CHAPTER 8

Summarizing discussion and future perspectives
The aim of this thesis was to technically validate several Positron Emission Tomography (PET) based quantitative imaging biomarkers for response evaluation in oncological setting. This thesis is divided into two sections: the first addressing the precision of $^{18}$F-FLT and $^{18}$F-FDG uptake metrics for response monitoring in non-small cell lung cancer (NSCLC) patients, the second evaluating the accuracy and precision of several simplified methods for quantification of $^{18}$F-FDHT uptake in metastatic castration resistant prostate cancer (mCRPC). In this chapter we will summarize the presented results and discuss future perspectives for the use of these tracers as imaging biomarkers of response in oncology.

PART 1

$^{18}$F-FLT and $^{18}$F-FDG PET/CT scans have high potential for minimally invasive measurements of response to anti-tumor therapy. Historically, $^{18}$F-FDG is the mostly used tracer as it is highly sensitive for detection of malignancies and their metastases through direct assessment of glucose metabolism. Yet, increased glucose metabolism of tumor cells is secondary to change in other intra-cellular processes requiring extra energy such as proliferation. Moreover, increased metabolism can also be caused by other mechanisms, e.g. inflammation, limiting the specificity of $^{18}$F-FDG. The latter may particularly hamper the use of $^{18}$F-FDG PET/CT in response evaluation setting because of treatment-induced inflammatory responses.

$^{18}$F-FLT PET/CT has a potential advantage over $^{18}$F-FDG because it is sensitive to proliferation and is therefore a more direct and specific characteristic of malignancy. Many response evaluation studies have been performed, assessing the performance of several $^{18}$F-FDG uptake metrics and several systematic reviews have been performed. Yet, only a limited amount of studies assessing the performance of $^{18}$F-FLT PET/CT as quantitative imaging biomarker of response have been performed, and no comprehensive overview of these studies was available. In Chapter 2 we therefore performed a systematic review, identifying 35 studies using $^{18}$F-FLT for...
SUMMARIZING DISCUSSION AND FUTURE PERSPECTIVES

The aim of this thesis was to technically validate several Positron Emission Tomography (PET) based quantitative imaging biomarkers for response evaluation in oncological setting. This thesis is divided into two sections: the first addressing the precision of $^{18}$F-FLT and $^{18}$F-FDG uptake metrics for response monitoring in non-small cell lung cancer (NSCLC) patients, the second evaluating the accuracy and precision of several simplified methods for quantification of $^{18}$F-FDHT uptake in metastatic castration resistant prostate cancer (mCRPC). In this chapter we will summarize the presented results and discuss future perspectives for the use of these tracers as imaging biomarkers of response in oncology.

PART 1

$^{18}$F-FLT and $^{18}$F-FDG PET/CT scans have high potential for minimally-invasive measurements of response to anti-tumor therapy. Historically, $^{18}$F-FDG is the mostly used tracer as it is highly sensitive for detection of malignancies and their metastases through direct assessment of glucose metabolism (1). Yet, increased glucose metabolism of tumor cells is secondary to change in other intra-cellular processes requiring extra energy such as proliferation (2). Moreover, increased metabolism can also be caused by other mechanisms, e.g. inflammation, limiting the specificity of $^{18}$F-FDG (3-5). The latter may particularly hamper the use of $^{18}$F-FDG PET/CT in response evaluation setting because of treatment-induced inflammatory responses. $^{18}$F-FLT PET/CT has a potential advantage over $^{18}$F-FDG because it is sensitive to proliferation and is therefore a more direct and specific characteristic of malignancy (6).

Many response evaluation studies have been performed, assessing the performance of several $^{18}$F-FDG uptake metrics and several systematic reviews have been performed (7-9). Yet, only a limited amount of studies assessing the performance of $^{18}$F-FLT PET/CT as quantitative imaging biomarker of response have been performed, and no comprehensive overview of these studies was available. In Chapter 2 we therefore performed a systematic review, identifying 35 studies using $^{18}$F-FLT for
response assessment of systemic-, radio- and chemoradiotherapy. Overall, $^{18}$F-FLT showed to be a good predictor of early response, because progression free survival and disease free survival showed good correlations with change in $^{18}$F-FLT uptake. However, the correlation with overall survival was less consistent. All studies were single-center observational studies with small sample sizes. Due to these small sample sizes, studies might be underpowered to confirm correlations with overall survival. Yet, a subset of studies in NSCLC patients also compared the performance of $^{18}$F-FLT with that of $^{18}$F-FDG, of which the latter often also correlated with overall survival.

Additionally, the review highlighted the importance of performing response assessment at an adequate time interval during treatment. If response is assessed too soon or too late after start of treatment, the discriminative power of $^{18}$F-FLT might be compromised. The ideal time interval remains to be determined and might differ based on treatment type. Moreover, most studies defined thresholds, discriminating responders from non-responders, retrospectively and did not take intrinsic variability of $^{18}$F-FLT PET/CT into account. Thus, before $^{18}$F-FLT can be validated as imaging biomarker of treatment response, prospective trials taking the intrinsic variability into account are needed.

Knowledge of day-to-day variability of quantitative $^{18}$F-FLT uptake metrics is not only essential for accurate evaluation of response but also for drug development. Repeatability of $^{18}$F-FLT uptake metrics in solid tumors has only been studied in a few small single center cohorts ($N \leq 10$ patients) (10-13). Here, only a limited number of uptake metrics were assessed and uptake intervals and tumor segmentation methods varied between studies. These variations between studies can significantly influence measured uptake metrics and their repeatability. To overcome these inconsistencies and determine repeatability of $^{18}$F-FLT we performed an individual patient data meta-analysis in Chapter 3 by combining all available $^{18}$F-FLT repeatability data on solid tumors from previously published studies using a semi-automatic segmentation method. Permission was obtained to re-analyze the original $^{18}$F-FLT repeatability data from four out of five cohorts, comprising data of 10 in breast cancer (10), 9 in head and neck squamous cell carcinoma (11) and 33 in NSCLC patients (11,12) with a total of 52 individual lesions. From the assessed quantitative $^{18}$F-FLT PET/CT uptake metrics,
SUV\textsubscript{peak} yielded best repeatability coefficients (23.1%). However, performance of other SUV measures was not significantly worse. In line with several other studies PET-based volumetric measures showed a higher variability compared to SUV (14,15). Repeatability of all uptake metrics improved when assessed on a per-patient basis instead of per-lesion. Moreover, if lesion selection criteria were applied and only lesions with SUV\textsubscript{max} ≥ 4.0 were included, repeatability of volumetric measures improved while SUV metrics remained unaffected. Also other lesion selection criteria, e.g. including only lesions with a volume ≥ 4.2 mL (to avoid partial volume related underestimation of tracer uptake), the hottest or primary lesions did not improve variability (16).

In this study we used semi-automatic segmentation methods, which require adequate tumor-to-background ratios for robust volume of interest (VOI) definition (17). This explains the improvement in performance of volumetric measures when lesions with low contrast between tumor and background radioactivity are excluded. Studies using ordered subset expectation maximization (OSEM) as well as studies using Filtered Back-projection (FBP) reconstructions were included in this meta-analysis. Semi-automatic segmentation improved repeatability of volumetric measurements of the OSEM reconstructed images compared to manual delineation in contrast to results found for the FBP dataset. In addition, the variability of SUV\textsubscript{mean} significantly increased in the latter dataset, as was also observed when other (semi-)automatic segmentation algorithms were used for lesion delineation. This raised the issue of appropriateness of semi-automatic segmentation in FBP reconstructed images.

Two studies validating simplified quantitative 18F-FLT uptake metrics in NSCLC showed that response evaluation using tissue-to-blood ratio (TBR) is quantitatively more precise compared to SUV (18,19). Moreover, normalization to blood uptake could correct for inaccuracies in dose calibration, weight measurements and variations of tracer supply to the tumor. However, TBR has been shown to be highly uptake time dependent for 18F-FLT, requiring very timely procedures, which are often not feasible in clinical practice (18,20). In our study static images were derived from dynamic 18F-FLT PET/CT scans, avoiding differences in uptake time between the test and retest scan that could possibly affect variability measurements. We found that normalizing SUV to blood pool radioactivity concentrations did not improve variability for the OSEM
reconstructed datasets, and even negatively affected the repeatability for $^{18}$F-FLT images reconstructed with FBP. This is similar to results seen in other studies (14,21).

In contrast to $^{18}$F-FLT, the variability of quantitative $^{18}$F-FDG uptake metrics has been widely investigated in single as well as multi-center setting (22-25), and integrated into the PET Response Criteria in Solid Tumors (PERCIST) (7). PERCIST proposes a threshold of 30% change in SUV (combined with a minimal absolute change of 0.9 SUV units) to discriminate partial response and progressive disease. Evidence underlying these thresholds consists of both stand-alone PET and PET/CT data with variable uptake intervals (60 vs. 90 min post-injection). The PERCIST criteria recommend measuring up to 5 lesions for response assessment in metastatic disease. However the impact of lesion selection strategies on repeatability of $^{18}$F-FDG uptake metrics has not been extensively assessed. Additionally, it could be hypothesized that overall response in metastatic disease is more predictive for response to anti-tumor therapy than that of individual lesions. The influence of this per patient analysis on repeatability has also been assessed very limitedly (22).

In Chapter 4 we comprehensively investigated the repeatability of various quantitative whole body $^{18}$F-FDG uptake measures in advanced NSCLC patients as a function of uptake interval and lesion selection strategy. Patients underwent double baseline $^{18}$F-FDG PET/CT scans, at 60 and 90 minutes post injection (p.i.), within 3 days. Without lesion selection criteria repeatability coefficients of $\text{SUV}_{\text{mean}}$ and $\text{SUV}_{\text{peak}}$ ranged from 15%-20%. When applying the PERCIST lesion selection criteria performance improved, producing repeatability coefficients of less than 15% and even less than 10% if repeatability was assessed per patient. Our results were consistent with other single center studies, however in multi-center setting often a slightly larger variability ($\pm 30\%$) is found (22-25). Scan execution according to strict adherence to protocol of $^{18}$F-FDG PET/CT scans might be more difficult in this setting and could affect test-retest performance of multicenter studies through an accumulation of small errors. It has been suggested that normalization to blood or liver uptake could improve repeatability by correcting for these small errors (26). However, in contrast to expectations, normalization of SUV to any of these factors did not improve repeatability.
In the PERCIST criteria only SUV measures are taken into account, while there is also an increasing interest in alternative $^{18}$F-FDG uptake measures such as metabolically active tumor volume (MATV) and total lesion glycolysis (TLG). There is only limited data on the repeatability of these quantitative uptake metrics, let alone that the influence of uptake time and lesions selection has been assessed (15,27-29). Variability of MATV and TLG was larger compared to SUV measures in general. Yet, similar to SUV a decrease in variability is seen when lesions selection criteria are applied. These results suggest that higher thresholds are required when assessing response using $^{18}$F-FDG volume estimates.

No significant difference in variability of quantitative uptake metrics was found at 60 and 90 minutes p.i.. Nevertheless, absolute SUV values and TLG were higher at 90 min p.i., stressing the importance of timely procedures in quantitative PET/CT imaging. This can be difficult in daily clinical practice and supports the need for methods to correct SUV for differences in uptake time in longitudinal setting (20,30). We applied two methods suggested by Van den Hoff et al. (30) and found an improvement of the correlation between the data acquired at 60 min p.i. from the test scan and the data acquired at 90 min p.i. from the retest scan corrected to 60 min. These methods take tracer kinetics into account and use the estimated change of bloodpool SUV over time to correct the tumor SUVs. In addition, repeatability of the quantitative uptake metrics was not affected by these correction methods and therefore they have potential as strategy for compensation of undesired variability in uptake times between scans.

As presented in the systematic review in chapter 2, early response prediction using quantitative $^{18}$F-FLT PET/CT uptake metrics in NSCLC often correlated well with regular response evaluation using RECIST and progression free survival, but no correlation was found with overall survival. Most studies investigating the performance of $^{18}$F-FLT as biomarker of response only assessed $^{18}$F-FLT uptake using a region-based method (e.g. SUV). A drawback of these methods is the loss of spatial information as only the maximum, peak or average value of the VOI is reported (31). Spatial heterogeneity or heterogeneous change in $^{18}$F-FLT uptake after start of treatment cannot be assessed using these region-based methods, potentially hampering accurate response evaluation. Voxel-based analysis using parametric methods preserves spatial

8
information and enables assessment of spatial features within regions of interest. In Chapter 5 we technically validated such parametric methods in advanced-stage NSCLC patients with an activating EGFR mutation before and during TKI treatment.

We assessed three different parametric methods: Logan graphic analysis (LGA) (32), a 2-tissue-compartment basis function model (BFM) (33) and spectral analysis (SA) (34). All generated qualitatively good parametric images after optimization of input parameter settings for use with $^{18}$F-FLT, and both BFM and SA showed a strong correlation with non-linear regression (NLR) based $V_T$ before as well as after start of treatment. LGA also showed a strong correlation however, it had a larger negative bias compared to the two other methods. Another advantage of BFM and SA over LGA is the possibility to obtain parametric $K_1$ images for assessment of changes in $^{18}$F-FLT influx. BFM yielded slightly more accurate parametric $K_1$ images than SA.

In addition, we studied the sensitivity of these parametric methods to noise. Two statistically and hemodynamically equivalent datasets were generated by count-wise splitting of the original list-mode data. Parametric images were regenerated and pharmacokinetic parameters were recalculated for both datasets containing 50% of the original counts. Correlation between $V_T$ of the split datasets was strong on voxel- and whole-tumor level for all methods. Furthermore difference with the original dataset containing 100% of the voxels was small ($< 5\%$). Variability of $V_T$ values on regional-level was smaller compared to assessment on voxel-level. For $K_1$ a similar pattern is seen, but variability at both voxel and regional level was larger in general. This was expected as $K_1$ is known to be more sensitive to small differences in input function and therefore more susceptible to increased noise (35,36). Based on these results BFM was recommended as preferred parametric method for analysis of dynamic $^{18}$F-FLT PET/CT studies. Nevertheless, SA can also be used for accurate generation of parametric $V_T$ and $K_1$ images, even when less activity is injected.

**PART 2**

The androgen receptor (AR) plays a pivotal role in all stages of prostate cancer, including that of the most advanced stage: metastasized castration-resistant prostate cancer (mCRPC). Many different options are available for treatment of mCPRC patients,
including several agents specifically targeting the AR, which have shown to improve survival and quality of life in the majority of mCPRC patients (37-40). Nonetheless, a subset of patients will not respond to these drugs, especially after first line AR-directed treatment (41). Direct assessment of AR on a lesion-by-lesion level using 18F-FDHT could potentially serve as a prognostic or even predictive imaging biomarker for AR-directed therapies and consequently limit unnecessary toxicities and costs because of ineffective treatment. In addition, it has the potential to aid AR-targeted drug development by determining the biologic dose, specific targeting, and AR avidity in early drug development.

The reference method for quantitative assessment of tracer uptake in PET is pharmacokinetic modeling (42). This method requires complex dynamic scanning protocols, is expensive and only feasible in highly specialized PET centers. Moreover, it is not compatible with the whole body acquisitions typically required in patients with metastasized prostate cancer. Therefore, to perform large multi-center trials and use 18F-FDHT in routine clinical practice for assessment of AR-status in mCPRC, more simplified methods for quantification of 18F-FDHT uptake have to be validated.

A previous report performed pharmacokinetic modeling of the 18F-FDHT tracer and compared it to SUV normalized to bodyweight (SUVAUW) (43). However, in this study no continuous arterial sampling was used, which could possibly have confounded results of the pharmacokinetic assessment. In Chapter 6 we therefore validated the pharmacokinetic modeling of 18F-FDHT uptake in mCPRC patients using continuous arterial sampling and found that tumor time-activity curves were best described by a 2-tissue irreversible compartment model, suggesting an irreversible uptake of the tracer. This model provides the net influx rate macroparameter Ki as kinetic rate constant of 18F-FDHT uptake. The Ki obtained using an IDIF calibrated with venous sample data correlated nearly perfect with those of the reference standard, eliminating the need of arterial sampling yet still requires dynamic scanning.

In a larger multicenter population we evaluated the use of simplified methods suitable for whole-body PET/CT assessment of 18F-FDHT and compared them to Ki obtained using venous sampling. SUVAUCPP (SUV normalized to the area under the parent-plasma input curve) showed a strong correlation with NLR, however requires an extra 30 min scan during the initial dynamic phase of the tracer and venous blood
samples to obtain the AUC of the parent plasma input function. Further simplifications, such as $\text{SUV}_{BW}$, $\text{SUV}_{LBM}$, $\text{SUV}_{WB}$, and $\text{SUV}_{PP}$ (SUV normalized to parent-plasma blood activity concentrations) and $\text{SUV}_{AUC,WB}$ (SUV normalized to the area under the parent-plasma input curve), lead to a decrease in accuracy, which is mainly caused by loss of information on tracer kinetics. This may particularly hamper the use of these simplified methods in longitudinal setting, where kinetics might change after start of or during treatment. In blood the majority of $^{18}$F-FDHT is bound to proteins, of which sex hormone-binding globulin (SHBG) is the most common (44,45). In principle, only non-protein bound tracer is able to bind to the AR and therefore correcting $^{18}$F-FDHT uptake for SHBG levels may partly compensate for the loss of information on tracer kinetics. This is supported by the fact that correcting $\text{SUV}_{BW}$ for plasma SHBG levels ($\text{SUV}_{SHBG}$) significantly improved correlation with NLR. In contrast to parent plasma fraction analysis, SHBG measurements are widely available and less cumbersome. Determining serum SHBG just prior to the $^{18}$F-FDHT scan could therefore potentially be used as a surrogate for parent plasma measurements.

Additionally, we found that quantitative $^{18}$F-FDHT uptake metrics were not affected by tumor perfusion. Despite $^{18}$F-FDHT being an irreversible tracer undergoing fast metabolism (< 30 min), relatively small $k_3$ values limit the influence of perfusion on tracer uptake. These results have to be validated for response evaluation studies, especially for drugs that may have an impact on tumor perfusion.

An essential step in technical validation of $^{18}$F-FDHT for evaluation and/or prediction of response to AR-directed therapies is determination of the repeatability of these quantitative parameters. In Chapter 6 we evaluated the repeatability of full kinetic modeling parameters and several simplified methods. The variability of simplified methods was smaller compared to that of $K_i$ and other microparameters obtained using a 2-tissue irreversible compartment model. This was expected, as parameters obtained using pharmacokinetic modeling are generally more susceptible to noise compared to simplified methods (46). In Chapter 7 we assessed the repeatability and reproducibility of several quantitative whole-body $^{18}$F-FDHT PET/CT uptake metrics. Similar to other studies in prostate cancer, SUV metrics provided the best repeatability (Repeatability coefficients 21-25%) with high interreader reproducibility (Coefficient of Variance < 2.5%) (21). Variability of volumetric
measures was larger in the current study compared to those seen in other studies; most likely due to the use of manual background definition. Uptake metrics that are dependent on tumor segmentation methods (e.g. volumetric measures) are influenced by inter- and intra-observer variability in the defined background. We therefore suggest the use of automatic background definition if available. Assessing repeatability of quantitative $^{18}$F-FDHT uptake metrics on patient level improved results, in particularly for the volumetric measures. Other lesion selection criteria based on size or absolute SUV$_{\text{max}}$ did not affect variability. The same applies for PSA levels, Gleason score, weight, age and differences in uptake time between both scans.

We also evaluated the repeatability of SUV$\text{AUC,PP}$ on whole body $^{18}$F-FDHT PET/CT scans. As shown in chapter 6 this metric showed near perfect correlation with $K_i$ obtained using pharmacokinetic modeling. We normalized SUV$_{\text{max}}$, SUV$_{\text{peak}}$ and SUV$_{\text{mean}}$ to the AUC of the parent plasma input function. This increased the variability of all parameters on lesion and patient level. In this study two outliers were found explaining a large part of the increase in the observed variability. In both subjects a 50% difference in the whole blood activity concentrations between the test and retest scans was found. As is also seen in previous studies, including an additional variable into uptake metric calculations can increase measurement uncertainty and therefore attention for errors is needed when using these uptake metrics in response evaluation setting (14,21).
Chapter 8

Response evaluation using the tracers described in this thesis could add valuable information to conventional imaging techniques assessing only morphological changes after start of therapy in clinical setting. Furthermore, they may aid the field of drug development by assessing the efficacy of novel drugs at an early stage. Over the course of recent years several new (targeted) therapies have been developed leading only to limited change in morphological features. Yet, glucose metabolism and proliferation rates are likely to decrease at an early stage and therefore both $^{18}$F-$^{\text{FLT}}$ and $^{18}$F-$^{\text{FDG}}$ have potential as imaging biomarkers of early response assessment. One major disadvantage of $^{18}$F-$^{\text{FLT PET/CT}}$ compared to $^{18}$F-$^{\text{FDG}}$ is its biodistribution. High $^{18}$F-$^{\text{FLT}}$ uptake in liver and bone marrow, both frequent locations for metastasis, limits its sensitivity and consequently its applicability for response evaluation in metastasized disease. Secondly, $^{18}$F-$^{\text{FLT}}$ uptake is dependent on the fraction of cells S-phase of the cell cycle and can be influenced by de novo synthesis of thymidine. In addition, $^{18}$F-$^{\text{FLT PET/CT}}$ imaging procedures have to be standardized as has been done for quantitative $^{18}$F-$^{\text{FDG}}$ imaging. Up to now, several imaging protocols have been used, which could have influenced results of quantitative analyses and hindered comparability and therefore pooling of results. PET studies tend to have small sample sizes. Yet, to clinically validate $^{18}$F-$^{\text{FLT}}$ or $^{18}$F-$^{\text{FDG}}$ as imaging biomarker of treatment response, large prospective trials evaluating a wide variety of quantitative uptake metrics and interventional studies with larger clinically representative patient populations are needed. Several studies have compared the performance of both tracers and found that $^{18}$F-$^{\text{FDG}}$ often outperformed $^{18}$F-$^{\text{FLT}}$ in case of systemic therapy. Additionally, there is more experience with $^{18}$F-$^{\text{FDG}}$ scanning and as already mentioned imaging procedures are more standardized. We therefore suggest that these larger prospective and interventional studies primarily focus on $^{18}$F-$^{\text{FDG}}$ as imaging biomarker of response to systemic therapy. Thresholds for response and progression based on $^{18}$F-$^{\text{FDG}}$ should be defined prospectively and take intrinsic variability of individual uptake metrics into account. Nevertheless, this does not mean that $^{18}$F-$^{\text{FLT PET/CT}}$ scans are redundant; on the contrary. In this era of increasing interest in personalized medicine, not only

Figure 8.1: An overview of the status of $^{18}$F-FDG, $^{18}$F-FLT and $^{18}$F-FDHT PET/CT in the process of clinical validation as quantitative imaging biomarker of response. Technical validation has been performed for all tracers. $^{18}$F-FLT and $^{18}$F-FDHT require further clinical and biological validation and only $^{18}$F-FDG PET/CT is ready to be evaluated in larger trials. (‡: Additional technical validation of $^{18}$F-FDHT has to be done to assess the accuracy of simplified methods after start of treatment, *: Multi-Center assessment of repeatability of $^{18}$F-FLT uptake metrics has only been performed in a meta-analysis; #: $^{18}$F-FLT PET/CT requires further correlation with tumour biology for some types of cancer)
FUTURE PERSPECTIVES

Response evaluation using the tracers described in this thesis could add valuable information to conventional imaging techniques assessing only morphological changes after start of therapy in clinical setting. Furthermore, they may aid the field of drug development by assessing the efficacy of novel drugs at an early stage.

Over the course of recent years several new (targeted) therapies have been developed leading only to limited change in morphological features. Yet, glucose metabolism and proliferation rates are likely to decrease at an early stage and therefore both $^{18}$F-FLT and $^{18}$F-FDG have potential as imaging biomarkers of early response assessment. One major disadvantage of $^{18}$F-FLT PET/CT compared to $^{18}$F-FDG is its biodistribution (47,48). High $^{18}$F-FLT uptake in liver and bone marrow, both frequent locations for metastasis, limits its sensitivity and consequently its applicability for response evaluation in metastasized disease. Secondly, $^{18}$F-FLT uptake is dependent on the fraction of cells S-phase of the cell-cycle and can be influenced by de novo synthesis of thymidine. In addition, $^{18}$F-FLT PET/CT imaging procedures have to be standardized as has been done for quantitative $^{18}$F-FDG imaging. Up to now, several imaging protocols have been used, which could have influenced results of quantitative analyses and hindered comparability and therefore pooling of results.

PET studies tend to have small sample sizes. Yet, to clinically validate $^{18}$F-FLT or $^{18}$F-FDG as imaging biomarker of treatment response, large prospective trials evaluating a wide variety of quantitative uptake metrics and interventional studies with larger clinically representative patient populations are needed. Several studies have compared the performance of both tracers and found that $^{18}$F-FDG often outperformed $^{18}$F-FLT in case of systemic therapy (3,49-52). Additionally, there is more experience with $^{18}$F-FDG scanning and as already mentioned imaging procedures are more standardized (1). We therefore suggest that these larger prospective and interventional studies primarily focus on $^{18}$F-FDG as imaging biomarker of response to systemic therapy. Thresholds for response and progression based on $^{18}$F-FDG should be defined prospectively and take intrinsic variability of individual uptake metrics into account.

Nevertheless, this does not mean that $^{18}$F-FLT PET scans are redundant; on the contrary. In this era of increasing interest in personalized medicine, not only...
personalized treatment but also personalized biomarkers are required. Response evaluation with \textsuperscript{18}F-FDG PET/CT can be hampered due to therapy-induced inflammation. This is often the case after radiotherapy, for example in head and neck cancer, impeding the identification of viable tumor tissue and evaluation of treatment response (53). Some small studies also showed the superiority of \textsuperscript{18}F-FLT compared to \textsuperscript{18}F-FDG in response assessment after radiotherapy but this remains to be further explored.

We already mentioned differences in biodistribution between \textsuperscript{18}F-FDG and \textsuperscript{18}F-FLT PET. Even though the high uptake in liver and bone marrow are drawbacks, there is no physiological \textsuperscript{18}F-FLT uptake in the brain in contrast to \textsuperscript{18}F-FDG. This potentially makes \textsuperscript{18}F-FLT PET a good candidate for response evaluation in brain tumors. Due to the relatively small area of interest not requiring whole body acquisitions, it also enables more complex techniques such as dynamic scanning and evaluation using parametric imaging. These techniques could potentially provide valuable extra information on several pharmacokinetic parameters and tumor heterogeneity. This does not only apply for primary brain tumors, but of course also for brain metastasis which occurs in about 20-40\% of the lung cancer patients (54,55). A general problem with many therapeutic drugs is that they do not easily to penetrate the blood-brain-barrier (BBB) and are thus less or even ineffective in treatment of brain metastasis even though the extra-cranial disease is responding (56). \textsuperscript{18}F-FLT could therefore provide a non-invasive method to assess the capability of drug to penetrate the BBB. No change in proliferation and consequently in \textsuperscript{18}F-FLT uptake after start of treatment suggests the inability of drugs to reach its target. This could be used to identify the feasibility of drugs to penetrate the BBB in routine clinical practice and adjust therapy, however also in an early stage of drug development. Moreover, \textsuperscript{18}F-FLT PET could be used to assess drug efficacy in an early stage of its development in general. It could be incorporated in phase 1 studies and if no change in proliferation is observed the likelihood of obtaining and effective treatment is very low. Rejecting a drug in an early phase would markedly save development costs, as large clinical trials involved in the later stages of drug development are highly expensive (57). Nevertheless, adequate biological validation remains to be assessed for some tumor types (58).
\[^{18}\text{F}-\text{FDHT PET/CT}\] is a promising imaging biomarker for response evaluation as well as prediction of efficacy of AR-targeted therapies. As an essential step in the development of potential biomarkers, we technically validated simplified methods for quantitative evaluation of \[^{18}\text{F}-\text{FDHT}\] uptake and assessed the repeatability of several whole-body \[^{18}\text{F}-\text{FDHT PET/CT}\] uptake metrics in a multi-institutional study. As a next step in the development of \[^{18}\text{F}-\text{FDHT}\] as biomarker of response, clinical validation studies have to be performed assessing the response to AR-targeted therapies such as abiraterone. Additional technical validation has to be performed to assess the accuracy of simplified methods after start of treatment. Some of these studies are currently ongoing and should apply thresholds for response assessment based on intrinsic variability presented in this thesis. In addition, it would be interesting to study the biological changes in mCPRC patients during first and second line AR-target therapies, as this is a period in which key resistance mechanisms develop (59,60). This does also apply for non-AR directed therapies (e.g. docetaxel), which are often initiated after failing first line AR-targeted therapy. It has been shown that these drugs also influence AR expression in prostate cancer cells, however mainly in \textit{in-vitro} assessment. \[^{18}\text{F}-\text{FDHT}\] could help to evaluate this \textit{in-vivo} (61-63).

We have assessed the quantitative accuracy of simplified methods for quantification of \[^{18}\text{F}-\text{FDHT}\] uptake. These results should be validated and quantitative \[^{18}\text{F}-\text{FDHT}\] measurements should be correlated against quantitative analysis of AR expression in histopathology. This ideally also requires a method able to adequately quantify AR expression in tissue specimens, rather than detecting AR presence or absence.

\[^{18}\text{F}-\text{FDHT PET/CT}\] aims at phenotyping lesions rather than mere lesion detection and therefore a second imaging modality is needed to assess the total disease burden. Vargas et al. (64) have demonstrated the accuracy of \[^{18}\text{F}-\text{FDHT}\] at identifying tumor localizations and a possible association between number of \[^{18}\text{F}-\text{FDHT}\] avid mCPRC lesions and survival. The former study used CT and \[^{18}\text{F}-\text{FDG PET/CT}\] scans to evaluate \[^{18}\text{F}-\text{FDHT}\] uptake in bone lesions only. However, recent studies have shown that new PSMA-ligand PET tracers have higher specificity and sensitivity for detection of prostate cancer metastasis in patients with high and low PSA levels (65-68). Moreover, it would enable better assessment of the total disease burden, because bone as well as
soft tissue lesions can be visualized simultaneously. Combining PSMA-ligand PET/CT with $^{18}$F-FDHT PET/CT imaging in response evaluation setting could not only provide a comprehensive overview of disease heterogeneity in mCPRC patients at baseline, but also possibly demonstrate heterogeneity of response and progression when receiving AR-targeted therapy. Additionally, a head-to-head comparison of PSMA-ligand and $^{18}$F-FDHT uptake could provide insight into the tumor biology as PSMA and AR expression are inversely correlated (69,70).

Besides PET, several other serum and tissue-based prognostic and predictive biomarkers are being investigated. The most prominent biomarker of resistance to AR-targeted therapy is the splice-variant AR-V7. AR-V7 positive tumor cells are constitutively activated, lack a ligand-binding domain and can be measured in circulating tumor cells (CTCs) (60,71,72). Several other biomarkers include genomic profiling of tumor tissue specimens and circulating cell-free DNA (73-77). Despite the safety and ease of serum biomarkers only the circulating component of the disease is assessed. Tissue biopsies only assess the genomic profile of one lesion, which may or may not represent the remaining disease and moreover it cannot be repeated. Future studies are needed to compare the performance of these biomarkers and assess whether they compete or complement each other.

It is important to perform technical validation of quantitative imaging biomarkers prior to or preferably simultaneously with clinical and biological validation studies. If the latter two are performed prior to technical validation studies, there is a high risk of selecting quantitatively inaccurate and imprecise uptake metric. This may mean that vast amounts of money may be wasted and ethical questions could be raised due to unnecessary and futile patient exposure to medical procedures and, in case of PET and CT, also to radiation. However, if performed in consecutive order, these studies may cost a lot of time and thus delay large clinical trials and implementation in routine clinical practice. In our opinion these various validation steps could therefore best be performed simultaneously in one multicenter validation study. The initial trial protocol should contain the steps presented in figure 8.2 to enable pharmacokinetic analysis, validation of simplified methods, assessment of perfusion dependency and accurate response assessment without causing delay. After determining the quantitatively most
accurate simplified method and assessing its perfusion dependency in part A, a more simplified scan protocol could be used to scan the remaining patients. In case of response assessment repeatability of the uptake measures should also be assessed and a dynamic PET protocol is required to evaluate the influence of changed pharmacokinetics, perfusion and blood volume on the accuracy of simplified methods. The results of these scans can subsequently be correlated biological and/or clinical outcome variables, which can be obtained as part of the previously described study.

**Quantitative imaging biomarker trial design for technical validation:**

**Part A. The first subset of patients scanned with a novel radiotracer:**
1. Dynamic PET/CT scans with (continuous) arterial and venous blood sampling.
2. Dynamic H₂¹⁵O PET/CT

**Part B. Extended cohort (Point 2 - 4 only in case of response assessment biomarkers):**
1. Simplified scan protocol
2. Repeated baseline scans within 1 week
3. Dynamic PET/CT scan with blood sampling after start of treatment
4. Dynamic H₂¹⁵O PET/CT after start of treatment

**Figure 8.2**

Furthermore, as part of biomarker development the cost effectiveness of imaging biomarkers has to be determined. Health-care costs in the Netherlands covered ± 14% of the gross domestic product (96 billion euro) in 2016 (78) and about 25% of these costs consisted of hospital care. Moreover, an increasing proportion of the hospital budgets (8.5% in 2015) is spent on “expensive” drugs, which mainly contains novel oncological therapies (79). Accurate prognostic and predictive biomarkers in the oncological setting could therefore help to diminish these costs by providing the right drugs to the right patients and minimizing costs related to ineffective treatment. Even though PET scans are a relatively expensive diagnostic tool for prediction and evaluation of response, it has potential to provide information needed to avoid clinical decisions that do not add value for patients (e.g. improve overall survival or quality of life) and increase healthcare costs. This potential savings can be directly calculated,
however the more indirect financial gains of minimizing ineffective treatment should also be taken into account in cost-effectiveness studies. These include costs related to treatment of unnecessary side effects and complications, such as drug use and hospital admissions and possibly even time of work. Yet, the latter may be of less importance in an oncological population.

**CONCLUSION**

\(^{18}\text{F-FDG and }^{18}\text{F-FLT PET/CT have been technically validated for use as imaging biomarkers of response in NSCLC patients. In addition, }^{18}\text{F-FDHT PET/CT has largely been technically validated as a tool for evaluation of response in CRPC patients. Both }^{18}\text{F-FLT and }^{18}\text{F-FDHT, however, require further biological and clinical validation studies before implementation in larger prospective and/or interventional trials evaluating treatment response. Furthermore, while }^{18}\text{F-FLT PET could be used to assess drug efficacy in an early stage, adequate biological validation still has to be performed for some tumor types and biodistribution of }^{18}\text{F-FLT should be taken into account.}

Moreover, we suggest that technical validation of any future quantitative PET imaging biomarkers should be performed simultaneously with clinical and biological validation in a multicenter setting. This design helps to avoid selection of quantitatively inaccurate and imprecise uptake metrics and delay of implementation in routine clinical practice.
REFERENCES


Chapter 8


Summarizing discussion and future perspectives


Chapter 8


