validated a non-invasive pharmacokinetic imaging biomarker to assess EGFR mutation status by radiolabeling erlotinib, an EGFR TKI. Validation was done by comparing $^{11}$C-erlotinib kinetics with EGFR mutation status in tumor biopsy samples and subsequent qualification as predictive biomarker was done by relating the results to patient outcome [18]. In this example, an invasive method was used to discover the target that predicts for treatment benefit. Next, it was used as a platform to validate the non-invasive biomarker $^{11}$C-erlotinib.

Although baseline biomarkers can be highly predictive of patient outcome to therapy, this is unfortunately not the case for all patients, despite biomarker homogeneity at baseline. Tumor-host interactions and additional therapeutic strategies (e.g. radiotherapy or adding a second drug) can modify the clinical response. Therefore, a read-out of response during treatment, preferably as soon as possible, is essential to detect treatment failure before clinical deterioration occurs, limiting further treatment options. Conceptually, imaging can perform this job. Early stopping rules can be applied when change in tumor biology is absent, thereby adding valuable information to baseline predictive biomarkers. Repetitive biopsy forms an alternative, but clinical feasibility is hampered by its invasive character. Also, in contrast to imaging, this method does not allow to assess intra- and inter-lesional heterogeneity.

Recently, the Radiological Society of North America (RSNA) initiated QIBA (Quantitative Imaging Biomarkers Alliance) to advance quantitative imaging and the use of imaging biomarkers in clinical trials and clinical practice by engaging researchers, healthcare professionals and industry (http://qibawiki.rsna.org). Committees for DCE-MRI, $^{18}$F-FDG PET and volumetric CT were formed to define basic standards and quality control measures that should ensure consistent, reliable and fit-for-purpose quantitative biomarkers. Because of the potential of biomarkers to improve drug develop-
ment and patient outcome on the individual level, QIBA aims to speed up qualification and to receive FDA approval for integral use of the biomarkers.

There are several sophisticated trial designs that facilitate this process by combining drug development with biomarker qualification in early phase studies (Figure 1) [19]. The results obtained from these studies can be used to adjust eligibility criteria for subsequent phase III trials to include only those patients that are likely to benefit from treatment.

In this thesis, PET and DCE-MRI derived biomarkers were used in an integrated manner (see Table 1 in chapter 1 for definition). Compared to perfusion measurements with PET and DCE-MRI, $^{18}$F-FDG SUV was associated with better clinical feasibility and a lower drop-out rate. Although the primary end-point (NPR of 75%) was met in this study, PFS was disappointing when compared to that of standard treatment. However, the subgroup of patients with a decrease in $^{18}$F-FDG SUV experienced longer PFS, qualifying it as a surrogate endpoint biomarker. Several other trials showed similar results of $^{18}$F-FDG SUV performance during EGFR TKI treatment [16, 20, 21]. However, none of these trials used the biomarker integrally (i.e. discontinuation of treatment when a decrease in $^{18}$F-FDG SUV was absent) and stopping rules were still based on RECIST (i.e. discontinuation of treatment in case of progressive disease). In theory, even patients with progressive disease on CT can benefit from treatment, as has been shown for gastrointestinal stroma tumors (GIST) where patients with PD on CT and a decrease in $^{18}$F-FDG SUV responded favorably to drug therapy with imatinib [22].

**Figure 1.** Examples of phase II biomarker and drug co-development [19].

A. Adaptive parallel design [32, 33]. Two two-stage phase II trials are conducted in parallel; one in the biomarker positive group (expected to benefit more from treatment) and one in the biomarker negative group of patients. After the first stage, the trial may continue in all patients or only in the biomarker positive group.

B. Tandem two-step design [33, 34]. All patients are entered in the first stage, regardless of the biomarker result. If the number of clinical responses that are observed in the first stage is large enough, the study proceeds to the second stage in the overall population. If the number of responses observed in the first stage is insufficient, the study accrues only patients in the subgroup predicted by the biomarker to be responders, and study termination is governed by a standard optimal two-stage phase II trial design in that subgroup of patients.
This warrants a study with integral use of $^{18}$F-FDG SUV. The same eligibility criteria can be applied. Based on the results of Zander et al [16], metabolic response can be assessed with a $^{18}$F-FDG PET scan after one week of treatment. Only patients that show a decrease in $^{18}$F-FDG SUV will continue treatment and PFS will be the primary endpoint. Of note, an increase in the summed size of target lesions of > 20% in combination with a decrease in $^{18}$F-FDG SUV should not be regarded as progressive disease. Patients without a decrease in $^{18}$F-FDG SUV after one week should be switched to alternative treatment.

Recent data show that resistance to EGFR TKIs might be reversed by hydroxychloroquine (HCQ), an antimalaria and antirheumatic drug, thereby (re)sensitizing tumor cells to EGFR TKIs. This encouraged us to initiate a trial to explore the efficacy of this combination together with validation of $^{18}$F-FDG SUV as surrogate endpoint biomarker for clinical benefit (Figure 2). Eligibility criteria will be based on previous treatment and EGFR mutation status, a baseline predictive biomarker. A tandem two-step design will be used with PFS as primary endpoint. All patients are entered in the first stage of a two-stage Simon’s design phase II study, regardless of the $^{18}$F-FDG SUV result after one week (Figure 2A). If PFS is comparable to that of standard treatment, the study proceeds to the second stage in all patients. If PFS is disappointing, a second two-stage Simon’s design phase II study will accrue patients that show a decrease in $^{18}$F-FDG SUV after one week of treatment (Figure 2B).

Figure 2. Tandem two-step design, modified for a trial evaluating drug efficacy of erlotinib and HCQ treatment, together with qualification of $^{18}$F-FDG SUV as surrogate endpoint biomarker for PFS benefit.

A. First step

Eligible patients for erlotinib and HCQ treatment

Start treatment

Stage I of a two-stage Simon’s design trial

Adequate PFS benefit

Continue drug trial

Inadequate PFS benefit

Stop drug trial in the general study population

Stage II

B. Second step

Eligible patients for erlotinib and HCQ treatment

Start treatment

Stage I of a two-stage Simon’s design trial

Lack of $^{18}$F-FDG SUV decrease

Discontinue treatment

Adequate PFS benefit

Continue drug trial

Inadequate PFS benefit

Drug combination is abandoned

Decrease in $^{18}$F-FDG SUV

Continue treatment

Stage II
In recent years, interesting progress has been made in scanner hardware and tracer development, providing new and potentially better parameters (Box 1). Several tracers are already available or are currently being developed that study numerous aspects of tumor biology and allow to evaluate target modulation to several classes of drugs. For example, PET-MRI combines two established techniques in one scanner (analogous to PET-CT).

The CT technique allows to perform volumetric analysis, a more sophisticated way to evaluate tumor size that might be more sensitive than the unidimensional RECIST method. Validation studies have been performed for volumetric analysis and the method shows better repeatability than RECIST. Qualification studies are currently being performed to evaluate its performance against RECIST.

Other techniques like diffusion weighted MRI (DW-MRI) and magnetic resonance spectroscopy (MRS) are of great interest and potential, but still lack appropriate validation and qualification. However, definition of basic standards and uniform application of these biomarkers in dedicated trials should reveal their value on short notice.

To conclude, it is the hope that this and future research on biomarker and drug development will lead to better treatment for cancer patients by selecting the best drug on the individual level and by applying early stopping rules based on changes in tumor biology, resulting in better survival.

Box 1. New developments in imaging biomarkers.

---

**Integrated PET-MRI** is a promising technique that performs sequential studies with PET and DCE-MRI within one imaging session and allows combination studies with minimal spatial mismatch. The MR technique can also be used to study other biological processes than perfusion and permeability by magnetic resonance spectroscopy (MRS) and diffusion weighted MRI (DW-MRI).

**MRS** relies on the fact that nuclei resonate at slightly different frequencies depending on the surrounding molecular environment. This allows to evaluate the molecular component of tissues. Several nuclei can be monitored (e.g. 1H, 13C, 14N, 19F, 23Na, 31P). The technique detects molecules containing the selected atom. Although it allows to detect several metabolites in one single measurement, limiting factors are sensitivity and resolution and the technique requires further development before biomarkers can be qualified for clinical use in oncology [23].

**DW-MRI** measures the mobility of water within tissues. The basic biological premise for the use of DW-MRI in cancer is that malignant tissues are generally more cellular and have more high-water content than normal tissue, both of which lead to high signal intensity. There are a number of features that affect tissue water diffusivity, including tissue perfusion, cell density and distribution, integrity of cellular membranes, and tissue organization. Signal change thus depends on complex interplay between several
biological processes in response to therapy. Validation on the biological (what aspects of tumor biology are reflected in the parameter) and analytical level (standardization and repeatability) needs to be optimised before these biomarkers can be used in clinical trials.

CT still has many advantages when compared to newer imaging techniques like MRI and PET. It is widely available, relatively inexpensive, quick, and requires little time for data analysis and a minimum amount of personnel.

**CT volumetric analysis** might overcome some of the disadvantages of unidimensional size measurements, because ill-defined and irregular lesions are subject to interobserver variation [24] and size change can be asymmetric. Volumetric analysis is associated with better repeatability than unidimensional analysis [25, 26] and response assessment by CT volumetry seems to be more discriminative [25-29]. However, despite the potential to complement or even replace RECIST, the superior value of this technique as surrogate endpoint for patient outcome in clinical practice remains to be proven [30]. The additional discriminative level is generally small and unidimensional measurement still represents an adequate alternative with the advantage of better clinical work flow [25].

As with MRI, CT can also be used to measure tumor perfusion by studying tracer (contrast) kinetics, a technique called dynamic contrast-enhanced computed tomography (DCE-CT). Kinetic modelling is similar to that of DCE-MRI. Sequential images are made before, during, and following the injection of a contrast agent. The extravasation of contrast agent from the vascular compartment to the extravascular extracellular interstitial space provides information on blood flow, blood volume and microvascular permeability. In potential, the technique offers an interesting alternative to PET perfusion studies because of the wide availability of CT scanners and the need for an on-site cyclotron to perform PET perfusion studies due to the short half-life of H$_2$^{15}$O (2 min). Although a comparison study of DCE-CT and H$_2$^{15}$O PET showed a moderate correlation for tumor perfusion using a static PET technique [31] providing limited validation, robust validation and clinical qualification using basic standards and quality control measures are still awaited for DCE-CT.
REFERENCE LIST


Discussion and future perspectives