Response evaluation during targeted therapy: early incorporation of imaging biomarkers

Joop de Langen

ISBN: 978-90-8659-609-6

Cover and lay-out: Marthe Noordijk

The printing of this thesis was financially supported by Amgen BV, Lilly Nederland BV, Roche Nederland BV, Bayer Healthcare, GlaxoSmithKline Pharmaceuticals en Consumer Healthcare, Pfizer BV, Novartis Pharma BV, Boehringer Ingelheim BV, Janssen-Cilag BV and Philips Healthcare Nederland.

VRIJE UNIVERSITEIT

Response evaluation during targeted therapy: early incorporation of imaging biomarkers

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op vrijdag 27 april 2012 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Adrianus Johannes de Langen

geboren te Alkmaar

promotoren:	prof.dr. E.F. Smit
	prof.dr. O.S. Hoekstra
copromotor:	dr. M. Lubberink

Contents

Chapter 1	Introduction and scope of the thesis	7
Chapter 2	Use of $H_2{}^{15}O$ PET and DCE-MRI to measure tumor blood flow *Oncologist. 2008; 13(6): 631-644*	23
Chapter 3	Reproducibility of tumor perfusion measurements using ^{15}O-labeled water and PET *J Nucl Med. 2008; 49(11): 1763-1768*	49
Chapter 4	Repeatability of ^{18}F-FDG uptake measurements in tumors: a meta-analysis *Accepted for publication in J Nucl Med*	63
Chapter 5	Reproducibility of quantitative ^{18}F-3'-deoxy-3'-fluorothymidine measurements using positron emission tomography *Eur J Nucl Med Mol Imaging. 2009; 36(3): 389-395*	83
Chapter 6	First-line erlotinib and bevacizumab in patients with locally advanced and/or metastatic non-small-cell lung cancer: a phase II study including molecular imaging *Ann Oncol. 2011; 22(3): 559-566*	97
Chapter 7	Monitoring response to antiangiogenic therapy in non-small-cell lung cancer using imaging markers derived from PET and dynamic contrast-enhanced MRI *J Nucl Med. 2011; 52(1): 48-55*	115
Chapter 8	Discussion and future perspectives	133
Nederlandse samenvatting		147
Dankwoord		157
Curriculum Vitae		163
Publications		167

1

Introduction and scope of the thesis

FROM "ONE SIZE FITS ALL" TO "PATIENT CUSTOMIZED" TREATMENT

Anticancer drug development attempts to translate understanding gained from basic research into improved clinical practice through cancer drug trials [1]. These trials aim to test new ways of cancer treatment. Traditionally, the specific goal of early therapeutic trials is to define the safety, tolerability and pharmacological properties (phase I) as well as antitumor efficacy of novel agents (phase II). Later stage therapeutic trials attempt to prove that a treatment imparts clinical benefit, usually compared to standard treatment (phase III and IV). The latter generally include hundreds to thousands of patients, can last several years and be very expensive; the often quoted price to generate one licensed drug is US$ 1 billion [2, 3]. Frustratingly, the majority of clinical cancer trials have little impact on either patient benefit or understanding of cancer biology, raising major concerns about the current anticancer drug development process [2, 3]. The overall concern is that it increases the pressure on the pharmaceutical industry to generate anticancer drugs with broad applicability in cancer patients and therefore large fiscal returns [4].

This 'one size fits all' approach may not be the best or most efficient way to develop anticancer drugs [2], as is evident from the high proportion of negative large randomized trials for common cancers as well as the very limited benefits achieved in terms of disease outcome for the small proportion of positive trials that lead to drug approval. These trials impart the risk that an active drug for a small subgroup is being masked by inactivity in the large population, resulting in disapproval of the drug for all. If the trial turns out positive, this will only define the best treatment for the average patient whereas it may not be the best treatment on the individual level.

The degree of benefit from a new treatment should preferably be recognized in early phase trials, i.e. before large randomized trials are pursued, resulting in a decrease in the proportion of large, late stage, negative trials. To do this, a paradigm shift in the process of drug development and response evaluation is needed, facilitated by sophisticated markers of treatment benefit, so-called biomarkers.

BIOMARKER FACILITATED DRUG EVALUATION

The discovery of increasingly numbers of novel molecularly targeted anticancer drugs and the many possible combinations of these agents necessitates "hypothesis-testing drug evaluation" in selected patients. This process starts with in vitro studies identifying a target presumed to be key to tumor survival. Subsequently, multiple drugs are tested in in vitro and animal models and those that modulate the target in a way that results in cytotoxic or cytostatic effects are selected. During this preclinical work, hypotheses can be generated on how and to what degree the drug alters the target and downstream pathways by using static and dynamic markers that reflect these changes (e.g. phosphorylation of proteins, apoptosis, proliferation, perfusion, metabolism). The dose range and schedule that results

in optimal target modulation is selected for clinical studies.

Next, in vivo validation is warranted. Although early clinical trials should still continue to focus on drug safety and tolerability in combination with pharmacokinetics ("what the body does to the drug"), pharmacodynamics ("what the drug does to the body/tumor") should be more emphasized before large phase III trials are designed and executed. Pharmacodynamics can identify which patients show the drug-induced tumor changes that were observed in the preclinical models (hypothesis-testing), thereby offering the opportunity to link tumor characteristics with the degree of treatment benefit. Eligibility criteria for subsequent phase III trials can then be adjusted to include only those patients that are expected to benefit from the treatment.

Because drugs generally trigger a cascade of tumor alterations, the direct target itself (e.g. a tumor growth factor) or more downstream effects (e.g. proliferation or cell viability) are available for pharmacodynamic measurements. Downstream effects are often the result of several upstream effects and thus less specific, but have the advantage that unexpected drug effects can be monitored and that the net result of interactions between several upstream molecular pathways can be evaluated.

BIOMARKER TYPES

Following the definition adopted by the Biomarkers Definitions Workgroup, a biomarker is being defined as "a characteristic that indicates normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention" [5]. Mechanistic understanding of tumor and drug biology can be obtained by biomarkers and their use can speed up drug development resulting in more patients being treated with the optimal drug regimen. Each step of drug development requires specific biomarker classes. Of note, these categories are not mutually exclusive.

A *pharmacodynamic* (PD) biomarker is a dynamic assessment that shows that a biological process has occurred or altered in a patient after treatment.

- Examples are a change in protein expression or phosphorylation or a decrease in tumor metabolic activity, perfusion or size.
- PD biomarkers can be used to confirm the proof of mechanism by identifying the intended change in tumor biology, also called "target modulation".
- PD biomarkers can guide selections of drug dose and regimen for testing in phase II/III studies (i.e. which regimen results in optimal target modulation).
- In contrast to predictive and prognostic biomarkers, PD biomarkers do not have to relate with patient or treatment outcome.

A *predictive* biomarker is a baseline characteristic that categorizes patients by their likelihood to respond to a therapeutic intervention.

- Examples are genetic alterations that sensitize patients to certain drugs.
- They may predict a favorable or unfavorable response (i.e. an adverse event).

A *prognostic* biomarker is a baseline patient or disease characteristic that informs about the natural history of the disorder in that particular patient in the absence of a therapeutic intervention.

Surrogate endpoint biomarkers are a subset of pharmacodynamic biomarkers that relate to patient or treatment outcome and are used as a substitute for a clinical efficacy endpoint.

- Examples are measurements of tumor size, metabolic activity, cell proliferation, or perfusion.
- They are expected to predict clinical benefit or harm (or lack of benefit or harm).

BIOMARKER DEVELOPMENT

Like drugs, biomarkers require several phases of development before they can be used for clinical decision making. This includes technical and methodological validation and clinical qualification. Although these steps are uniform for each biomarker (e.g. genomic, immunohistochemical [IHC], protein level, receptor expression, or imaging), we will focus on quantitative imaging biomarkers in this thesis.

Ideally, the development of quantitative imaging biomarkers is hypothesis driven. This can be illustrated by the use of radiolabeled water ($H_2{}^{15}O$), a PET tracer. Drugs have been designed to target tumor vasculature, thereby reducing tumor perfusion with the aim to starve the tumor resulting in growth arrest.

$H_2{}^{15}O$ PET non-invasively measures tumor perfusion and in theory measuring perfusion with $H_2{}^{15}O$ PET allows monitoring drug effects on tumor vasculature. This pharmacodynamic measurement might predict patient benefit, qualifying it as a surrogate endpoint biomarker.

After discovery, the biomarker has to be standardized to obtain identical results in different laboratories and clinics. This includes standardization of acquisition, reconstruction, data processing and data analytical protocols. After standardization, repeatability and reproducibility must be studied to discriminate noise from true biological change. Validation is the often used term to describe this process. A critical distinction should be made between analytical validation and clinical qualification. Although validation, qualification and evaluation have been used interchangeably in the literature, the distinction should be made to properly describe the particular phase the biomarker is transitioning through.

Following analytical validation, a stepwise process of qualification and verification should be followed [6-8]. Qualification describes the clinical evaluation process of a biomarker. It ensures that the biomarker is used in the correct context and that it reliably performs the job it is intended or expected to [9]. That context may be selecting or deselecting people for a clinical trial, monitoring drug-induced toxicity, or other purposes. The amount of evidence needed to qualify a biomarker for a given purpose is related to the consequences of using the result to make decisions, such as to pursue the development of a drug or whether to withhold a drug from individuals in a clinical trial. The use of poorly validated and qualified biomarkers will yield misleading results and can lead to subsequent disregard of a potentially useful biomarker or an inappropriate decision being made about the drug and patient care. Thus the biomarker should be "fit-for-purpose".

DRUG AND BIOMARKER CO-DEVELOPMENT

When known and measurable using validated and qualified biomarkers, molecular characteristics can guide patient stratification in clinical trials. However, most commonly the molecular characteristics that drive a patient's response to treatment are not fully known and might only be discovered during the clinical development of drugs. Also, in case the mechanism is known, validated and qualified biomarkers are rarely available at the start of the trial. This necessitates co-development trials that combine biomarker qualification and drug development [10].

Table 1. Biomarker role in clinical trials, depending on the level of qualification.

Role	Example
Integral = used for clinical decisions.	Prospectively defined populations (selection or stratification factor).
Integrated = intended to validate hypotheses for establishing further trials, not used for clinical decisions in the current trial.	Pharmacodynamic markers in phase I trials; prespecified predictive biomarker analysis in phase II or phase III study, not used for eligibility criteria or patient management.
Exploratory	Retrospective hypothesis generating biomarker identification; pilot studies.

Depending on the available evidence of the predictive value of the biomarker under investigation, combination trials can be designed. The role of a biomarker in clinical trials can be integral, integrated or exploratory (Table 1). An integral role is only justified when the biomarker is already qualified as such. Patients that are sensitive to the drug are identified upfront or early after treatment initiation by the biomarker and the information is used for stratification, thereby directly influencing trial design. Integrated use tests hypotheses on drug action that were raised in preclinical or phase I/II studies. This information can be used to initiate subsequent trials with integral use of the biomarker, but the biomarker is not used for clinical decision making in the current trial. Exploratory use can generate hypotheses on drug effects and how to evaluate these. Biomarkers with an exploratory role should be implemented as early as possible during drug development. If explored in preclinical and phase I studies, they can be validated and qualified simultaneously with drug development in subsequent phase II and phase III studies.

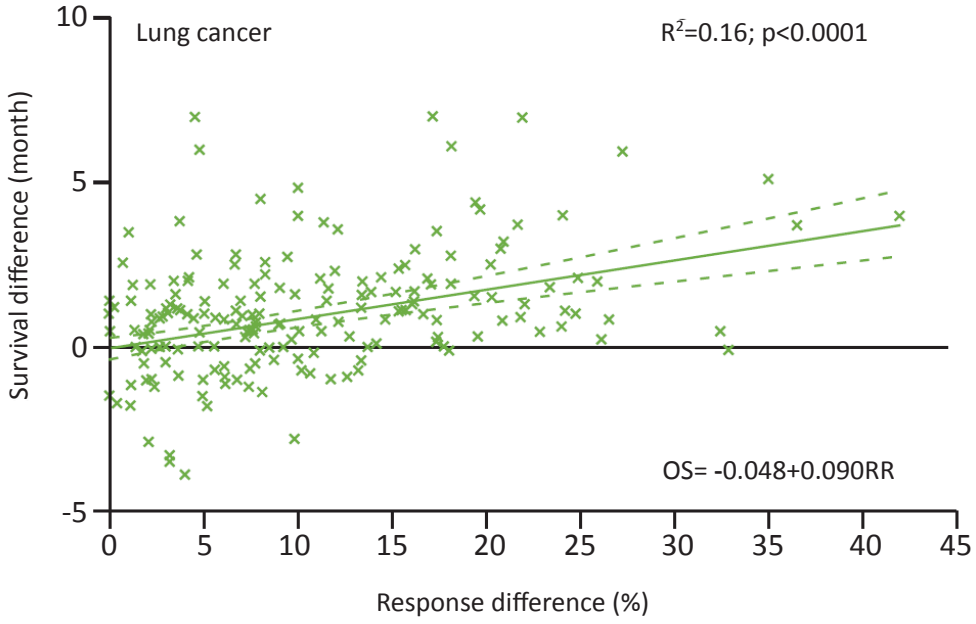
RATIONALE FOR THE INCORPORATION OF FUNCTIONAL IMAGING BIOMARKERS IN THIS THESIS

Response monitoring in advanced non-small-cell lung cancer (NSCLC) is complex. Tumor shrinkage usually does not occur until several cycles of chemotherapy and is difficult to evaluate because of inter- and intra-observer variation [11]. With targeted agents, size change can be absent despite survival benefit [12-15]. Also, the association between anatomical response and survival is weak at best [16] (Figure 1). Early use of the "response evaluation criteria in solid tumors" (RECIST) as surrogate endpoint biomarker might underestimate treatment efficacy because size change is a slow process and stable disease reflects a heterogeneous group of patients with both good and bad outcome. Therefore, alternative response criteria are warranted.

In this thesis we evaluated combined treatment with bevacizumab (B), an antibody against the vascular endothelial growth factor (VEGF), and erlotinib (E), a tyrosine kinase inhibitor (TKI) against the epidermal growth factor receptor (EGFR), aiming to target tumor vasculature and growth signaling pathways. Their supposed effects are a decrease in vascular permeability and vascular density, and an increase in apoptosis [17, 18]. Positron emission tomography (PET) and dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) derived parameters were selected to evaluate target modulation to BE treatment and to qualify the parameters as surrogate endpoint biomarkers.

Because the antivascular activity of combined EGFR and VEGF treatment might result in a decrease in tumor perfusion, $H_2{}^{15}O$ PET and DCE-MRI were selected as pharmacodynamic biomarkers. They were used side-by-side because their signals reflect different aspects of tumor vasculature. In contrast to $H_2{}^{15}O$, gadolinium-based contrast agents are not freely diffusible, but partially leak through the vessel wall into the interstitial space and leak more through pathological tumor vessels as a result of greater permeability. Therefore, DCE-MRI derived

Figure 1. The relation between tumor size change and and the difference in survival for metastatic lung cancer. Although the surrogate endpoint is a significant predictor of survival, much less than half the variability in the survival difference was explained by the variability in the difference in tumor size [40].



parameters do not selectively reflect tumor perfusion, but rather a combination of perfusion, vessel permeability and vascular surface area. DCE-MRI signal approaches true tumor perfusion in case permeability and vessel surface area are not limiting factors for tracer kinetics. However, if so, the signal reflects a combination of perfusion, vessel surface area and permeability.

^{18}F-FDG PET was selected to monitor the EGFR signaling pathway. Akt is a downstream protein kinase of the EGFR signaling pathway and plays a central role in transducing oncogenic signals from EGFR activation to metabolic as well as cell survival and proliferative effects. It is activated by phosphatidylinositol 3-kinase (PI3 kinase), negatively regulated by the dual-specificity phosphatase and tensin homologue (PTEN), and phosphorylates mammalian target of rapamycin (mTOR). PI3 kinase–dependent Akt stimulation regulates glucose metabolism in response to growth factor stimuli [19]. Therefore, Akt seems to be a key mediator of the establishment and maintenance of glycolysis in cancer cells. The hypoxia-inducible factor-1α (HIF-1α) is a downstream mediator of Akt that contributes to the regulation of glucose transport and metabolism [20, 21]. As depicted in Figure 2, emerging data suggest that targeted as well as classical chemotherapeutic agents directly or indirectly affect these signaling pathways, glucose transporters, and metabolic enzymes that control glycolysis. Therefore, glu-

cose metabolism measured with ^{18}F-FDG PET seems to be an attractive pharmacodynamic biomarker to study EGFR targeted therapy. Because activation of EGFR results in accelerated tumor proliferation, the PET proliferation tracer ^{18}F-FLT is another biomarker of interest. Inhibition of the EGFR axis as well as tumor starvation (by impairment of perfusion) might result in a decrease in tumor proliferation, measurable with ^{18}F-FLT PET.

IMAGING BIOMARKERS USED IN THIS THESIS

Ideal biomarkers are accurate, reproducible, minimally invasive, inexpensive, entail negligible or no risk when measured and have a minimal drop-out rate. These requirements can be met by non-invasive methods like PET and DCE-MRI [22]. The techniques can be repeated several times during treatment without interfering with tumor biology and allow to study inter- and intra-lesional heterogeneity. They in-vivo measure tumor pharmacokinetics (by labeling drugs) and dynamics (by monitoring biological change). A main advantage to alternatives like tumor biopsy is that functional imaging parameters reflect tumor biology as a complex of biological processes in both tumor and supporting cells that

Figure 2. Molecular targets of cytotoxic and cytostatic drugs in the pathways controlling glucose metabolism [42].

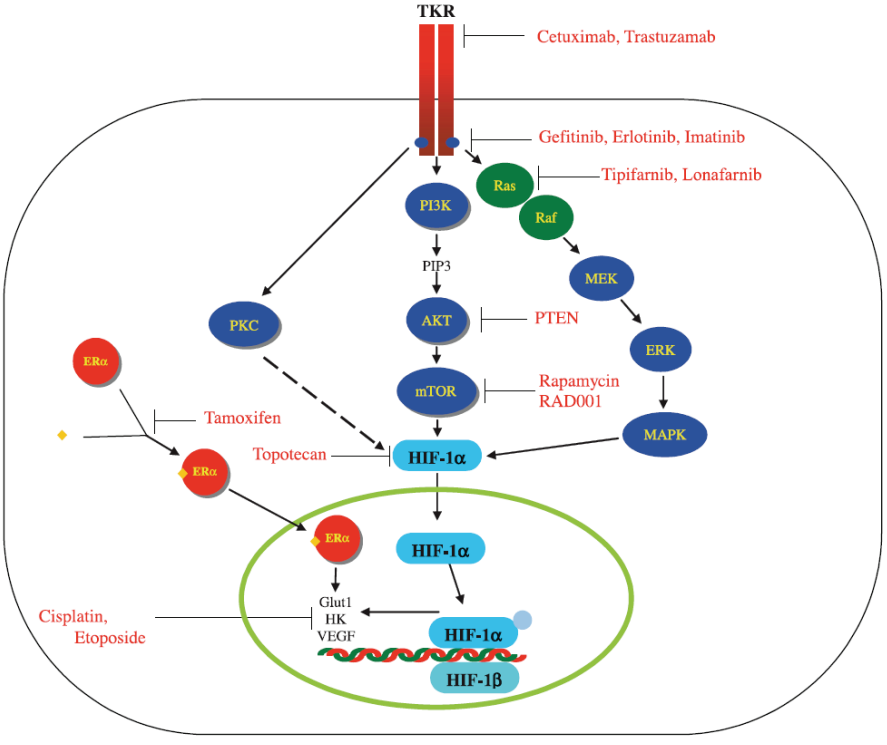


Table 2. Categories of response, according to RECIST *v* 1.1.

Response	Definition
Partial response (PR)	> 30% reduction in the sum of tumor diameter
Complete response (CR)	Total tumor disappearance
Stable disease (SD)	< 30% decrease and < 20% increase in the sum of tumor diameter
Progressive disease (PD)	> 20% increase in the sum of tumor diameter with a minimum of 5 mm absolute increase or appearance of any new lesion

interact with each other. This is an important advantage because clinical failure of drugs that show target modulation and antitumor activity in preclinical models are not an exception [4]. Therefore in-vitro and xenograft results cannot be extrapolated to the clinical setting without in-vivo validation of the results (hypothesis-testing drug evaluation).

Two main categories of imaging biomarkers are distinguished; anatomical and functional. Anatomical imaging biomarkers allow uni-, bi-, or three-dimensional tumor size measurements using static techniques. Functional ones combine anatomical with biological information and allow in-vivo dynamic studies of tumor characteristics without interfering with it.

Anatomical Imaging Biomarkers

Computed Tomography (CT)

Response assessment with CT is based on the response criteria in solid tumors (RECIST). This method uses unidimensional measurements of the sum of the longest diameter of selected target lesions. The criteria were first published in 2000 by an international collaboration including the European Organisation for Research and Treatment of Cancer, the National Cancer Institute of the United States, and the National Cancer Institute of Canada Clinical Trials Group (RECIST *v* 1.0) [23] and updated in 2009 (RECIST *v* 1.1) [24]. Current RECIST guidelines limit size measurements to a maximum of two lesions per organ and five in total. Lesions should have a minimal diameter of 1 cm. Response is categorized as shown in Table 2. RECIST *v* 1.1 also includes a functional imaging biomarker to optimize the detection of tumor progression; any new lesion present on ^{18}F-FDG PET that is compatible with tumor metastasis is regarded as progressive disease.

Functional Imaging Biomarkers

Positron Emission Tomography (PET)

PET allows non-invasive measurements of the 3-dimensional distribution of a positron labelled compound within a living subject. Positron emitting radionuclides can be used to synthesize radiopharmaceuticals that act as substrates for endogenous pathways, thereby enabling the in-vivo tracking of biological processes. In the body, the radiopharmaceuticals emit positrons that undergo annihilation with nearby electrons, resulting in the release of two photons. These so-called annihilation photons leave at an angle of 180 degrees and are detected by coincidence imaging as they strike scintillation crystals. The resulting data can be reconstructed to reveal the distribution of the radiotracer within the subject. Currently, many positron emitters with different characteristics are available, including ^{18}F, ^{15}O, ^{11}C, ^{13}N, ^{124}I, ^{68}Ga and ^{86}Y, which can be used to label virtually every chemical compound. With PET, visualization of biological processes like tumor metabolism, cellular proliferation, specific cell surface receptors, angiogenesis, and tumor hypoxia is possible.

^{18}F-FDG

Today, the most widely used PET tracer is ^{18}F-fluorodeoxyglucose (^{18}F-FDG), a glucose analogue that allows the quantification of glucose metabolism. Tumor cells exhibit high rates of glycolysis to fulfill their increased energy demand. One of the key alterations associated with the high glycolytic rate of cancer cells is increased cellular glucose uptake. The uptake of ^{18}F-FDG is similarly accelerated in tumor cells [25-28].

The most accurate method to analyze tracer kinetics is to quantitatively assess the ^{18}F-FDG uptake rate over time. The metabolic rate for glucose is calculated from the time course of radiotracer concentration in tissue and arterial blood. Dynamic scanning following injection as well as an arterial input function (AIF) are required. For thoracic studies, the latter can be derived from vascular structures within the field of view.

Other methods of kinetic analysis include visual and semi-quantitative assessment of ^{18}F-FDG metabolism. These approaches assume that ^{18}F-FDG uptake is virtually complete and that the dephosphorylation rate is negligible. The standardized uptake value (SUV) is a semi-quantitative index of tumor uptake normalized to the injected dose and a measure of the total volume of distribution, e.g. body weight.

The SUV is a function of patient size, time interval between injection and scanning (usually 60 min), plasma glucose level and image reconstruction settings [29]. SUVs strongly correlate with the metabolic rate of ^{18}F-FDG. Because it does not require dynamic data acquisition and arterial blood sampling, the SUV has frequently been used as a measure of ^{18}F-FDG uptake to assess differences between scans. The main disadvantage of SUV, however, is the inherent assumption that plasma clearance is

always the same. If plasma clearance of ^{18}F-FDG changes as a result of therapy (i.e. due to differences in uptake or clearance in other tissues), the relationship between uptake at a certain time and injected dose will also change. This cannot be accounted for in the SUV calculation and, consequently, comparison of pre- and post-therapy scans might be misleading. This can be tested by validating SUV against the metabolic rate in a subset of patients [30].

Guidelines for ^{18}F-FDG PET response assessment were first published in 1999 by the European Organization for Research and Treatment of Cancer PET Study Group and recently updated by the group of Wahl et al who proposed a new concept of PET response evaluation called "PERCIST" (PET Response Criteria in Solid Tumors) that uses a combination of absolute and relative change to detect treatment effects [31]. This concept was used as a template to validate and qualify ^{18}F-FDG PET as imaging biomarker in this thesis.

^{18}F-FLT

^{18}F-fluorothymidine (^{18}F-FLT) is a PET tracer reflecting tissue proliferation. FLT is a substrate for thymidine kinase 1 (TK1), which is a key enzyme in the salvage pathway of thymidine DNA synthesis. Several studies have shown good correlation between ^{18}F-FLT uptake and other markers of cellular proliferation, including proliferating cell nuclear antigen, flow cytometry and Ki-67 nuclear staining [32-35]. Kinetics are analyzed the same way as for ^{18}F-FDG. Guidelines on ^{18}F-FLT PET response assessment are not yet published.

$H_2{}^{15}O$

Radiolabeled water ($H_2{}^{15}O$) allows the quantification of tumor perfusion. The tracer is freely diffusible and regional uptake therefore reflects tissue perfusion. One of the hallmarks of cancer is neoangiogenesis, the sprout of new blood vessels that help to sustain expanding neoplastic growth. Tissue perfusion is thought to reflect tumor vascular status and $H_2{}^{15}O$ PET therefore offers an interesting method for the non-invasive evaluation of drugs that target tumor vasculature.

Kinetic modeling is based on the Fick principle [36] and analyzed according to the Kety-Schmidt model [37]. In a dynamic scan, $dC_{tissue}(t)/dt$ (the change in tissue concentration of $H_2{}^{15}O$ at a certain time point) is equal to K_1 (the plasma to tissue transport rate constant) times $C_{plasma}(t)$ (the tracer concentration in plasma at that point in time) minus k_2 (the tissue to plasma rate constant) times $C_{tissue}(t)$ (the tracer concentration in tissue at that point in time):

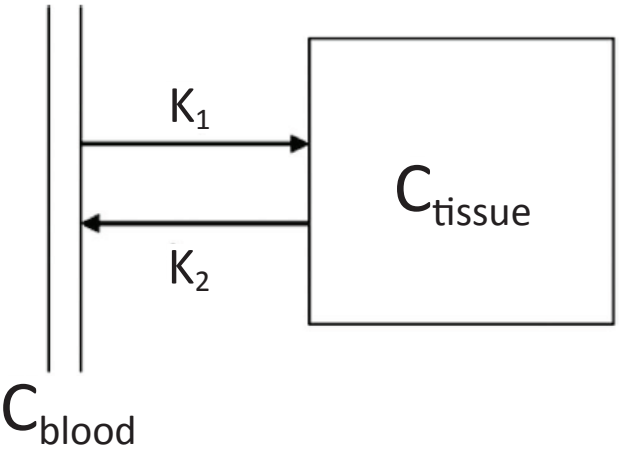
$$\frac{dC_{tissue}(t)}{dt} = K_1 C_{plasma}(t) - k_2 C_{tissue}(t)$$

The compartment model corresponding to this equation is shown in Figure 3. Dynamic scanning as well as an AIF are required. Including a fractional blood volume (V_b) and the volume of distribution (V_T), the equation corresponds to:

$$C_{tissue}(t) = (1 - v_b) \cdot F \cdot \exp(-\frac{F}{V_T}t) \otimes C_{blood}(t) + v_b \cdot C_{blood}(t)$$

V_T, or distribution volume, describes the tissue to plasma concentration ratio in equilibrium, which corresponds to the partition coefficient of water. V_T is equal to one if there is full exchange of tracer between blood and tissue. A nonlinear least-squares fit of this function to the measured time–activity curve in the tissue of interest yields perfusion F and the distribution volume of water V_T [38].

Figure 3. Single–tissue compartment model [41].



Abbreviations: C_{blood}, concentration of tracer in arterial whole blood; C_{tissue}, concentration of tracer in tissue; K_1, rate constant from plasma to tissue; k_2, rate constant for tissue clearance.

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI)

DCE-MRI allows the observation of contrast agent extravasation from the vascular compartment to the extravascular extracellular interstitial space (EES), providing parameters reflecting a combination of blood flow, vascular surface area and vascular permeability. Sequential magnetic resonance images are made before, during, and following the injection of a paramagnetic contrast agent, generally a small molecular weight gadolinium (Gd) containing compound such as gadopentetate dimeglumine.

Kinetic modelling is also based on the Fick principle [36] and analyzed according to the Kety-Schmidt

model [37]. In DCE-MRI studies, K_1 is a function of blood flow (F), permeability (P), and the vascular surface (S) and is called K^{trans}. Tofts et al [39] proposed the terms K^{trans}, k_{ep}, fPV, and v_e as outcome parameters derived from the Kety-Schmidt single–tissue compartment model. The model describes the change in contrast agent concentration in the interstitium of the tumor tissue $C_{tissue}(t)$, with $C_{plasma}(t)$ being the concentration of contrast agent in the plasma space of the tumor tissue. K^{trans} is the product of the endothelial transfer coefficient and surface area, which is the transport rate constant from plasma to tumor tissue. k_{ep} is the reflux coefficient or the transfer of contrast agent from tissue back to the blood. The equation is similar to that of given for PET:

$$C_{tissue}(t) = (1 - fPV)K^{trans} \cdot \exp(-k_{ep}t) \otimes C_{plasma}(t) + fPV \cdot C_{plasma}(t)$$

The two compartments within the tumor are the blood plasma and the EES. fPV is the fraction of plasma volume related to whole tissue volume. v_e, which equals K^{trans}/k_{ep}, is a measure for the EES fraction. Note that the definitions of F and K^{trans}, and V_T and v_e, are not similar because water is freely diffusible in tissue, including cells, whereas Gd is only transferred into the extracellular space.

SCOPE AND OUTLINE OF THE THESIS

The aim of the present thesis was to validate and qualify PET and DCE-MRI derived imaging parameters in patients with NSCLC that were treated with erlotinib and bevacizumab.

In **chapter two** we discuss the use of blood flow measurements with $H_2{}^{15}O$ PET and DCE-MRI in an oncological setting. The review was designed to discuss $H_2{}^{15}O$ PET and DCE-MRI side-by-side to address differences and synergistic value of flow measurements with these two techniques.

In **chapter three to five** the repeatability of three imaging biomarkers was assessed: $H_2{}^{15}O$, ^{18}F-FDG and ^{18}F-FLT. The aim of these studies was to provide results that can be used for binary decision making in future clinical trials.

Chapter six and seven describe the results of a multicenter efficacy study of combined BE treatment in advanced stage NSCLC. Patients were scanned with CT, ^{18}F-FDG PET, $H_2{}^{15}O$ PET and DCE-MRI at baseline and after three weeks of treatment to evaluate target modulation and to qualify the imaging biomarkers during BE treatment.

REFERENCE LIST

1. Workman P, de BJ. Targeted therapeutics for cancer treatment: major progress towards personalised molecular medicine. Curr Opin Pharmacol 2008;8(4):359-362.
2. Carden CP, Sarker D, Postel-Vinay S, Yap TA, Attard G, Banerji U, Garrett MD, Thomas GV, Workman P, Kaye SB, de Bono JS. Can molecular biomarker-based patient selection in Phase I trials accelerate anticancer drug development? Drug Discov Today 2010;15(3-4):88-97.
3. DiMasi JA, Grabowski HG. Economics of new oncology drug development. J Clin Oncol 2007;25(2):209-216.
4. de Bono JS, Ashworth A. Translating cancer research into targeted therapeutics. Nature 2010;467(7315):543-549.
5. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69(3):89-95.
6. Gutman S, Kessler LG. The US Food and Drug Administration perspective on cancer biomarker development. Nat Rev Cancer 2006;6(7):565-571.
7. Goodsaid F, Frueh F. Biomarker qualification pilot process at the US Food and Drug Administration. AAPS J 2007;9(1):E105-E108.
8. Tan DS, Thomas GV, Garrett MD, Banerji U, de Bono JS, Kaye SB, Workman P. Biomarker-driven early clinical trials in oncology: a paradigm shift in drug development. Cancer J 2009;15(5):406-420.
9. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Aerts HJ, Bendriem B, Bendtsen C, Boellaard R, Boone JM, Cole PE, Conklin JJ, Dorfman GS, Douglas PS, Eidsaunet W, Elsinger C, Frank RA, Gatsonis C, Giger ML, Gupta SN, Gustafson D, Hoekstra OS, Jackson EF, Karam L, Kelloff GJ, Kinahan PE, McLennan G, Miller CG, Mozley PD, Muller KE, Patt R, Raunig D, Rosen M, Rupani H, Schwartz LH, Siegel BA, Sorensen AG, Wahl RL, Waterton JC, Wolf W, Zahlmann G, Zimmerman B. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 2011;259(3):875-884.
10. Buyse M, Michiels S, Sargent DJ, Grothey A, Matheson A, de GA. Integrating biomarkers in clinical trials. Expert Rev Mol Diagn 2011;11(2):171-182.
11. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, Munden RF. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol 2003;21(13):2574-2582.
12. Shepherd FA, Rodrigues PJ, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R, van KM, Dediu M, Findlay B, Tu D, Johnston D, Bezjak A, Clark G, Santabarbara P, Seymour L. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005;353(2):123-132.
13. Hotta K, Matsuo K, Ueoka H, Kiura K, Tabata M, Harita S, Gemba K, Yonei T, Bessho A, Tanimoto M. Continued gefitinib treatment after disease stabilisation prolongs survival of Japanese patients with non-small-cell lung cancer: Okayama Lung Cancer Study Group experience. Ann Oncol 2005;16(11):1817-1823.
14. Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, Nishiwaki Y, Vansteenkiste J, Kudoh S, Rischin D, Eek R, Horai T, Noda K, Takata I, Smit E, Averbuch S, Macleod A, Feyereislova A, Dong RP, Baselga J. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. J Clin Oncol 2003;21(12):2237-2246.
15. Kris MG, Natale RB, Herbst RS, Lynch TJ, Jr., Prager D, Belani CP, Schiller JH, Kelly K, Spiridonidis H, Sandler A, Albain KS, Cella D, Wolf MK, Averbuch SD, Ochs JJ, Kay AC. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA 2003;290(16):2149-2158.
16. Birchard KR, Hoang JK, Herndon JE, Jr., Patz EF, Jr. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. Cancer 2009;115(3):581-586.
17. Normanno N, Bianco C, De LA, Maiello MR, Salomon DS. Target-based agents against ErbB receptors and their ligands: a novel approach to cancer treatment. Endocr Relat Cancer 2003;10(1):1-21.
18. Kim KJ, Li B, Winer J, Armanini M, Gillett N, Phillips HS, Ferrara N. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. Nature 1993;362(6423):841-844.
19. Frauwirth KA, Riley JL, Harris MH, Parry RV, Rathmell JC, Plas DR, Elstrom RL, June CH, Thompson CB. The CD28 signaling pathway regulates glucose metabolism. Immunity 2002;16(6):769-777.
20. Semenza GL, Artemov D, Bedi A, Bhujwalla Z, Chiles K, Feldser D, Laughner E, Ravi R, Simons J, Taghavi P, Zhong H. 'The metabolism of tumours': 70 years later. Novartis Found Symp 2001;240:251-260.
21. Minchenko A, Leshchinsky I, Opentanova I, Sang N, Srinivas V, Armstead V, Caro J. Hypoxia-inducible factor-1-mediated expression of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3 (PFKFB3) gene. Its possible role in the Warburg effect. J Biol Chem 2002;277(8):6183-6187.
22. De GU, Valero V, Rohren E, Dawood S, Ueno NT, Miller MC, Doyle GV, Jackson S, Andreopoulou E, Handy BC, Reuben JM, Fritsche HA, Macapinlac HA, Hortobagyi GN, Cristofanilli M. Circulating tumor cells and [18F]fluorode oxyglucose positron emission tomography/computed tomography for outcome prediction in metastatic breast cancer. J Clin Oncol 2009;27(20):3303-3311.

23. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van GM, van Oosterom AT, Christian MC, Gwyther SG. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92(3):205-216.
24. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):228-247.
25. Nolop KB, Rhodes CG, Brudin LH, Beaney RP, Krausz T, Jones T, Hughes JM. Glucose utilization in vivo by human pulmonary neoplasms. Cancer 1987;60(11):2682-2689.
26. Wahl RL, Hutchins GD, Buchsbaum DJ, Liebert M, Grossman HB, Fisher S. 18F-2-deoxy-2-fluoro-D-glucose uptake into human tumor xenografts. Feasibility studies for cancer imaging with positron-emission tomography. Cancer 1991;67(6):1544-1550.
27. Hatanaka M. Transport of sugars in tumor cell membranes. Biochim Biophys Acta 1974;355(1):77-104.
28. Langen KJ, Braun U, Rota KE, Herzog H, Kuwert T, Nebeling B, Feinendegen LE. The influence of plasma glucose levels on fluorine-18-fluorodeoxyglucose uptake in bronchial carcinomas. J Nucl Med 1993;34(3):355-359.
29. Westerterp M, Pruim J, Oyen W, Hoekstra O, Paans A, Visser E, van LJ, Sloof G, Boellaard R. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. Eur J Nucl Med Mol Imaging 2007;34(3):392-404.
30. Lammertsma AA, Hoekstra CJ, Giaccone G, Hoekstra OS. How should we analyse FDG PET studies for monitoring tumour response? Eur J Nucl Med Mol Imaging 2006;33 Suppl 1:16-21.
31. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 2009;50 Suppl 1:122S-150S.
32. Yamamoto Y, Nishiyama Y, Ishikawa S, Nakano J, Chang SS, Bandoh S, Kanaji N, Haba R, Kushida Y, Ohkawa M. Correlation of 18F-FLT and 18F-FDG uptake on PET with Ki-67 immunohistochemistry in non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2007;34(10):1610-1616.
33. Troost EG, Vogel WV, Merkx MA, Slootweg PJ, Marres HA, Peeters WJ, Bussink J, van der Kogel AJ, Oyen WJ, Kaanders JH. 18F-FLT PET does not discriminate between reactive and metastatic lymph nodes in primary head and neck cancer patients. J Nucl Med 2007;48(5):726-735.
34. Kenny LM, Vigushin DM, Al-Nahhas A, Osman S, Luthra SK, Shousha S, Coombes RC, Aboagye EO. Quantification of cellular proliferation in tumor and normal tissues of patients with breast cancer by [18F]fluorothymidine-positron emission tomography imaging: evaluation of analytical methods. Cancer Res 2005;65(21):10104-10112.
35. Barthel H, Cleij MC, Collingridge DR, Hutchinson OC, Osman S, He Q, Luthra SK, Brady F, Price PM, Aboagye EO. 3'-deoxy-3'-[18F]fluorothymidine as a new marker for monitoring tumor response to antiproliferative therapy in vivo with positron emission tomography. Cancer Res 2003;63(13):3791-3798.
36. Fick A. Uber die Messung des Blutquantums in den Herzventrinkeln. Sitzber Physik Med Ges Wurzburg 1870.
37. Kety SS. Measurement of local contribution within the brain by means of inert, diffusible. Acta Neurol Scand Suppl 1965;14:20-23.
38. Wilson CB, Lammertsma AA, McKenzie CG, Sikora K, Jones T. Measurements of blood flow and exchanging water space in breast tumors using positron emission tomography: a rapid and noninvasive dynamic method. Cancer Res 1992;52(6):1592-1597.
39. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HB, Lee TY, Mayr NA, Parker GJ, Port RE, Taylor J, Weisskoff RM. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999;10(3):223-232.
40. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, Hill SR, Ward RL. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncol 2006;7(9):741-746.
41. de Langen AJ, van den Boogaart VE, Marcus JT, Lubberink M. Use of H2(15)O-PET and DCE-MRI to measure tumor blood flow. Oncologist 2008;13(6):631-644.
42. Kelloff GJ, Hoffman JM, Johnson B, Scher HI, Siegel BA, Cheng EY, Cheson BD, O'shaughnessy J, Guyton KZ, Mankoff DA, Shankar L, Larson SM, Sigman CC, Schilsky RL, Sullivan DC. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clin Cancer Res 2005;11(8):2785-2808.

2

Use of $H_2{}^{15}O$ PET and DCE-MRI to measure tumor blood flow

Adrianus Johannes de Langen[1]
Vivian E.M. van den Boogaart[2]
J. Tim Marcus[3]
Mark Lubberink[4]

Departments of [1]Pulmonary Diseases, [3]Physics and Medical Technology, and [4]Nuclear Medicine & PET Research, VU University Medical Center, Amsterdam, The Netherlands; [2]Department of Pulmonary Diseases, Maastricht University Hospital, Maastricht, The Netherlands

Oncologist. 2008; 13(6): 631-644

ABSTRACT

Positron emission tomography (PET) with $H_2{}^{15}O$ and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) provide noninvasive measurements of tumor blood flow. Both tools offer the ability to monitor the direct target of antiangiogenic treatment, and their use is increasingly being studied in trials evaluating such drugs. Antiangiogenic therapy offers great potential and, to an increasing extent, benefit for oncological patients in a variety of palliative and curative settings. Because this type of targeted therapy frequently results in consolidation of the tumor mass instead of regression, monitoring treatment response with the standard volumetric approach (Response Evaluation Criteria in Solid Tumors) leads to underestimation of the response rate. Monitoring direct targets of anticancer therapy might be superior to indirect size changes. In addition, measures of tumor blood flow contribute to a better understanding of tumor biology.

This review shows that DCE-MRI and $H_2{}^{15}O$ PET provide reliable measures of tumor perfusion, provided that a certain level of standardization is applied. Heterogeneity in scan acquisition and data analysis complicates the interpretation of study results. Also, limitations inherent to both techniques must be considered when interpreting DCE-MRI and $H_2{}^{15}O$ PET results. This review focuses on the technical and physiological aspects of both techniques and aims to provide the essential information necessary to critically evaluate the use of DCE-MRI and $H_2{}^{15}O$ PET in an oncological setting.

INTRODUCTION

Angiogenesis, the formation of new blood vessels from the endothelium of pre-existing vasculature, plays a central role in tumor growth and metastasis [1].

Once a tumor grows beyond 1–2 mm in size, passive diffusion of nutrients and oxygen is insufficient and neovascularization becomes necessary [2]. These newly formed blood vessels are highly abnormal and heterogeneous, even in tumors of equal histology and grade [3–5]. Areas of dilated, tortuous, and leaky vessels exist together with less abnormal areas [6, 7]. Interstitial hypertension, caused by vascular hyperpermeability and mechanical stress of lymphatics by tumor cell growth, can lead to blood flow stasis or even retrograde flow in tumors [8, 9]. This results in hypoxia and acidosis, which in turn is a major cause of resistance to radiation and chemotherapy and associated with a poorer prognosis [10]. Recently, drugs have been designed to specifically inhibit angiogenesis. Their effect typically results in consolidation of the tumor mass instead of regression.

Therefore, the standard volumetric approach with computed tomography (CT) (Response Evaluation Criteria in Solid Tumors) may not be sufficient to assess treatment response. Measuring change in tumor blood flow (perfusion) during antiangiogenic therapy might allow for the discrimination of responders from nonresponders irrespective of volumetric response.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and positron emission tomography using radiolabeled water ($H_2{}^{15}O$ PET) are sensitive techniques that aim to noninvasively study and quantify the physiology of tumor microcirculation. Both techniques are based on the continuous acquisition of two-dimensional (2D) or 3D images during the uptake and clearance of an administered tracer. DCE-MRI makes use of paramagnetic tracers, mostly consisting of a low-molecular-weight gadolinium (Gd)-based agent. PET uses positron-emitting tracers, of which $H_2{}^{15}O$ can be used to study tumor blood flow. Both techniques are now being used in several phase I, II, and III clinical trials evaluating tumor vascular response to antiangiogenic drugs. However, acquisition protocols and study designs are far from uniform and may affect results. This complicates the comparison of imaging studies and possibly results in an underestimation of the abilities of DCE-MRI and $H_2{}^{15}O$ PET. Both PET and MRI have specific advantages and disadvantages. Some limitations are inherent to the techniques, but others can be overcome by study design. These issues need to be addressed when designing a study and must be kept in mind when interpreting the results of such studies.

Mutual understanding between those using $H_2{}^{15}O$ PET and those using DCE-MRI in oncology is generally very limited. Both techniques have great potential, but limitations still exist and must not be neglected. This is the first review to discuss both $H_2{}^{15}O$ PET and DCE-MRI in monitoring tumor vascular response side-by-side. We focus on the technical and physiological aspects of both techniques and aim to supply the reader with essential information necessary to critically evaluate the use of DCE-MRI and

$H_2{}^{15}O$ PET in an oncological setting. First, the technical and physiological background of PET and MRI are discussed. Then, the methodology of flow measurements with both techniques is presented, and their validation and reproducibility are addressed. Finally, the application of DCE-MRI and $H_2{}^{15}O$ PET in monitoring response to anticancer treatment and the methodological considerations influencing quantitative measurements of tumor perfusion are discussed.

BLOOD FLOW MEASUREMENTS

Background

Technical Background of PET/MRI Signal

There is a fundamental difference between imaging an MRI contrast agent and imaging a PET tracer. Paramagnetic contrast agents are not by themselves detectable with MRI, but are visible because they shorten T1 and T2 relaxation times of the nearby hydrogen nuclei. In clinical practice, Gd compounds are most commonly used. The relation between signal intensity and Gd concentration is complicated. Gd concentration cannot be measured directly, because the signal enhancement (T1 relaxation) is a characteristic of the tissue being studied. There is no linear relationship between the degree of enhancement and the Gd concentration, particularly at high concentrations. Therefore, absolute quantification with DCE-MRI is difficult, and most often the proportional change in signal enhancement between the baseline and post-treatment measurements is used.

In contrast, $H_2{}^{15}O$ concentration shows a linear relation with signal intensity as measured with PET. Therefore, this technique is able to absolutely quantify tumor perfusion.

Physiological Background of PET/MRI Signal

$H_2{}^{15}O$ is a freely diffusible positron-emitting tracer. Therefore, regional uptake directly and specifically reflects tissue perfusion. Gd-based contrast agents are never freely diffusible. Therefore, the degree of signal enhancement with DCE-MRI depends on several physiological and physical characteristics, including contrast concentration, tissue perfusion, permeability, and volume of the extracellular extravascular space (EES) [11]. Thus, $H_2{}^{15}O$ PET and DCE-MRI are both able to monitor tumor microvasculature, but the first specifically measures tissue perfusion whereas the latter measures a combination of processes, mainly perfusion and permeability.

Modeling Signal Intensity

General Model

PET and DCE-MRI perfusion measurements are both based on the Fick principle (Fick, 1870) and analyzed according to the Kety-Schmidt model [12], although the nomenclatures used in PET and DCE-MRI are slightly different. In a dynamic scan, $dC_{tissue}(t)/dt$ (the change in tissue concentration of a flow tracer or contrast agent at a certain time point) is equal to K_1 (the plasma to tissue transport rate constant) times $C_{plasma}(t)$ (the tracer concentration) minus k_2 (the tissue to plasma rate constant) times the tracer concentration in tissue at that point in time:

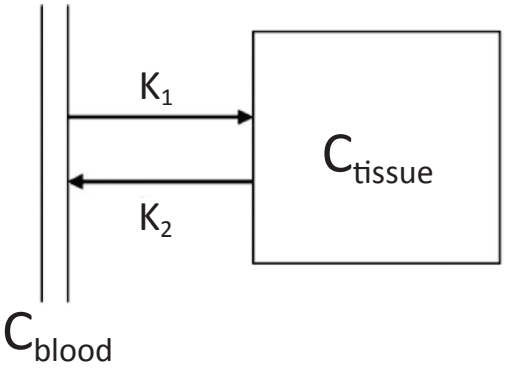
$$\frac{dC_{tissue}(t)}{dt} = K_1 C_{plasma}(t) - k_2 C_{tissue}(t) \quad (1)$$

The compartment model corresponding to this equation is shown in Figure 1. K_1 equals flow F multiplied by extraction fraction E, which in its turn is a function of F and the permeability surface product PS (the latter only being relevant in the case of nonfreely diffusible tracers):

$$K_1 = E \cdot F = \left(1 - e^{-PS/F}\right) \cdot F \quad (2)$$

In $H_2{}^{15}O$ PET studies, K_1 is equal to F and thus represents true blood flow, whereas in DCE-MRI studies, K_1 is a function of blood flow (F), permeability (P), and the vascular surface (S) and is also called K^{trans} or K^{ps}. The term K^{trans} is generally used to describe the kinetics of contrast agents with a high PS product, thus mainly representing perfusion, whereas K^{ps} is used for contrast agents with a lower PS product and thus represents a mixed effect of perfusion and permeability as shown in equation 2 [13, 14].
With $H_2{}^{15}O$ PET, C_{plasma} is effectively equal to C_{blood} because of the rapid exchange between blood cells and plasma. With DCE-MRI, C_{plasma} is calculated as C_{blood} divided by one minus the hematocrit, because Gd contrast agents are only present in plasma. For both techniques, a large vessel within the imaged volume can be used for definition of an image-derived arterial input function (IDIF). In order to use an IDIF, a feeding input vessel of substantial size (such as the aorta or left ventricle) must be included in the imaged volume, together with the target lesion. With PET, C_{blood} or the arterial input function (AIF) can be measured using (continuous) arterial blood sampling during the scan.

Figure 1. Single-tissue compartment model.



For a freely diffusible tracer with 100% extraction, such as water in the range of flow values relevant for tumor imaging, equation 1 corresponds to:

$$\frac{dC_{tissue}(t)}{dt} = FC_{blood}(t) - \frac{F}{V_T}C_{tissue}(t) \quad (3)$$

Here, V_T is the tissue to plasma concentration ratio in equilibrium, which corresponds to the partition coefficient of water. V_T is equal to one if there is full exchange of tracer between blood and tissue. The solution of this equation, including a fractional blood volume v_b, is:

$$C_{tissue}(t) = (1 - v_b)\cdot F \cdot \exp(-\frac{F}{V_T}t) \otimes C_{blood}(t) + v_b \cdot C_{blood}(t) \quad (4)$$

A nonlinear least-squares fit of this function to the measured time–activity curve in the tissue of interest yields the perfusion F and distribution volume of water V_T [15]. Therefore, this method also allows for the quantification of tissue perfusion when V_T does not equal one.

In the case of MRI, a different nomenclature is used. Low-molecular-weight Gd is not freely diffusible, but partially leaks through the vessel wall into the interstitial space and leaks more through pathological tumor vessels as a result of greater permeability. DCE-MRI applying Gd contrast agents exploits this hyperpermeable nature of tumor vessels. The same kinetic model as above can be applied to fit DCE-MRI data [16]. Tofts et al [11] proposed the terms K^{trans}, k_{ep}, fPV, and v_e as outcome parameters derived from this single–tissue compartment model. The equation is similar to equation 3:

$$\frac{dC_{tissue}(t)}{dt} = K^{trans}C_{plasma}(t) - k_{ep}C_{tissue}(t) \quad (5)$$

This equation describes the change in contrast agent concentration in the interstitium of the tumor tissue $C_{tissue}(t)$, where $C_{plasma}(t)$ is the concentration of contrast agent in the plasma space of the tumor tissue. K^{trans} is the endothelial transfer coefficient surface area product, which is the transport rate constant from plasma to tumor tissue. k_{ep} is the reflux coefficient or the transfer of contrast agent from tissue back to the blood. The solution of this equation is similar to the solution given for PET:

$$C_{tissue}(t) = (1 - fPV) \cdot K^{trans} \cdot \exp(-k_{ep} t) \otimes C_{plasma}(t) + fPV \cdot C_{plasma}(t) \quad (6)$$

The two compartments within the tumor are the blood plasma and the EES. fPV is the fraction of plasma volume related to whole tissue volume. v_e, which equals K^{trans}/k_{ep}, is a measure for the EES fraction. Note that the definitions of F and K^{trans}, and V_T and v_e, are not similar because water is freely diffused in tissue, including cells, whereas Gd is only transferred into the extracellular space. Table 1 lists all the parameters involved in modeling PET and MRI signaling.

Specific Aspects of $H_2{}^{15}O$ PET

A dynamic PET scan following a bolus injection of $H_2{}^{15}O$ consists of a series of consecutive short scans (frames), with typical durations starting with 5 seconds per frame and increasing to 30 seconds per frame (totaling 10 minutes), and starts simultaneously with the bolus injection. Although frame dura-

Table 1. PET and MRI nomenclature.

Symbol	Definition	Unit
F	Perfusion	$ml \cdot cm^{-3}\ min^{-1}$
E	Extraction	unitless
K_1	$F \cdot E$, rate constant from plasma to tissue	$ml \cdot cm^{-3}\ min^{-1}$
k_2	Rate constant for tissue clearance	min^{-1}
V_T	Distribution volume	$ml\ cm^{-3}$
V_b, fPV	Fractional blood or plasma volume	$ml\ cm^{-3}$
K^{trans}, K^{ps}	Volume transfer constant between plasma and EES	min^{-1}
k^{ep}	Rate constant between EES and plasma	min^{-1}
V_e	Fractional extravascular, extracellular leakage volume	unitless
AUGC	Initial area under the Gd concentration curve	AU
C_{plasma}, AIF	Concentration of tracer or Gd contrast agent in arterial plasma	$AU \cdot ml^{-1}$
C_{tissue}	Concentration of tracer in tissue	$AU \cdot ml^{-1}$
C_{blood}	Concentration of tracer in arterial whole blood	$AU \cdot ml^{-1}$

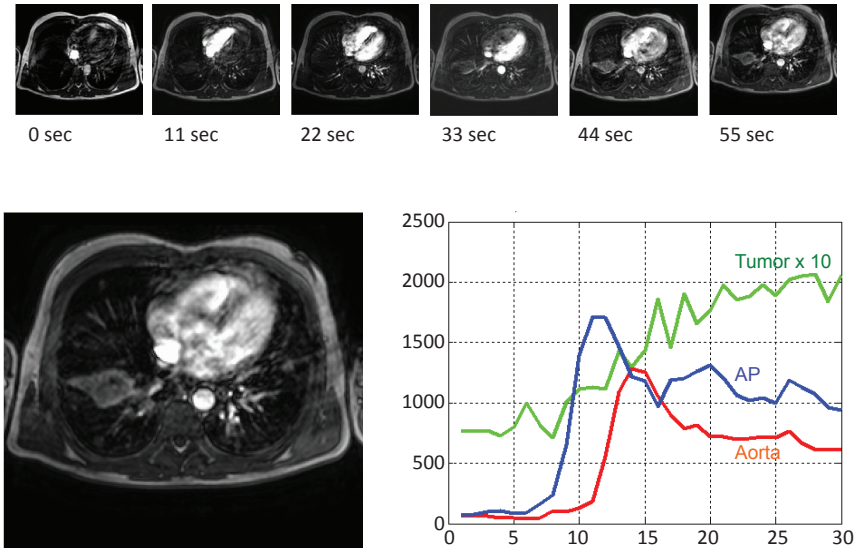
Abbreviations: AU, arbitrary units; EES, extracellular extravascular space; Gd, gadolinium; MRI, magnetic resonance imaging; PET, positron emission tomography.

tions shorter than 5 seconds are possible, they would result in very noisy images because of the low number of counts per frame and this would not contribute to more reliable flow measurements. The amount of injected $H_2{}^{15}O$ is about 1 GBq and depends on the high count rate capabilities of the scanner. Prior to the emission scan, a short (several minutes) transmission scan using rotating ^{68}Ge rod sources, analogous to a CT scan but acquired with 511 keV gamma radiation instead of x-rays, has to be performed to correct for photon attenuation in tissue. In the case of combined PET-CT scanners, a low-dose (20 –50 mAs) CT scan can be used for this purpose. A conventional PET scan after administration of 1 GBq $H_2{}^{15}O$ results in an effective radiation dose of approximately 1 mSv, or about half the average annual radiation load per person from the environment [17]. A combined PET-CT scan with 500 MBq $H_2{}^{15}O$ and a 30-mAs low-dose CT of, for example, the thorax results in a higher effective dose of 1.4 mSv. In the latter case, the lower amount of administered $H_2{}^{15}O$ is feasible because of the improved signal-to-noise ratio of state-of-the-art scanners. The spatial resolution of PET images is on the order of 6 mm in all three directions in clinical PET scanners. Because of the short radioactive half-life of ^{15}O (2 minutes), a scan can be repeated within 10 minutes of the previous scan, which allows for the measurement of direct short-term therapy effects within one imaging session. Also, a scan can be followed by a second scan with a different tracer to study other aspects of tumor biology. Because of the short half-life, $H_2{}^{15}O$ PET studies require the availability of an on-site cyclotron for the production of ^{15}O.

Specific Aspects of DCE-MRI

For DCE-MRI, a clinical MRI scanner can be used with field strength usually in the range of 1–3 Tesla. The scanning protocol starts with high-resolution 3D imaging of the tumor and its environment, with both T1- and T2-weighted sequences. These images are used to delineate the tumor volume. The protocol continues with five precontrast T1-weighted measurements with different flip angles to determine the T1 relaxation time in the blood and tissue before contrast arrival. This T1 value is required for the model-based quantification [16]. Then, the contrast is given by i.v. injection. Upon injection, the dynamic acquisition starts using the same 3D T1-weighted pulse sequence and the same slice positions, but with a fixed flip angle (e.g. 35 degrees), containing 30–35 scans of about 1 second each. Thus, the temporal resolution is 1 second. The spatial resolution inplane is about 1.5 x 1.5 mm, and through-plane is about 10 mm. The use of thick image planes, however, leads to a degradation of in-plane resolution, depending on heterogeneity in the axial direction, because each plane represents an average image of a 10-mm thick volume. For all tumors that are subject to respiratory motion, the patient is asked to hold his breath as long as possible. Renal clearance of contrast occurs with an elimination phase half-life of 100 minutes [18]. Because of the relatively slow clearance of contrast agent, DCE-MRI scans cannot be repeated within the same imaging session. Figure 2 shows images of a dynamic MR sequence following a contrast bolus injection.

Figure 2. Representative DCE-MR images from a patient with NSCLC. The sequential images show the tumor at different time points after contrast bolus. The signal intensity curve over time was measured in the tumor, the aorta, and the pulmonary artery (AP).



Despite the many research papers in this field, consensus is lacking on the exact kinetic model to be used for DCE-MRI. Therefore, model-free quantification of DCE-MRI data is still widely applied. This approach calculates basic properties of the tissue enhancement curve, such as time to onset and peak and end, time intervals of rising and transit, amplitude, and the mean and maximum upslope of the DCE-MRI tissue enhancement curve. The upslope values can be normalized for the area-under-the-curve and upslope of the AIF. Another frequently used semiquantitative measurement is the initial area under the Gd concentration curve (AUGC).

Parametric Images

In principle, it is possible to apply the above equations on a pixel-by-pixel basis to obtain parametric images, a graphical representation showing F, V_T, K^{trans}, or v_e, for each pixel [19]. Parametric images retain the original image resolution, allowing for better assessment of heterogeneity than region of interest (ROI)-based methods. Heterogeneity analyses can add to ROI analyses, and DCE-MRI seems especially promising because of its high spatial and temporal resolution. Nonlinear least-squares fitting of equations 4 or 6 on a pixel-by-pixel basis is very time-consuming, which is not a limiting factor for MRI because of the small number of image planes, but is impractical for the number of voxels of about one million in a typical clinical PET image volume. In addition, it yields noisy parametric images.

Although a subregion of the total imaged volume can be selected in order to limit the computation time, faster methods are available and preferable. Basically, there are two fast options for creating parametric images. First, the differential equation 1 can be integrated on both sides, resulting in the following equation:

$$\frac{C_{tissue}(t)}{\int_0^t C_{plasma}(t)d\tau} = K_1 - k_2 \frac{\int_0^t C_{tissue}(t)d\tau}{\int_0^t C_{plasma}(t)d\tau} \quad (7)$$

When evaluating this equation and plotting, for each frame, the left-hand side of the equation versus the integral ratio on the right-hand part, this equation describes a straight line, intercepting the y-axis at K_1 (F or K^{trans}) and the x-axis at K_1/k_2 (V_T or v_e). Using this linearization, F and V_T (or K^{trans} and v_e) can be obtained for each pixel using a simple and fast linear least-squares fit [20]. The second option to create parametric images is by the use of basis functions:

$$C_{tissue}(t) = K_1 \cdot \exp(-\beta_i t) \otimes C_{plasma}(t) + v_b C_{blood}(t) \quad (8)$$

A set of basis functions $\exp(-\beta_i t) \otimes C_{plasma}(t)$ is created by convolution of the plasma input function with a set of single exponential functions, with exponential clearance rate constants β_i, for example, between 0.01 and 1 min^{-1}. For each voxel, the basis function that, multiplied by K_1 (F or K^{trans}), best fits the measured data is selected. Each of these iterations involves a simple linear fit procedure, which is much faster than a nonlinear fit and limits the time necessary to construct parametric images from several hours to several minutes. In addition, this method reduces image noise while allowing for inclusion of a blood volume compartment and retaining the quantitative accuracy of the resulting parametric images. The basis function method also allows for fixation of V_T or v_e to further reduce noise [19, 21].

Figure 3. Coronal images of $H_2{}^{15}O$ uptake (A), perfusion (B), and, for reference, ^{18}F-3'-deoxy-3'-fluorothymidine (^{18}F-FLT) uptake (C) of a patient with a mediastinal relapse of non-small-cell lung cancer. The perfusion image was calculated using the basis function method, with the distribution volume V_T fixed to 1, and corrected for both arterial and venous blood volume.

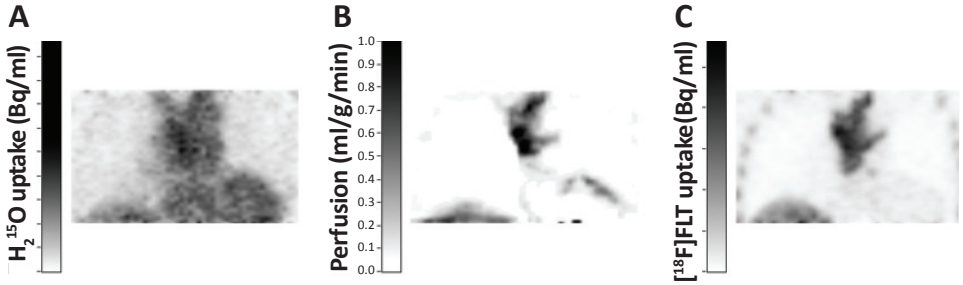


Figure 3 shows an $H_2{}^{15}O$ perfusion image calculated using the basis function method, compared with typical images of ^{18}F-3'-deoxy-3'-fluorothymidine (^{18}FLT), a proliferation tracer, and images of $H_2{}^{15}O$ radioactivity concentration. Figure 4 shows a DCE-MRI perfusion image calculated by nonlinear least-squares fits of equation 6 to each pixel's tissue enhancement curve.

VALIDATION AND REPRODUCIBILITY

$H_2{}^{15}O$-PET

Validation

Dynamic $H_2{}^{15}O$ PET has been well validated in the brain and myocardium [22–27]. Some validation studies have been performed in other tissues, including the lung [28, 29], kidney [30], and skeletal muscle [31, 32]. To date, no validation studies have been performed in tumors. In theory, the model should also fit tumor perfusion measurements because $H_2{}^{15}O$ is freely diffusible throughout the body and the range of reported values (0.15–1.29 mL blood/mL tissue per minute) lies well within the range of validated perfusion values [33, 34]. On the other hand, typical vascular abnormalities in tumor tissue (vascular shunts, large vessels situated within a lesion) can result in overestimation of perfusion if models are used without correction for these effects [35, 36]. Therefore, validation studies in tumors are still needed and awaited.

Reproducibility

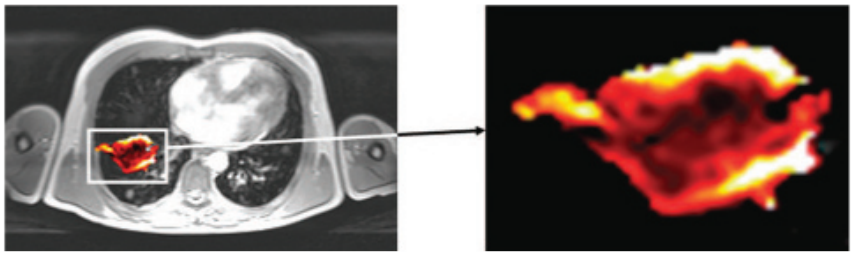
In brain and myocardium [37–40], as well as in some other tissues [34, 41, 42], $H_2{}^{15}O$ PET has been shown to be reproducible with within-subject coefficients of variation (wCV) in the range of 9%–14%. The reproducibility value (defined as the range within which 95% of measurements fall) was in the range of 22%–34%. Two oncological studies have been performed [15, 34]. Although the study groups were small, reproducibility results were in the same range as those observed in the brain and myocardium, with wCV of 11% for abdominal tumors [34] and errors up to 10% reported for breast tumors [15].

DCE-MRI

Validation

The first approach for validating DCE-MRI perfusion is to verify whether an MRI-derived parameter does indeed deliver absolute perfusion measurements. No such validation study has been performed in tumors. In the human myocardium, K^{trans} was compared with myocardial perfusion as measured us-

Figure 4. T1-weighted magnetic resonance image (left) and dynamic contrast-enhanced magnetic resonance imaging K^{trans} perfusion image (right), also shown as overlay in the T1 image, of a patient with non-small-cell lung cancer, calculated by least-squares fits to equation 6. K^{trans} is the volume transfer constant between plasma and the extravascular extracellular space.



ing $H_2{}^{15}O$ PET [43]. There was a strong correlation for dipyridamole-induced flow (r = 0.70; p = .001) and a moderate correlation for myocardial perfusion reserve (r = 0.48; p = .04) between MRI and PET. In a second approach, it was validated against immunohistochemistry as the gold standard. Correlation of DCE-MRI signal enhancement with microvascular density (MVD) was found to be inconsistent [44]. One reason for this may be that tissue enhancement is not only determined by MVD but also by the status of the capillary–interstitial space barrier (vessel wall permeability), which is not accounted for by immunohistochemistry. In breast cancer lesions, tissue expression of vascular endothelial growth factor (VEGF) was closely correlated to k_{ep} [45]. Because VEGF expression shows a spatial association with vessel permeability [46, 47], this finding further supports the theory that enhancement of the DCE-MRI signal partially reflects vessel permeability.

Because DCE-MRI measures a combination of flow and permeability, it is difficult to validate. Attempts have shown that signal enhancement reflects vascular status, but validation studies in tumors, as have been done for the myocardium, are still awaited.

Reproducibility

Two studies have evaluated reproducibility in an oncological setting. The first evaluated the reproducibility of quantitative and semiquantitative kinetic parameters in 21 patients with solid tumors [48]. AUGC and v_e were found to be reproducible with wCVs of 12% and 9%, respectively.

K^{trans} and k_{ep} showed greater variability (with wCVs of 29% and 24%, respectively). The latter are more sensitive to changes in the AIF, which could explain the lower reproducibility. Comparable results were found in a study of 11 patients with advanced cancers [49]. AUGC and K^{trans} showed wCVs of 16% and 19%, respectively. The reproducibility value was 30% for AUGC and 36% for K^{trans}. With serial imaging of patients, a decrease in K^{trans} of > 40%–45% can be addressed as a treatment effect [48–51].

MONITORING RESPONSE TO ANTICANCER TREATMENT

Predicting response to therapy as early as possible creates the opportunity to customize patient treatment. There is a need for this "customization" because response rates to intensive treatment with serious, and sometimes lethal, side effects are often low. To continue treatment with limited or no effect only provides the patient with side effects, while depriving the patient of potential benefit from other treatment. The ultimate goal is to customize treatment prior to initiation. Encouraging results have been published on gene-expression profiling, where correlations between gene expression and drug sensitivity were found [52–54]. Some success has been achieved in measuring the expression level of tumor markers to which the drug is targeted [55]. However, before these tools can be used in a clinical setting, further research is needed. Until then, predicting response early in treatment is still a major step forward when compared with size-based response evaluation after several cycles of systemic therapy.

MONITORING RESPONSE TO ANTIANGIOGENIC THERAPY

Several components of tumor vasculature have been targeted by systemic drugs. The main differentiation can be made between antiangiogenic and antivascular drugs. The first inhibit neovascularization. The most striking inhibition has been accomplished by targeting VEGF, via either monoclonal antibodies (MAbs) directed against circulating VEGF or by targeting the VEGF receptor (VEGFR) with tyrosine kinase inhibitors (TKIs). Bevacizumab (a VEGF MAb) has been shown to decrease tumor perfusion, vascular permeability, MVD, interstitial fluid pressure (IFP), and circulating endothelial and progenitor cells [56]. This suggests a direct antivascular effect. Receptor TKIs generally target not only the tyrosine kinase of the VEGFR but also the tyrosine kinases of other receptors, such as the platelet-derived growth factor receptor, fibroblast growth factor receptor, and epidermal growth factor receptor (EGFR), which suggests additional effects when compared with MAbs. Examples are imatinib, sorafenib, and sunitinib. Selective blocking of EGFR is another inhibitory pathway of angiogenesis [57]. Examples of EGFR TKIs are erlotinib and gefitinib. Antivascular or vascular disruptive agents (VDAs) damage existing blood vessels, aiming at depriving the tumor from oxygen and nutrients. Combretastatin, the major VDA under investigation, is able to rapidly shut down blood vessels and can result in pronounced tumor necrosis in patients with solid tumors [58, 59].

Knowledge of the nature of a drug action, and its influence on PET and MRI derived parameters, can aid in choosing the best imaging tool and time interval to monitor response. $H_2{}^{15}O$ PET measures perfusion and V_T. Because water is freely diffusible, a change in vascular permeability cannot be monitored accurately, although this could be measured with PET using other tracers such as [^{68}Ga]transferrin or [^{11}C]methylalbumin [23, 60]. A decrease in IFP (resulting from a decrease in vascular permeability or

lowering of mechanical stress by reducing tumor cell density) results in an increase in perfusion. A decrease in MVD results in a decrease in perfusion. V_T (the ratio of $H_2{}^{15}O$ in tissue to that in plasma at equilibrium) is expected to be stable with antiangiogenic therapy and to decrease with antivascular therapy because of induced necrosis or increased intercapillary distance.

DCE-MRI–derived K^{trans} has several physiological interpretations, depending on the balance between capillary permeability and blood flow. In situations in which capillary permeability is very high, the flux of contrast agent into the EES is flow limited and K^{trans} will approximate tissue perfusion. On the other hand, when permeability is low and blood flow is high, K^{trans} gives an indication of capillary permeability. However K^{trans} frequently indicates a combination of flow and capillary permeability [11]. A decrease in vascular permeability and/or MVD results in a decrease in K^{trans} because of a lower extraction rate and lower blood vessel surface area, respectively. A decrease in IFP, however,
results in an increase in K^{trans} because of lower interstitial resistance.

As indicated, antiangiogenic therapy promotes several morphological and functional changes in a tumor with sometimes opposite effects on MRI- and PET-derived parameters. The net effect is measured, and therefore it is of no surprise that the relation between treatment effects and imaging parameters is not straightforward. In spite of the above limitations, $H_2{}^{15}O$ PET and DCE-MRI have shown their potential in response assessment early in anticancer treatment. In renal cell cancer, paired ^{18}F-FDG and $H_2{}^{15}O$ PET scans were performed in five patients before and after SU5416 (a VEGFR TKI) and interferon-α. ^{18}FDG and $H_2{}^{15}O$ PET both showed a decrease in one patient with stable disease, an increase in one with progressive disease, and no change in three patients [61]. Razoxane (a cytostatic and antiangiogenic agent) was shown to reduce tumor perfusion in patients with advanced renal cell cancer [62]. Statistical significance, however, was not reached, and perfusion did not correlate with tumor progression, possibly because of the small sample size. Response monitoring to endostatin and combretastatin therapy revealed a reduction in tumor perfusion, as measured with $H_2{}^{15}O$ PET [59, 63]. Both studies showed a dose-dependent reduction in perfusion, with a nonlinear relation in the endostatin study. Combretastatin caused a rapid perfusion reduction in solid tumors after 30 minutes, which remained significant after 24 hours. Interestingly V_T also was determined, which showed a reduction 30 minutes after injection with recovery after 24 hours. This trend might be a result of partial reversal of the vascular response, indicating that part of the tissue was targeted but still viable.

DCE-MRI has been incorporated into several phase I and II trials. PTK/ZK (a VEGFR TKI) has been tested in a phase I trial in patients with advanced cancers [50, 51, 64]. DCE-MRI was performed at baseline, day 2, and the end of each 28-day cycle. K^{trans} decreased in a dose-dependent way. A reduction of > 40% in the baseline value at day 2 was predictive of stable disease. In a recent phase II study, patients with advanced breast cancer were treated with bevacizumab monotherapy for one cycle, followed by six

cycles of bevacizumab, doxorubicin, and docetaxel. Significant reductions in K^{trans}, k_{ep}, and v_e were seen after one cycle, which persisted when cytotoxic chemotherapy was added. The decreases in K^{trans}, k_{ep}, and v_e were correlated with increased apoptosis and a reduction in VEGFR expression [65]. Several targeted agents have been shown to decrease DCE-MRI–derived parameters in a dose-dependent manner, making DCE-MRI a useful indicator in drug pharmacology [66, 67]. A recent study of DCE-MRI in patients with progressive multiple myeloma treated with thalidomide found amplitude A, which is proportional to the relative signal enhancement, to be a good prognostic indicator of progression-free survival [68].

Most PET and MRI results are derived from small studies. Together with heterogeneity in acquisition protocols and study design, these factors might limit their value. Therefore, prospective phase III trials with predefined cutoff values for PET and MRI results are awaited.

Potential Synergistic Value of DCE-MRI and $H_2{}^{15}O$ PET

Because PET measures true perfusion and MRI measures a combination of perfusion and vascular permeability, the two modalities can complement each other. In theory, both perfusion and permeability can be isolated when F and K^{trans} are known. An alternative is the use of a PET tracer such as [^{68}Ga]transferrin or [^{11}C]methylalbumin to measure permeability directly. So far, no oncological study has been performed that incorporated both DCE-MRI and $H_2{}^{15}O$ PET.

Combining Information on Tumor Vasculature and Metabolism

Tumor metabolism can be measured with ^{18}F-FDG PET. ^{18}F-FDG uptake is the product of perfusion and extraction. It can be hypothesized that poorly perfused and hypometabolic areas reflect necrosis, while poorly perfused and hypermetabolic areas reflect hypoxia (because of anaerobic glucose hypermetabolism under hypoxic circumstances). It is possible to measure oxygen consumption directly using $^{15}O_2$ and PET. Knowledge of the regional tumor microenvironment can supply relevant information on mechanisms of treatment failure, but can also aid in treatment planning (e.g. radiotherapy). Recently, the synergistic value of $H_2{}^{15}O$ PET and ^{18}F-FDG PET was shown in breast cancer, where a pretherapeutic low ratio of ^{18}F-FDG uptake to perfusion was the best predictor of complete response to neoadjuvant chemotherapy and also predictive of disease-free survival [69]. ^{18}F-FDG metabolism and perfusion were positively correlated, although highly variable. Zasadny et al [70] also found a positive correlation between ^{18}F-FDG metabolism and perfusion, with the additional finding of a higher slope of the curve at lower flow rates, possibly indicating hypoxia in areas with low perfusion and high metabolism. In contrast, ^{18}F-FDG metabolism was found to be negatively correlated with perfusion in non-small-cell lung cancer (NSCLC), liver tumors, and head-and-neck cancer. Perfusion in the latter

two was assessed with contrast-enhanced CT [71–73]. In NSCLC and head-and-neck cancer, the ratio was found to positively correlate with tumor size, suggesting the presence of hypoxia in larger tumors where tumor growth exceeds angiogenesis [71, 73]. In contrast, a high tumor perfusion rate predicted poor response to radiotherapy in head-and-neck cancer [74]. This reflects the complexity of tumor biology. High tumor blood flow does not immediately indicate sufficient oxygenation and thus favorable radiotherapy outcome. Hypoxia stimulates angiogenesis via upregulation of specific transcription and growth factors, like VEGF [75]. This results in a more complex, but abnormal, chaotic vascular network with increased blood flow, but inadequate oxygen and nutrient supply. Thus, hypoxic areas can exist irrespective of high blood flow, a finding confirmed in a PET study in brain tumors using fluorine-18 fluoromisonidazole (a hypoxia tracer) and $H_2{}^{15}O$ [76].

Timing of Response Evaluation

Jain [77, 78] introduced the term "normalization" to describe the (transient) reversal of vessel abnormalities his group observed after antiangiogenic treatment. The time span of normalization, the "normalization window," describes the period in which abnormal vessels either regress or normalize. In this period, the nutrient, drug, and oxygen supply is thought to be improved [77, 78]. Eventually, all tumors show resistance, after which vessels become abnormal again. The time span of this window is unknown for most drugs and, in addition to drug characteristics, may depend on the tumor type and possibly on individual tumor and patient characteristics. Because of these time-dependent tumor changes, it seems extremely important to know when to scan in order to adequately measure treatment effect.

In most protocols, time to response evaluation is based on CT response protocols instead of considerations of tumor biology. This might be one of the reasons why $H_2{}^{15}O$ PET and DCE-MRI failed to demonstrate a correlation with outcome in some trials.

The reported normalization windows after treatment with VEGF TKIs or MAbs have been heterogeneous [79, 80], which makes designing future trials difficult but challenging. Serial measurements can aid in the definition of this time window for new drugs [79]. Therefore, if financially and logistically possible, serial PET and MRI measurements should be planned for new drug trials in order to assess the best time point for response monitoring after antiangiogenic treatment. Although most patients in such trials have a limited life expectancy, this strategy can result in a high cumulative radiation dose, especially with combined PET-CT scanners.

METHODOLOGICAL CONSIDERATIONS — FACTORS INFLUENCING QUANTITATIVE MEASUREMENTS OF TUMOR PERFUSION AND PERMEABILITY

AIF Definition

Independently of the parameter used, several considerations have to be made about the ROI and AIF. The latter describes the concentration of MRI contrast or PET tracer in plasma over time, which has a great influence on measurements of (semi)quantitative parameters. It is generally accepted that the AIF varies between patients and between visits within each patient because of variations in cardiac output, renal function, and vascular tone. In principle, the AIF should be accurately measured for each scan. With PET, accurate measurement of the AIF can be obtained by arterial blood sampling with an online system, measuring arterial radioactivity concentration independently of tumor site. Because arterial sampling is invasive, an image-derived AIF may be more convenient. However, in many cases it is not possible to measure an image-derived AIF because of the absence of a suitable artery within the imaging field of view.

With MRI, image-derived measurement of the AIF is the only option, because measurement of the contrast agent concentration by arterial sampling is not possible. AIF measurements have shown that DCE-MRI reproducibility is worse with suboptimal AIF definition than with using a standardized AIF [49, 81]. In addition, the relation between tracer concentration and signal intensity becomes progressively nonlinear with increasing tracer concentration. This is especially the case in the feeding artery during the first passage of the administered bolus. Therefore, it has been proposed that the AIF should be measured with a small prebolus, and then the true AIF for the main bolus can be reconstructed from this [82]. Several other methods have been developed to meet this problem. Most of them are based on reference region models or a functional form of a population-derived AIF [81, 83]. Whichever technique is being used, it has a great impact on the results.

Tissue ROI Definition

The tissue ROI should be a well-defined part of the tumor, consistent throughout follow-up and preferably throughout the whole study to avoid sampling errors. This could be the entire tumor or a subregion. ROI can be defined on CT, MRI, PET emission, or PET transmission images. DCE-MRI obviously uses the high-resolution MR images, but PET has no well-defined anatomic landmarks available. CT can be used, but coregistration incorporates spatial error. Integrated PET-CT overcomes much of this problem, although respiratory motion still challenges accurate coregistration in the thorax and upper

abdomen. Definition of anatomic ROIs using CT and MRI has its own disadvantages. Metabolically inactive areas (reflecting necrosis) are difficult to identify and might be incorporated, causing underestimation of the perfusion in viable tumor tissue. In addition, atelectasis and adjacent infectious consolidation complicate ROI definition.

With PET, ROI can be defined either on the $H_2{}^{15}O$ image or on a subsequently performed PET scan with a tissue-accumulating tracer. Because of the limited spatial resolution of PET and the "noisy" data produced by a diffusible tracer like $H_2{}^{15}O$, ROI definition on an $H_2{}^{15}O$ PET image is difficult (Figure 3A). Peritumoral areas of high flow (resulting from infection or adjacent large vessels) and small or highly irregular tumors complicate delineation. Parametric images, depicting flow instead of radioactivity concentration, show better tumor-to-background contrast [19], but are still in need of improvement (Figures 3B and 4). In the case of heterogeneity, areas of low or absent flow are not accounted for (not visible) and only perfused areas are selected. Therefore, the ROIs defined on parametric images probably do not reflect the true tumor mass. Adding a tissue-accumulating tracer (allowing for better tumor-to-background contrast, e.g. ^{18}F-FDG or ^{18}F-FLT) to the scan protocol creates the opportunity to define an ROI reflecting metabolically active tissue, which can then be copied to the perfusion image (Figure 3). Preferably, a 3D ROI is constructed to ensure all tumor tissue is incorporated. When 2D regions are used, it is recommended to have a minimum of three sections through the tumor in order to account for spatial heterogeneity within a lesion [84]. ROI definition is not easy and straightforward, but should be based on the desired information and knowledge of treatment effects, and, most importantly, it should be performed in a standardized, reproducible manner [85].

Organ and Tumor Position

Quantification of tumor perfusion is more complicated in the lung and upper abdominal lesions. In the liver, AIF definition is difficult because of the dual hepatic blood supply (portal vein and hepatic artery). The hepatic artery accounts for the greatest part of the blood supply to liver tumors, but the portal vein contributes to a small amount in small metastases and at the surface of large metastases [86–89]. A model including the dual hepatic blood supply, addressing these issues, has been suggested [90, 91]. Tumor motion influences delineation and quantification. Thoracic lesions at the base of the lung and close to the thoracic wall, as well as upper abdominal lesions, are most vulnerable to respiratory motion [92]. A dynamic DCE-MRI sequence can be performed within a single breathhold, thereby avoiding substantial motion error. PET, with its lower temporal resolution, is especially affected. Pulmonary gated PET, using only the data required during a certain phase of the respiratory motion cycle, is a promising tool to compensate for this type of error [93, 94], although difficult to implement for dynamic $H_2{}^{15}O$ scans because of limited count rates.

Spatial and Temporal Resolution

Heterogeneity occurs at the microscopic level. DCE-MRI with its high spatial resolution, typically 1.5 x 1.5 mm in-plane and 10 mm through-plane, offers detailed pictures, whereas PET has a spatial resolution of approximately 6 mm, resulting in an averaged perfusion picture [95]. MRI also has better temporal resolution, which is on the order of 0.1 seconds for a 2D acquisition and 1 second for a 3D acquisition. This ensures that the upslope of the contrast enhancement can be measured with adequate temporal resolution, mostly within one breathhold. This allows for detailed heterogeneity analyses that might add to perfusion measures. Tumor vasculature is highly heterogeneous throughout lesions, both in place and time. Microvascular blood flow fluctuates markedly within seconds, even without treatment [96, 97]. Excellent spatial and temporal resolution only give a quick snapshot and might not be representative of the overall vascular status within a given tumor region. Therefore, it can be hypothesized that an average picture (time and place) might provide more practical information than a detailed one.

Partial Volume Effects

Partial volume effects typically occur in small lesions. This effect refers to two distinct phenomena that negatively influence quantitative measurements, image blurring and averaging signal at the voxel level [98]. Because of the finite resolution, the activity of a small source is projected as a larger but less active source (part of the signal spills out). Averaging at the voxel level occurs with all image modalities and is explained by the fact that every image is built up by pixels (2D) or voxels (volumetric pixels, 3D). Lesion boundaries do not conform themselves to this tight framework and thus voxels can exhibit both tumor and adjacent tissue, resulting in an over- or underestimation of the true pixel/voxel activity.
In this way, spill in can also be explained. Nearby large vessels, obviously with high blood flow, can cause spill in of activity that can be accounted for in compartment models (equation 4). In a similar way, hypoperfused lesions surrounded by physiologically highly perfused tissue (e.g. liver or spleen) are also affected.

Response Definition in Clinical Trials

Response evaluation is usually done by calculating the percentage change between baseline and post-treatment values. This delta function substantially depends on the value of the divisor. With a small divisor, a small difference results in an unrealistic response rate. At the other end of the spectrum, a response could be missed because of a relatively large divisor. Thus, simply calculating the percentage change might not be useful for very small or large tumors. Reproducibility studies can shed light on this subject [99]. Quantification is highly dependent on both hardware and software. Image resolution, re-

construction methods, ROI strategy, and quality control (i.e. calibration) all account for fluctuations in results. Differences of up to 40% are described in the literature for standardized uptake values (SUVs) with ^{18}F-FDG PET [100]. SUV measurements, however, require crosscalibration of a dose calibrator and the PET scanner, whereas flow measurements using $H_2{}^{15}O$ do not, removing uncertainties resulting from calibration errors in the case of flow studies. In general, most errors can be overcome when only one scanner is used, together with uniform reconstruction methods and data analysis. However, today most clinical studies are multicenter designed. If absolute values are necessary for within-study evaluation and when general recommendations for cutoff values are desired, standardization is required. Until that time, it is recommended that at least each patient be scanned on the same scanner pre- and post-treatment with equal methods of scan acquisition, data reconstruction, and analysis. In this way, the proportional change in flow values can be used safely.

CONCLUSION

DCE-MRI and $H_2{}^{15}O$ PET are able to reliably measure tumor perfusion, provided that a certain level of standardization of both the scan acquisition and data analysis is applied. Their value and potential have been shown in several phase I, II, and III drug trials. Both tools can be used to (a) assess anticancer drug activity early in drug development, (b) evaluate response to anticancer drugs early in treatment, and (c) study biological processes in tumors, which increases knowledge on the tumor microenvironment and might lead to a better understanding of drug failure patterns.

There are still some issues that need to be reconsidered to allow for routine clinical use of DCE-MRI and $H_2{}^{15}O$ PET. $H_2{}^{15}O$ PET studies require the availability of an onsite cyclotron, while DCE-MRI data analysis is not straightforward, and can be complex and time-consuming. In addition, the predictive value of $H_2{}^{15}O$ PET and DCE-MRI has to be confirmed in phase III trials with predefined cutoff values for response definition before recommendations on clinical use can be made.

However, their noninvasive nature in combination with promising results must be regarded as a stimulus to incorporate DCE-MRI and $H_2{}^{15}O$ PET in future trials. This will contribute to a better understanding of the value of both techniques and might one day serve patients in the clinical setting.

ACKNOWLEDGMENTS

The authors are grateful to Dr. E. F. Smit and Dr. A. C. Dingemans for their comments and to Dr. A. A. Lammertsma for critically reviewing the manuscript.

AUTHOR CONTRIBUTIONS

Conception/design: Adrianus J. de Langen, J. Tim Marcus, Mark Lubberink

Provision of study materials or patients: Adrianus J. de Langen, Vivian E. M. van den Boogaart, Mark Lubberink

Manuscript writing: Adrianus J. de Langen, Vivian E. M. van den Boogaart, J. Tim Marcus, Mark Lubberink

Final approval of manuscript: Adrianus J. de Langen, Vivian E. M. van den Boogaart, J. Tim Marcus, Mark Lubberink

REFERENCE LIST

1. Folkman J. Role of angiogenesis in tumor growth and metastasis. Semin Oncol 2002;29(Suppl 16):15-18.
2. Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med 1971;285:1182-1186.
3. Mineura K, Sasajima T, Itoh Y et al. Blood flow and metabolism of central neurocytoma: a positron emission tomography study. Cancer 1995;76:1224-1232.
4. Mineura K, Shioya H, Kowada M et al. Blood flow and metabolism of oligodendrogliomas: a positron emission tomography study with kinetic analysis of 18F-fluorodeoxyglucose. J Neurooncol 1999;43:49-57.
5. Lassen U, Andersen P, Daugaard G et al. Metabolic and hemodynamic evaluation of brain metastases from small cell lung cancer with positron emission tomography. Clin Cancer Res 1998;4:2591-2597.
6. Jain RK. Determinants of tumor blood flow: a review. Cancer Res 1988;48:2641-2658.
7. Hobbs SK, Monsky WL, Yuan F et al. Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment. Proc Natl Acad Sci U S A 1998;95:4607-4612.
8. Padera TP, Stoll BR, Tooredman JB et al. Pathology: cancer cells compress intratumour vessels. Nature 2004;427:695.
9. Jain RK, Tong RT, Munn LL. Effect of vascular normalization by antiangiogenic therapy on interstitial hypertension, peritumor edema, and lymphatic metastasis: insights from a mathematical model. Cancer Res 2007;67:2729-2735.
10. Fukumura D, Jain RK. Tumor microenvironment abnormalities: causes, consequences, and strategies to normalize. J Cell Biochem 2007;101:937-949.
11. Tofts PS, Brix G, Buckley DL et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999;10:223-232.
12. Kety SS. Measurement of local contribution within the brain by means of inert, diffusible tracers; examination of the theory, assumptions and possible sources of error. Acta Neurol Scand Suppl 1965;14:20-23.
13. Padhani AR, Hayes C, Landau S et al. Reproducibility of quantitative dynamic MRI of normal human tissues. NMR Biomed 2002;15:143-153.
14. Taylor JS, Tofts PS, Port R et al. MR imaging of tumor microcirculation: promise for the new millennium. J Magn Reson Imaging 1999;10:903-907.
15. Wilson CB, Lammertsma AA, McKenzie CG et al. Measurements of blood flow and exchanging water space in breast tumors using positron emission tomography: a rapid and noninvasive dynamic method. Cancer Res 1992;52:1592-1597.
16. de Lussanet QG, Backes WH, Griffioen AW et al. Gadopentetate dimeglumine versus ultrasmall superparamagnetic iron oxide for dynamic contrast-enhanced MR imaging of tumor angiogenesis in human colon carcinoma in mice. Radiology 2003;229:429-438.
17. Smith T, Tong C, Lammertsma AA et al. Dosimetry of intravenously administered oxygen-15 labelled water in man: a model based on experimental human data from 21 subjects. Eur J Nucl med 1994;21:1126-1134.
18. Brix G, Semmler W, Port R et al. Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. J Comput Assist Tomogr 1991;15:621–628.
19. Lodge MA, Carson RE, Carrasquillo JA et al. Parametric images of blood flow in oncology PET studies using [15O]water. J Nucl Med 2000;41:1784-1792.
20. Blomqvist G. On the construction of functional maps in positron emission tomography. J Cereb Blood Flow Metab 1984;4:629-632.
21. Boellaard R, Knaapen P, Rijbroek A et al. Evaluation of basis function and linear least squares methods for generating parametric blood flow images using 15O-water and positron emission tomography. Mol Imaging Biol 2005;7:273-285.
22. Raichle ME, Martin WR, Herscovitch P et al. Brain blood flow measured with intravenous H2(15)O. II. Implementation and validation. J Nucl Med 1983;24:790-798.
23. Brooks DJ, Frackowiak RS, Lammertsma AA et al. A comparison between regional cerebral blood flow measurements obtained in human subjects using 11C-methylalbumin microspheres, the C15O2 steady-state method, and positron emission tomography. Acta Neurol Scand 1986;73:415-422.
24. Bergmann SR, Fox KA, Rand AL et al. Quantification of regional myocardial blood flow in vivo with H215O. Circulation 1984;70:724-733.
25. Araujo LI, Lammertsma AA, Rhodes CG et al. Noninvasive quantification of regional myocardial blood flow in coronary artery disease with oxygen-15-labeled carbon dioxide inhalation and positron emission tomography. Circulation 1991;83:875-885.
26. Schafers KP, Spinks TJ, Camici PG et al. Absolute quantification of myocardial blood flow with H(2)(15)O and 3-dimensional PET: an experimental validation. J Nucl Med 2002;43:1031-1040.
27. Rimoldi O, Schafers KP, Boellaard R et al. Quantification of subendocardial and subepicardial blood flow using 15O-labeled water and PET: experimental validation. J Nucl Med 2006;47:163-172.
28. Richard JC, Janier M, Decailliot F et al. Comparison of PET with radioactive microspheres to assess pulmonary blood flow. J Nucl Med 2002;43:1063-1071.
29. Mintun MA, Ter Pogossian MM, Green MA et al. Quantitative measurement of regional pulmonary blood flow with

positron emission tomography. J Appl Physiol 1986;60:317-326.
30. Juillard L, Janier MF, Fouque D et al. Renal blood flow measurement by positron emission tomography using 15O-labeled water. Kidney Int 2000;57:2511-2518.
31. Fischman AJ, Hsu H, Carter EA et al. Regional measurement of canine skeletal muscle blood flow by positron emission tomography with H2(15)O. J Appl Physiol 2002;92:1709-1716.
32. Ruotsalainen U, Raitakari M, Nuutila P et al. Quantitative blood flow measurement of skeletal muscle using oxygen-15-water and PET. J Nucl Med 1997;38:314-319.
33. Hoekstra CJ, Stroobants SG, Hoekstra OS et al. Measurement of perfusion in stage IIIA-N2 non-small cell lung cancer using H(2)(15)O and positron emission tomography. Clin Cancer Res 2002;8:2109-2115.
34. Wells P, Jones T, Price P. Assessment of inter- and intrapatient variability in C15O2 positron emission tomography measurements of blood flow in patients with intra-abdominal cancers. Clin Cancer Res 2003;9:6350-6356.
35. Eskey CJ, Wolmark N, McDowell CL et al. Residence time distributions of various tracers in tumors: implications for drug delivery and blood flow measurement. J Natl Cancer Inst 1994;86:293-299.
36. Ponto LL, Madsen MT, Hichwa RD et al. Assessment of Blood Flow in Solid Tumors Using PET. Clin Positron Imaging 1998;1:117-121.
37. Lammertsma AA, Jones T. Low oxygen extraction fraction in tumours measured with the oxygen-15 steady state technique: effect of tissue heterogeneity. Br J Radiol 1992;65:697-700.
38. Coles JP, Fryer TD, Bradley PG et al. Intersubject variability and reproducibility of 15O PET studies. J Cereb Blood Flow Metab 2006;26:48-57.
39. Carroll TJ, Teneggi V, Jobin M et al. Absolute quantification of cerebral blood flow with magnetic resonance, reproducibility of the method, and comparison with H2(15)O positron emission tomography. J Cereb Blood Flow Metab 2002;22:1149-1156.
40. Kaufmann PA, Gnecchi-Ruscone T, Yap JT et al. Assessment of the reproducibility of baseline and hyperemic myocardial blood flow measurements with 15O-labeled water and PET. J Nucl Med 1999;40:1848-1856.
41. Taniguchi H, Kunishima S, Koh T et al. Reproducibility of repeated human regional splenic blood flow measurements using [15O] water and positron emission tomography. Nucl Med Commun 2001;22:755-757.
42. Taniguchi H, Kunishima S, Koh T. The reproducibility of independently measuring human regional hepatic arterial, portal and total hepatic blood flow using [15O]water and positron emission tomography. Nucl Med Commun 2003;24:497-501.
43. Pärkkä JP, Niemi P, Saraste A et al. Comparison of MRI and positron emission tomography for measuring myocardial perfusion reserve in healthy humans. Magn Reson Med 2006;55:772-779.
44. Brasch R, Turetschek K. MRI characterization of tumors and grading angiogenesis using macromolecular contrast media: status report. Eur J Radiol 2000;34:148-155.
45. Knopp MV, Weiss E, Sinn HP et al. Pathophysiologic basis of contrast enhancement in breast tumors. J Magn Reson Imaging 1999;10:260-266.
46. Furman-Haran E, Margalit R, Grobgeld D et al. Dynamic contrast-enhanced magnetic resonance imaging reveals stress-induced angiogenesis in MCF7 human breast tumors. Proc Natl Acad Sci U S A 1996;93:6247-6251.
47. Bhujwalla ZM, Artemov D, Glockner J. Tumor angiogenesis, vascularization, and contrast-enhanced magnetic resonance imaging. Top Magn Reson Imaging 1999;10:92-103.
48. Galbraith S, Lodge MA, Taylor NJ et al. Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. NMR in Biomedicine 2002;15:132-142.
49. Morgan B, Utting JF, Higginson A et al. A simple reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging. British Journal of Cancer. 2006;94:1420-1427.
50. Morgan B, Thomas AL, Drevs J et al. Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ZK222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastasis: results from two phase I studies. Journal of Clinical Oncology 2003;21:3955-3964.
51. Thomas AL, Morgan B, Horsfield MA et al. Phase I study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of PTK787/ZK222584 administered twice daily in patients with advanced cancer. Journal of Clinical Oncology 2005;23:4162-4171.
52. Robert J, Vekris A, Pourquier P et al. Predicting drug response based on gene expression. Crit Rev Oncol Hematol 2004;51:205-227.
53. Quintieri L, Fantin M, Vizler C. Identification of molecular determinants of tumor sensitivity and resistance to anticancer drugs. Adv Exp Med Biol 2007;593:95-104.
54. Jain KK. Personalised medicine for cancer: from drug development into clinical practice. Expert Opin Pharmacother 2005;6:1463-1476.
55. Jubb AM, Oates AJ, Holden S et al. Predicting benefit from anti-angiogenic agents in malignancy. Nat Rev Cancer 2006;6:626-635.
56. Willett CG, Boucher Y, di Tomaso E et al. Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer. Nat Med 2004;10:145-147.

57. Hirata A, Ogawa S, Kometani T et al. ZD1839 (Iressa) induces antiangiogenic effects through inhibition of epidermal growth factor receptor tyrosine kinase. Cancer Res 2002;62:2554-2560.
58. Hinnen P, Eskens FA. Vascular disrupting agents in clinical development. Br J Cancer 2007;96:1159-1165.
59. Anderson HL, Yap JT, Miller MP et al. Assessment of pharmacodynamic vascular response in a phase I trial of combretastatin A4 phosphate. J Clin Oncol 2003;21:2823-2830.
60. Schuster DP, Markham J, Welch MJ. Positron emission tomography measurements of pulmonary vascular permeability with Ga-68 transferrin or C-11 methylalbumin. Crit Care Med 1998;26:518-525.
61. Lara PN, Jr., Quinn DI, Margolin K et al. SU5416 plus interferon alpha in advanced renal cell carcinoma: a phase II California Cancer Consortium Study with biological and imaging correlates of angiogenesis inhibition. Clin Cancer Res 2003;9:4772-4781.
62. Anderson H, Yap JT, Wells P et al. Measurement of renal tumour and normal tissue perfusion using positron emission tomography in a phase II clinical trial of razoxane. Br J Cancer 2003;89:262-267.
63. Herbst RS, Mullani NA, Davis DW et al. Development of biologic markers of response and assessment of antiangiogenic activity in a clinical trial of human recombinant endostatin. J Clin Oncol 2002;20:3804-3814.
64. Mross K, Drevs J, Müller M et al. Phase I clinical and pharmacokinetic study of PTK/ZK, a multiple VEGF receptor inhibitor, in patients with liver metastases from solid tumours. European Journal of Cancer 2005;41:1291-1299.
65. Wedam SB, Low JA, Yang SX et al. Antiangiogenic and antitumor effects of Bevacizumab in patients with inflammatory and locally advanced breast cancer. J Clinical Oncology 2006;5:769-777.
66. Liu G, Rugo HS, Wilding G et al. Dynamic contrast-enhanced magnetic resonance imaging as a pharmacodynamic measure of response after acute dosing of AG-013736, an oral angiogenesis inhibitor, in patients with advanced solid tumors: results from a phase I study. J Clinical Oncology 2005;23:5464-5473.
67. O'Donnell A, Padhani A, Hayes C et al. A phase I study of the angiogenesis inhibitor SU5416 (semaxanib) in solid tumours, incorporating dynamic contrast MR pharmacodynamic end points. British Journal of Cancer 2005;93:876-883.
68. Hillengass J, Wasser K, Delorme S. Lumbar bone marrow microcirculation measurements from dynamic contrast-enhanced magnetic resonance imagingis a predictor of event-free survival in progressive multiple myeloma. Clinical Cancer Research 2007;13:475-481.
69. Mankoff DA, Dunnwald LK, Gralow JR et al. Blood flow and metabolism in locally advanced breast cancer: relationship to response to therapy. J Nucl Med 2002;43:500-509.
70. Zasadny KR, Tatsumi M, Wahl RL. FDG metabolism and uptake versus blood flow in women with untreated primary breast cancers. Eur J Nucl Med Mol Imaging 2003;30:274-280.
71. Miles KA, Griffiths MR, Keith CJ. Blood flow-metabolic relationships are dependent on tumour size in non-small cell lung cancer: a study using quantitative contrast-enhanced computer tomography and positron emission tomography. Eur J Nucl Med Mol Imaging 2006;33:22-28.
72. Fukuda K, Taniguchi H, Koh T et al. Relationships between oxygen and glucose metabolism in human liver tumours: positron emission tomography using (15)O and (18)F-deoxyglucose. Nucl Med Commun 2004;25:577-583.
73. Hirasawa S, Tsushima Y, Takei H et al. Inverse correlation between tumor perfusion and glucose uptake in human head and neck tumors. Acad Radiol 2007;14:312-318.
74. Lehtio K, Eskola O, Viljanen T et al. Imaging perfusion and hypoxia with PET to predict radiotherapy response in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2004;59:971-982.
75. Dachs GU, Tozer GM. Hypoxia modulated gene expression: angiogenesis, metastasis and therapeutic exploitation. Eur J Cancer 2000;36:1649-1660.
76. Bruehlmeier M, Roelcke U, Schubiger PA et al. Assessment of hypoxia and perfusion in human brain tumors using PET with 18F-fluoromisonidazole and 15O-H2O. J Nucl Med 2004;45:1851-1859.
77. Jain RK. Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. Nat Med 2001;7:987-989.
78. Jain RK. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science 2005;307:58-62.
79. Batchelor TT, Sorensen AG, di Tomaso E et al. AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. Cancer Cell 2007;11:83-95.
80. Winkler F, Kozin SV, Tong RT et al. Kinetics of vascular normalization by VEGFR2 blockade governs brain tumor response to radiation: role of oxygenation, angiopoietin-1, and matrix metalloproteinases. Cancer Cell 2004;6:553-563.
81. Parker GJ, Roberts C, Macdonald A et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magn Reson Med 2006;56:993-1000.
82. Kostler H, Ritter C, Lipp M et al. Prebolus quantitative MR heart perfusion imaging. Magn Reson Med 2004;52:296-299.
83. Yankeelov TE, Cron GO, Addison CL et al. Comparison of a reference region model with direct measurement of an AIF in the analysis of DCE-MRI data. Magn Reson Med 2007;57:353-361.
84. NCI, CIP, MR. Workshop on translational research in cancer: recommendations for MR measurement methods at 1.5-Tesla and endpoints for use in phase 1/2a trials of anti-cancer therapeutics affecting tumor vascular function, 2006. http://imaging.cancer.gov/reportsandpublications/reportsandpresentations/magneticresonance, accessed November 01, 2007.

85. Krak NC, Boellaard R, Hoekstra OS et al. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging 2005;32:294-301.
86. Breedis C, Young G. The blood supply of neoplasms in the liver. Am J Pathol 1954;30:969-977.
87. Ackerman NB. Experimental studies on the circulation dynamics of intrahepatic tumor blood supply Cancer 1972;29:435-439.
88. Archer SG, Gray BN. Vascularization of small liver metastases. Br J Surg 1989;76:545-548.
89. Voboril R. Blood supply of metastatic liver tumors: an experimental study. Int Surg 2005;90:71-77.
90. Taniguchi H, Oguro A, Koyama H et al. Analysis of models for quantification of arterial and portal blood flow in the human liver using PET. J Comput Assist Tomogr 1996;20:135-144.
91. Ziegler SI, Haberkorn U, Byrne H et al. Measurement of liver blood flow using oxygen-15 labelled water and dynamic positron emission tomography: limitations of model description. Eur J Nucl med 1996;23:169-177.
92. Goerres GW, Kamel E, Seifert B et al. Accuracy of image coregistration of pulmonary lesions in patients with non-small cell lung cancer using an integrated PET/CT system. J Nucl Med 2002;43:1469-1475.
93. Erdi YE, Nehmeh SA, Pan T et al. The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. J Nucl Med 2004;45:1287-1292.
94. Nehmeh SA, Erdi YE, Ling CC et al. Effect of respiratory gating on quantifying PET images of lung cancer. J Nucl Med 2002;43:876-881.
95. Schauwecker DS, Siddiqui AR, Wagner JD et al. Melanoma patients evaluated by four different positron emission tomography reconstruction techniques. Nucl Med Commun 2003;24:281-289.
96. Vaupel P, Kallinowski F, Okunieff P. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. Cancer Res 1989;49:6449-6465.
97. Mathias CJ, Green MA, Morrison WB et al. Evaluation of Cu-PTSM as a tracer of tumor perfusion: comparison with labeled microspheres in spontaneous canine neoplasms. Nucl Med Biol 1994;21:83-87.
98. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med 2007;48:932-945.
99. Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med. 1999 Nov;40(11):1771-7.
100. Westerterp M, Pruim J, Oyen W et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. Eur J Nucl Med Mol Imaging 2007;34:392-404.

3

Reproducibility of tumor perfusion measurements using ^{15}O-labeled water and PET

Adrianus J. de Langen[1]
Mark Lubberink[2]
Ronald Boellaard[2]
Marieke D. Spreeuwenberg[3]
Egbert F. Smit[1]
Otto S. Hoekstra[2]
and Adriaan A. Lammertsma[2]

[1]Department of Respiratory Medicine, VU University Medical Center, Amsterdam, The Netherlands; [2]Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam, The Netherlands; and [3]Department of Clinical Epidemiology and Biostatistics VU University Medical Center, Amsterdam, The Netherlands

J Nucl Med. 2008; 49(11): 1763-1768

ABSTRACT

Purpose

PET and ^{15}O-labeled water ($H_2{}^{15}O$) can be used to noninvasively monitor tumor perfusion. This allows evaluation of the direct target of antiangiogenic drugs, that is, tumor vasculature. Because these drugs often result in consolidation rather than regression of the tumor mass, a change in perfusion might be a more sensitive way to evaluate response than are indirect size measures on a CT scan. However, to use the technique for serial imaging of individual patients, good reproducibility is essential. The purpose of the present study was to evaluate the reproducibility of quantitative $H_2{}^{15}O$ measurements.

Methods

Nine patients with non-small-cell lung cancer (NSCLC) were scanned twice within 7 d and before any therapy. All $H_2{}^{15}O$ scans were followed by an ^{18}F-fluorothymidine scan to allow for adequate volume-of-interest (VOI) definition. VOIs were defined using a 3-dimensional threshold technique. Tumor perfusion and the volume of distribution (V_T) were obtained using a 1-tissue-compartment model including an arterial blood volume component and an image-derived input function. The level of agreement between test and retest values was assessed using the intraclass correlation coefficient (ICC) and Bland–Altman analyses. Possible dependency on absolute values and lesion size was assessed by linear regression.

Results

All primary tumors and more than 90% of clinically suspected locoregional metastases could be delineated. In total, 14 lesions in 9 patients were analyzed. Tumor perfusion showed excellent reproducibility, with an ICC of 0.95 and SD of 9%. The V_T was only moderately reproducible, with an ICC of 0.52 and SD of 16%. No dependency was found on absolute values of perfusion ($p = 0.14$) and V_T ($p = 0.15$). In addition, tumor volume did not influence the reproducibility of perfusion ($p = 0.46$) and V_T ($p = 0.25$).

Conclusion

Quantitative measurements of tumor perfusion using $H_2{}^{15}O$ and PET are reproducible in NSCLC. When patients are repeatedly being scanned during therapy, changes of more than 18% in tumor perfusion and 32% in V_T (> 1.96 x SD) are likely to represent treatment effects.

INTRODUCTION

Because of the limited effects of cytotoxic chemotherapy, interest is increasing in more specific molecular targeted therapeutics for cancer treatment. At present, the most promising therapy for non-small-cell lung cancer (NSCLC) is provided by antivascular or antiangiogenic agents, which target the tumor vascular system. In tumor size and the associated standard size–based approach for monitoring response (Response Evaluation Criteria in Solid Tumors) [1], the effects of antivascular or antiangiogenic agents are not as dramatic as those with classic chemotherapeutics, because antiangiogenic agents often result in consolidation rather than regression of the tumor mass. Because the effect of these agents is not major cell kill, ^{18}F-FDG might not be the ideal tracer for monitoring treatment response with PET. $H_2{}^{15}O$, on the other hand, offers the opportunity to measure tumor perfusion, reflecting vascular status. This method, by allowing for monitoring of the direct target of antiangiogenic drugs, may be superior to indirect measures such as tumor size. Early reports (phase I and phase II trials) have supported this hypothesis [2–4]. Another advantage of $H_2{}^{15}O$ is the short half-life of ^{15}O (2 min), which allows for serial measurements within a single scan session, enabling the evaluation not only of long-term effects but also of immediate effects of an intervention.

To interpret the effects of treatment, however, test–retest variability needs to be known. Although it has been determined for $H_2{}^{15}O$ studies of cerebral and myocardial blood flow [5–7], only little information is available for tumor perfusion studies, in which flow might be much more variable [8–10]. Therefore, the purpose of the present study was to assess the reproducibility of quantitative tumor blood flow measurements in patients with NSCLC using $H_2{}^{15}O$ and PET.

MATERIALS AND METHODS

Eligible patients were prospectively included after informed consent was obtained in accordance with approval by the Medical Ethics Committee of VU University Medical Center. Nine patients with advanced-stage NSCLC were scanned twice within 7 d (mean, 2.4 d) and before any treatment, using an ECAT EXACT HR+ scanner (Siemens/CTI). This scanner has an axial field of view of 15 cm, divided into 63 contiguous planes. On the basis of prior CT data, the patient was positioned supine on the scanner bed with the tumor in the center of the axial field of view. All patients received a venous catheter for radiotracer injection. Acquisition started with a 10- to 15-min transmission scan to correct for photon attenuation [11]. Next, a bolus injection of 1,100 MBq of $H_2{}^{15}O$ was administered through an injector (Medrad International) at 0.8 mL/s, after which the line was flushed with 42 mL of saline (2.0 mL/s). Simultaneous with the injection of $H_2{}^{15}O$, a dynamic emission scan (in 2-dimensional acquisition mode) was started, which lasted 10 min and had variable frame lengths (12 x 5 s, 12 x 10 s, 6 x 20 s, and 10 x 30 s). After an interval of 10 min to allow for the decay of ^{15}O, a bolus of 370 MBq of ^{18}F-FLT in 5 mL

of saline was injected through the same injector at the same speed and with the same flushing protocol. Simultaneous with the injection of ^{18}F-FLT, a dynamic emission scan (in 2-dimensional acquisition mode) was started. It had a total duration of 60 min and variable frame lengths (6 x 5 s, 6 x 10 s, 3 x 20 s, 5 x 30 s, 5 x 60 s, 8 x 150 s, and 6 x 300 s). All dynamic scan data were normalized; corrected for dead time, decay, scatter, randoms, and photon attenuation; and reconstructed as 128 x 128 matrices using filtered backprojection (FBP) with a Hanning filter (cutoff, 0.5 cycle per pixel). This resulted in a transaxial spatial resolution of around 7 mm in full width at half maximum.

Image Analysis

Volumes of interest (VOIs) were defined separately for each scan using a semiautomatic threshold technique (41% of the maximum pixel value with correction for local background) for any lesion with adequate focal uptake [12]. VOIs were defined on ^{18}F-FLT images because of their superior contrast between tumor and background. To assess whether VOIs could be defined without the use of an additional scan, VOIs were also defined on parametric flow images, which were generated using a basis-function approach [13]. For ^{18}F-FLT, the last 3 frames of the ^{18}F-FLT sinograms (45–60 min after injection) were summed and reconstructed using ordered-subset expectation maximization with 2 iterations and 16 subsets followed by postsmoothing of the reconstructed image using a gaussian filter of 5 mm in full width at half maximum to obtain the same resolution as the one obtained for the FBP images. For $H_2{}^{15}O$, parametric images were constructed by the use of basis functions (Appendix, Eq. 4A). Delineation of lesions with limited tumor-to-background contrast (due to low focal uptake or high background activity) is difficult and hampered by observer variation. Therefore, only lesions that required minor or no manual delineation were included. In the latter case, background values were set to zero for voxels adjacent to the VOI that had physiologically high ^{18}F-FLT uptake (hematopoietic bone marrow, liver tissue, and hypervascular areas such as the mediastinum) to prevent inclusion of such voxels in the VOI. For VOIs defined on parametric flow images, the same procedure was followed for adjacent hypervascular or well-perfused areas. After this procedure, the threshold technique was applied, resulting in volumes specifically containing tumor. Tumor VOIs were transferred to FBP-reconstructed dynamic $H_2{}^{15}O$ images for generating time–activity curves. An image-derived input function (IDIF) was obtained by manually drawing multiple 2-dimensional regions of interest over the aortic arch using FBP data [14, 15]. Regions of interest were drawn in the appropriate frame with optimal aorta-to-background contrast (first pass of $H_2{}^{15}O$ bolus) and were then applied to all frames to generate input time–activity curves. The ^{18}F-FLT scans were obtained to allow for adequate VOI definition. The focus of the present study was, however, on the reproducibility of perfusion measurements, and therefore, ^{18}F-FLT results will be reported elsewhere.

Data Analysis

Tumor perfusion and volume of distribution (V_T) were estimated by fitting tumor time–activity curves using standard nonlinear regression techniques and a single-tissue-compartment model together with an IDIF and weighting data for acquired counts and frame duration [16]. Both a model with and a model without inclusion of an arterial blood volume component (V_A) were explored. Model equations are provided in the Appendix. The presence of V_A, and the need to include this in the compartment model, were assessed by comparing the residual sum of the squares of fits with and without V_A, using Akaike and Schwarz criteria [17–19].

If the SE of either fitted perfusion or V_A exceeded 25%, the lesion was excluded from further analysis. The level of agreement between test and retest values was assessed using the intraclass correlation coefficient (ICC) with a 2-way random model with absolute agreement and Bland–Altman analysis [20]. In the latter case, the percentage difference (Δ) values between 2 measures was plotted against the mean of both measures and the mean threshold-defined volume. In this way, possible dependency on both absolute values and tumor size could be visualized. Statistical dependency was analyzed using linear regression. Finally, a 1-sample *t* test was applied to the Δ values to assess systematic bias.

RESULTS

All primary lesions and more than 90% of all clinically suspected locoregional metastases could be delineated, including mediastinal lymph nodes, using ^{18}F-FLT images for VOI definition. In total, 14 lesions in 9 patients were analyzed. One hilar lymph node of 1.8 cm^3 was excluded from analysis because of an SE of greater than 25% (49%) of the estimated perfusion value in 1 of the 2 scans, possibly due to patient movement.

Only 8 lesions in 8 patients could be defined using VOIs defined on parametric flow images, as a result of less contrast. In addition, for this limited number of lesions, reproducibility of perfusion for these VOIs was significantly poorer than that for corresponding VOIs defined on ^{18}F-FLT images (data not shown). Consequently, in the remainder only results for ^{18}F-FLT–defined VOIs are shown.

Individual data are presented in Table 1. Median lesion size (threshold volume as defined on the ^{18}F-FLT images) was 12.9 cm^3 (range, 2.0–89.8 cm^3), median perfusion was 0.43 mL x min^{-1} x g^{-1} (range, 0.28–0.75 mL x min^{-1} x g^{-1}), and median V_T was 0.79 (range, 0.59–1.13). The single-tissue-compartment model including V_A provided better fits than did the model without V_A in 22 (79%) and 25 (89%) of 28 lesions, respectively, according to Akaike and Schwarz criteria. Therefore, all analyses were performed for data obtained with the model including an arterial blood component.

Perfusion showed excellent reproducibility, with an ICC of 0.95 (95% confidence interval, 0.86–0.99) and SD of 9% for percentage and 0.037 for absolute change. The V_T was moderately reproducible, with an ICC of 0.52 (95% confidence interval, -0.21 to 0.82) and SD of 16% for percentage and 0.136 for

Table 1. Absolute values of VOI size, perfusion, volume of distribution of water (V_T) and arterial blood volume fraction (V_A) for both test and retest scans of all patients.

		Test				Retest			
Patient	Lesion	VOI (cm^3)	Perfusion	V_T	V_A	VOI (cm^3)	Perfusion	V_T	V_A
1	Primary	13.1	0.28	0.62	0.06	14.0	0.37	0.59	0.04
	Naruke 5	10.9	0.28	0.70	0.12	13.6	0.32	0.69	0.11
2	Primary	12.6	0.39	0.95	0.08	31.1	0.38	0.70	0.05
	Naruke 4	10.0	0.51	0.93	0.16	11.3	0.56	0.81	0.11
3	Primary	5.9	0.55	0.75	0.10	9.6	0.57	0.66	0.08
4	Primary	29.8	0.63	0.79	0.03	30.6	0.69	0.85	0.07
5	Primary	31.5	0.41	0.98	0.15	42.3	0.37	0.86	0.09
6	Primary	6.6	0.42	0.69	0.08	9.1	0.38	0.60	0.00
	Naruke 4	47.0	0.42	0.76	0.07	53.8	0.41	0.80	0.13
	Naruke 7	8.2	0.56	0.66	0.08	8.1	0.49	0.86	0.01
7	Naruke 7	3.4	0.47	0.90	0.03	3.0	0.43	1.13	0.11
	Naruke 5	2.0	0.46	0.65	0.18	2.0	0.45	0.78	0.25
8	Primary	82.4	0.41	0.89	0.03	89.8	0.38	0.90	0.03
9	Mediastinal mass	58.0	0.74	0.85	0.12	53.5	0.75	0.99	0.08

Naruke numbers are based on the mediastinal lymph node classification, as described by Naruke et al [32].

absolute change. These data show that a change of less than 18% in tumor perfusion and 32% in V_T (< 1.96 x SD) is likely to be due to test–retest variability.

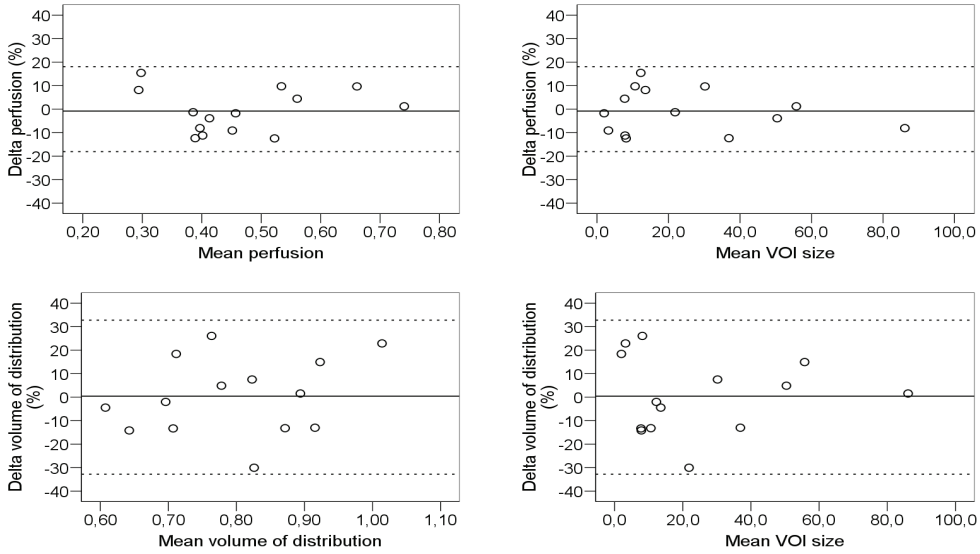
Both visual assessment of Bland–Altman plots and linear regression analysis showed no dependency on the absolute level of perfusion ($p = 0.14$) and V_T ($p = 0.15$) or on tumor volume as measured with PET (perfusion, $p = 0.46$; V_T, $p = 0.25$) (Figure 1). In addition, no evidence of systematic bias was found ($p > 0.80$ for both). Finally, variability in ^{18}F-FLT–defined VOI size had no significant effect on the reproducibility of perfusion ($p = 0.39$) or V_T ($p = 0.16$).

DISCUSSION

The main finding of the present study is that in patients with NSCLC, tumor blood flow can be measured reproducibly using $H_2{}^{15}O$ and PET. When patients are followed over time, a change of more than 18% in tumor perfusion and more than 32% in V_T is likely to represent a biologic effect. No significant dependency on absolute values or lesion size was found, which implies that the same threshold can be used for all tumor lesions. Only 1 (small) lesion was excluded because of an SE of more than 25% for the estimated perfusion value. Because this was the case in only 1 of the 2 scans, patient movement seems to be the most likely cause.

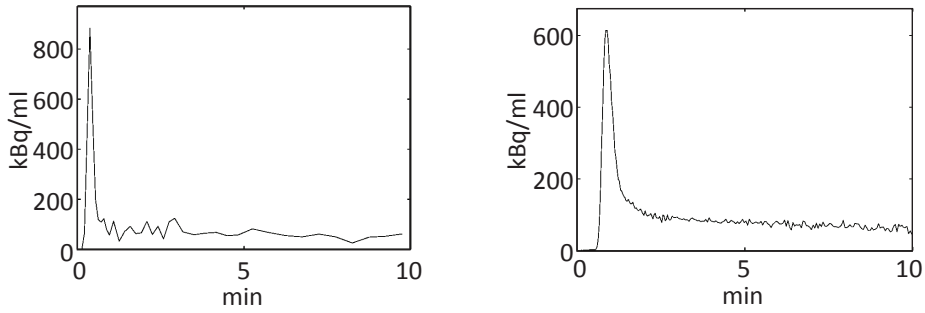
There have been several reports on the reproducibility of $H_2{}^{15}O$ PET measurements in brain [5, 6] and myocardium [7] and some in other nontumor tissues [21, 22]. Although studies show excellent reproducibility, these results cannot easily be translated to an oncologic setting because of differences

Figure 1. Bland–Altman plots for tumor perfusion and V_T. Delta is percentage change between test and retest scans. Mean VOI size is mean tumor volume of test and retest scans in cubic centimeters. Straight lines represent mean values and dotted lines ± 1.96 x SD values.



in VOI definition, tissue characteristics, and perhaps most important, heterogeneity, which can be substantial in tumors. The latter can cause underestimation of perfusion measurements [23], which in turn might result in decreased reproducibility. Two ^{15}O inhalation PET studies have been performed to explore reproducibility of perfusion measurements in a limited range of tumors (mainly hepatic lesions) within a small number of patients [8, 9]. The dual hepatic blood supply (portal vein and hepatic artery) complicates quantification of perfusion in the liver [24]. Consequently, it is difficult to compare blood flow measurements in hepatic lesions with those in tumors of other origin. Wells et al [8] analyzed 7 lesions (6 liver, 1 lung) in 5 patients and found ICCs and SDs of 0.89 and 29% for perfusion and 0.96 and 12% for V_T, respectively. Wilson et al [9] reported on 5 breast tumors in 5 patients and found differences of up to ± 10% for perfusion measurements. The relatively low reproducibility of perfusion measurements in the study by Wells et al [8] might be because of an inadequate definition of the input function, caused by either the dual hepatic input or delay and dispersion of the measured input function (arterial blood sampling). In contrast to IDIF, arterial blood sampling requires corrections for delay and dispersion [9, 25], and this could have affected accuracy [26]. On the other hand, reproducibility of V_T was excellent, a finding that could not be reproduced in the present study. This might be due to the relatively high amount of noise in the tail of IDIF curves as compared with the much smoother input curves derived from arterial blood sampling (Figure 2).

Figure 2. Time–activity curves of arterial input function for 2 different patients demonstrated by IDIF (A) and arterial blood sampling (B). IDIF peak is sharper, because it does not suffer from dispersion. On the other hand, tail is noisier because of poorer count statistics.

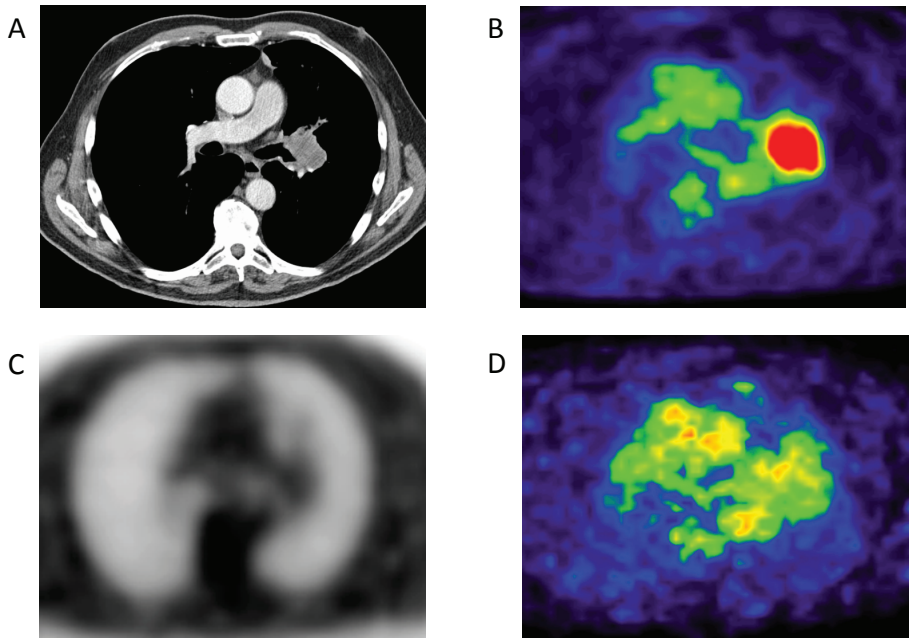


Definition of tumor VOI is difficult for $H_2{}^{15}O$ studies because of low tumor-to-background contrast. Many methods have been proposed for VOI definition, including the use of CT [27], $H_2{}^{15}O$ itself [9], ^{18}F-FDG [19, 28, 29], or a combination of anatomic (CT or PET transmission) and functional ($H_2{}^{15}O$) images [3, 8, 30] (Figure 3).

It is likely that the lack of $H_2{}^{15}O$ test–retest and response monitoring studies is partly because of the VOI definition problem. In the present study a threshold technique was applied to a subsequent ^{18}F-FLT scan. A lower limit of 41% of the maximum pixel value was chosen, because this best reflects true tumor volume [12]. That definition of VOI was not restricted to ^{18}F-FLT–abundant voxels within the tumor regions. This could have led to bias, because the relationship between perfusion and proliferation in tumors is not known. As a result of the use of a separate ^{18}F-FLT scan for VOI definition, total scan duration was 90 min. The addition of an ^{18}F-FLT scan to the acquisition protocol also led to an increase in radiation load. The total amount per patient in this study was approximately 20 mSv. Although all patients had advanced-stage disease with a limited estimated life expectancy of 8–10 mo [31], limitation of radiation load would become important when patients with early-stage disease are being scanned or when better treatment becomes available for those with advanced-stage disease.

If VOIs could be defined without the use of an ^{18}F-FLT scan, total scan time could be reduced to about 20 min and radiation dose to approximately 2.5 mSv. Therefore, an additional analysis was performed, defining VOIs on parametric flow images. Parametric images rather than summed $H_2{}^{15}O$ images were used, because the latter suffer from contamination of (high) intraarterial signals. Nevertheless, because of perfusion of the surrounding tissue, 6 of 14 lesions could not be defined using these parametric images. In addition, reproducibility was inferior to that for ^{18}F-FLT–defined VOIs (test–retest variability 15% for parametric VOIs, compared with 9% for ^{18}F-FLT–defined VOIs). Consequently, this

Figure 3. CT (A), ^{18}F-FDG PET emission (B), PET transmission (C), and $H_2{}^{15}O$ PET emission (1–4 min after injection) (D) images that can be used for VOI definition. All images are from same patient.



approach does not seem to be a viable option for routine response-monitoring studies. The use of a separate CT scan for VOI definition was not attempted, as even a small misalignment between CT and PET scans could result in erroneous flow values, especially for smaller lesions.

With the introduction of integrated PET-CT, however, it is likely that the CT scan can be used for VOI definition, making the additional ^{18}F-FLT scan (or that of another tissue-accumulating tracer) obsolete. This would reduce scan time to less than 15 min. radiation dose of $H_2{}^{15}O$ PET-CT studies would primarily be determined by the CT scan protocol (amount of milliamperes applied and body parts scanned), because $H_2{}^{15}O$ studies themselves only account for a small dose because of the short half-life of ^{15}O. Nevertheless, further studies using a PET-CT scanner are required to assess reproducibility for CT-defined VOI.

Another reason for the limited number of oncologic $H_2{}^{15}O$ studies might be the relatively limited availability of on-site cyclotrons, which are needed because of the short half-life of ^{15}O (2 min). However, when both PET-CT and an on-site cyclotron are available, perfusion studies can be performed with relative ease for both patient and physician. The production of $H_2{}^{15}O$ requires minimal staff, scan time is short, and, at least for patients with NSCLC, there is no need for arterial cannulation. In addition, the present results indicate that tumor perfusion measurements are reproducible.

CONCLUSION

With the increasing use of antiangiogenic drugs and other targeted agents, suitable biomarkers are desired to evaluate response to such treatment. Biologic imaging offers the unique opportunity to noninvasively monitor treatment effects, irrespective of size changes. In this study we have shown that tumor perfusion measurements are reproducible in patients with NSCLC using $H_2{}^{15}O$ PET and an IDIF. Our data suggest that future intervention studies should apply a 20% cutoff value to define response to treatment.

APPENDIX

Modeling Perfusion Data.

For the present study, perfusion measurements were based on the original tracer kinetic model proposed by Kety [33], using the implementation presented by Hoekstra et al [16]. In brief, in a dynamic scan, $dC_T(t)/dt$ (the change in tissue concentration) of a tracer at a certain time point is equal to K_1 (the plasma-to-tissue transport rate constant) multiplied by $C_p(t)$ (the tracer concentration in plasma) minus k_2 (the tissue-to-plasma rate constant) multiplied by the tracer concentration in tissue (C_T) at that time point:

$$\frac{dC_T(t)}{dt} = K_1 C_P(t) - k_2 C_T(t) \qquad \text{Eq. 1A}$$

Because water is freely diffusible (i.e. extraction is 100%), K_1 is equal to perfusion (F). In addition, in case of blood flow, the delivery is not determined by the plasma concentration but rather by the whole-blood concentration, C_B. Thus, for $H_2{}^{15}O$, Equation 1A can be rewritten as:

$$\frac{dC_T(t)}{dt} = F C_B(t) - \frac{F}{V_T} C_T(t) \qquad \text{Eq. 2A}$$

The solution of this equation, including a fractional arterial blood volume (V_A), is:

$$C_T(t) = (1 - V_A) \cdot F \cdot \exp(-\frac{F}{V_T} t) \otimes C_B(t) + V_A \cdot C_B(t) \qquad \text{Eq. 3A}$$

A nonlinear, least-squares fit of this function to the measured time–activity curve in the tissue of interest ($C_T(t)$) yields perfusion and V_T [9]. If V_A is negligible, it can be set to zero before fitting, thereby increasing precision of perfusion and V_T estimates.

Parametric Images.

Parametric images can be obtained using a basis-function implementation of the same single-tissue-compartment model:

$$C_T(t) = (1 - V_A) \cdot F \cdot \exp(-\beta_i t) \otimes C_B(t) + V_A \cdot C_B(t) \qquad \text{Eq. 4A}$$

In this approach, first a set of basis functions $\exp(-ß_i t) \otimes C_B(t)$ is created by convolution of the whole-blood input function with a set of single exponential functions. In the present study, a set of exponential constants, $ß_i$, was used, ranging from 0.01 to 1 min^{-1}. Next, for each basis function all voxels were fitted to Equation 4A using simple linear regression. Finally, for each voxel, the linear fit (i.e. basis function) was selected that provided the best fit. Because parametric images were generated only for VOI definition, in the present study V_T was fixed to 0.75 to reduce noise [13, 19, 21].

REFERENCE LIST

1. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst. 2000;92:205–216.
2. Anderson H, Yap JT, Wells P, et al. Measurement of renal tumour and normal tissue perfusion using positron emission tomography in a phase II clinical trial of razoxane. Br J Cancer. 2003;89:262-267.
3. Anderson HL, Yap JT, Miller MP, Robbins A, Jones T, Price PM. Assessment of pharmacodynamic vascular response in a phase I trial of combretastatin A4 phosphate. J Clin Oncol. 2003;21:2823-2830.
4. Herbst RS, Mullani NA, Davis DW, et al. Development of biologic markers of response and assessment of antiangiogenic activity in a clinical trial of human recombinant endostatin. J Clin Oncol. 2002;20:3804-3814.
5. Carroll TJ, Teneggi V, Jobin M, et al. Absolute quantification of cerebral blood flow with magnetic resonance, reproducibility of the method, and comparison with H2(15)O positron emission tomography. J Cereb Blood Flow Metab. 2002;22:1149-1156.
6. Coles JP, Fryer TD, Bradley PG, et al. Intersubject variability and reproducibility of 15O PET studies. J Cereb Blood Flow Metab. 2006;26:48-57.
7. Kaufmann PA, Gnecchi-Ruscone T, Yap JT, Rimoldi O, Camici PG. Assessment of the reproducibility of baseline and hyperemic myocardial blood flow measurements with 15O-labeled water and PET. J Nucl Med. 1999;40:1848-1856.
8. Wells P, Jones T, Price P. Assessment of inter- and intrapatient variability in C15O2 positron emission tomography measurements of blood flow in patients with intra-abdominal cancers. Clin Cancer Res. 2003;9:6350-6356.
9. Wilson CB, Lammertsma AA, McKenzie CG, Sikora K, Jones T. Measurements of blood flow and exchanging water space in breast tumours using positron emission tomography: a rapid and noninvasive dynamic method. Cancer Res. 1992;52:1592-1597.
10. Vaupel P, Kallinowski F, Okunieff P. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. Cancer Res. 1989;49:6449-6465.
11. Boellaard R, van Lingen A, van Balen SC, Lammertsma AA. Optimization of attenuation correction for positron emission tomography studies of thorax and pelvis using count-based transmission scans. Phys Med Biol. 2004;49:N31-38.
12. Erdi YE, Mawlawi O, Larson SM, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997;80:2505-2509.
13. Boellaard R, Knaapen P, Rijbroek A, Luurtsema GJ, Lammertsma AA. Evaluation of basis function and linear least squares methods for generating parametric blood flow images using 15O-water and Positron Emission Tomography. Mol Imaging Biol. 2005;7:273-85.
14. Hoekstra CJ, Hoekstra OS, Lammertsma AA. On the use of image-derived input functions in oncological fluorine-18 fluorodeoxyglucose positron emission tomography studies. Eur J Nucl Med. 1999;26:1489-1492.
15. van der Weerdt AP, Klein LJ, Boellaard R, Visser CA, Visser FC, Lammertsma AA. Image-derived input functions for determination of MRGlu in cardiac (18)F-FDG PET scans. J Nucl Med. 2001;42:1622-1629.
16. Hoekstra CJ, Stroobants SG, Hoekstra OS, Smit EF, Vansteenkiste JF, Lammertsma AA. Measurement of perfusion in stage IIIA-N2 non-small cell lung cancer using H(2)(15)O and positron emission tomography. Clin Cancer Res. 2002;8:2109-15.
17. Akaike H. A new look at the statistical identification. IEEE Trans Automat Contr. 1978;19:716-723
18. Schwarz G. Estimating the dimension of a model. Ann Statist. 1978;6:461–464.
19. Hoekstra CJ, Hoekstra OS, Stroobants SG, et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. J Nucl Med. 2002;43:1304-1309.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307-310.
21. Taniguchi H, Kunishima S, Koh T. The reproducibility of independently measuring human regional hepatic arterial, portal and total hepatic blood flow using [15O]water and positron emission tomography. Nucl Med Commun. 2003;24:497-501.
22. Taniguchi H, Kunishima S, Koh T, Oguro A, Yamagishi H. Reproducibility of repeated human regional splenic blood flow measurements using [15O] water and positron emission tomography. Nucl Med Commun. 2001;22:755-7.
23. Blomqvist G, Lammertsma AA, Mazoyer B, Wienhard K. Effect of tissue heterogeneity on quantification in positron emission tomography. Eur J Nucl Med. 1995;22:652-663.
24. Ziegler SI, Haberkorn U, Byrne H, et al. Measurement of liver blood flow using oxygen-15 labelled water and dynamic positron emission tomography: limitations of model description. Eur J Nucl Med. 1996;23:169-177.
25. Lammertsma AA, Cunningham VJ, Deiber MP, et al. Combination of dynamic and integral methods for generating reproducible functional CBF images. J Cereb Blood Flow Metab. 1990;10:675-686.
26. Rajasekharan S, Boellaard R, Lubberink M, Lammertsma AA, Jones T. Measurement of abdominal tumour blood flow: independent assessment of delay and dispersion of the arterial input function. IEEE Medical Imaging Conference, Rome, 2004, conference record

27. Inaba T. Quantitative measurements of prostatic blood flow and blood volume by positron emission tomography. J Urol. 1992;148:1457-1460.
28. Kurdziel KA, Figg WD, Carrasquillo JA, et al. Using positron emission tomography 2-deoxy-2-[18F]fluoro-D-glucose, 11CO, and 15O-water for monitoring androgen independent prostate cancer. Mol Imaging Biol. 2003;5:86-93.
29. Lehtio K, Oikonen V, Gronroos T, et al. Imaging of blood flow and hypoxia in head and neck cancer: initial evaluation with [(15)O]H(2)O and [(18)F]fluoroerythronitroimidazole PET. J Nucl Med. 2001;42:1643-1652.
30. Logan TF, Jadali F, Egorin MJ, et al. Decreased tumour blood flow as measured by positron emission tomography in cancer patients treated with interleukin-1 and carboplatin on a phase I trial. Cancer Chemother Pharmacol. 2002;50:433-444.
31. Schiller JH, Harrington D, Belani CP, et al. Eastern Cooperative Oncology Group. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med. 2002;346:92-98.
32. Naruke T, Suemasu K, Ishikawa S. Lymph node mapping and curability at various levels of metastasis in resected lung cancer. J Thorac Cardiovasc Surg. 1978;76:832-839.
33. Kety SS. The theory and applications of the exchange of inert gas at the lungs and tissues. Pharmacol Rev. 1951;3:1-41.

4

Repeatability of ^{18}F-FDG uptake measurements in tumors: a meta-analysis

Adrianus J. de Langen[1]
Andrew Vincent[2]
Linda M. Velasquez[3]
Harm van Tinteren[2]
Ronald Boellaard[4]
Lalitha K. Shankar[5]
Maarten Boers[6]
Egbert F. Smit[1]
Sigrid Stroobants[7]
Wolfgang A. Weber[8]
Otto S. Hoekstra[5]

Departments of [1]Pulmonary Diseases, VU University Medical Center, Amsterdam, The Netherlands, [2]Biostatistics, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [3]Bristol-Myers Squibb Co., [4]Nuclear Medicine & PET research, VU University Medical Center, Amsterdam, The Netherlands, [5]Cancer Imaging Program, National Cancer Institute, Bethesda, Maryland, USA, [6]Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands, [7]Nuclear Medicine, University Hospital Antwerpen, Antwerpen, Belgium, [8]Nuklearmedizinische Klinik, Universitätsklinikum Freiburg, Freiburg, Germany

Accepted for publication in J Nucl Med

ABSTRACT

Purpose

Positron emission tomography with the glucose analogue ^{18}F-fluorodeoxyglucose (^{18}F-FDG PET) is increasingly used to monitor tumor response to therapy. In order to use quantitative measurements of tumor ^{18}F-FDG uptake for assessment of tumor response, the repeatability of this quantitative metabolic imaging method needs to be established. Therefore we determined the repeatability of different SUV measurements based on the available data.

Methods

A systematic literature search was performed to identify studies addressing ^{18}F-FDG repeatability in malignant tumors. The level of agreement between test and retest values of two PET uptake measures, SUV_{max} and SUV_{mean}, was assessed with the coefficient of repeatability using generalized linear mixed effects models. In addition the influence of tumor volume on repeatability was assessed. Principal component transformation was used to compare the reproducibility of the two different uptake measures.

Results

Five cohorts were identified for this meta-analysis. For SUV_{max} and SUV_{mean}, datasets of 86 and 102 patients were available, respectively. Percentage repeatability is a function of the level of uptake. SUV_{mean} had the best repeatability characteristics; for serial PET scans, a threshold of a combination of 20% as well as 1.2 SUV_{mean} units was most appropriate. After adjusting for uptake rate, tumor volume had minimal influence on repeatability.

Conclusion

SUV_{mean} had better repeatability performance than SUV_{max}. Both measures showed poor repeatability for lesions with low ^{18}F-FDG uptake. We recommend to evaluate biological effects in PET imaging by reporting a combination of minimal relative and absolute changes to account for test-retest variability.

INTRODUCTION

^{18}F-FDG PET has gained an important role in the clinical setting to detect and stage malignancies and assess treatment response [1-6]. In the research setting, PET is increasingly being used to study early changes of biological effects during and after anticancer treatment [7-10]. The non-invasive nature of PET allows multiple serial measurements without interfering with biological processes within the tumor and might obviate more invasive procedures, like biopsy.

Even though PET clinical practice is still dominated by qualitative (visual) image analysis, several potential indications require quantification, e.g. when prognostic and predictive information is required beyond the level of TNM staging. A decrease of ^{18}F-FDG uptake after therapy is associated with favorable clinical outcome [1-6, 11] However, to date only qualitative, and not quantitative, PET measures have been incorporated in response classification systems for solid tumors and lymphoma [12, 13].

The European Organization for Research and Treatment of Cancer (EORTC) PET study group published recommendations, as far back as 1999, for response monitoring using quantitative PET data to promote consistency in the reporting of studies [14]. The proposed system was based on the results of both drug evaluation and a few repeatability studies.

To discriminate true signal change from noise and to be able to stratify patients based on changes in ^{18}F-FDG uptake values, the repeatability of the measurement and the error of the determination need to be known. The present meta-analysis intends to determine the repeatability of different SUV measurements based on the available data and to evaluate potential sources of heterogeneity.

MATERIALS AND METHODS

Study Design

We performed a systematic literature search of Medline and Embase databases to identify studies addressing ^{18}F-FDG repeatability in malignant tumors by using the following search terms: positron emission tomography, FDG, repeatability and test-retest. Additionally, extensive cross-referencing was done, review articles were screened, and experts in the field were consulted. Studies were included when the following criteria were met: 1) assessment of repeatability with ^{18}F-FDG PET in malignant tumors; 2) use of standardized uptake values (SUV); 3) uniform acquisition and reconstruction protocols; 4) application of the same scanner for the test and retest scan for each patient (i.e. no within patient scanner variation).

For dynamic studies, SUVs were calculated using the last frame of the dynamic acquisition. Because the method of tumor delineation (e.g. maximum pixel value, or threshold based or fixed diameter volume of interest [VOI]) can affect ^{18}F-FDG uptake measures, we attempted to obtain uniformly de-

fined tumor volumes between studies for each uptake measure. This volume was defined by a 3-dimensional threshold based volume (isocontour defined, with a cutoff of 50% of the maximum ^{18}F-FDG concentration within the tumor). If we were unable to extract the required data from the original publications, authors were asked to provide it or to reanalyse their data with the isocontour technique (using in-house developed software provided by us).

Statistical Analysis

For both PET uptake measures, SUV_{max} and SUV_{mean}, the level of agreement between test and retest measurements was assessed using the intraclass correlation coefficient (ICC) calculated using a random-effects model with random intercepts for published study, patient and tumor location. Kruskal-Wallis tests were applied to the uptake measures to assess systematic bias between studies. One sample Anderson-Darling tests were used to assess the distribution of the means of the test and retest observations.

Variance-mean and Variance-volume Relations

To assess repeatability we determined the relation between the mean and variance of the test and retest scans, where the variance in the test-retest measurements is assumed solely due to measurement error. To account for differences in this relation between published studies we assessed this association using generalized linear mixed effects models (with published study as a random factor) and generalized linear models (with published study as a fixed effect). The outcome variable was the square of the difference in test-retest measurements with the log transformed test-retest mean as a fixed effect. Given the assumption of normality, the square of the difference is chi-square distributed, and therefore Gamma error distributions were employed [15]. The log-link function was used to relate the estimated variance to the test-retest mean resulting in an allometric mean-variance relation. Differences between published studies were assessed using the generalized linear models with published study as a fixed effect. The influence of tumor volume on ^{18}F-FDG uptake test-retest repeatability was assessed by including the log transformed tumor volumes as fixed effects. To avoid extrapolation, we limited the estimated variance-mean associations to values between the 5th smallest and 5th largest observed PET measurements values.

Coefficient of 95% Repeatability

Once the variance-mean relation had been estimated, the relation between the Coefficient of 95% Repeatability (CR95) and the mean was calculated as 1.96 times the standard deviation [16, 17]. The CR95 is the variation solely due to measurement error. If the difference of two measurements exceeds

the CR95, then this difference is 95% likely to be due to a true change in tumor ^{18}F-FDG uptake rate and not measurement error.

Single PET Observations

The test-retest CR95 corresponds to a two observation setting, such as serial baseline and post-treatment assessments. In some clinical trials however, a single PET scan result is used, e.g. for patient stratification during randomization. The equivalent single observation CR95 can be calculated by dividing the test-retest CR95 by √2 (i.e. half the variance), with the assumption that all measurements are performed on the same PET scanner using the same acquisition and reconstruction protocols.

Comparison of the Different Quantitative ^{18}F-FDG Measures

To compare the repeatability of the two ^{18}F-FDG uptake measurements, a principal component analysis was performed using data from the four studies which analysed both PET measures. This resulted in a transformation of each of the measures onto the first principal component, and thus allowing their CR95s to be compared on the same scale. The significance level of all tests was set at 0.05. The generalized linear mixed effects models were fitted using the GLIMMIX procedure in SAS (SAS Institute Inc., Cary, NC, USA). All other analyses were performed in R v2.9.2 (The R Foundation for Statistical Computing, 2009).

RESULTS

Eight repeatability studies were identified [16, 18-24]. One study was excluded due to the use of different scanners for the test and retest scan [24]. All authors were contacted to provide patient based data on tumor volume, location, and ^{18}F-FDG uptake. The study characteristics are summarized in Table 1. Extended analyses of two cohorts were published separately from the original ones; Minn et al [20] and Nakamoto et al [22]; Hoekstra et al [18] and Krak et al [19]. Thus, five cohorts were available for this meta-analysis. All repeat scans were performed using the same scanner as previously for each patient. Three studies used a dynamic scan protocol [16, 18, 20], while the remaining two applied a static protocol [21, 23]. All but one [23] were single center studies. Tumor types were gastro-intestinal [23], lung [18, 20], and miscellaneous primary, predominantly located in the chest [16, 21]. Two studies [21, 23] used integrated PET-CT scanners, the others a PET only system. Nahmias et al [21] applied a 90 min. time interval between radiotracer injection and scanning, while all other studies applied a 60 min. time interval. The multicenter study by Velasquez et al [23] reported on two datasets, one before and one after quality assurance assessment. The latter dataset was used for the present study. Finally, Minn et al [20] excluded lesions with a diameter lower than 2 cm, while the other studies had no clear restrictions regarding lesion size.

Table 1. Individual study characteristics.

		Hoekstra	Minn	Nahmias (50%-iso-contour)	Nahmias (manual)	Velasquez	Weber
Patients	N	10	10	21	21	45	16
Tumors	N	27	10	21	21	105	50
Tumors per patient	Median (range)	2 (1 - 7)	1 (1 - 1)	1 (1 - 1)	1 (1 - 1)	3 (1 - 4)	2.5 (1 - 8)
Time*	Minutes	60	60	90	90	60	70
Bed position time	Minutes	Dynamic scan	Dynamic scan	3	3	Not available	Dynamic scan
Time (days)†	Mean (max)	1 (1)	1.8 (7)	2.8 (5)	2.8 (5)	4.1 (7)	3 (10)
Scanner (patients)	PET alone	10	10	0	0	7	16
	PET-CT	0	0	2	2	45	3
Location (patients)	Thoracical	10	10	19	19	0	13
	Abdominal	0	0	2	2	45	3
Tumor volume (cm^3)	Median	6.2	42.6	4.9	4.9	6.4	5.1
	(range)	(0.7 - 111.4)	(18.6–231.0)	(1.0 - 79.4)	(1.0 - 79.4)	(0.4 - 491.4)	(0.6 - 86.9)
Threshold technique	4x4 voxels around the max voxel		X				
	50% of the max voxel	X		X			X
	70% of the max voxel					X	
	Manual delineation				X		
SUV_{mean}‡	Median	5.5	8.1	6.6	5.1	6.8	4.5
	(range)	(2.3 - 11.3)	(3.9 - 16.4)	(2.6 - 17.3)	(1.6 - 17.4)	(2.5 - 24.4)	(1.3 - 10.5)
SUV_{max}§	Median	8.3	9.2	10	8.9	7.3	
	(range)	(3.8 - 18.5)	(4.6 - 19.5)	(4.1 - 24.1)	(4.0 - 23.8)	(2.5 - 29.7)	

* Time between ^{18}F-FDG injection and the start of the scan for static scans and total scan time for dynamic scans.

† Time between the test and retest scan.

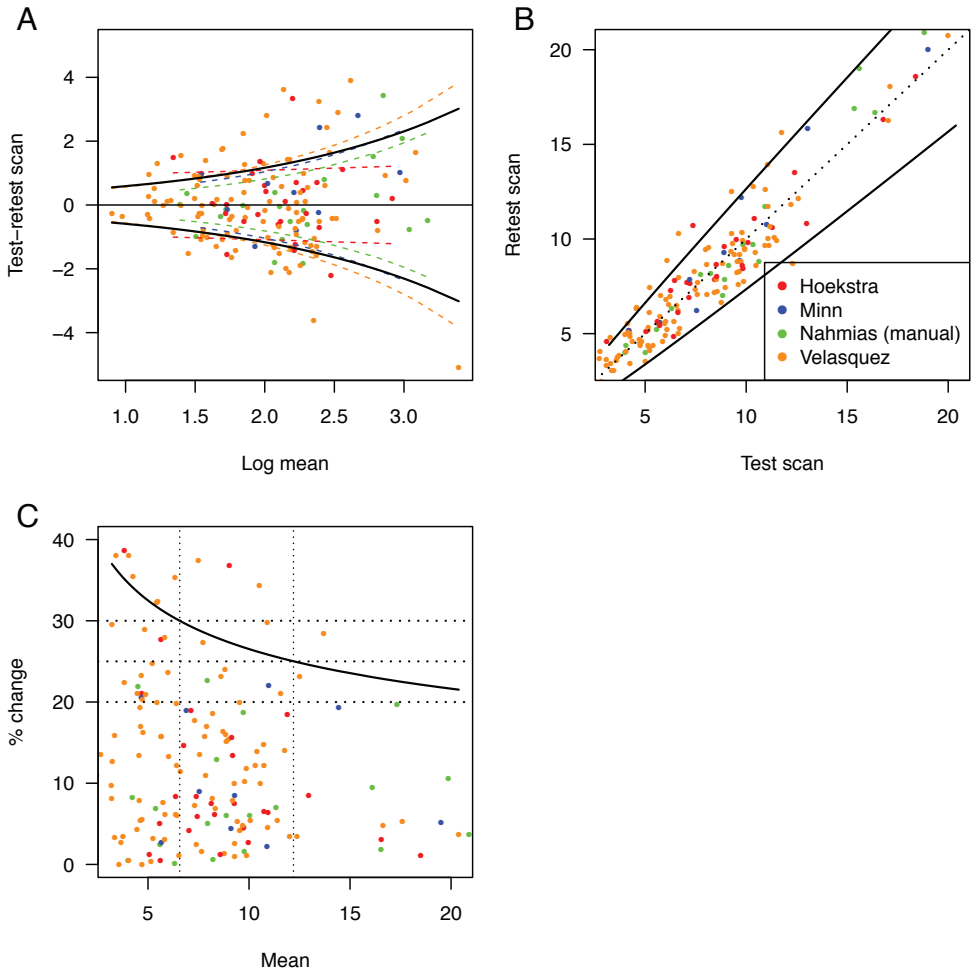
‡ SUV_{mean} values were lower in the Weber study than in the other four studies ($p = 0.0006$).

§ SUV_{max} values were lower in the Velasquez study than in the other three studies ($p = 0.04$).

Repeatability of Standardized Uptake Values

SUV_{max}. The four available datasets for SUV_{max} [18, 20, 21, 23] included 163 tumor lesions in 86 patients. The ICC of SUV_{max} was 0.90. The goodness-of-fit tests for normality indicated that SUV_{max} on the original scale was significantly different from normal ($p < 0.0001$). After log-transformation the distribution was normalized ($p = 0.37$). Figure 1A presents the overall and study-specific relations between the mean and the standard deviation (i.e. variance) of the test and retest scans. The difference in

Figure 1. SUV_{max} repeatability. (A) The estimated study-specific standard deviations (coloured dashed lines; study as a fixed effect) and the overall standard deviation (black solid line; study as a random effect). (B) The test and retest scan values plotted on the original scale. The solid line is the coefficient of repeatability (CR95). (C) The relation between the CR95, as percentage change, with the level of SUV_{max}. The dotted lines indicate the absolute and relative differences, as presented in Table 2.



test-retest variability between studies was not significant (p = 0.25). Test-retest variability in absolute terms increases with higher SUV_{max} values (Figures 1A and 1B), while the percentage change decreases with higher SUV_{max} values (Figure 1C). As a consequence, it is not possible to generate just one value that defines the limits of test-retest variability for the full range of SUV_{max} values. However, this can be accounted for by combining an absolute and a percentage change of SUV_{max}. From Figure 1C these

values can be extracted. The required absolute difference is calculated as the relative difference multiplied by the mean value where the CR95 curve crosses the relative difference value. For SUV_{max}, the 30% relative difference crosses the CR95 at a SUV_{max} value of 6.7. Thus the combination of an absolute change in SUV_{max} of more than 2 units (0.3 x 6.7) and a relative change of more than 30% is more than 95% likely to be true change rather than measurement error for all SUV_{max} values that were encountered in this study. From the same figure it can be calculated that a change of more than 3.1 units and 25% relative change also exceeds 95% test-retest variability. The combinations of the required minimal absolute and percentage changes that are needed to exceed 95% test-retest variability for the SUV parameters are presented in Table 2.

SUV_{mean}. The five available datasets for SUV_{mean} [16, 18, 20, 21, 23] included 213 tumor lesions in 102 patients. The tumor delineation methods differed between the five available studies. Hoekstra et al [18] and Weber et al [16] delineated isocontour defined volumes with a 50% threshold of the maximum voxel value. Nahmias et al [21] originally published manually defined volumes, but reanalysed their data on our request with the 50% isocontour technique. Velasquez et al [23] used a 70% threshold technique because it was not possible to define 50% volumes without contaminating the tumor VOI with background tissue (due to ^{18}F-FDG uptake heterogeneity and a relatively low tumor-to-background contrast). Minn et al [20] delineated tumor volumes as 4x4 voxels around the voxel with maximum ^{18}F-FDG uptake.

Table 2. Required relative and absolute differences to exceed test-retest variability.*

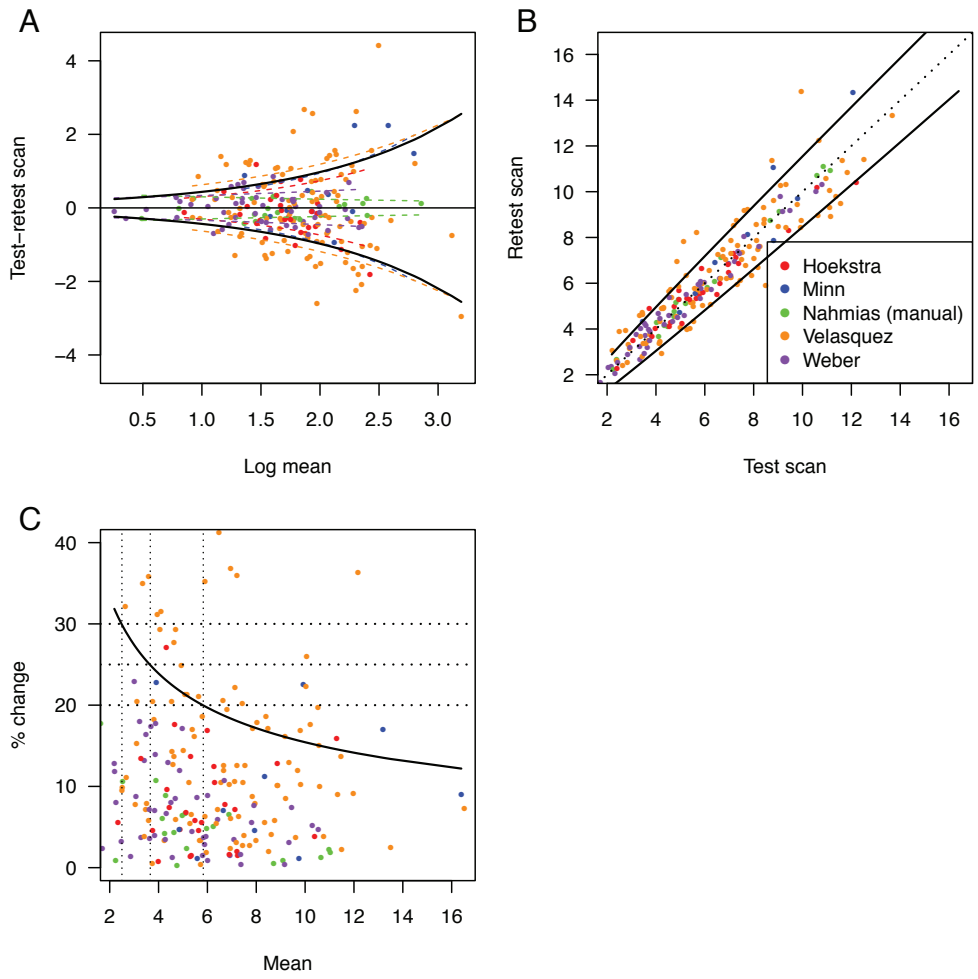
	Relative difference	Absolute difference
SUV_{max}	-	-
	25%	3.1
	30%	2.0
SUV_{mean} H, N, M, W, V	20%	1.17
	25%	0.96
	30%	0.75
SUV_{mean} H, N†, W	20%	0.96
	25%	0.78
	30%	0.66

* Thresholds can be extracted from Figures 1-3 as explained for SUV_{max} in the text.
Differences have to exceed both the relative and absolute thresholds.
H = Hoekstra et al, N = Nahmias et al manual analysis, N† = Nahmias et al 50% isocontour analysis, M = Minn et al,
V = Velasquez et al, W = Weber et al.

The ICC of SUV_{mean} was 0.91. The goodness-of-fit tests for normality indicated that SUV_{mean} on the original scale was significantly different from normal (p = 0.002). After log-transformation this was normalized (p = 0.77).

Figure 2A presents the overall and study-specific relations between the mean and the standard deviation (i.e. variance) of the test and retest scans. These associations differed significantly across studies (p < 0.0001), with the greatest variance observed in the study of Velasquez et al [23] and the lowest in the study of Nahmias et al [21] (manual VOI delineation). When only studies applying the 50% isocontour technique were included (Hoekstra et al [18], Weber et al [16], 50% isocontour data of Nahmias et al [21]), no difference in test-retest variability was apparent between studies (p = 0.13; Figure 3).

Figure 2. Same as in Figure 1 for SUV_{mean} measurements.



In the dataset of Nahmias et al [21], the manual delineation method performed better in terms of repeatability than the automated one (p = 0.0001).

Figure 2 for SUV_{mean} shows the same trend as was observed for SUV_{max}. The absolute change in test-re-test variability increases with higher SUV_{mean} values (Figures 2A and 2B), while the percentage change decreases with higher SUV_{mean} values (Figure 2C). As a consequence, a combination of an absolute and a percentage change is necessary to cover the limits of test-retest variability for the full range of SUV_{mean} values. From Figure 2 it can be extracted that the combined change in SUV_{mean} of more than 1.2 units and 20% exceeds 95% test-retest variability for all SUV_{mean} values that were encountered in this study, irrespective of the tumor delineation method (values are presented in Table 2). Homogeneous tumor delineation (50% threshold technique) resulted in lower test-retest variability (Figure 3, Table 2).

Figure 3. Same as in Figure 1 for the three studies with SUV_{mean} measurements assessed using 50%-isocontouring.

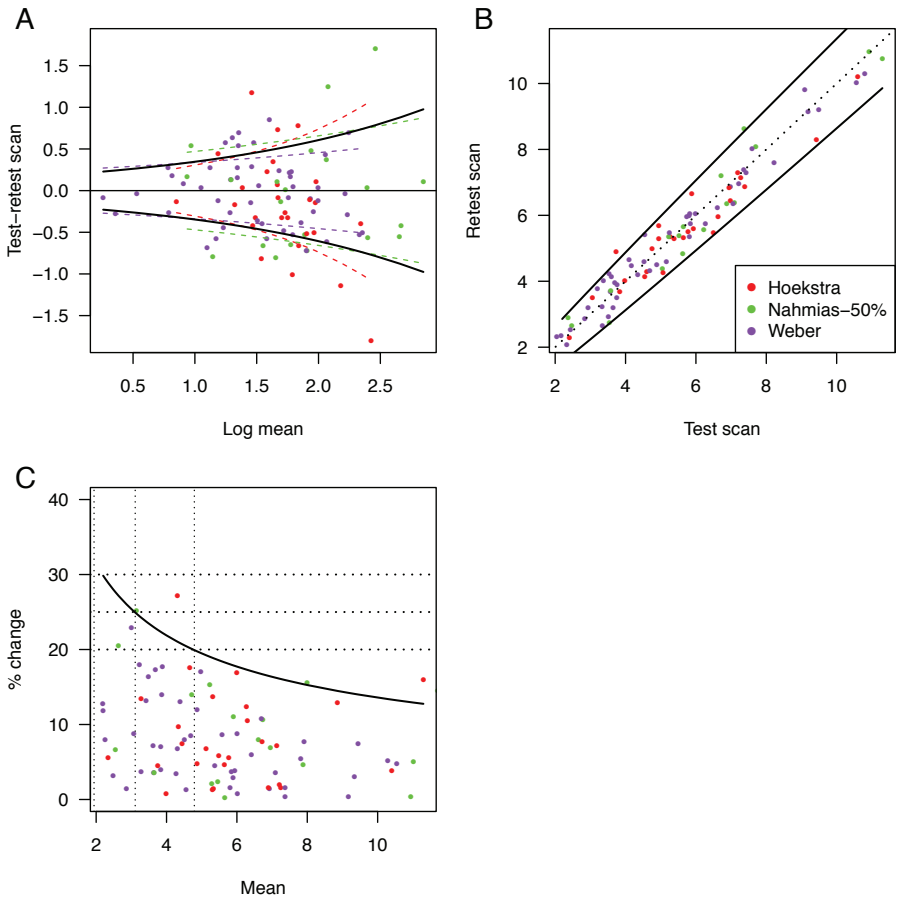


Table 3. Correlations between ^{18}F-FDG uptake and tumor volume and their influence on test-retest variability.

		Correlations	Log likelihood ratio test *p* values			
	Studies	Uptake vs. Vol	Uptake (alone)	Uptake (adj. vol)	Vol (alone)	Vol (adj. mean)
SUV_{max}	H, M, N, V	0.38	< 0.0001	< 0.0001	0.03	0.26
SUV_{mean}	H, M, N, W, V	0.39	< 0.0001	< 0.0001	0.03	0.71
SUV_{mean}	H, N*, W	0.45	0.002	0.003	0.32	0.78

H = Hoekstra et al, N = Nahmias et al (manual analysis), N* = Nahmias et al (50% isocontour analysis),

M = Minn et al, V = Velasquez et al, W = Weber et al.

We found no statistically significant difference between studies using static (Nahmias et al [21] and Velasquez et al [23]) as compared with dynamic PET techniques (Hoekstra et al [18], Minn et al [20] and Weber et al [16]) (p = 0.90). The principal component analysis indicated that the test-retest variability of SUV_{max} is larger than that of the SUV_{mean} (Figures S1 and S2).

Single Assessment Setting

By definition the CR95 for one observation is smaller than the equivalent two observation value due to fewer sources of measurement error. The single assessment CR95s were within 25% change for SUV_{max} and within 20% for SUV_{mean} (Table S1, Figure S2).

Impact of Tumor Volume on Repeatability

Tumor volume was moderately correlated with the ^{18}F-FDG uptake level (Table 3, Figure S3A). However, uptake level was a better indicator for changes in test-retest variability than tumor volume for both parameters (Table 3). Repeatability of SUV_{mean} seems to be relatively unaffected by tumor volume, irrespective of the delineation method (Table 3, Figure S3B).

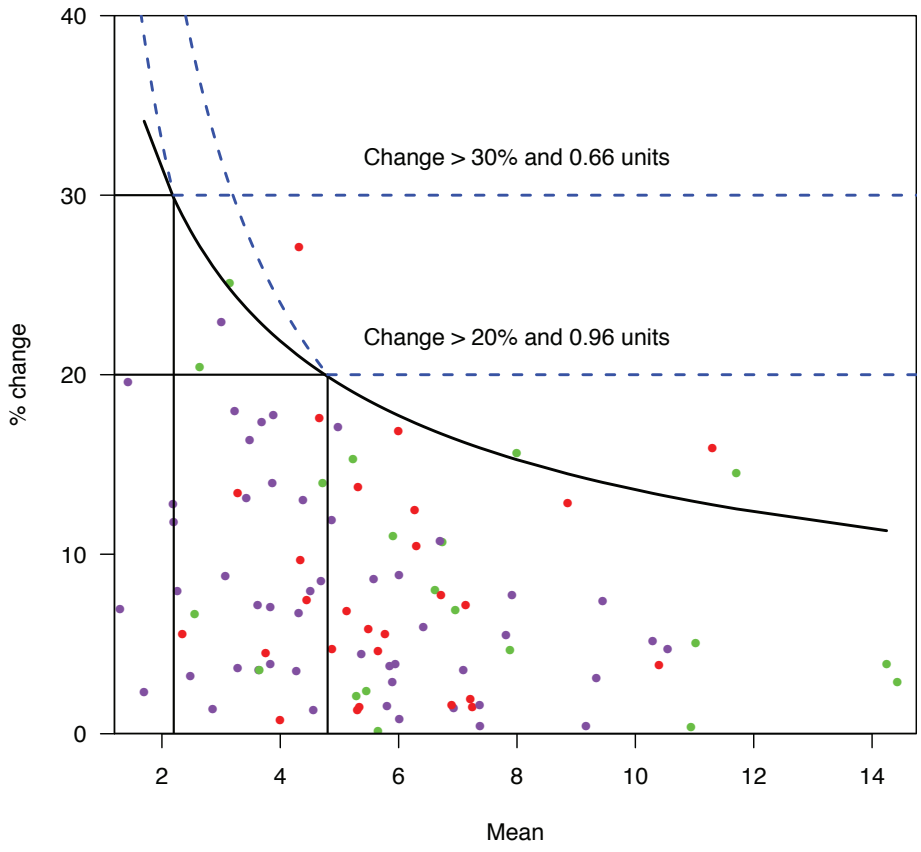
DISCUSSION

This meta-analysis summarizes the published evidence on the repeatability of commonly used quantitative ^{18}F-FDG measurements in oncology. Our results apply to the use of PET in serial as well as in single assessment settings. Compared to SUV_{max}, SUV_{mean} had the better repeatability.

For both measures, the percentage change in test-retest variability is not constant across the range of parameter values and negatively related to the level of ^{18}F-FDG uptake. In a clinical setting, combining a minimal relative and absolute change is sufficient to define a (biological) effect that cannot be explained by measurement error only (see Table 2 for thresholds). In the context of analysing serial

^{18}F-FDG uptake changes, the PERCIST response classification system assumes a biological change with SUV_{peak} changes > 30% in combination with 0.8 unit change of absolute SUV_{peak} [25]. In the present meta-analysis, there was insufficient data to explore SUV_{peak} values. Numerical values of SUV_{peak} typically vary between SUV_{max} and SUV_{mean} (if VOI volumes of SUV_{mean} exceed the SUV_{peak} ones). Figure 4 corroborates the PERCIST assumption of combining 30% with 0.8 unit SUV_{peak} (for SUV_{mean} we found 30% and 0.75). The same Figure shows that, when using SUV_{mean}, a minimal relative change of 20% in combination with 1.2 unit change will also represent a biological change.

Figure 4. Same as Figure 3C, now presenting the cutoff rules at 20% and 30% relative change with associated absolute changes for the three studies with SUV_{mean} measurements assessed using 50%-iso-contouring. The required absolute difference is calculated as the relative difference multiplied by the mean value where the CR95 curve crosses the relative difference value. For example, the CR95 equals 30% at 2.2, hence the associated absolute difference is 0.66 (2.2 x 0.3). Therefore a change of at least 30% and more than 0.66 units will be less than 5% likely to be due to measurement error in ^{18}F-FDG uptake rate. The blue dashed lines represent the areas of combined coverage for each rule.



The first combination is advantageous for tumors with low uptake values where small differences in absolute values correspond to large differences in percentage change. The latter combination is advantageous for high uptake values where relatively large differences in absolute values correspond to relatively small differences in percentage change. It is probable that for tumors with extremely high uptake values, an even smaller relative-change threshold would be appropriate. We are unable to suggest corresponding absolute change values as these uptake values are outside the range of our data.
Alternatively, to select lesions for evaluation with ^{18}F-FDG PET, a minimal SUV could be applied (see figures for trade off values). This simplifies the situation by excluding low uptake values with large (relative change) test-retest variability, leaving the remaining values with a consistent relative test-retest change that can be used as a (one-value) cutoff for response assessment. The disadvantage of this approach is that increases in ^{18}F-FDG uptake (e.g. due to disease progression) can be measured reliably, but decreases to values below the threshold (e.g. due to treatment response) cannot.
Even though methodologies were not fully consistent throughout the studies, we found no important interstudy differences in parameter repeatability. For SUV_{max} no difference was found at all, while for SUV_{mean} this was only present when all studies were included, irrespective of the delineation method used. This difference was caused by a lower performance of the Velasquez data, and good performance of the manually delineated data of Nahmias et al. Velasquez et al [23] applied a multicenter design with a large number of centers. Also, the patient population was substantially different from that of the other studies, because only patients with advanced gastrointestinal malignancies were included by Velasquez et al whereas the other studies predominantly evaluated lung lesions (Table 1). Since the physiologic ^{18}F-FDG uptake by the liver and gastrointestinal tract is higher and more variable than in the lung, this may have challenged tumor delineation. Furthermore, spillover of activity from normal organs with variable ^{18}F-FDG uptake (e.g. kidneys, bowel loops) may have affected the test-retest variability. The authors reported that they were forced to apply a higher threshold of the maximum voxel value (70%) to enable tumor tissue discrimination from the normal background. Earlier work from our center showed that test-retest variability increases with higher thresholds of the maximum tumor volume [19]. Possibly these factors resulted in lower repeatability.
Therefore we performed a subset analysis of homogeneously delineated lesions using the 50% isocontour technique. This inherently excluded the data of Velasquez et al [23], resulting in interstudy homogeneity. Better repeatability was found for this subset as presented in Figure 3 and Table 2. Whether the worse repeatability in the overall dataset is due the delineation technique (70% threshold) or the location of the lesions (abdominal) remains elusive. Therefore the cutoff values found for the 50% threshold subset only hold for 50% isocontoured extra-abdominal lesions, while the overall results can be used for all organ sites and histology and irrespective of the SUV_{mean} tumor delineation method.

Interestingly, the manually delineated data of Nahmias et al performed better than the threshold analysis of the same data set. This was also reported earlier by our group [19]. However, both studies were single-observer, while the strength of the semi-automated technique is the high intra- and inter-observer repeatability for lesion delineation. Therefore, superiority can only be evaluated by performing a head-to-head analysis in a multi-observer setting.

In previous studies it was suggested that repeatability is a function of ^{18}F-FDG uptake value and tumor volume. Weber et al [16] reported that the absolute difference between the test and retest values remains constant across the range of parameter value, but that the relative change increases with decreasing uptake value. Nahmias et al [21] also reported that the absolute difference of the mean uptake value was constant across the range of SUV_{mean} with increasing relative change for lower SUV_{mean} values. However, this group reported that for SUV_{max} the absolute difference increased with increasing SUV_{max} value, while the relative change remained constant.

In the pooled dataset of the current study, mean uptake was a better indicator for changes in test-retest variability than tumor volume (Table 3). SUV repeatability seems to be relatively unaffected by VOI size.

We observed a moderately positive correlation between the uptake value and tumor volume (Table 3, Figure S3A). Hypothetically this can, at least in part, be explained by the “partial volume” phenomenon, which results in an underestimation of ^{18}F-FDG uptake in smaller lesions, which in turn results in more difficult tumor delineation due to lower tumor-to-background contrast [26].

Our results apply to the setting where the same scanner is being used for both the baseline, as well as the post-treatment scan with the use of the same acquisition, reconstruction, and data processing and analytical protocol. These inclusion criteria were mandatory as variation in these parameters can result in parameter differences of up to 50%, only representing noise and not indicative of true biological change [27, 28]. Although interstudy differences in image contrast and resolution remained, their effects on repeatability are negligible [29]. If this is also true for newer systems with better spatial resolution cannot be forecasted based on the current data.

The study by Kamibayashi [24] was excluded because of the use of a different PET scanner for the test and retest scans. In this study, tumor SUV test-retest variability was evaluated in patients that were scanned once on a PET only scanner and once on a PET-CT scanner without treatment intervention between the two scans. The scanners were from the same vendor and the same acquisition and analytical protocol was used for both scans. The authors applied a manually drawn 2D ROI for tumor delineation. The standard deviation was only slightly worse than that of the other singlecenter repeatability studies and comparable to the multicenter study of Velasquez et al [23] with a SD of 12 ± 10.2% for SUV_{mean} and 16.1 ± 10.5% for SUV_{max}. This might indicate that different scanners can be used at baseline

and post-treatment, provided that standardisation of acquisition, reconstruction, data processing and analysis is applied. If this also holds for semi-automatic VOI definition and the use of PET scanners from different vendors remains to be further examined.

Although no statistical interstudy difference was present after pooling the data and correcting for VOI definition heterogeneity, the cutoff values for definition of test-retest variability should be assessed with care because some methodological and patient spectrum heterogeneity between studies prevailed.

However, we believe that this meta-analysis provides the most critical and robust view at ^{18}F-FDG PET repeatability in the oncological setting to date. It enabled the evaluation of dependency on mean uptake value and tumor volume, issues that were open for discussion since the publication of Weber's study in 1999 [16]. Ideally, these results should be confirmed prospectively, preferably in a large multicenter study. Possibly, studies such as the ACRIN (American College of Radiology Imaging Network) study 6678 - *FDG-PET/CT as a predictive marker of tumor response and patient outcome: Prospective Validation in Non-Small Cell Lung Carcinoma* -, incorporating a test-retest study evaluating ^{18}F-FDG PET-CT with a static protocol at 60 min. post-injection, can provide the opportunity for this.

CONCLUSION

This meta-analysis shows the repeatability of different ^{18}F-FDG uptake measurements, based on the available data. SUV_{mean} performed better than SUV_{max}. Importantly, both parameters showed worse repeatability for lesions with low ^{18}F-FDG uptake. This can be accounted for by combining relative and absolute differences. For SUV_{mean} a 30% and 0.75 unit change or 20% and 1.2 unit change exceeds 95% test-retest variability. Homogeneous delineation and exclusion of abdominal lesions improved repeatability. For SUV_{max} the equivalent values are 30% and 2 units change. These cutoff values can serve as a guide for future clinical trials. Given the limited data and as some interstudy heterogeneity prevailed, our results should preferably be confirmed in a prospective repeatability study, preferentially including SUV_{peak}. For serial (e.g. baseline and post-intervention scan) and single PET measurements, different thresholds should be used since each scan is affected by test-retest noise.

ACKNOWLEDGMENTS

The authors are grateful to dr. C. Nahmias and dr. L.M. Wahl for performing reanalyses of their published study results.

SUPPLEMENTAL FIGURES AND TABLES

Figure S1. A comparison of published study-specific first principal components (dashed lines) with the first principal component calculated from pooled data (black solid line).

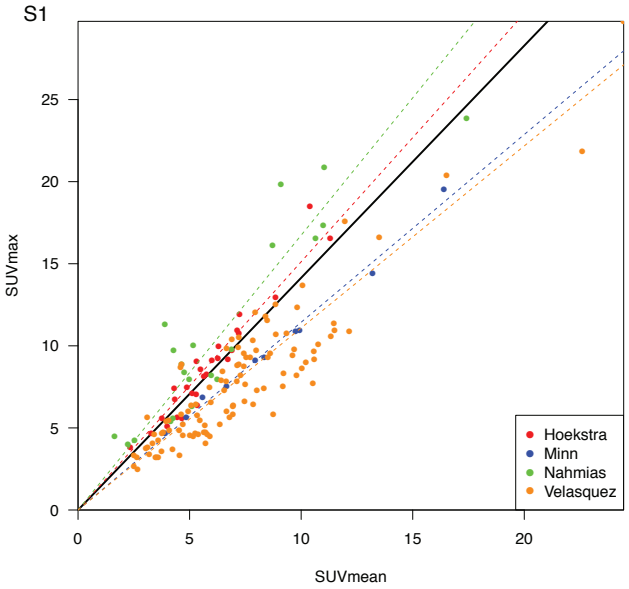


Figure S2. A comparison of CR95s (i.e. mean-variance relations) for SUV_{max} (blue), SUV_{mean} (all data; red) and SUV_{mean} (50% isocontour; orange). The scales are matched via the first principal component (Figure S1). Solid lines represent the test-retest CR95s, while the dotted lines are the one observation CR95s (i.e. variance divided by √2).

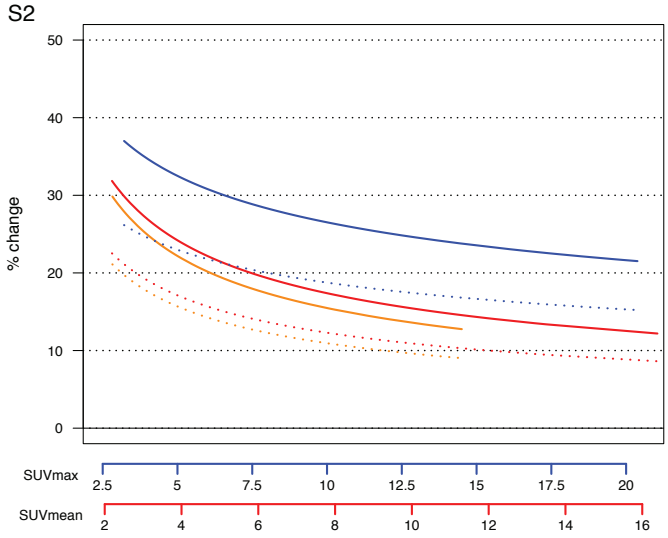


Figure S3. SUV_{mean} repeatability with respect to tumor volume. (A) The log of the tumor uptake plotted against the log of the tumor volume. (B) The estimated published study-specific standard deviations (dashed lines) and the model fit with published study as a random factor (solid line).

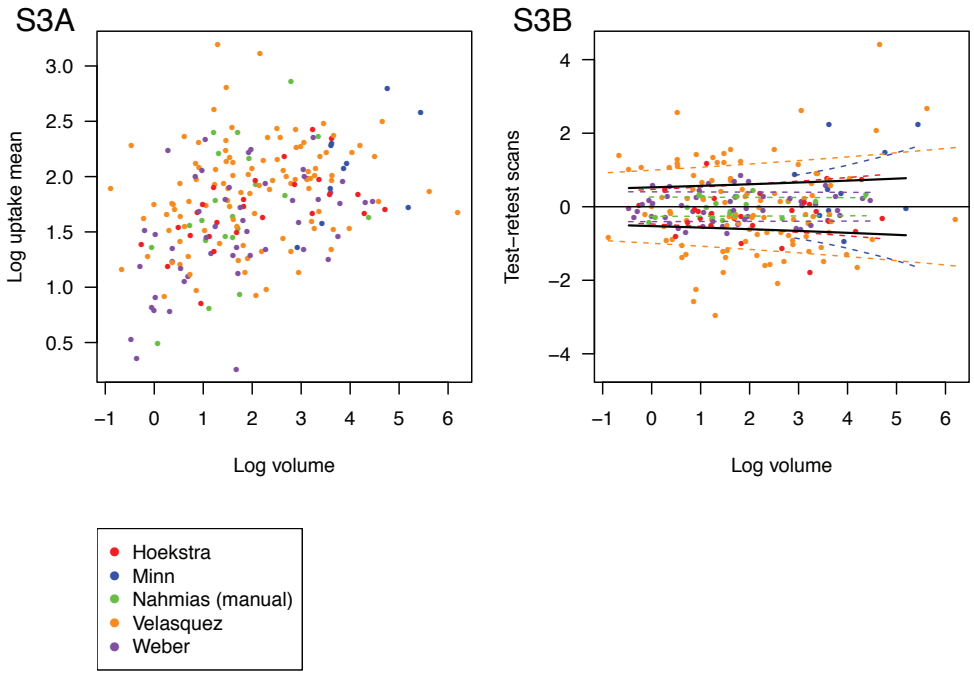


Table S1. Required relative and absolute differences to exceed test-retest variability.*

	Relative difference	Absolute difference
SUV_{max}	15%	3.2
	20%	1.6
	25%	0.9
SUV_{mean} H, N, M, W, V	10%	1.21
	15%	0.77
	20%	0.55
SUV_{mean} H, N†, W	10%	0.93
	15%	0.62
	20%	0.48

* Listed are the relative and absolute differences required for a difference to have less than 5% probability of being due to measurement error, when comparing a single PET observation to a fixed value. Differences have to exceed both the relative and absolute thresholds.

REFERENCE LIST

1. Avril N, Sassen S, Roylance R. Response to therapy in breast cancer. J Nucl Med. 2009;50:55S-63S.
2. de Geus-Oei LF, Vriens D, van Laarhoven HW, van der Graaf WT, Oyen WJ. Monitoring and predicting response to therapy with 18F-FDG PET in colorectal cancer: a systematic review. J Nucl Med. 2009;50:43S-54S.
3. Hicks RJ. Role of 18F-FDG PET in assessment of response in non-small cell lung cancer. J Nucl Med. 2009;50:31S-42S.
4. Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. J Nucl Med. 2009;50:21S-30S.
5. Krause BJ, Herrmann K, Wieder H, zum Buschenfelde CM. 18F-FDG PET and 18F-FDG PET/CT for assessing response to therapy in esophageal cancer. J Nucl Med. 2009;50:89S-96S.
6. Schoder H, Fury M, Lee N, Kraus D. PET monitoring of therapy response in head and neck squamous cell carcinoma. J Nucl Med. 2009;50:74S-88S.
7. Martoni AA, Zamagni C, Quercia S, et al. Early (18)F-2-fluoro-2-deoxy-d-glucose positron emission tomography may identify a subset of patients with estrogen receptor-positive breast cancer who will not respond optimally to preoperative chemotherapy. Cancer. 2010;116:805-813.
8. McLarty K, Fasih A, Scollard DA, et al. 18F-FDG small-animal PET/CT differentiates trastuzumab-responsive from unresponsive human breast cancer xenografts in athymic mice. J Nucl Med. 2009;50:1848-1856.
9. Prior JO, Montemurro M, Orcurto MV, et al. Early prediction of response to sunitinib after imatinib failure by 18F-fluorodeoxyglucose positron emission tomography in patients with gastrointestinal stromal tumor. J Clin Oncol. 2009;27:439-445.
10. Storto G, De RA, Pellegrino T, et al. Assessment of metabolic response to radioimmunotherapy with 90Y-ibritumomab tiuxetan in patients with relapsed or refractory B-cell non-Hodgkin lymphoma. Radiology. 2010;254:245-252.
11. Herrmann K, Krause BJ, Bundschuh RA, Dechow T, Schwaiger M. Monitoring response to therapeutic interventions in patients with cancer. Semin Nucl Med. 2009;39:210-232.
12. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-247.
13. Juweid ME, Stroobants S, Hoekstra OS, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. J Clin Oncol. 2007;25:571-578.
14. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. Eur J Cancer. 1999;35:1773-1782.
15. Nevill AM, Copas JB. Using generalized linear models (GLMs) to model errors in motor performance. J Mot Behav. 1991;23:241-250.
16. Weber WA, Ziegler SI, Thodtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med. 1999;40:1771-1777.
17. Daly LE, Bourke GJ. Interpretation and uses of medical statistics. Oxford, U.K.: Blackwell Science; 2000.
18. Hoekstra CJ, Hoekstra OS, Stroobants SG, et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. J Nucl Med. 2002;43:1304-1309.
19. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging. 2005;32:294-301.
20. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. Radiology. 1995;196:167-173.
21. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. J Nucl Med. 2008;49:1804-1808.
22. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F]fluoro-D-glucose. Mol Imaging Biol. 2002;4:171-178.
23. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. J Nucl Med. 2009;50:1646-1654.
24. Kamibayashi T, Tsuchida T, Demura Y, et al. Reproducibility of semi-quantitative parameters in FDG-PET using two different PET scanners: influence of attenuation correction method and examination interval. Mol Imaging Biol. 2008;10:162-166.
25. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50:122S-150S.
26. de Langen AJ, van den Boogaart VE, Marcus JT, Lubberink M. Use of H2(15)O-PET and DCE-MRI to measure tumor blood flow. Oncologist. 2008;13:631-644.
27. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med. 2004;45:1519-1527.

28. Westerterp M, Pruim J, Oyen W, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. Eur J Nucl Med Mol Imaging. 2007;34:392-404.
29. Cheebsumon P, van Velden FH, Yaqub M, et al. Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment. J Nucl Med. 2011;52:1550-1558.

5

Reproducibility of quantitative ^{18}F-3'-deoxy-3'-fluorothymidine measurements using positron emission tomography

Adrianus J. de Langen
Bianca Klabbers
Mark Lubberink
Ronald Boellaard
Marieke D. Spreeuwenberg
Ben J. Slotman
Remco de Bree
Egbert F. Smit
Otto S. Hoekstra
Adriaan A. Lammertsma

A. J. de Langen . E. F. Smit, Department of Respiratory Medicine, VU University Medical Center, Amsterdam, The Netherlands; M. Lubberink . R. Boellaard . O. S. Hoekstra . A. A. Lammertsma, Department of Nuclear Medicine & PET Research, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands, e-mail: aa.lammertsma@vumc.nl; M. D. Spreeuwenberg, Department of Clinical Epidemiology & Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; B. Klabbers . B. J. Slotman, Department of Radiotherapy, VU University Medical Center, Amsterdam, The Netherlands; R. de Bree, Department of Otolaryngology/Head and Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands

Eur J Nucl Med Mol Imaging. 2009; 36(3): 389-395

ABSTRACT

Purpose

Positron emission tomography (PET) using ^{18}F-3'-deoxy-3'-fluorothymidine ^{18}F-FLT allows noninvasive monitoring of tumor proliferation. For serial imaging in individual patients, good reproducibility is essential. The purpose of the present study was to evaluate the reproducibility of quantitative ^{18}F-FLT measurements.

Methods

Nine patients with non-small-cell lung cancer (NSCLC) and six with head-and-neck cancer (HNC) underwent ^{18}F-FLT PET twice within 7 days prior to therapy. The maximum pixel value (SUV_{max}) and a threshold defined volume ($SUV_{41\%}$) were defined for all delineated lesions. The plasma-to-tumor transfer constant (K_i) was estimated using both Patlak graphical analysis and nonlinear regression (NLR). NLR was also used to estimate k_3, which, at least in theory, selectively reflects thymidine kinase 1 activity. The level of agreement between test and retest values was assessed using the intraclass correlation coefficient (ICC) and Bland-Altman analysis.

Results

All primary tumors and > 90% of clinically suspected locoregional metastases could be delineated. In total, 24 lesions were defined. NLR-derived K_i, Patlak-derived K_i, $SUV_{41\%}$ and SUV_{max} showed excellent reproducibility with ICCs of 0.92, 0.95, 0.98 and 0.93, and SDs of 16%, 12%, 7% and 11%, respectively. Reproducibility was poor for k_3 with an ICC of 0.43 and SD of 38%.

Conclusion

Quantitative ^{18}F-FLT measurements are reproducible in both NSCLC and HNC patients. When monitoring response in individual patients, changes of more than 15% in $SUV_{41\%}$, 20–25% in SUV_{max} and Patlak-derived K_i, and 32% in NLR3k-derived K_i are likely to represent treatment effects.

INTRODUCTION

^{18}F-3'-deoxy-3'-fluorothymidine ^{18}F-FLT has been proposed as a positron emission tomography (PET) tracer of proliferation. FLT is a substrate for thymidine kinase 1 (TK1), which is a key enzyme in the salvage pathway of thymidine DNA synthesis. Several studies have shown good correlations between ^{18}F-FLT uptake and other markers of cellular proliferation, including proliferating cell nuclear antigen, flow cytometry and Ki-67 nuclear staining [1–5].

In most tumors ^{18}F-FLT uptake is lower than ^{18}F-FDG uptake [6]. In addition, its biodistribution and metabolic profile are not ideal [7]. Consequently, ^{18}F-FDG remains the method of choice for diagnosing and staging of tumors. It has been suggested that ^{18}F-FLT may have additional value in combination with ^{18}F-FDG because the combination has been reported to result in fewer false-positive findings [8, 9], but results have not been consistent [2, 10, 11].

Based on its TK1-related signal, ^{18}F-FLT PET has been proposed as a biomarker for predicting (early) response to systemic or locoregional treatment [12–14]. Recent studies have shown that ^{18}F-FLT PET can predict response to therapy as early as 1 week after treatment [15], and a decrease in ^{18}F-FLT uptake seems to be correlated with prolonged overall survival [16, 17].

In addition, ^{18}F-FLT might also be a specific tracer for monitoring the effects of agents targeting thymidylate synthase (TS), an enzyme that plays a central role in the de novo pathway of DNA synthesis. Downregulation of this pathway results in an upregulation of the salvage pathway and thus of ^{18}F-FLT uptake. Recent results have indicated that anti-TS effects can be monitored as soon as 2 hours after administration [18].

To evaluate response in individual patients, reproducibility of the parameter of interest needs to be known. The purpose of the present study was to determine reproducibility of quantitative ^{18}F-FLT measurements in a prospective study of patients with non-small-cell lung cancer (NSCLC) or head-and-neck cancer (HNC).

MATERIALS AND METHODS

Eligible patients were included prospectively after providing written informed consent in accordance with institutional review board approval. In total, nine patients with NSCLC and six with HNC were scanned twice within 7 days (mean 1.9 days, median 1 day) prior to any therapy using an ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, TN). This scanner has an axial field of view (FOV) of 15 cm, divided into 63 contiguous planes. Each patient was positioned supine on the scanner bed with the primary tumor in the centre of the axial FOV. The primary tumor and all other evaluable lesions in the FOV were analysed. A venous catheter was placed in all patients, which was used for injection of ^{18}F-FLT and venous blood sampling [19]. In addition, in HNC patients an arterial catheter was placed in

the radial artery for arterial blood sampling. Each acquisition started with a 10 to 15-min transmission scan in order to acquire a fixed number of 85 million counts [20], which was used for attenuation correction of the subsequent emission scan. This was followed by a bolus injection of 370 MBq ^{18}F-FLT in 5 mL saline through an injector (Medrad International, Maastricht, The Netherlands) at 0.8 mL/s, after which the line was flushed with 42 mL saline (2.0 mL/s). Simultaneously with the injection of ^{18}F-FLT, a dynamic emission scan (in 2-D acquisition mode) was started with a total duration of 60 min and with variable frame lengths (6 x 5 s, 6 x 10 s, 3 x 20 s, 5 x 30 s, 5 x 60 s, 8 x 150 s, and 6 x 300 s). All dynamic scan data were corrected for dead time, decay, scatter, randoms and photon attenuation, and were reconstructed as 128 x 128 matrices using filtered back projection (FBP) with a Hanning filter (cutoff, 0.5 cycles per pixel). This resulted in a transaxial spatial resolution of around 7 mm full-width at half-maximum (FWHM). As FBP reconstructed images suffer from streak artefacts, volume of interest (VOI) definition may be inaccurate, especially in lesions with low tumor-to-background contrast [21, 22]. Therefore, for VOI definition purposes only, the last three frames of the sinograms (45–60 min after injection) were summed and reconstructed using ordered-subset expectation maximization (OSEM) with two iterations and 16 subsets followed by postsmoothing of the reconstructed images using a 5-mm FWHM gaussian filter to obtain the same resolution as for the FBP images [21].

In the HNC patients, arterial ^{18}F concentrations were monitored continuously using a fully automated blood sampling device (Veenstra Instruments, Joure, The Netherlands) [23]. The withdrawal rate was 5 mL/min during the first 10 min and 2.5 mL/min thereafter. In all patients, venous blood samples were drawn at 5, 10, 20, 30, 40 and 60 min after injection to correct for plasma/whole blood ratios and plasma metabolite fractions, and as a quality control procedure for defining an image-derived input function (IDIF) for NSCLC patients as described for ^{18}F-FDG [24]. To avoid contamination, 3–5 mL blood was withdrawn prior to each sample and the line was flushed with 1.5 mL saline after sampling, as described previously [19].

Venous blood samples were analysed using solid-phase extraction chromatography for separation of ^{18}F-FLT from ^{18}F-FLT-glucuronide. For this procedure 0.3 mL plasma was dissolved in 2 mL water. This solution was placed onto a SepPak Vac 6cc (1 g) C18 cartridge (Waters Corporation, Milford, MA). The eluate was collected, after which the cartridge was rinsed with 5 mL water to collect the polar metabolites, being primarily ^{18}F-FLT-glucuronide. The cartridge was then rinsed with 5 mL 96% ethanol to collect the parent compound. All fractions and the cartridge were counted using a Wallac 1480 Wizard well counter (Perkin-Elmer Life Science, Zaventem, Belgium), and the percentage parent within each plasma sample calculated. Metabolite data were interpolated by fitting to an exponential function.

Image analysis

The maximum pixel value within the tumor and a threshold defined volume (41% of the maximum pixel value with correction for local background) were defined semiautomatically for any lesion with adequate focal uptake [25, 26]. Boundaries of lesions without adequate tumor-to-background contrast (low focal uptake and/or high level of background) are difficult to define and, consequently, delineation is hampered by increased observer variation. Therefore, only lesions which required no or only minor manual delineation were included. In the latter case background values were set to zero for voxels directly adjacent to the VOI that had a physiologically high radiotracer concentration (hypervascular areas, haematopoietic bone marrow and liver). After this procedure the threshold technique was applied, resulting in volumes specifically containing tumor. Tumor VOIs were defined on OSEM reconstructed images and transferred to FBP reconstructed dynamic images, thereby generating time-activity curves (TACs). An input function was obtained by continuous arterial blood sampling in HNC patients and derived from the dynamic FBP images in NSCLC patients. IDIFs were defined by multiple manually drawn 2-D regions of interest (ROIs) over the aortic arch (elliptical ROI, 8 × 30 mm) and ascending aorta (spherical ROI, 15 mm) [27]. Again, ROIs were then projected onto all frames to generate input TACs.

Data analysis

Full kinetic analysis to derive values of the individual rate constants and K_i, the plasma-to-tumor transfer constant, was performed for threshold defined VOIs using the following methods: (1) Patlak graphical analysis (time interval 10–60 min after injection; K_i only) [28] and (2) nonlinear regression (NLR) using the standard two-tissue compartment model with both three (NLR3k) and four (NLR4k) rate constants and an additional blood volume parameter (V_b). In both cases, the metabolite-corrected plasma curve was used as input function. The presence of a fourth rate constant (representing dephosphorylation of phosphorylated ^{18}F-FLT) and the need to include this in the model were assessed by comparing fits with and without a k_4 parameter using Akaike [29] and Schwarz [30] criteria.

Kinetic analysis of regional tissue TACs was performed using dedicated software developed within Matlab (The Mathworks, Natick, MA). In general, fits are rejected when the (fitted) standard error of any parameter exceeds 25%. This was, however, never the case in the present study. In addition, the goodness of fit was checked visually for all TACs. SUV was derived for both the maximum pixel value (SUV_{max}) and the threshold defined VOI ($SUV_{41\%}$) for the time intervals 40–60 min and 50–60 min after injection (with multiple correction factors [31]). The level of agreement between test and retest values was assessed using the intraclass correlation coefficient (ICC) with a two-way random model with absolute agreement, and Bland-Altman analysis [32]. In the latter case, the percentage difference

in values (Δ) between two measures was plotted against (1) the mean of both measures and (2) the mean threshold defined volume. In this way, possible dependency on both absolute ^{18}F-FLT uptake and tumor size could be evaluated. Additionally, dependency was statistically analysed using linear regression. Finally, a one-sample *t* test was applied to the Δ values to assess systematic bias and the parameter values of both test and retest scans were compared using the Wilcoxon signed ranks test to evaluate the distribution.

RESULTS

All primary tumors and more than 90% of clinically suspected locoregional metastases could be delineated. In total, 24 lesions (15 NSCLC and 9 HNC) were defined. A small subset of lesions (5/24) required manual intervention during delineation (Figure 1). Full kinetic analysis was possible for 23 lesions; blood analyses failed in one patient. The median lesion size (threshold defined PET volume) was 8.2 cm^3 (range 1.7–86.1 cm^3), the median NLR3k-derived K_i was 0.047 $mL \cdot min^{-1} \cdot mL^{-1}$ (range 0.021–0.120 $mL \cdot min^{-1} \cdot mL^{-1}$), the median Patlak-derived K_i was 0.037 $mL \cdot min^{-1} \cdot mL^{-1}$ (range 0.017–0.074 $mL \cdot min^{-1} \cdot mL^{-1}$), the median $SUV_{41\%}$ was 3.3 (range 1.3–6.4) and the median SUV_{max} was 4.8 (range 2.4–9.3). The 3k model provided better fits than the 4k model in 26 (57%) and 28 (61%) of 46 measured lesions according to Akaike and Schwarz criteria, respectively. In other words, the majority of the data were best fitted with the 3k model. Visual check of the TACs revealed good fits for all lesions.

Descriptive statistics for all parameters of ^{18}F-FLT kinetics for both test and retest scans, as well as *p* values for the Wilcoxon signed ranks test, are presented in Table 1. No systematic bias (one-sample *t* test: $SUV_{41\%}$, $p = 0.98$; Patlak K_i, $p = 0.25$), or a significant difference in the distribution of paired differences (Wilcoxon signed ranks test; $p > 0.08$ for all) was found. The threshold VOI definition technique resulted in consistent volumes for most tumors throughout both scans. The median change in tumor volume between test and retest scans was 10%. Only the delineation of one lesion resulted in an ex-

Figure 1. Two coronal ^{18}F-FLT images of tumor lesions that required manual intervention: **a**. lesion in right upper lobe close to costal bone, **b**. large intrapulmonary lesion with inferior part close to right diaphragm.

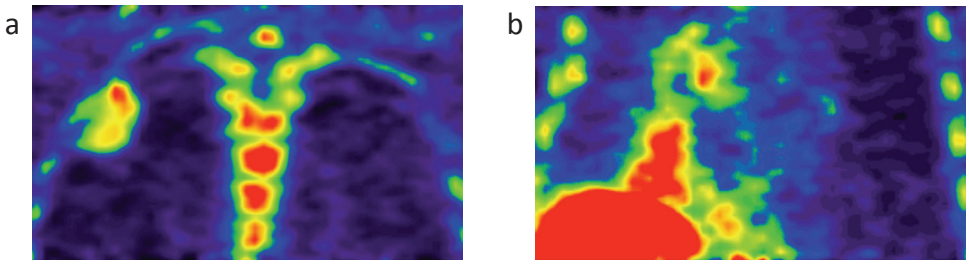


Table 1. Descriptive statistics of mean and median values, standard deviation and range for all parameters of ^{18}F-FLT kinetics of both scans.

	Test scan					Retest scan					
Parameter	Mean	Median	SD	Range	No of Measurements	Mean	Median	SD	Range	No of Measurements	*p* value (Wilcoxon signed ranks test)
NLR3k K_i	0.048	0.042	0.021	0.022 – 0.120	24	0.047	0.042	0.018	0.021 – 0.101	23	0.73
NLR3k K_1	0.177	0.139	0.128	0.076 – 0.608	24	0.159	0.142	0.076	0.087 – 0.393	23	0.86
NLR3k k_2	0.118	0.109	0.067	0.030 – 0.360	24	0.102	0.087	0.052	0.041 – 0.257	23	0.65
NLR3k k_3	0.047	0.044	0.019	0.017 – 0.104	24	0.043	0.042	0.015	0.020 – 0.079	23	0.33
NLR3k V_b	0.153	0.109	0.143	0.013 – 0.608	24	0.123	0.114	0.100	0.001 – 0.394	23	0.17
Patlak K_i[a]	0.038	0.036	0.013	0.019 – 0.072	24	0.039	0.037	0.015	0.017 – 0.074	23	0.32
$SUV_{41\%}$[b]	3.35	3.28	1.11	1.42 – 6.42	24	3.35	3.20	1.12	1.34 – 6.23	24	0.79
SUV_{max}[b]	5.17	5.14	1.63	2.66 – 9.32	24	5.04	4.78	1.66	2.40 – 9.22	24	0.17
VOI size[c]	16.5	8.4	20.1	1.6 – 82.4	24	18.5	9.4	21.8	1.8 – 89.8	24	0.08

a Patlak analysis was performed using the time interval 10–60 min after injection.

b SUV values presented were corrected for body weight and measured between 40 and 60 min after injection.

c VOI size is tumor volume in cubic centimetres, defined by a 41% threshold technique as described in the text.

ceptional difference in volume of 150% between the two scans. This was probably due to inaccurate definition of the maximum pixel value, possibly caused by patient motion in one or both scans.

Only very small variations were found between different $SUV_{41\%}$ normalization methods (body weight, body surface area and lean body mass) and time intervals (40–60 min and 50–60 min after injection) with ICCs ranging from 0.97 to 0.98. Therefore, the SUV results are presented for only one correction method (body weight) and a single time interval (40–60 min after injection). NLR3k-derived K_i, Patlak-derived K_i, $SUV_{41\%}$ and SUV_{max} showed excellent reproducibility with ICCs of 0.92, 0.95, 0.98 and 0.93, and SDs of 16%, 12%, 7% and 11%, respectively (Table 2). In contrast, NLR3k-derived k_3, which, at least in theory, selectively reflects TK1 activity, showed poor reproducibility with an ICC of 0.43 and an SD of 38%. As expected from Akaike and Schwarz analyses, inclusion of a fourth rate constant resulted in decreased reproducibility with an ICC of 0.75 and an SD of 25% for NLR4k-derived K_i, and an ICC of 0.33 and an SD of 70% for NLR4k-derived k_3 (Table 2).

These data show that, when monitoring therapy effects, changes of less than 32% in NLR3k-derived K_i, 24% in Patlak-derived K_i, 14% in $SUV_{41\%}$ and 22% in SUV_{max} (< 1.96 × SD) cannot be distinguished from normal test–retest variability.

Table 2. Reproducibility results of simplified and full kinetic analyses of ^{18}F-FLT uptake.

	Patients	ICC (95% CI)	SD % change	SD abs. change
NLR3k K_i	All	0.92 (0.83–0.97)	16%	0.008
NLR3k k_3	All	0.43 (0.30–0.71)	38%	0.019
NLR4k K_i	All	0.75 (0.49–0.88)	25%	0.016
NLR4k k_3	All	0.33 (-0.95–0.65)	70%	0.073
Patlak K_i	All	0.95 (0.89–0.98)	12%	0.004
$SUV_{41\%}$	All	0.98 (0.95–0.99)	7%	0.2
SUV_{max}	All	0.93 (0.85–0.97)	11%	0.6
NLR3k K_i	NSCLC	0.92 (0.78–0.98)	11%	0.004
Patlak K_i	NSCLC	0.92 (0.75–0.97)	12%	0.004
NLR3k K_i	HNC	0.92 (0.72–0.98)	22%	0.011
Patlak K_i	HNC	0.96 (0.85–0.99)	12%	0.005

Subgroup analysis for NSCLC and HNC patients showed similar results for Patlak-derived K_i and an even better SD result for NLR3k-derived K_i in NSCLC patients as compared to HNC patients, suggesting no error due to the use of an IDIF (Table 2).

Bland-Altman plots of $SUV_{41\%}$ and Patlak-derived K_i showed no dependency on absolute ^{18}F-FLT uptake, but a trend of reduced reproducibility for smaller lesions (Figure 2). This trend, however, was not statistically significant ($SUV_{41\%}$, $p = 0.35$; Patlak K_i, $p = 0.51$).

DISCUSSION

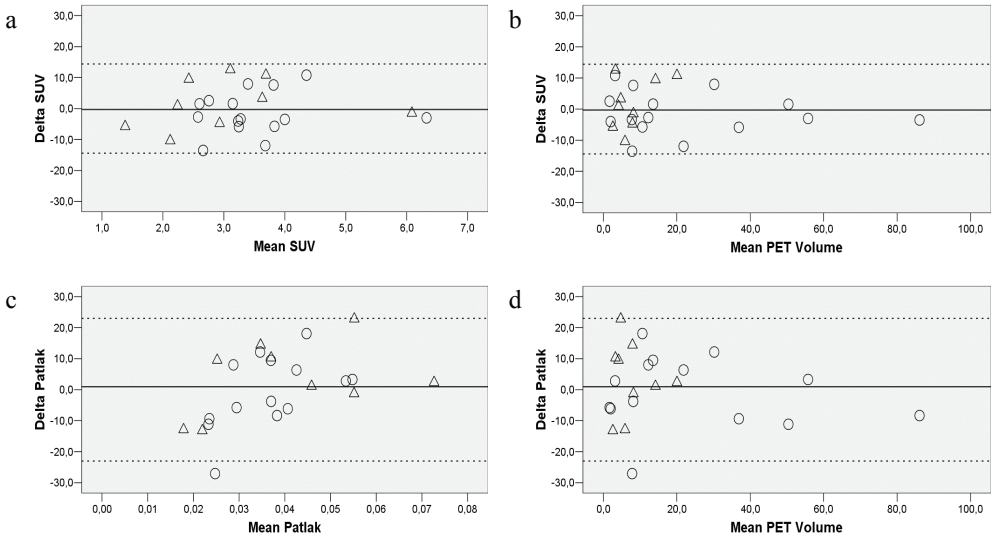
Our results show that quantitative ^{18}F-FLT measurements are reproducible in both NSCLC and HNC patients. When serial measurements in patients are performed, changes of more than 15% in $SUV_{41\%}$, 20–25% in Patlak-derived K_i and SUV_{max}, and 30–35% in NLR3k-derived K_i are likely to represent biological effects. In addition, no significant dependency between absolute ^{18}F-FLT uptake and lesion size was found, implying that the same threshold can be used for all tumor lesions. Although one tumor lesion showed a change of 150% in tumor volume between the two scans, the threshold technique resulted in reproducible tumor volumes for all other lesions.

Akaike and Schwarz analyses showed that for most tumor lesions data were best fitted with the 3k model, indicating that ^{18}F-FLT phophorylation was irreversible, at least within the time frame of the measurements. Therefore, the fourth rate constant should be set to zero. In a recent study it was

shown that significant dephosphorylation starts at ± 50 min after injection and that k_4 can reliably be estimated with a scan time of 120 min [33]. Furthermore, a strong correlation was found between NLR3k K_i derived from a 60-min scan and NLR4k K_i derived from a 120-min scan [33]. A scan protocol of 120 min, however, is too long for routine (serial) clinical studies, especially in critically ill patients. In addition, the risk of movement artefacts increases with longer scan times. Our results indicate that with a scan time of 60 min, a fourth rate constant can be neglected.

To the best of our knowledge, reproducibility of ^{18}F-FLT measurements in malignant tumors has only been assessed in eight patients (nine lesions) with breast cancer [15] and in six patients with NSCLC [34]. Shields et al [34] evaluated an unknown number of lesions in six patients using SUV_{max} and SUV_{mean} (50% threshold technique) and carried out full kinetic analyses (NLR with unknown number of rate constants and Patlak) with the use of an IDIF. Unfortunately, full details of their study (e.g. the number of rate constants in the NLR analyses, location and number of lesions, whether mean or median for absolute difference of test and retest scans) were not reported. Nevertheless, reproducibility results were comparable with the present results (around 20% test–retest variability in SUV and kinetic parameters).

Figure 2. Bland-Altman plots for $SUV_{41\%}$ and Patlak K_i (*circles* NSCLC lesions, *triangles* NHC lesions, *delta values* percentage change between test and retest scans, $SUV_{41\%}$ and SUV_{max} for the time interval 40–60 min after injection and corrected for body weight, *Patlak* Patlak-derived K_i, *mean PET* volume mean threshold defined volume of the test and retest scan in cm^3, *straight lines* mean values, *dotted lines* ± 1.96 x SD).



Results in breast cancer [15] are also consistent with the present results with reported ICCs of 0.99 and 0.97, and SDs of 11% and 15% for SUV and Patlak K_i, respectively. Differences in study design and methods, however, make it difficult to perform a direct comparison. The absence of IDIF analyses, the relatively low number of lesions included with no locoregional (lymph node) metastases, and the very high tumor-to-background ratio in the breast (due to low uptake of ^{18}F-FLT in fatty tissue) [35, 36] imply that results for breast tumors do not necessarily hold for tumors of other origins (with lower tumor-to-background ratio). In the present study IDIF was used for NSCLC lesions, thereby greatly simplifying the acquisition protocol for thoracic tumors. Clearly, this is much more convenient for patients and better suited for routine clinical use. Subgroup analyses for NLR3k and Patlak-derived K_i did not reveal inferior results for NSCLC as compared to HNC, where arterial blood sampling was performed, suggesting that the use of an IDIF does not significantly contribute to test–retest variability.

In the present analyses, locoregional lymph nodes were also included, and thereby all possible scenarios encountered in clinical trials were evaluated. In addition, this provided a means for evaluating possible dependency on absolute ^{18}F-FLT value and tumor size.

SUV_{max} is the most frequently used (semi)quantitative measure in clinical PET studies due to its ease of definition and low observer variation. Nevertheless, results indicate that better reproducibility can be achieved by using threshold defined $SUV_{41\%}$. Although pulmonary lesions are not easy to delineate due to physiologically high uptake in surrounding mediastinum, liver, vertebrae and costal bones, reproducibility was not reduced, indicating that the applied threshold technique is reliable.

Since radiotracer uptake depends on perfusion and extraction, SUV and K_i are not specific measurements of TK1 activity. ^{18}F-FLT uptake mainly depends on the extracellular ATP concentration, which has greatest impact on the turnover of ^{18}F-FLT due to induction of a change in structure of TK1 from a dimer to a tetramer, which has about a 20-fold greater effect on the turnover of the tracer [37]. However, previous data have shown that ^{18}F-FLT uptake may also depend on perfusion (blood flow) rather than on TK1 activity [38]. Therefore, at least in theory, k_3 appears to be the parameter of interest for studying TK1 activity. Unfortunately, NLR3k-derived k_3 showed poor reproducibility. However, both K_i and SUV have shown good correlations with other proliferation markers, with K_i having the strongest correlation [4, 15, 33]. The present results indicate that both K_i and SUV are reproducible.

Since K_i has the best correlation with other proliferation markers, dynamic scanning should, whenever possible, be applied in response monitoring studies. If static scans are acquired, however, the present results support the use of a threshold defined VOI instead of the maximum pixel value. In the present study only therapy-naive tumors were analysed. It should be kept in mind, however, that an intervention may affect the various rate constants in a different manner. For example, a certain therapy could result in a reduction in delivery (K_1) with only a minor change in TK1 activity (k_3). In that case, both K_i

and SUV would decrease more than the actual decrease in proliferation rate. In addition, SUV and K_i might not be similarly affected.

CONCLUSION

Our results show that both simplified and full kinetic analyses of ^{18}F-FLT data have excellent reproducibility in NSCLC and HNC patients. Furthermore, the data support the use of an IDIF for kinetic analyses of intrathoracic lesions. Future response monitoring studies, correlating ^{18}F-FLT response with pathological and clinical outcome, should be performed to assess which ^{18}F-FLT parameter is best for predicting response to therapy.

REFERENCE LIST

1. Yamamoto Y, Nishiyama Y, Ishikawa S, Nakano J, Chang SS, Bandoh S, et al. Correlation of 18F-FLT and 18F-FDG uptake on PET with Ki-67 immunohistochemistry in non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2007;34:1610–6.
2. Troost EG, Vogel WV, Merkx MA, Slootweg PJ, Marres HA, Peeters WJ, et al. 18F-FLT PET does not discriminate between reactive and metastatic lymph nodes in primary head and neck cancer patients. J Nucl Med 2007;48:726–35.
3. Kenny LM, Vigushin DM, Al-Nahhas A, Osman S, Luthra SK, Shousha S, et al. Quantification of cellular proliferation in tumor and normal tissues of patients with breast cancer by [18F]fluorothymidine-positron emission tomography imaging: evaluation of analytical methods. Cancer Res 2005;65:10104–12.
4. Vesselle H, Grierson J, Muzi M, Pugsley JM, Schmidt RA, Rabinowitz P, et al. In vivo validation of 3'deoxy-3'-[(18) F]fluorothymidine ([(18)F]FLT) as a proliferation imaging tracer in humans: correlation of [(18)F]FLT uptake by positron emission tomography with Ki-67 immunohistochemistry and flow cytometry in human lung tumors. Clin Cancer Res 2002;8:3315–23.
5. Barthel H, Cleij MC, Collingridge DR, Hutchinson OC, Osman S, He Q, et al. 3'-deoxy-3'-[18F]fluorothymidine as a new marker for monitoring tumor response to antiproliferative therapy in vivo with positron emission tomography. Cancer Res 2003;63:3791–8.
6. Salskov A, Tammisetti VS, Grierson J, Vesselle H. FLT: measuring tumor cell proliferation in vivo with positron emission tomography and 3'-deoxy-3'-[18F]fluorothymidine. Semin Nucl Med 2007;37:429–39.
7. Dittmann H, Dohmen BM, Paulsen F, Eichhorn K, Eschmann SM, Horger M, et al. [18F]FLT PET for diagnosis and staging of thoracic tumours. Eur J Nucl Med Mol Imaging 2003;30:1407–12.
8. Yamamoto Y, Nishiyama Y, Kimura N, Ishikawa S, Okuda M, Bandoh S, et al. Comparison of (18)F-FLT PET and (18)F-FDG PET for preoperative staging in non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2008;35:236–45.
9. Buck AK, Halter G, Schirrmeister H, Kotzerke J, Wurziger I, Glatting G, et al. Imaging proliferation in lung tumors with PET: 18F-FLT versus 18F-FDG. J Nucl Med 2003;44:1426–31.
10. Cobben DC, Elsinga PH, Hoekstra HJ, Suurmeijer AJ, VaalburgW, Maas B, et al. Is 18F-3'-fluoro-3'-deoxy-L-thymidine useful for the staging and restaging of non-small cell lung cancer? J Nucl Med 2004;45:1677–82.
11. Shields AF. PET imaging with 18F-FLT and thymidine analogs: promise and pitfalls. J Nucl Med 2003;44:1432–4.
12. Buck AK, Kratochwil C, Glatting G, Juweid M, Bommer M, Tepsic D, et al. Early assessment of therapy response in malignant lymphoma with the thymidine analogue [18F]FLT. Eur J Nucl Med Mol Imaging 2007;34:1775–82.
13. Herrmann K,Wieder HA, Buck AK, Schöffel M, Krause BJ, Fend F, et al. Early response assessment using 3'-deoxy-3'-[18F]fluorothymidine-positron emission tomography in high-grade non-Hodgkin's lymphoma. Clin Cancer Res 2007;13:3552–8.
14. Apisarnthanarax S, Alauddin MM, Mourtada F, Ariga H, Raju U, Mawlawi O, et al. Early detection of chemoradioresponse in esophageal carcinoma by 3'-deoxy-3'-3H-fluorothymidine using preclinical tumor models. Clin Cancer Res 2006;12:4590–7.
15. Kenny L, Coombes RC, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3'-deoxy-3'-[18F]fluorothymidine positron emission tomography. Eur J Nucl Med Mol Imaging 2007;34:1339–47.
16. Chen W, Delaloye S, Silverman DH, Geist C, Czernin J, Sayre J, et al. Predicting treatment response of malignant gliomas t to bevacizumab and irinotecan by imaging proliferation with [18F]fluorothymidine positron emission tomography: a pilot study. J Clin Oncol 2007;25:4714–21.
17. Kasper B, Egerer G, Gronkowski M, Haufe S, Lehnert T, Eisenhut M, et al. Functional diagnosis of residual lymphomas after radiochemotherapy with positron emission tomography comparing FDG- and FLT-PET. Leuk Lymphoma 2007;48:746–53.
18. Perumal M, Pillai RG, Barthel H, Leyton J, Latigo JR, Forster M, et al. Redistribution of nucleoside transporters to the cell membrane provides a novel approach for imaging thymidylate synthase inhibition by positron emission tomography. Cancer Res 2006;66:8558–64.
19. Hoekstra CJ, Hoekstra OS, Lammertsma AA. On the use of the injection catheter for venous blood sampling in quantitative FDG PET studies. Eur J Nucl Med 2000;27:1579.
20. Boellaard R, van Lingen A, van Balen SCM, Lammertsma AA. Optimization of attenuation correction for positron emission tomography studies of thorax and pelvis using count-based transmission scans. Phys Med Biol 2004;49:N31–8.
21. Boellaard R, van Lingen A, Lammertsma AA. Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. J Nucl Med 2001;42:808–17.
22. Riddell C, Carson RE, Carrasquillo JA, Libutti SK, Danforth DN, Whatley M, et al. Noise reduction in oncology FDG PET images by iterative reconstruction: a quantitative assessment. J Nucl Med 2001;42:1316–23.
23. Boellaard R, van Lingen A, van Balen SC, Hoving BG, Lammertsma AA. Characteristics of a new fully programmable blood sampling device for monitoring blood radioactivity during PET. Eur J Nucl Med 2001;28:81–9.
24. Hoekstra CJ, Hoekstra OS, Lammertsma AA. On the use of image-derived input functions in oncological fluorine-18 fluorodeoxyglucose positron emission tomography studies. Eur J Nucl Med 1999;26:1489–92.

25. Boellaard R, Oyen WJ, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardization and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008; [Epub ahead of print]. doi:10.1007/s00259-008-0954-3
26. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer 1997;80:2505–9.
27. van der Weerdt AP, Klein LJ, Boellaard R, Visser CA, Visser FC, Lammertsma AA. Image-derived input functions for determination of MRGlu in cardiac (18)F-FDG PET scans. J Nucl Med 2001;42:1622–9.
28. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J Cereb Blood Flow Metab 1983;3:1–7.
29. Akaike H. A new look at the statistical identification. IEEE Trans Automat Contr 1978;19:716–23.
30. Schwarz G. Estimating the dimension of a model. Ann Statist 1978;6:461–4.
31. Hoekstra CJ, Paglianiti I, Hoekstra OS, Smit EF, Postmus PE, Teule GJ, et al. Monitoring response to therapy in cancer using [18F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. Eur J Nucl Med 2000;27:731–43.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
33. Muzi M, Vesselle H, Grierson JR, Mankoff DA, Schmidt RA, Peterson L, et al. Kinetic analysis of 3'-deoxy-3'-fluorothymidine PET studies: validation studies in patients with lung cancer. J Nucl Med 2005;46:274–82.
34. Shields AF, Lawhorn-Crews JM, Briston DA, Douglas KA, Mangner TJ, Muzik O. The reproducibility of FLT PET in patients with untreated non-small cell lung cancer. J Nucl Med 2005;46:426P.
35. Smyczek-Gargya B, Fersis N, Dittmann H, Vogel U, Reischl G, Machulla HJ, et al. PET with [18F]fluorothymidine for imaging of primary breast cancer: a pilot study. Eur J Nucl Med Mol Imaging 2004;31:720–4.
36. Been LB, Elsinga PH, de Vries J, Cobben DC, Jager PL, Hoekstra HJ, et al. Positron emission tomography in patients with breast cancer using (18)F-3'-deoxy-3'-fluoro-l-thymidine ((18)F-FLT) – a pilot study. Eur J Surg Oncol 2006;32:39–43.
37. Dimitrakopoulou-Strauss A, Strauss LG. The role of 18F-FLT in cancer imaging: does it really reflect proliferation? Eur J Nucl Med Mol Imaging 2008;35:523–6.
38. Muzi M, Spence AM, O'Sullivan F, Mankoff DA, Wells JM, Grierson JR, et al. Kinetic analysis of 3'-deoxy-3'-18F-fluorothymidine in patients with gliomas. J Nucl Med 2006;47:1612–21.

6

First-line erlotinib and bevacizumab in patients with locally advanced and/or metastatic non-small-cell lung cancer: a phase II study including molecular imaging

A.-M. C. Dingemans[1,2]*
A. J. de Langen[3]
V. van den Boogaart[1,2]
J. T. Marcus[4]
W. H. Backes[5]
H. T. G. M. Scholtens[6]
H. van Tinteren[7]
O. S. Hoekstra[8]
J. Pruim[9]
B. Brans[2,10]
F. B. Thunnissen[11]
E. F. Smit[3]
H. J. M. Groen[6]

[1]Department of Pulmonary Diseases, Maastricht University Medical Center, Maastricht; [2]GROW—School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht; Departments of [3]Pulmonary Diseases; [4]Physics and Medical Technology, VU University Medical Center, Amsterdam; [5]Department of Radiology, Maastricht University Medical Center, Maastricht; [6]Department of Pulmonary Diseases, University Medical Center Groningen, University of Groningen, Groningen; [7]Department of Medical Statistics, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam; [8]Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam; [9]Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Groningen; [10]Department of Nuclear Medicine, Maastricht University Medical Center, Maastricht; [11]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands

Ann Oncol. 2011; 22(3): 559-566

ABSTRACT

Purpose

Both bevacizumab and erlotinib have clinical activity in non-small-cell lung cancer (NSCLC). Preclinical data suggest synergistic activity.

Patients and methods

Chemonaive patients with stage IIIb or IV non-squamous NSCLC were treated with bevacizumab 15 mg/kg every 3 weeks and erlotinib 150 mg daily until progression. Primary end point was nonprogression rate (NPR) at 6 weeks. Tumor response was measured with computed tomography, 2-[fluorine-18] fluoro-2-deoxy-D-glucose (^{18}F-FDG PET) and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). KRAS and EGFR mutations were assessed in tumor samples.

Results

Forty-seven patients were included. Median follow-up was 15.2 months. NPR at 6 weeks was 75%. Median progression-free survival (PFS) was 3.8 [95% confidence interval (CI) 2.3–5.4] months and median overall survival (OS) was 6.9 (95% CI 5.5–8.4) months. Toxicity was mainly mild. The presence of KRAS (n = 10) or EGFR mutations (n = 5) did not influence outcome. After 3 weeks of treatment, > 20% decrease in standard uptake value as measured with positron emission tomography predicted for longer PFS (9.7 versus 2.8 months; p = 0.01) and > 40% decrease in K^{trans} as assessed by DCE-MRI did not predict for longer PFS.

Conclusions

First-line treatment with bevacizumab and erlotinib in stage IIIb/IV NSCLC resulted in an NPR of 75%. OS was however disappointing. Early response evaluation with ^{18}F-FDG PET is the best predictive test for PFS.

INTRODUCTION

The prognosis for patients with advanced non-small-cell lung cancer (NSCLC) remains poor. Platinum-based combination chemotherapy has reached an efficacy plateau from which it will not ascend by its own [1]. For several reasons, the combination of bevacizumab and erlotinib is of interest in NSCLC. First, preclinical and clinical data indicate that antivascular endothelial growth factor (VEGF) and anti-epidermal growth factor receptor (EGFR) therapies have at least an additive antitumor effect [2, 3]. Secondly, the combination has proven to be well tolerated even when both are administered at their recommended phase II dose [4]. Finally, both erlotinib as single agent and bevacizumab in combination with cytotoxic chemotherapy are of proven benefit in the clinical management of advanced NSCLC [4–6].

Objective response rates (ORRs), as determined by RECIST, obtained in phase II studies define whether a drug is moved toward phase III testing [7]. In contrast to traditional anticancer agents, the effects of these agents, especially bevacizumab, are cytostatic rather than cytotoxic [8–10]. Therefore, determining the activity of both agents by anatomical criteria alone probably underestimates their activity. The 2-[fluorine-18]fluoro-2-deoxy-D-glucose (^{18}F-FDG PET) studies may find evidence of anti-neoplastic activity within a week after initiation of treatment before changes in tumor volume can be detected [11]. Effects of drugs on tumor (micro)vasculature may be determined by dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) [12, 13].

The aim of this study was to evaluate the efficacy of erlotinib and bevacizumab in first-line treatment of advanced NSCLC as determined by the non-progression rate (NPR) at 6 weeks. Secondary end points included ORR, safety, overall survival (OS) and progression-free survival (PFS) and the relation of molecular imaging-derived parameters with PFS and response.

PATIENTS AND METHODS

Patient Eligibility

Key eligibility criteria were stage IIIB (pleural effusion) or IV non-squamous NSCLC; no prior antitumor therapy; measurable disease (RECIST); Eastern Cooperative Oncology Group performance status (PS) of two or less and adequate hematologic, renal and hepatic function.

Key exclusion criteria were evidence of tumor invading major blood vessels; presence of a cavitating lesion; major surgical procedure, open biopsy, significant injury or radiotherapy within 28 days; serious non-healing wound or ulcer; active peptic ulcer disease; evidence of bleeding diathesis, coagulopathy or history of ≥ grade 2 hemoptysis; proteinuria ≥ 2+; brain metastasis or spinal cord compression unless previously treated with evidence of stable disease (SD) for > 2 months; treatment with anticoagu-

lants for therapeutic purposes; ongoing treatment with aspirin (> 325 mg/day) or other medications known to predispose to gastrointestinal ulceration and any uneven systemic disease.
Written informed consent was obtained from all patients before inclusion.

Study Design and Treatment

The study is a non-randomized phase II study, carried out at three university medical centers in The Netherlands. The study was approved by the institutional medical ethics review board of each participating center.
Patients were treated with bevacizumab 15 mg/kg as an i.v. infusion every 3 weeks and erlotinib 150 mg orally daily. No dose reductions of bevacizumab were allowed. Dose reductions of erlotinib were according to the label. Patients remained on treatment until disease progression, unacceptable toxicity and/or patient refusal. In case of documented tumor progression, patients received further treatment as per investigator's decision.

Study assessments

Toxicity was scored every 3 weeks during treatment according to the common toxicity criteria adverse events version 3.0. Tumor response was measured with computed tomography (CT) (RECIST) every 6 weeks until disease progression. Objective responses were confirmed after at least 4 weeks.

Exploratory end points

Imaging with CT, DCE-MRI and ^{18}F-FDG PET was carried out at baseline and after 3 weeks of treatment (just before bevacizumab infusion). Size was measured with CT and response defined by RECIST 1.0, tumor metabolism (SUV) with ^{18}F-FDG PET and a combined measure of perfusion and permeability (the endothelial transfer constant, K^{trans}) with DCE-MRI. Dynamic ^{18}F-FDG PET studies were carried out selectively at VU University Medical Center, whereas the other centers applied a static whole body ^{18}F-FDG PET protocol. (For imaging acquisition: see supplementary material available in Annals of Oncology online.)

Tissue biomarkers

Mutation analysis of EGFR (exon 19-21) and KRAS (exon 1) was carried out on extracted tumor DNA. Formalin-fixed and paraffin-embedded tumor tissue sections were manually microdissected, and total genomic DNA was isolated using QIAamp DNA extraction kits (Qiagen, Venlo, The Netherlands). Nested polymerase chain reaction sequencing was carried out using the BigDye Terminator *v* 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), and the ABI PRISM 310 Genetic Analyzer (Applied Biosystems) [14].

Statistical Considerations and Analytical Plan

The primary end point of the study was NPR (RECIST) at 6 weeks. This end point was used previously and it was shown that it is predictive for clinical benefit [15, 16]. To be able to discontinue the trial early if treatment showed insufficient activity, a two-stage design was implemented (Simon's optimal design; p0 = 40%, p1 = 60%, alpha = 0.05, beta = 0.20). A total of 46 patients would be entered and the treatment was declared to have sufficient activity to deserve further evaluation if at least 24 patients did not progress. A first analysis was made after 16 patients were included; if nine or more patients of them progressed within 6 weeks, the study would be stopped. Patients without post baseline tumor assessments were regarded as progressive. The analysis for the primary efficacy end points was based upon all registered subjects who received at least one dose of study treatment.

PFS was defined as the time from the start of the treatment to the date of the first documented progression in terms of RECIST or the date of death. OS was determined from the date of the start of the treatment to the date of death irrespective of the cause of death. Patients who had not progressed or died at the time of the analysis were censored at the date of last contact. The Kaplan–Meier method was used to plot PFS and OS. Predefined cutoff points were used and curves compared by log-rank testing.

Imaging results are presented as percentage change from the baseline value. Predefined cut points for imaging studies were based on reproducibility data. A 20% cut point was used for positron emission tomography (PET) (i.e. values outside 1.96 x standard deviation) [17], while a 40% cut point was used for DCE-MRI-derived K^{trans} [18, 19].

Statistical analyses were carried out using SPSS (version 15.0; SPSS Inc., Chicago, IL). Results are presented with 95% confidence intervals (CIs) and *p* values. A *p* value of < 0.05 was considered to be significant.

RESULTS

Patient Characteristics

From January 2006 till March 2007, 50 patients were registered; 47 patients received at least one dose of study treatment. Three patients were ineligible after inclusion but before start of treatment (one pulmonary embolism, one tumor involvement of large blood vessel, one trachea–esophageal fistula). Patient characteristics are provided in Table 1.

As at the time of analysis, 11 patients were alive, 3 patients were on treatment; 35 patients stopped treatment due to progressive disease (PD), 5 due to toxicity, 2 because of surgical intervention and 2 patients died before progression was estimated. Patients received a median of four cycles of bevacizumab (range 1–21). Median follow-up for all patients was 15.2 months.

Table 1. Patient characteristics: total patient population and patients in imaging study.

Characteristics	No. of patients (n = 47)	No. of patients in imaging study (n = 40)
Age, years		
Median	59	58
Range	34-40	34-78
Sex		
Male/female	23/24	18/22
ECOG performance status		
0	22	19
1	19	17
2	6	4
Stage		
IIIb	11	9
IV	36	31
Histology		
Adeno	31	28
Other	16	12
Smoking status		
Never	8	7
Former	23	21
Current	16	12

Efficacy

Of the first 16 patients, 11 were non-progressive at 6 weeks. Therefore, the study continued for further accrual. For all patients at 6 weeks, 35 of 47 patients (75%) showed no progression. Twelve patients (25%) fulfilled the criteria of an objective response with one complete response (CR). Progression was more frequently observed in patients without rash (50% of patients without rash versus 15% with any grade of rash, Pearson X^2 6.78; p = 0.012); no correlation was found with gender, PS or smoking history. At progression, 27 patients received further chemotherapeutic treatment with a platinum doublet. Twenty patients received palliative care only of which six at their own request (one patient underwent complementary treatment). The major clinical characteristic that determined whether or not patients received further chemotherapy was clinical deterioration due to the progression; nine patients died within 1 month after progression.

Secondary End points

Median PFS was 3.8 (95% CI 2.3–5.4) months. No significant differences were observed between males and females or (former) smokers and never smokers. Neither rash nor development of hypertension was related to PFS, but patients that developed proteinuria had a significant longer PFS (5.7, 95% CI

1.3–10.2, months versus 2.3, 95% CI 0.4–4.2, months; $p = 0.04$). Median OS was 6.9 (95% CI 5.5–8.4) months, with 11 patients censored. No survival difference was observed between females and males, patients with versus without rash, ever smokers versus never smokers and between patients with PS 0–1 or 2. Median duration of tumor response was 10.9 (95% CI 5.0–16.8) months.

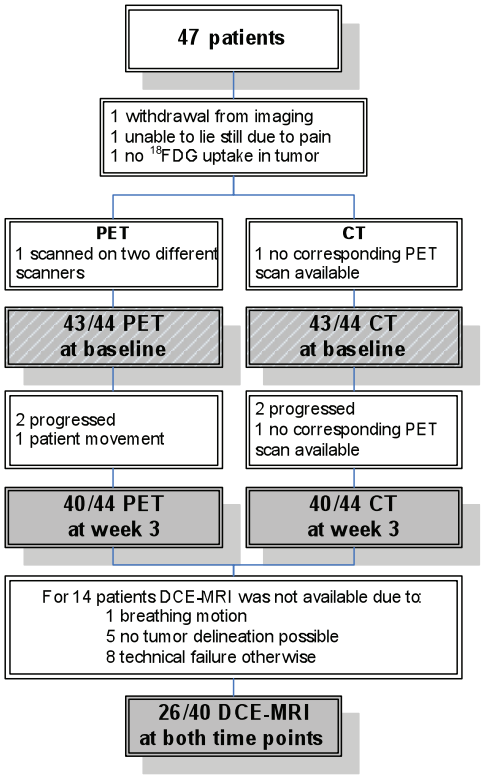
Safety

Toxicity was mainly as expected (Table 2). But two patients discontinued treatment because of toxicity that was not prespecified as per protocol. One patient developed sudden increase of liver enzymes, thrombocytopenic purpura and renal function disturbances, mimicking the renal thrombotic microangiopathy syndrome caused by bevacizumab [20]. Ultimately, the patient died at home at day 16. Autopsy was not carried out. One patient had a sigmoid perforation that was successfully treated by surgery. The resection specimen showed no tumor on microscopic examination. No severe bleeding episodes were observed. Three grade 3/4 thromboembolic events were encountered, one pulmonary embolism and one cerebrovascular accident. The latter patient was on study for 8 months and had as best response SD. The CT scan of the brain at the time of the cerebrovascular accident did not reveal brain metastasis.

Table 2. Related adverse events occurring in > 10% of CTC grade 3/4 patients or having CTC grade 3/4.

Category	Adverse event	All grades	Grade 3/4
Cardiac	Hypertension	8	1
Constitutional	Myalgia	6	-
Dermatology	Dry skin	14	-
	Pruritus	7	-
	Rash-acneiform	29	11
	Rash-desquamation	5	-
Gastrointestinal	Bilirubine	12	2
	Diarrhea	27	2
	GGT	1	1
	GI-perforation	1	1
	Mucositis	8	-
	Nausea	6	-
Hemorrhage	All	13	0
	Nose	6	-
	Pulmonary	2	-
Metabolic/laboratory	Increased ASAT/ALAT	16	3
	Hyperbilirubinemia	12	2
	Increased GGT	1	1
	Increased LDH	1	1
	Proteinuria	23	-
	Thrombocytopenia	1	1
Vascular	Thrombosis	5	3
	Pulmonary embolism	1	1
	Cerebrovascular accident	1	1

Figure 1. Flowchart of all patients. Transparent boxes provide information on scan failure. Results of baseline and week 3 can be found in the shaded white/gray and even gray boxes, respectively.



Exploratory end points

imaging. Patients were included for imaging evaluation when at least CT and ^{18}F-FDG PET were available at baseline and after 3 weeks of treatment. Forty patients were found eligible (Figure 1). The baseline characteristics of the patients selected for imaging evaluation were comparable to the intention-to-treat group as depicted in Table 1. Figure 2 shows an example of response assessment by the three imaging modalities. Nine of the 40 patients had a CR or partial response (PR) as best response according to RECIST. After 3 weeks of treatment, five of these nine patients were classified as responder by CT scan (RECIST) and six of nine by ^{18}F-FDG PET. DCE-MRI was available for 26 of the 40 patients. Six of these 26 patients had a CR or PR as best response according to RECIST, and with DCE-MRI two of six patients were identified at week 3.

At week 3, both SUV and K^{trans} values decreased but this was not significant [mean SUV and K^{trans} decrease 5.1% [interquartile range (IQR) -17% to +12%; $p = 0.19$] and 15.1% (IQR -57% to -3%; $p = 0.19$), respectively].

Patients with a metabolic response (SUV decrease > 20%) at week 3 had longer PFS (9.7, 95% CI 1.8–17.6, months) than those without (2.8, 95% CI 2.0–3.5, months; p = 0.01) (Table 3 and Figure 3). A response as assessed by DCE-MRI at week 3 (K^{trans} decrease > 40%) did not predict for longer PFS (26 patients; 4.3, 95% CI 3.5–5.1, months versus 2.7, 95% CI 0–9.1, months; p = 0.63). Although patients with a response on CT at week 3 had a longer PFS, this was not significant (4.6, 95% CI 1.3–7.9, months versus 2.9, 95% CI 1.3–4.5, months; p = 0.08) (Figure 3).

tissue biomarkers. Sufficient paraffin-embedded tumor material was available for EGFR (exons 19–21) (n = 24) and KRAS mutation (n = 27) analysis. Five EGFR mutations were detected: four in exon 20 and one in exon 21. None of these mutations were known to be predictive for response to erlotinib. In 10 of 27 patients, a KRAS mutation was detected. No differences in ORR, or PFS, were observed between wild-type patients and those with a mutation in either EGFR or KRAS (Table 4).

Figure 2. Representative image of baseline (A–C) and post-treatment (D–F) 2-[fluorine-18]fluoro-2-deoxy-D-glucose positron emission tomography (A, D), dynamic contrast-enhanced magnetic resonance imaging (B, E) and computed tomography (C, F) images showing a decrease in 2-fluoro-2-deoxy-D- glucose uptake, tumor perfusion and size after 3 weeks of treatment.

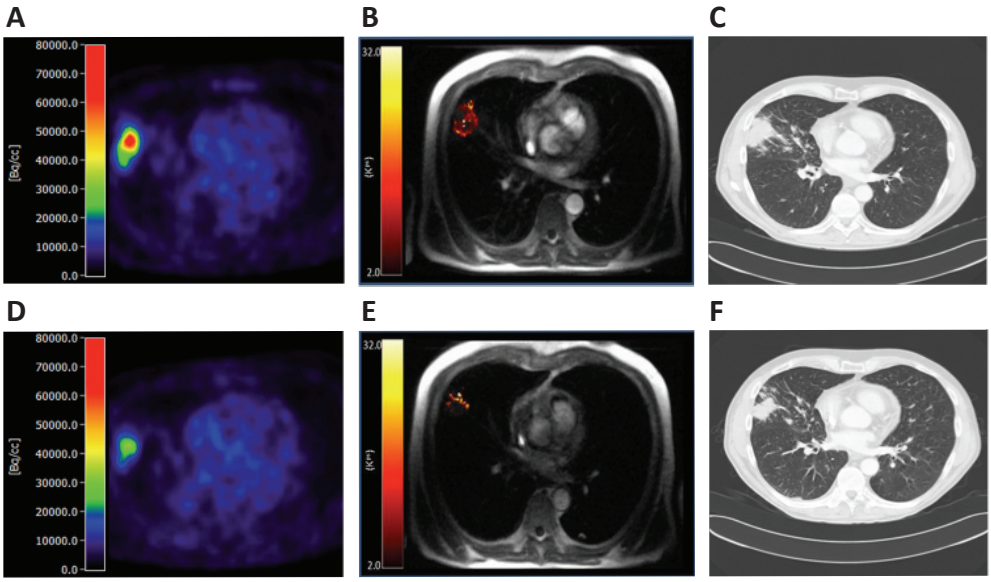


Table 3. Response assessment by ^{18}F-FDG PET and clinical characteristics.

	PET responder	PET non-responder	HR, (95% CI) *p* value
	n	*n*	
	7	33	0.27 (0.01–0.78), $p = 0.002$
Female	4/3	14/19	0.15 (0.02–1.14)
WHO PS 0	3	16	0.41 (0.09–1.84), NS
Adeno/BAC	5	23	0.27 (0.08–0.93), $p = 0.04$
Never smokers	3	4	0.01 (<0.0001–37), NS
CR/PR	6	3	0.58 (0.11–2.93), NS
Rash	7	23	0.28 (0.1–0.084), $p = 0.02$
Hypertension	1	4	0.02 (<0.0001–200), NS
Proteinuria	3	16	0.65 (0.15–2.86), NS

^{18}F-FDG PET, 2-[fluorine-18]fluoro-2-deoxy-D-glucose; PET, positron emission tomography; HR, hazard ratio; CI, confidence interval; WHO, World Health Organization; PS, performance status; NS, not significant; CR, complete response; PR, partial response; BAC, bronchoalveolar carcinoma.

Figure 3. Kaplan–Meier curves [progression-freesurvival (PFS)] stratified for response after 3 weeks of treatment according to computed tomography (CT) (A), 2-[fluorine-18]fluoro-2-deoxy-D-glucose positron emission tomography (B) and dynamic contrast-enhanced magnetic resonance imaging (C).

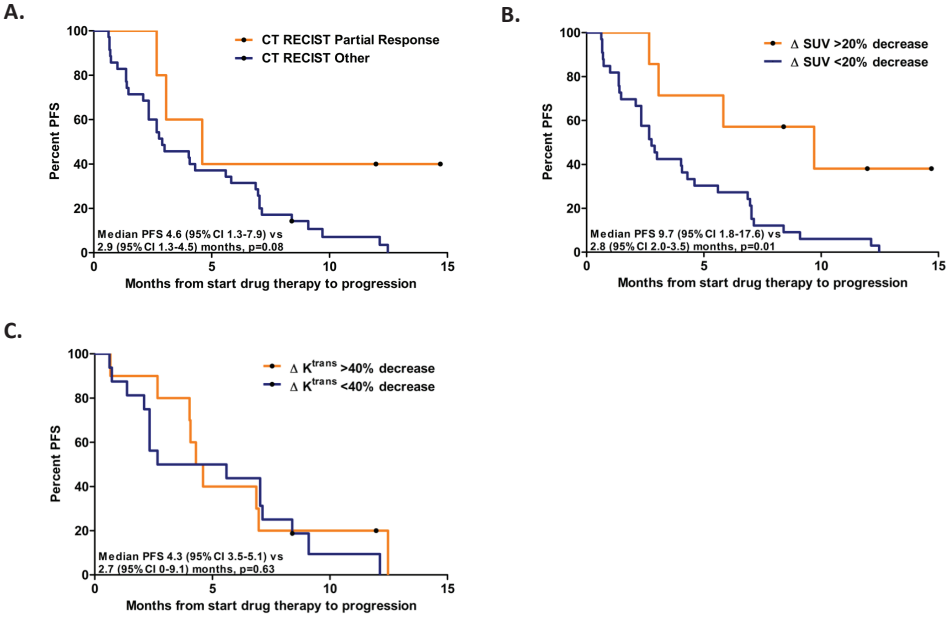


Table 4. Mutation analysis.

	Smoking history (*n*)		Response (*n*)			PFS
	Former/ current	Never	PR	SD	PD	Median, months (95% CI), *p* value
KRAS						
Mutant	9	1	1[a]	6	3	2.3 (2.0–2.6)
Wild type	14	3	3	10	4	4.0 (2.6–5.4), *p* = 0.46
EGFR						
Muntant	5	0	0	4	1	4.0 (1.1–7.0)
Wild type	19	4	4	14	5	3.9 (1.9–5.8), *p* = 0.88

[a]One patient (female, adenocarcinoma, current smoker) with KRAS mutation developed a PR on bevacizumab and erlotinib.
NSCLC, non-small-cell lung cancer; PR, partial response; SD, stable disease; PD, progressive disease; PFS, progression-free survival; CI, confidence interval.

DISCUSSION

This is the first study reporting on the activity of the combination of erlotinib and bevacizumab in chemonaive patients with advanced NSCLC. The rationale for the clinical evaluation of this combination in NSCLC is provided by the antitumor activity of dual blockade of the VEGF and EGFR pathway both in vitro and in vivo. The primary end point, an NPR of 75%, was met. At the time of design of this study, several considerations led to the choice of this primary end point. First, a study investigating the efficacy of single-agent erlotinib employed a similar primary end point and demonstrated the feasibility of our approach [16]. Instead of the NPR at 8 weeks, the NPR at 6 weeks was chosen for the present study because of the three-weekly administration of bevacizumab. Secondly, large cooperative phase III trials investigating different cytotoxic regimen all report a PD rate in the order of 20%, suggesting that an NPR of 75% is a reasonable primary end point for biological treatment [21]. In addition, as these agents are believed to be cytostatic rather than cytotoxic, ORR may be a less suitable end point for phase II studies with these agents [15]. Two studies evaluated erlotinib–bevacizumab in randomized phase II trials in the relapse setting of NSCLC and observed median times to progression of 4.4 and 3.4 months, respectively [22, 23]. These figures are not different from our results and from that obtained in the latest phase III studies in NSCLC in the first-line setting using cytotoxic chemotherapy alone [1, 5, 24]. But recent phase III trials in comparable patient groups as our study, i.e. untreated non-squamous NSCLC, showed increased PFS (5.3–6.7 months) and OS (11.9–13.6 months) when these patients were treated with bevacizumab- or pemetrexed-containing chemotherapy [5, 6, 24].

Although the combination of bevacizumab with erlotinib did significantly improve PFS when compared with erlotinib and placebo in a phase III study in pretreated NSCLC patients, this was not translated in a survival benefit [23]. Recent data show that adding erlotinib to bevacizumab in the maintenance treatment of advanced NSCLC significantly prolongs PFS when compared with placebo [25]. The ORR of 24.5% was not different from that what can be obtained with classical chemotherapy. Disappointingly however, in the study reported here, median OS was 6.9 months only. This disturbing figure calls into question the so-called 'window of opportunity studies' of novel agents in untreated patients with advanced NSCLC. Noteworthy is the fact that only half of the patients received further platinum-based chemotherapy upon progression of disease. This was mainly due to deterioration of PS due to disease progression and patient refusal. Although disappointing, this is comparable to recent maintenance trials in which 50%–60% of patients receive second-line chemotherapy [26–28]. Careful selection of patients, possibly by biomarkers [29] and PS [16], for entry into this type of study is mandatory as are early stopping criteria (e.g. molecular imaging).

None of the 27 patients where tissue was available had 'classical' EGFR mutations that would have predicted for response to erlotinib alone. The relatively high number of exon 20 mutations in this study is likely to be a chance phenomenon as the mutation frequency of >200 samples tested with the same, internationally accepted, method in our laboratory is in line with previous reported data (E. Thunnissen, D. Heideman, unpublished data). In the relapse setting, ORRs of 13%–17% are reported for this combination in unselected patients [22, 23], higher than may be expected with single-agent erlotinib [4]. The Iressa Pan-Asia Study (IPASS) study reported a 1.1% ORR in chemonaive, EGFR mutation-negative, advanced NSCLC patients when treated with single-agent gefitinib [30] and showed that only patients with activating EGFR mutations benefit from first-line treatment with EGFR tyrosine-kinase inhibitor (TKI). All patients in the IPASS study were selected on clinical characteristics and it was clearly shown that when mutation status is not known or when EGFR is wild type, patients should be offered standard chemotherapy instead of first-line EGFR TKI as median PFS was only 2 months for EGFR wild-type patients treated with gefitinib. Therefore, one may assume that dual inhibition of the EGFR and VEGF pathway exerts an antitumor effect in EGFR mutation-negative patients and this strategy may be applicable to a broader population of patients compared with EGFR TKIs alone. Remarkably, one PR was obtained in a patient harboring a KRAS mutation. Objective responses in patients with KRAS-mutated tumors upon treatment with EGFR TKIs alone are rare [31] and this finding illustrates the complexity of the effect size of mutations in different pathways on tumor response. These data suggest that simultaneous inhibition of both the EGFR and the VEGF pathway might overcome primary resistance to EGFR inhibitors. A hypothesis that is supported by recent results obtained in an in vitro model [32].

A phase I/II study of bevacizumab–erlotinib showed that these agents can each be administered at their phase III dose and schedule [33]. Toxicity in our study was mainly as expected, but some serious adverse events occurred. One patient developed a colon perforation during treatment and another developed a thrombotic microangiopathy syndrome. Ten patients experienced grades 3–4 thrombotic events, including three pulmonary embolisms and one cerebro vascular accident. Other than these, the rate of grades 3–4 toxicity was < 5%. These results are comparable with other studies that evaluated the bevacizumab and erlotinib combination in NSCLC [22, 33]. Rash was the predominant form of toxicity and correlated with PFS and OS, although not significantly.

In this study, efforts were made to correlate molecular imaging with clinical outcome. In phase II studies using cytotoxic chemotherapy, response rate as determined by anatomical criteria, i.e. RECIST, is frequently used as the primary end point. Anatomical imaging has been subject to criticism among others by a recent review that showed that < 50% of the difference in survival in 191 phase III studies conducted in NSCLC could be explained by variability in response rate [34]. We and others have shown that clinical outcome to chemotherapy can be predicted early [35] and more robust by using response monitoring consisting of serial ^{18}F-FDG PET scanning [36]. Indeed, in this study we were able to show that response to targeted treatment as assessed by serial ^{18}F-FDG PET scanning early during treatment was significantly associated with PFS, whereas tumor response as assessed by CT was not. In particular, the finding that response assessment by ^{18}F-FDG PET scanning at 3 weeks is predictive for PFS, even in patients with clinical good prognostic characteristics, is of importance as this opens the possibility to halt experimental treatment in the window of opportunity studies in chemonaive NSCLC patients early in order not to jeopardize survival by ineffective treatment. CT, in contrast, showed SD at week 3 for 45% of patients with PR/CR as their best CT response, reflecting the underestimation of treatment effects at early response assessments. Second, three patients with PR at ^{18}F-FDG PET and SD at CT had a median PFS of 9.7 months, suggesting that ^{18}F-FDG PET is more discriminative. Antiangiogenic agents such as bevacizumab are thought to exert their antitumor effects by normalization of tumor vasculature, inhibition of lymphangiogenesis and reduction of intratumoral interstitial pressure, thereby allowing for improved delivery of therapeutics [37]. Several of these consequences of antiangiogenic treatment may be monitored with currently available techniques such as DCE-MRI. This is the first report in lung cancer showing that most tumors have a decrease in K^{trans} value after 3 weeks of treatment, indicating an effect on the tumor vasculature. A similar finding was reported by Willet et al in a study in 12 patients with rectal carcinoma treated with bevacizumab preoperatively [9]. However, in our study a decrease in K^{trans} was not predictive for prolonged PFS.

In conclusion, in this phase II study in chemonaive advanced NSCLC, we observed a 75% NPR at 6 weeks following treatment with a combination of bevacizumab and erlotinib. In contrast, OS was disappointing and some serious toxic effects occurred. Tumor effects with these biologicals are readily measurable with PET.

FUNDING

Roche (Woerden, The Netherlands) for data management and free study drug supply.

ACKNOWLEDGEMENTS

Presented as a poster at the Annual meeting of the American Society of Clinical Oncology 2007, as oral presentation at the World Conference on Lung Cancer 2007 and as oral presentation at the European Conference on Clinical Oncology 2007.

REFERENCE LIST

1. Schiller JH, Harrington D, Belani CP et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 2002; 346: 92–98.
2. Perrotte P, Matsumoto T, Inoue K et al. Anti-epidermal growth factor receptor antibody C225 inhibits angiogenesis in human transitional cell carcinoma growing orthotopically in nude mice. Clin Cancer Res 1999; 5: 257–265.
3. Jung YD, Mansfield PF, Akagi M et al. Effects of combination anti-vascular endothelial growth factor receptor and anti-epidermal growth factor receptor therapies on the growth of gastric cancer in a nude mouse model. Eur J Cancer 2002; 38: 1133–1140.
4. Shepherd FA, Rodrigues Pereira J, Ciuleanu T et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005; 353: 123–132.
5. Sandler A, Gray R, Perry MC et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 2006; 355: 2542–2550.
6. Reck M, von Pawel J, Zatloukal P et al. Phase III trial of cisplatin plus gemcitabine with either placebo or bevacizumab as first-line therapy for nonsquamous non-small-cell lung cancer: AVAil. J Clin Oncol 2009; 27: 1227–1234.
7. O'Connor JP, Jackson A, Asselin MC et al. Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives. Lancet Oncol 2008; 9: 766–776.
8. Parulekar WR, Eisenhauer EA. Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: theory and practice. J Natl Cancer Inst 2004; 96: 990–997.
9. Willett CG, Boucher Y, di Tomaso E et al. Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer. Nat Med 2004; 10: 145–147.
10. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. J Clin Oncol 2004; 22: 4442–4445.
11. Sunaga N, Oriuchi N, Kaira K et al. Usefulness of FDG-PET for early prediction of the response to gefitinib in non-small cell lung cancer. Lung Cancer 2008; 59: 203–210.
12. de Lussanet QG, Backes WH, Griffioen AW et al. Gadopentetate dimeglumine versus ultrasmall superparamagnetic iron oxide for dynamic contrast-enhanced MR imaging of tumor angiogenesis in human colon carcinoma in mice. Radiology 2003; 229: 429–438.
13. Hoekstra CJ, Stroobants SG, Hoekstra OS et al. Measurement of perfusion in stage IIIA-N2 non-small cell lung cancer using H(2)(15)O and positron emission tomography. Clin Cancer Res 2002; 8: 2109–2115.
14. Janmaat ML, Rodriguez JA, Gallegos-Ruiz M et al. Enhanced cytotoxicity induced by gefitinib and specific inhibitors of the Ras or phosphatidyl inositol-3 kinase pathways in non-small cell lung cancer cells. Int J Cancer 2006; 118: 209–214.
15. Lara PN Jr, Redman MW, Kelly K et al. Disease control rate at 8 weeks predicts clinical benefit in advanced non-small-cell lung cancer: results from Southwest Oncology Group randomized trials. J Clin Oncol 2008; 26: 463–467.
16. Giaccone G, Gallegos Ruiz M, Le Chevalier T et al. Erlotinib for frontline treatment of advanced non-small cell lung cancer: a phase II study. Clin Cancer Res 2006; 12: 6049–6055.
17. Hoekstra CJ, Hoekstra OS, Stroobants SG et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. J Nucl Med 2002; 43: 1304–1309.
18. Thomas AL, Morgan B, Horsfield MA et al. Phase I study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of PTK787/ZK 222584 administered twice daily in patients with advanced cancer. J Clin Oncol 2005; 23: 4162–4171.
19. Morgan B, Thomas AL, Drevs J et al. Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ ZK 222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase I studies. J Clin Oncol 2003; 21: 3955–3964.
20. Eremina V, Jefferson JA, Kowalewska J et al. VEGF inhibition and renal thrombotic microangiopathy. N Engl J Med 2008; 358: 1129–1136.
21. Smit EF, van Meerbeeck JP, Lianes P et al. Three-arm randomized study of two cisplatin-based regimens and paclitaxel plus gemcitabine in advanced nonsmall-cell lung cancer: a phase III trial of the European Organization for Research and Treatment of Cancer Lung Cancer Group–EORTC 08975. J Clin Oncol 2003; 21: 3909–3917.
22. Herbst RS, O'Neill VJ, Fehrenbacher L et al. Phase II study of efficacy and safety of bevacizumab in combination with chemotherapy or erlotinib compared with chemotherapy alone for treatment of recurrent or refractory non small-cell lung cancer. J Clin Oncol 2007; 25: 4743–4750.
23. Hainsworth J, Herbst RS. A phase III, multicenter, placebo-controlled, doubleblind, randomized, clinical trial to evaluate the efficacy of bevacizumab (Avastin) in combination with erlotinib (Tarceva) compared with erlotinib alone for treatment of advanced non-small cell lung cancer (NSCLC) after failure of standard first line chemotherapy (BETA). J Thorac Oncol 2008; 3: S302.
24. Scagliotti GV, Parikh P, von Pawel J et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 2008; 26: 3543–3551.

25. Miller VA, O'Connor P, Soh C et al. A randomized, double-blind, placebocontrolled, phase IIIb trial (ATLAS) comparing bevacizumab (B) therapy with or without erlotinib (E) after completion of chemotherapy with B for first-line treatment of locally advanced, recurrent, or metastatic non-small cell lung cancer (NSCLC). J Clin Oncol (Meeting Abstracts) 2009; 27: LBA8002.
26. Fidias PM, Dakhil SR, Lyss AP et al. Phase III study of immediate compared with delayed docetaxel after front-line therapy with gemcitabine plus carboplatin in advanced non-small-cell lung cancer. J Clin Oncol 2009; 27: 591–598.
27. Ciuleanu T, Brodowicz T, Zielinski C et al. Maintenance pemetrexed plus best supportive care versus placebo plus best supportive care for non-small-cell lung cancer: a randomised, double-blind, phase 3 study. Lancet 2009; 374: 1432–1440.
28. Cappuzzo F, Ciuleanu T, Stelmakh L et al. Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomised, placebo-controlled phase 3 study. Lancet Oncol. 2010; 6: 521-529
29. Bunn PA Jr, Dziadziuszko R, Varella-Garcia M et al. Biological markers for non-small cell lung cancer patient selection for epidermal growth factor receptor tyrosine kinase inhibitor therapy. Clin Cancer Res 2006; 12: 3652–3656.
30. Mok TS, Wu YL, Thongprasert S et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361: 947–957.
31. Jackman DM, Sequist LV, Cioffredi L et al. Impact of EGFR and KRAS genotype on outcomes in a clinical trial registry of NSCLC patients initially treated with erlotinib or gefitinib. J Clin Oncol (Meeting Abstracts) 2008; 26: 8035.
32. Naumov GN, Nilsson MB, Cascone T et al. Combined vascular endothelial growth factor receptor and epidermal growth factor receptor (EGFR) blockade inhibits tumor growth in xenograft models of EGFR inhibitor resistance. Clin Cancer Res 2009; 15: 3484–3494.
33. Herbst RS, Johnson DH, Mininberg E et al. Phase I/II trial evaluating the antivascular endothelial growth factor monoclonal antibody bevacizumab in combination with the HER-1/epidermal growth factor receptor tyrosine kinase inhibitor erlotinib for patients with recurrent non-small-cell lung cancer. J Clin Oncol 2005; 23: 2544–2555.
34. Johnson KR, Ringland C, Stokes BJ et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncol 2006; 7: 741–746.
35. Hoekstra CJ, Stroobants SG, Smit EF et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol 2005; 23: 8362–8370.
36. Weber WA, Petersen V, Schmidt B et al. Positron emission tomography in nonsmall-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. J Clin Oncol 2003; 21: 2651–2657.
37. Jain RK, Tong RT, Munn LL. Effect of vascular normalization by antiangiogenic therapy on interstitial hypertension, peritumor edema, and lymphatic metastasis: insights from a mathematical model. Cancer Res 2007; 67: 2729–2735.
38. Boellaard R, van Lingen A, van Balen SC, Lammertsma AA. Optimization of attenuation correction for positron emission tomography studies of thorax and pelvis using count-based transmission scans. Phys Med Biol 2004; 49: N31–38.
39. Boellaard R, Oyen WJ, Hoekstra CJ et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008; 35: 2320–2333.
40. de Langen AJ, Lubberink M, Boellaard R et al. Reproducibility of tumor perfusion measurements using 15O-labeled water and PET. J Nucl Med 2008; 49: 1763–1768.
41. Hoekstra CJ, Hoekstra OS, Lammertsma AA. On the use of image-derived input functions in oncological fluorine-18 fluorodeoxyglucose positron emission tomography studies. Eur J Nucl Med 1999; 26: 1489–1492.
42. Krak NC, Boellaard R, Hoekstra OS et al. Effects of ROI definition and reconstruction method on quantitative outcome a and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging 2005; 32: 294–301.
43. Parker GJ, Roberts C, Macdonald A et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magn Reson Med 2006; 56: 993–1000.
44. Tofts PS, Brix G, Buckley DL et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999; 10: 223–232.

7

Monitoring response to antiangiogenic therapy in non-small-cell lung cancer using imaging markers derived from PET and dynamic contrast-enhanced MRI

Adrianus J. de Langen*[1]
Vivian van den Boogaart*[2]
Mark Lubberink[3]
Walter H. Backes[4]
Johannes T. Marcus[5]
Harm van Tinteren[6]
Jan Pruim[7]
Boudewijn Brans[8]
Pieter Leffers[9]
Anne-Marie C. Dingemans[2]
Egbert F. Smit[1]
Harry J.M. Groen[10]
Otto S. Hoekstra[3]

[1]Department of Pulmonary Diseases, VU University Medical Center, Amsterdam, The Netherlands; [2]Department of Pulmonary Diseases and GROW, School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands; [3]Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam, The Netherlands; [4]Department of Radiology, Maastricht University Medical Center, Maastricht, The Netherlands; [5]Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands; [6]Department of Biostatistics, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; [7]Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, Groningen, The Netherlands; [8]Department of Nuclear Medicine, Maastricht University Medical Center, Maastricht, The Netherlands; [9]Department of Epidemiology, Maastricht University Medical Center, Maastricht, The Netherlands; and [10]Department of Pulmonary Diseases, University Medical Center Groningen, Groningen, The Netherlands

*Contributed equally to this work.

J Nucl Med. 2011; 52(1): 48-55

ABSTRACT

Purpose

With antiangiogenic agents, tumor shrinkage may be absent, despite survival benefit. The present study assessed the predictive value of molecular imaging for the identification of survival benefit during antiangiogenic treatment with bevacizumab and erlotinib in patients with advanced non-small-cell lung cancer.

Methods

Patients were evaluated using an imaging protocol including CT, ^{18}F-FDG PET, $H_2{}^{15}O$ PET, and dynamic contrast-enhanced MRI to derive measurements on tumor size, glucose metabolism, perfusion, and microvascular permeability. The percentage change in imaging parameters after 3 wk of treatment as compared with baseline was calculated and correlated with progression-free survival (PFS).

Results

Forty-four patients were included, and 40 underwent CT and ^{18}F-FDG PET at both time points. Complete datasets, containing all imaging modalities, were available for 14 patients. Bevacizumab and erlotinib treatment resulted in decreased metabolism, perfusion, and tumor size. A decrease in standardized uptake value or tumor perfusion of more than 20% at week 3 was associated with longer PFS (9.7 vs. 2.8 mo, $p = 0.01$, and 12.5 vs. 2.9 mo, $p = 0.009$, respectively). Whole-tumor K^{trans} (the endothelial transfer constant) was not associated with PFS, but patients with an increase of more than 15% in the SD of tumor K^{trans} values—that is, an increase in regions with low or high K^{trans} values—after 3 wk had shorter PFS (2.3 vs. 7.0 mo, $p = 0.008$). A partial response, according to the response evaluation criteria in solid tumors (RECIST), at week 3 was also associated with prolonged PFS (4.6 vs. 2.9 mo, $p = 0.017$). However, 40% of patients with a partial response as their best RECIST response still had stable disease at week 3. In these cases tumor perfusion was already decreased and K^{trans} heterogeneity showed no increase, indicating that the latter parameters seem to be more discriminative than RECIST at the 3-wk time point.

Conclusion

PET and dynamic contrast-enhanced MRI were able to identify patients who benefit from bevacizumab and erlotinib treatment. Molecular imaging seems to allow earlier response evaluation than CT.

INTRODUCTION

Response monitoring in advanced non-small-cell lung cancer (NSCLC) is complex. Tumor shrinkage usually does not occur until several cycles of chemotherapy and is difficult to evaluate because of inter- and intraobserver variation [1]. With targeted agents, volumetric change might not happen at all, despite survival benefit [2–5]. Also, the association between response and survival is weak at best [6]. Therefore, alternative criteria for response assessment are warranted. Monitoring tumor biology during therapy offers an interesting alternative and might add to size-based criteria.

PET and dynamic contrast-enhanced (DCE) MRI allow for the noninvasive quantification of aspects of tumor biology, depending on tracer characteristics [7]. ^{18}F-FDG, the most frequently and widely used tracer in the clinical setting, enables the quantification of glucose metabolism. Tumor perfusion can be measured with the short-lived PET tracer $H_2{}^{15}O$ (radiolabeled water), which is a freely diffusible and accurate perfusion tracer. DCE-MRI uses a paramagnetic gadolinium-based contrast agent to measure tumor perfusion by a combination of microvascular flow, permeability, and surface area. Particularly DCE-MRI, and to a lesser extent $H_2{}^{15}O$ PET, have been studied in clinical trials evaluating antiangiogenic drugs [8–10], but few included patients with NSCLC.

The drug combination used in the present study, bevacizumab and erlotinib (BE), targets the vascular network of tumors. Erlotinib inhibits the function of the epidermal growth factor receptor, and bevacizumab targets circulating vascular endothelial growth factor. Their supposed effects are a decrease in vascular permeability, microvascular density, and cell density [11, 12]. These effects may eventually result in tumor necrosis. This ultimate goal would result in a decrease in tumor size (measured by CT). However, antiangiogenic drug therapy often results in consolidation rather than mass regression. Despite the absence of size change, such therapy could result in decreased glucose metabolism because of lower ^{18}F-FDG availability (by antivascular effects) or a lower metabolic rate of glucose in tumor tissue. More subtle effects on tumor perfusion and permeability can be visualized with $H_2{}^{15}O$ PET and DCE-MRI.

The aim of the present study was to assess the predictive value of DCE-MRI– and PET-derived biomarkers during antiangiogenic treatment in patients with NSCLC. The combined use of DCE-MRI, $H_2{}^{15}O$ PET, and ^{18}F-FDG PET in this study offers the opportunity to evaluate the individual parameters side by side.

MATERIALS AND METHODS

Study Design and Eligibility Criteria

All subjects included in a prospective multicenter phase II trial evaluating the efficacy of BE treatment in chemonaïve patients with advanced NSCLC were studied using an extensive imaging protocol including CT, DCE-MRI, $H_2{}^{15}O$ PET, and ^{18}F-FDG PET. Patients were included when at least 2 imaging studies (including CT) were available at baseline. The protocol was approved by the institutional medical ethics review board of each participating center. Written informed consent was obtained from all patients. Inclusion criteria were histo- or cytologically documented stage IIIB (malignant pleural effusion) or IV nonsquamous NSCLC; no prior systemic therapy; measurable disease, as defined by the response evaluation criteria in solid tumors (RECIST); and an Eastern Cooperative Oncology Group performance status of 0–2. Major exclusion criteria were evidence of the tumor invading major blood vessels; the presence of a cavitating lesion; radiotherapy within 28 d before registration; evidence of bleeding diathesis, coagulopathy, or history of grade 2 or greater hemoptysis; and brain metastasis or spinal cord compression, unless previously treated, with evidence of stable disease for at least 2 mo.

Study Treatment

Patients were treated with bevacizumab (15 mg/kg) as an intravenous infusion every 3 wk and erlotinib (150 mg orally) daily. No dose reductions were made for bevacizumab. In the case of severe side effects, the dose of erlotinib was reduced according to the label. Patients remained on treatment until disease progression, unacceptable toxicity, or patient refusal. In the case of documented tumor progression (as assessed by RECIST), patients received further treatment as per investigator decision. Study medication and imaging were discontinued in these cases.

Imaging Schedule and Parameters

Imaging with CT, DCE-MRI, $H_2{}^{15}O$ PET, and ^{18}F-FDG PET was performed at baseline and after 3 wk of treatment (just before bevacizumab infusion) to derive measures on tumor size, vascular permeability, perfusion, and metabolism. Additional CT scans were made every 6 wk from baseline until disease progression. Size was measured with CT and response defined by RECIST. Perfusion (F) was determined with $H_2{}^{15}O$ PET; glucose metabolism (metabolic rate of glucose [MR_{glu}] and the standardized uptake value [SUV]) with ^{18}F-FDG PET; and a combined measure of microvascular flow, permeability, and surface area (the endothelial transfer constant [K^{trans}]) with DCE-MRI.

Dynamic ^{18}F-FDG and $H_2{}^{15}O$ PET studies were performed selectively at 1 center, whereas the other centers applied a static whole-body ^{18}F-FDG PET protocol without perfusion measurements. CT and

DCE-MRI were available for all centers.

Imaging Acquisition

PET was performed with 4 scanners: 2 Siemens ECAT EXACT HR+ scanners, 1 Philips Gemini TF-64 PET/CT scanner, and 1 Siemens Biograph bismuth germanate PET/CT scanner. Dynamic scans were obtained on a single Siemens ECAT EXACT HR+ scanner and static ones on the other systems. All patients were asked to fast for 6 h before scanning. Patients received a venous injection of ^{18}F-FDG (179–458 MBq), depending on individual patient and scanner characteristics.

Dynamic acquisition started with a 10- to 15-min transmission scan to correct for photon attenuation [13]. After a bolus injection of $H_2{}^{15}O$ (1,100 MBq), a dynamic emission scan (in 2-dimensional acquisition mode; total duration, 10 min) was started. After an interval of 10 min to allow for decay of $H_2{}^{15}O$, a bolus injection of ^{18}F-FDG (370 MBq) was administered, and a dynamic emission scan (2-dimensional; total duration, 60 min) was started. Data were reconstructed as 128 x 128 matrices using filtered backprojection with a Hanning filter (cutoff, 0.5 cycle/pixel). For volume-of-interest (VOI) definition purposes, the last 3 frames of the ^{18}F-FDG sinograms (45–60 min after injection) were summed. This summed frame was reconstructed using ordered-subset expectation maximization (OSEM) with 2 iterations and 16 subsets, followed by postsmoothing of the reconstructed image using a gaussian filter (5 mm in full width at half maximum) to obtain the same resolution as for the filtered backprojection images.

Static acquisition was performed with the following settings: Siemens HR+: 2-dimensional mode, 4-min emission and 3-min transmission per bed position, and OSEM reconstruction (2 iterations, 16 subsets); Philips Gemini: 3-dimensional mode, 135-s emission per bed position, and time-of-flight OSEM reconstruction using default settings; and Siemens Biograph: 3-dimensional mode, 5-min emission per bed position, and Fourier rebinning plus OSEM reconstruction (4 iterations, 8 subsets). All appropriate corrections for normalization, dead time, random coincidences, scatter, and attenuation were applied. Attenuation correction for the PET/CT systems was based on a low-dose CT scan acquired during tidal breathing and on transmission scans with rotating ^{68}Ge rod sources for the Siemens HR+ system. All reconstructions resulted in an image resolution of 7 mm at the center of the field of view (FOV).

DCE-MRI was performed on three 1.5-T clinical MRI systems: 2 Siemens Sonata systems and an Intera Philips system. All DCE-MR images were acquired in a transverse plane (5 slices; slice thickness, 10 mm), with the patient using a breath-hold technique. The acquisition protocol included 5 precontrast T1-weighted (3-dimensional spoiled gradient-echo sequences) measurements with different flip angles (35°, 25°, 15°, 10°, 8°, 4°, and 2°) to determine the T1 relaxation time in the blood and tissue before contrast arrival. Next, gadolinium–diethylenetriaminepentaacetic acid (0.1 mmol/kg of body weight) (Magnevist, 0.5 mol/L; Bayer Schering Pharma) was intravenously injected (3.0 mL/s) and

flushed with 15 mL of saline. This injection was followed by the DCE series, using the same sequence as the 5 precontrast T1-weighted measurements but with a flip angle of 35°, containing 30–35 scans of 2 s each. Images were acquired with the following parameters: first Siemens scanner: repetition time/echo time, 2.84/1 ms; FOV, 350 mm; and matrix, 263 x 350, and second Siemens scanner: repetition time/echo time, 2.05/0.75 ms; FOV, 350 mm; and matrix, 160 x 256. For the Philips scanner, parameters were repetition time/echo time, 4.5/2 ms; FOV, 350 mm; and matrix, 144 x 256. A fast data acquisition period (interscan interval, 2 s) was started before the bolus arrival, with the patient under breath-hold instructions to minimize motion artifacts during the first passage of the contrast agent.

Data Processing

During the study, clinicians were unaware of the imaging results, and those who analyzed the images were unaware of clinical outcome and the alternative modalities (except for appropriate FOV positioning of initial PET using the baseline CT).

PET

Patients were evaluable only when all scans were obtained on the same scanner, with a consistent protocol with respect to injected ^{18}F-FDG dose, interval from injection to scanning, and acquisition settings (i.e. exclusion of within-patient variability) [14]. Because individual medical centers used different PET scanners, only the relative change of imaging parameters was used. In this way, heterogeneity due to the use of multiple scanners was nearly eliminated [15]. All images were converted to ECAT 7 format and analyzed using software developed in-house. Threshold-defined VOIs of the primary tumor (41% of the maximum pixel value, with correction for physiologic uptake in the local background) were defined semiautomatically [15, 16]. For dynamic scans, the ^{18}F-FDG tumor VOIs were applied to both ^{18}F-FDG and $H_2{}^{15}O$ dynamic data to generate time–activity curves. An image-derived input function was created by multiple 2-dimensional regions of interest (ROIs) manually drawn over the aortic arch and ascending aorta. Three venous blood samples were taken as a quality control for the image-derived input function [17]. ROIs were then applied to all frames to generate an input time–activity curve. Full kinetic analysis to derive values of the MR_{glu} was performed using a 3-compartment model and Patlak graphical analysis [18]. For all scans, simplified semiquantitative measures were derived by calculating the mean SUV for primary tumor VOIs (50–60 min after injection for dynamic scans and 80 min after injection for static ones) and corrected for lean body mass. The correlation between simplified (SUV) and full kinetic (MR_{glu}) analysis was explored to validate the use of SUV during this intervention [19]. Mean SUV was chosen above maximum because of superior reproducibility [20] and because it allowed the validation of SUV against MR_{glu}, which cannot be done for a single voxel.

Tumor perfusion was estimated using dynamic $H_2{}^{15}O$ data and a 1-tissue-compartment model, considering an arterial blood volume component, and an image-derived input function [21, 22].

DCE-MRI

A single central slice containing the primary tumor was selected, and an ROI was manually drawn around the whole tumor. Before an ROI was drawn, the unenhanced and enhanced images were both viewed so that nontumor tissue and large vessels were avoided. Care was taken to ensure that tumor ROIs of subsequent scans were drawn at the same anatomic position. To obtain an arterial input function in every patient, an ROI was manually drawn in a major blood vessel using 1 transverse plane in the peak arterial enhancement phase of imaging. Unfortunately, it appeared impossible to derive an adequate arterial input function in every patient, mainly because of the anatomic location of some tumors, the absence of a large vessel in the field of view, and strong inflow (signal saturation) effects. Therefore, a standardized arterial input function was used [23]. All DCE-MRI data were analyzed with software developed in-house. A pharmacokinetic 2-compartment bidirectional exchange model was used to determine K^{trans} [24, 25]. Pharmacokinetic analysis was done on a pixel-per-pixel basis. K^{trans} values of individual pixels were collected and tabulated in a K^{trans} pixel histogram. The SD was calculated, representing the heterogeneity of tumor K^{trans} distribution (K^{trans} SD) [25–27]. Analysis was performed after ruling out zero values of K^{trans} to exclude nonperfused regions, for which the pharmacokinetic model is not valid.

Statistical Analysis

To investigate the agreement of SUV and MR_{glu} measurements at baseline and after 3 wk of treatment, the r^2 value was calculated together with the slope of the standardized values of the difference of both paired measurements. The 95% confidence interval (CI) of the slope was obtained using the adjusted bootstrap percentile method. Baseline patient characteristics were compared with ANOVA for continuous variables and with the Fisher exact and X^2 tests where appropriate for qualitative variables.

The predictive value of CT-, DCE-MRI–, and PET-derived measures for PFS was assessed using the percentage change from the baseline value (Δ value). PFS was defined as the time from start of treatment to the date of first documented disease progression in terms of RECIST or the date of death. Patients who had not progressed or died at the time of analysis were censored at the date of last contact. Predefined cutoff points were used and curves were compared by log rank testing. On the basis of reproducibility data, a 20% cutoff point was used for PET (i.e. values outside 1.96 x SD) [16, 28]. For DCE-MRI, a 40% cutoff point was used for K^{trans} and 15% for K^{trans} SD [10, 29]. Statistical analyses were performed using SPSS (version 15.0; SPSS Inc.). Results are presented with hazard ratios (HRs), including 95% CIs and *p* values. A *p* value of less than 0.05 was considered to be significant.

RESULTS

Forty-seven patients were enrolled from 3 centers in The Netherlands. One patient received thoracic radiotherapy, but this was 6 Gy before inclusion. Three patients were excluded because no PET and MRI scans were available, resulting in the inclusion of 44 patients (Figure 1).

Individual patient results are presented in a supplemental table, available in Journal of Nuclear Medicine online. As indicated in Figure 1, complete imaging datasets were available for 14 patients, mainly because of the single-center availability of $H_2{}^{15}O$ PET and, to a lesser extent, technical failures of DCE-MRI. At the time of analysis, 35 patients (80%) had died and 41 (93%) showed disease progression. Median follow-up was 15.6 mo. An example of an imaging profile for a patient with a response on all modalities is shown in Figure 2.

Figure 1. Flow chart of patients. Transparent boxes provide information on scan failure. Baseline and week 3 time points can be found in hatched boxes and shaded boxes, respectively.

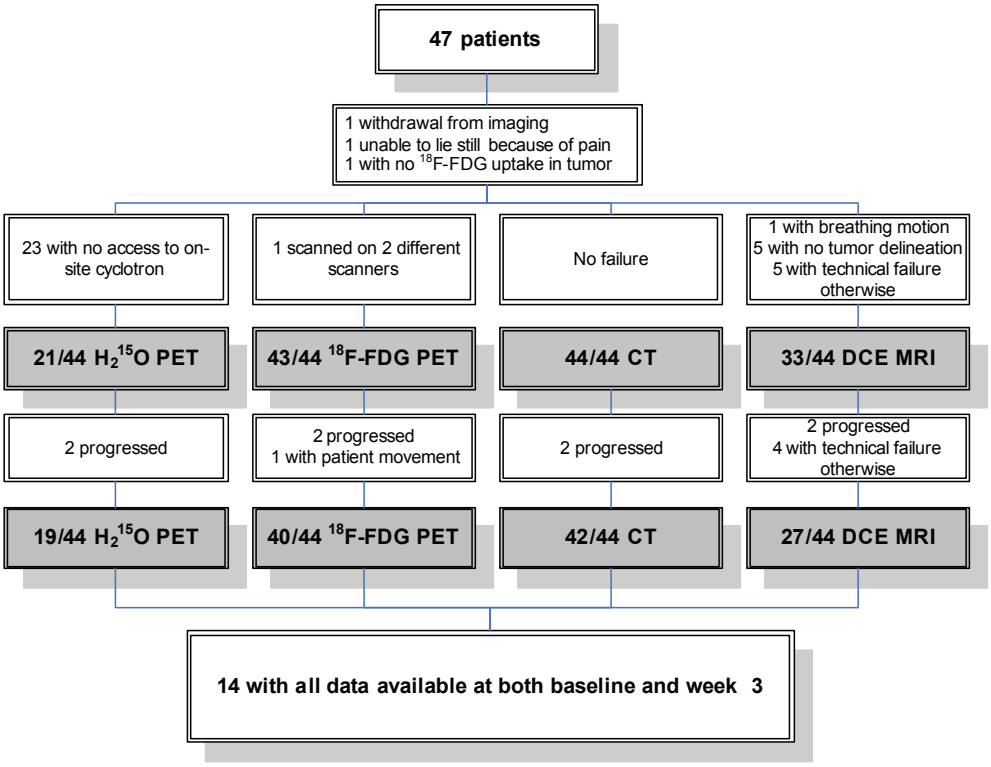
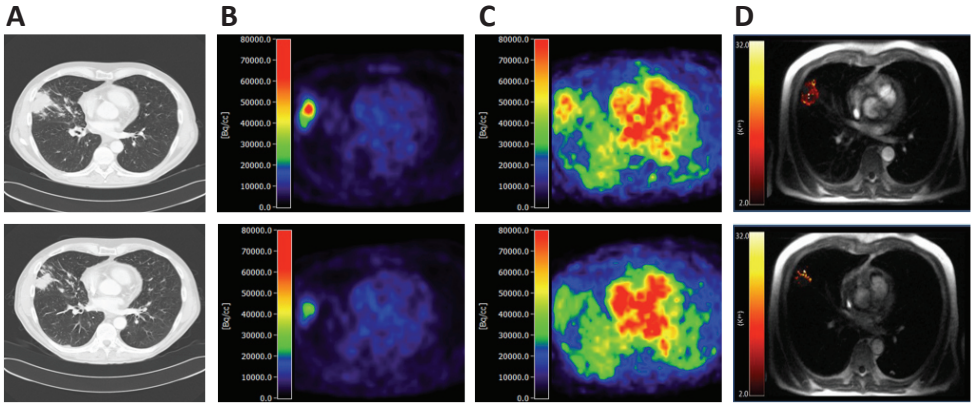


Figure 2. Baseline (top) and week 3 (bottom) scans of patient with a response on all modalities: CT (A), ^{18}F-FDG PET (B), $H_2{}^{15}O$ PET (C), and DCE-MRI (D).



Predictive Value of CT

The mean change in the sum of diameters for all target lesions after 3 wk of treatment was a decrease of 8% (interquartile range [IQR], -22% to 4%). After 3 wk, RECIST results were as follows: complete response (CR), 0 (0%); partial response (PR), 6 (14%); stable disease, 31 (74%); and progressive disease (PD), 5 (12%). At week 6, 3 patients changed from stable disease to PR (7%) and 5 from stable disease to PD (12%). Best overall response rates at the time of analysis were CR, 1 (2%); PR, 9 (21%); stable disease, 22 (50%); and PD, 12 (27%). Forty percent of patients with PR as their best response on CT had stable disease at week 3. Patients with PR at week 3 had a median PFS of 4.6 mo as compared with 2.9 mo for all other patients (HR, 0.39; 95% CI, 0.18–0.84; $p = 0.017$) (Figure 3A).

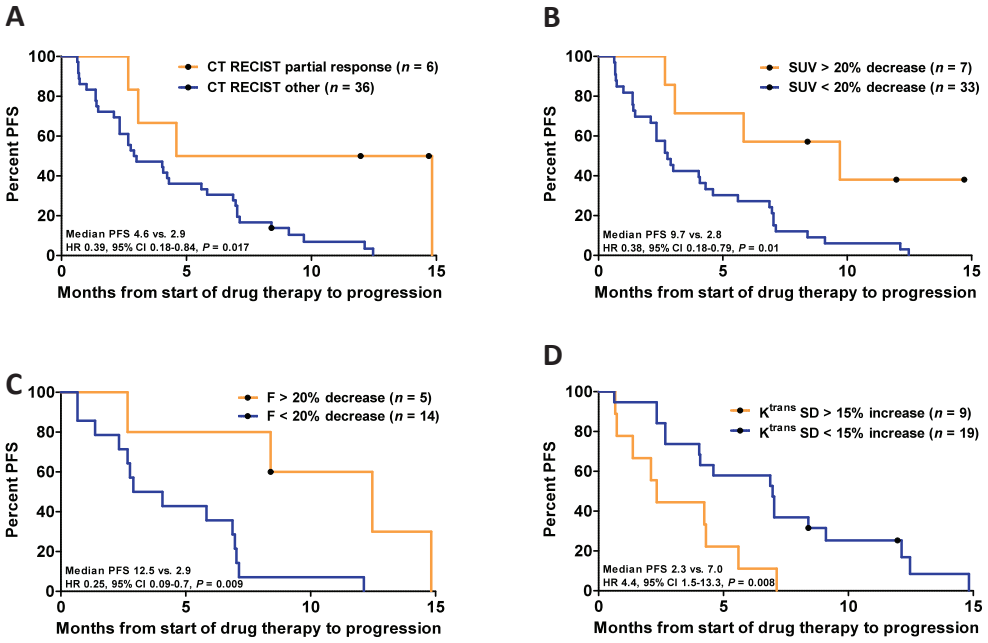
PET: Validation of SUV Against MR_{glu}

The results of 17 patients were available at both baseline and after 3 wk. At both time points, there was a strong correlation between SUV and MR_{glu} ($r^2 = 0.96$). The slope of the actual regression line was 0.98, with a CI (bootstrapped, $n = 1{,}000$) of 0.82%–1.05%, indicating that SUV can be safely used as an alternative for the fully quantitative Patlak-derived MR_{glu} during BE treatment.

Predictive Value of Glucose Metabolism (^{18}F-FDG PET)

Forty patients were scanned both at baseline and after 3 wk. Mean SUV decreased by 5% (IQR, -16% to 12%). The median PFS for patients with a metabolic response (SUV decrease > 20%) at week 3 was 9.7 mo as compared with 2.8 for other patients (HR, 0.38; 95% CI, 0.18–0.79; $p = 0.01$) (Figure 3B). Direct comparison of ^{18}F-FDG PET with CT was available for 40 patients. Although a response according

Figure 3. Kaplan–Meier curves (PFS) stratified for response after 3 wk of treatment according to CT (A), ^{18}F-FDG PET (B), $H_2{}^{15}O$ PET (C), and DCE-MRI (D).



to RECIST was predictive for longer PFS (4.6 vs. 2.9 mo), this was no longer significant (HR, 0.47; 95% CI, 0.2–1.1; $p = 0.08$).

All patients with PR as their best RECIST score were scanned with ^{18}F-FDG PET, and 7 of 10 showed a significant reduction in tumor metabolism, including 2 patients with stable disease at their week 3 CT scan.

Correlation of SUV and RECIST Response at Week 3

Discrepancy was seen in both directions, but there were no patients with PD using 1 imaging modality and PR with the other. Three of 7 patients with PR using ^{18}F-FDG PET showed stable disease at CT. Median PFS for these patients was 9.7 mo. One patient showed a PR according to RECIST and the absence of a significant metabolic response and had a PFS of 4.6 mo. Remarkably, not all patients with PR using both modalities ($n = 4$) performed well. Two patients had a relatively short PFS (2.7 and 3.1 mo), as compared with 12 and 14.7 mo for the other 2 patients. One had progression due to brain metastases and the other due to pleural effusion with malignant characteristics.

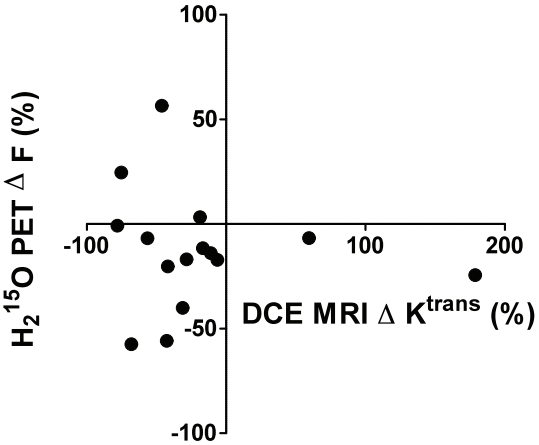
Predictive Value of Perfusion ($H_2{}^{15}O$ PET)

Tumor blood flow was measured for 19 patients both at baseline and after 3 wk. Mean blood flow decreased by 11% (IQR, -20% to -1%). Patient characteristics were not statistically different from the entire cohort (data not shown, $p > 0.34$). Patients with a greater than 20% decrease in tumor perfusion had a median PFS of 12.5 mo, as compared with 2.9 mo for the others (HR, 0.25; 95% CI, 0.09–0.70; $p = 0.009$) (Figure 3C). Direct comparison of $H_2{}^{15}O$ PET with CT showed that patients with PR on CT had longer PFS (median, 5.8 vs. 4.1 mo), but this lost significance (HR, 0.31; 95% CI, 0.04–2.49; $p = 0.59$). Five of 10 patients with PR as their best RECIST score were scanned with $H_2{}^{15}O$ PET, and all showed a significant flow reduction, including 3 patients with stable disease at their week 3 CT scan.

Predictive Value of DCE-MRI

Twenty-eight patients were scanned both at baseline and after 3 wk. Mean K^{trans} decreased by 17% (IQR, -55% to -7%). Patient characteristics were not statistically different from the entire cohort (data not shown, $p > 0.52$). K^{trans} was not associated with PFS (40% cutoff point, log rank, $p = 0.39$). Histogram analysis of tumor K^{trans} on a pixel-per-pixel basis by K^{trans} SD showed that an increase of 15% in intratumor heterogeneity was predictive for treatment failure (median PFS, 2.3 vs. 7.0 mo; HR, 4.4; 95% CI, 1.5–13.3; $p = 0.008$) (Figure 3D). Direct comparison with CT revealed that CT was not discriminative in this subset (median PFS, 4.6 vs. 4.2 mo; HR, 0.43; 95% CI, 0.16–1.16; $p = 0.1$). Seven of 10 patients with PR as their best RECIST score were scanned with DCE-MRI, and none showed an increase in K^{trans} SD, including 3 patients with stable disease at their week 3 CT scan.

Figure 4. Scatter plot of correlation between Δ K^{trans} and Δ F.



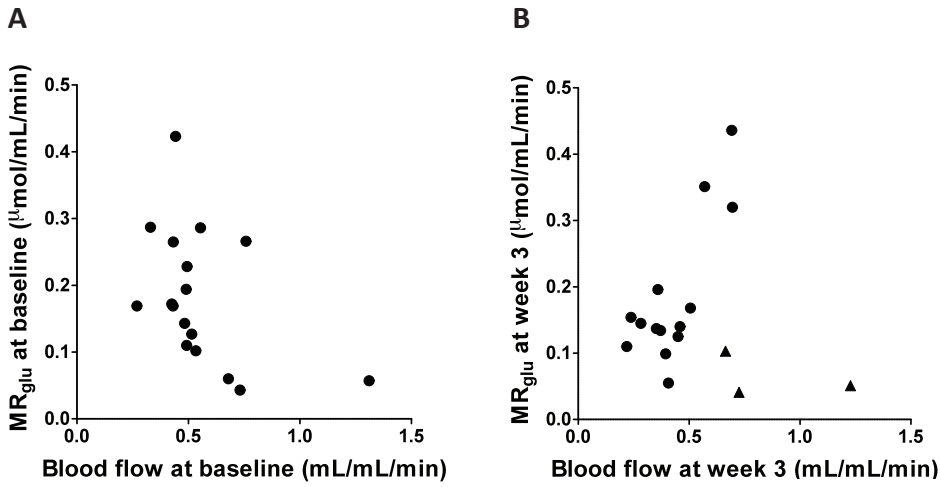
Correlative Studies

Although both K^{trans} and F decreased during treatment, there was no correlation between the change in both parameters (Figure 4). Also, tumor blood flow was not correlated to glucose metabolism at baseline and after 3 wk of treatment, although a nonsignificant trend toward a positive correlation was observed after 3 wk (Figure 5). Three patients had a low MR_{glu}-to-F ratio after 3 wk of treatment, of which 2 already had a low baseline MR_{glu}-to-F ratio (Figure 5). However, PFS for these patients was heterogeneous (range, 2.9–6.9 mo), such that no prognostic value could be attributed to this phenomenon.

DISCUSSION

This study shows that molecular imaging allows early response evaluation and seems superior to static imaging with CT in patients with NSCLC treated with combined anti–epidermal growth factor receptor and vascular endothelial growth factor therapy. Size-based response criteria are known for their inconsistent relation with PFS [2]. With targeted therapy, tumor consolidation might also represent therapeutic efficacy, resulting in prolonged survival [3–5]. However, early after treatment initiation, patients with tumor consolidation form a heterogeneous group because both tumor shrinkage and growth are relatively slow processes. As a result, response and progression can be underestimated when tumor size is used as an early predictive marker. In this study, we explored molecular imaging techniques and used prospectively defined cutoff values that were based on reproducibility data.

Figure 5. Scatter plot of correlation between PET-derived tumor perfusion and metabolism at baseline (A) and after 3 wk of treatment (B). Three data points in B with low MR_{glu}:F ratio are marked as triangles.



The latter allow the identification of any biologic effect due to the therapeutic intervention that is outside the area of test–retest variability.

Regardless of whether tumor size changed after 3 wk, patients with a more than 20% reduction in tumor glucose metabolism had a favorable PFS time. These data suggest that response may occur in patients with stable disease at CT and that ^{18}F-FDG PET is able to identify these patients. We used SUV for all patients and validated it against the fully quantitative Patlak method in a subpopulation [19]. The data do not show a significant effect of BE treatment on the pharmacokinetics of ^{18}F-FDG PET, and therefore SUV can be safely used as a substitute for the more advanced Patlak method during BE treatment in NSCLC.

Today, most trials rely on size criteria (RECIST) to base decisions on whether to discontinue drug therapy. However, it can be questioned whether this approach is applicable for trials with targeted agents, looking at our and others' results [2]. Early discontinuation of ineffective therapy can prevent harmful adverse events. In addition, patients can deteriorate soon after treatment failure, depriving them of potentially beneficial second-line therapy. Therefore, there is a need for more accurate response evaluation early in the course of treatment. In our study, all imaging modalities were able to show longer PFS for responders than for nonresponders after 1 BE cycle, but head-to-head comparison favored molecular imaging.

Four of 10 patients with PR or CR as their best RECIST result had stable disease at week 3. $H_2{}^{15}O$ PET was performed in half these patients, and all showed a major flow reduction, irrespective of size change. DCE-MRI was performed in most cases [7/10], and an increase in K^{trans} SD, associated with treatment failure, was not seen. SUV decreased in all but 3 patients. One patient had a borderline-insignificant SUV decrease, whereas 2 others had a 10% decrease and relatively low baseline uptake (SUV, 2.9 and 3.0). These 2 patients had stable disease according to RECIST at week 3. Tumor perfusion, however, was strongly reduced in these patients, indicating that blood flow might be a sensitive response predictor.

Blood flow measurements were performed in a subpopulation in this study. A decrease of 20% in tumor perfusion was associated with longer PFS (Figure 3C). Even though DCE-MRI–derived K^{trans} was not predictive of PFS, the pixel-by-pixel histogram analysis of tumor K^{trans} suggested that the absence of an increase in intratumor K^{trans} heterogeneity after 3 wk was associated with a therapy benefit. An increase in heterogeneity indicates a less even distribution of gadolinium contrast extravasation into the extravascular extracellular space, reflecting a pathologic vessel bed with areas of leaky vessels and increased perfusion and areas with severe hypoperfusion and thus absence of normalization.

The potential of combining data on perfusion and metabolism has been evaluated by Mankoff et al [30] and Tseng et al [31] in patients with breast cancer. They observed that glucose metabolism and

perfusion exhibit a weak positive correlation at baseline, which increases after (conventional) chemotherapy. In addition, patients with high pretherapy MR_{glu} relative to perfusion (MR_{glu}-to-F ratio) were more likely to experience therapy resistance and relapse. We did not observe this in our dataset. A low MR_{glu}:F ratio is caused by a tumor that has low metabolic activity or is highly vascularized. Anti–vascular endothelial growth factor treatment may result in a decrease in microvessel density, vascular permeability, and interstitial fluid pressure [32]—that is, a combination of an overall reduction in tumor perfusion and normalization of the remaining vessel bed—resulting in less hypoxia in remaining tumor tissue. Our results support this hypothesis to some extent because overall tumor blood flow and K^{trans} decreased, reflecting a reduction in microvascular flow, permeability, and/or surface area. K^{trans} SD showed that an increase in pathologic vessel bed was associated with worse outcome in terms of PFS. Similar results were obtained in rectal cancer 12 d after bevacizumab monotherapy [33].

Our study is the first, to our knowledge, to examine the correlation of blood flow measurement between DCE-MRI and PET during targeted anticancer therapy. Although both K^{trans} and F decreased during treatment, reflecting a reduction in blood flow and/or vessel permeability, no correlation was found (Figure 4), possibly reflecting the different nature of the signals. Although F is a selective measure of blood flow, K^{trans} is influenced by microvascular flow, permeability, and surface area. At baseline, the extravasation of gadolinium contrast agent is not restricted by the permeability of the numerous leaky blood vessels but is primarily flowdriven. After therapy, the extravasation of gadolinium becomes more restricted by the normalized vessel permeability and surface area. This varying pathophysiologic meaning of K^{trans} at baseline and after therapy might explain the difference in Δ F and Δ K^{trans}. Another explanation could be the semiquantitative nature of the DCE-MRI technique due to the nonlinear relationship between the degree of signal enhancement and gadolinium concentration.

CONCLUSION

This study explored the opportunities of molecular imaging as an early response-evaluation tool. The results show that SUV, F, and K^{trans} SD all predict for PFS benefit at a significant level and that these parameters seem to perform better than RECIST, whereas mean K^{trans} did not. The study included relatively few patients, and few were eligible for head-to-head comparison of all techniques (only 14 patients were scanned using all imaging modalities at both baseline and after 3 wk, reflecting the intensity of the protocol). Nevertheless, these results should stimulate similar trials, in part to investigate mechanisms of action by integrating data from different imaging modalities (with PET/MRI as the logical modality, provided that such a device would not compromise the quantitative potential of either technique) and in part to explore whether molecular imaging improves patient management or go-no-go decision making in drug development.

REFERENCE LIST

1. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol. 2003;21:2574–2582.
2. Shepherd FA, Rodrigues PJ, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med. 2005;353:123–132.
3. Hotta K, Matsuo K, Ueoka H, et al. Continued gefitinib treatment after disease stabilisation prolongs survival of Japanese patients with non-small-cell lung cancer: Okayama Lung Cancer Study Group experience. Ann Oncol. 2005;16:1817–1823.
4. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial). J Clin Oncol. 2003;21:2237–2246.
5. Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA. 2003;290:2149–2158.
6. Birchard KR, Hoang JK, Herndon JE Jr, Patz EF Jr. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. Cancer. 2009;115:581–586.
7. de Langen AJ. van dB, V, Marcus JT, Lubberink M. Use of H215O-PET and DCEMRI to measure tumor blood flow. Oncologist. 2008;13:631–644.
8. Baar J, Silverman P, Lyons J, et al. A vasculature-targeting regimen of preoperative docetaxel with or without bevacizumab for locally advanced breast cancer: impact on angiogenic biomarkers. Clin Cancer Res. 2009;15:3583–3590.
9. Hahn OM, Yang C, Medved M, et al. Dynamic contrast-enhanced magnetic resonance imaging pharmacodynamic biomarker study of sorafenib in metastatic renal carcinoma. J Clin Oncol. 2008;26:4572–4578.
10. Thomas AL,Morgan B, HorsfieldMA, et al. Phase I study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of PTK787/ZK 222584 administered twice daily in patients with advanced cancer. J Clin Oncol. 2005;23:4162–4171.
11. Normanno N, Bianco C, De Luca A, Maiello MR, Salomon DS. Target-based agents against ErbB receptors and their ligands: a novel approach to cancer treatment. Endocr Relat Cancer. 2003;10:1–21.
12. Kim KJ, Li B, Winer J, et al. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. Nature. 1993;362:841–844.
13. Boellaard R, van Lingen A, van Balen SC, Lammertsma AA. Optimization of attenuation correction for positron emission tomography studies of thorax and pelvis using count-based transmission scans. Phys Med Biol. 2004;49:N31–N38.
14. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50(suppl 1):11S–20S.
15. Boellaard R, Oyen WJ, Hoekstra CJ, et al. The Netherlands protocol for standardization and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging. 2008;35:2320–2333.
16. de Langen AJ, Klabbers B, Lubberink M, et al. Reproducibility of quantitative 18F-3'-deoxy-3'-fluorothymidine measurements using positron emission tomography. Eur J Nucl Med Mol Imaging. 2009;36:389–395.
17. Hoekstra CJ, Hoekstra OS, Lammertsma AA. On the use of image-derived input functions in oncological fluorine-18 fluorodeoxyglucose positron emission tomography studies. Eur J Nucl Med. 1999;26:1489–1492.
18. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J Cereb Blood Flow Metab. 1983;3:1–7.
19. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. J Nucl Med. 2006;47:1059–1066.
20. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. J Nucl Med. 2008; 49:1804–1808.
21. de Langen AJ, Lubberink M, Boellaard R, et al. Reproducibility of tumor perfusion measurements using 15O-labeled water and PET. J Nucl Med. 2008;49: 1763–1768.
22. Hoekstra CJ, Stroobants SG, Hoekstra OS, Smit EF, Vansteenkiste JF, Lammertsma AA. Measurement of perfusion in stage IIIA-N2 non-small cell lung cancer using H215O and positron emission tomography. Clin Cancer Res. 2002;8:2109–2115.
23. Parker GJ, Roberts C, Macdonald A, et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magn Reson Med. 2006;56:993–1000.
24. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10:223–232.
25. de Lussanet QG, Backes WH, Griffioen AW, van Engelshoven JM, Beets-Tan RG. Gadopentetate dimeglumine versus ultrasmall superparamagnetic iron oxide for dynamic contrast-enhanced MR imaging of tumor angiogenesis in human colon carcinoma in mice. Radiology. 2003;229:429–438.
26. de Lussanet QG, Langereis S, Beets-Tan RG, et al. Dynamic contrast-enhanced MR imaging kinetic parameters and molecular weight of dendritic contrast agents in tumor angiogenesis in mice. Radiology. 2005;235:65–72.
27. Mayr NA, Yuh WT, Arnholt JC, et al. Pixel analysis of MR perfusion imaging in predicting radiation therapy outcome in cervical cancer. J Magn Reson Imaging. 2000;12:1027–1033.

28. Hoekstra CJ, Hoekstra OS, Stroobants SG, et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. J Nucl Med. 2002;43:1304–1309.
29. Morgan B, Thomas AL, Drevs J, et al. Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ZK 222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase I studies. J Clin Oncol. 2003;21:3955–3964.
30. Mankoff DA, Dunnwald LK, Gralow JR, et al. Changes in blood flow and metabolism in locally advanced breast cancer treated with neoadjuvant chemotherapy. J Nucl Med. 2003;44:1806–1814.
31 Tseng J, Dunnwald LK, Schubert EK, et al. 18F-FDG kinetics in locally advanced breast cancer: correlation with tumor blood flow and changes in response to neoadjuvant chemotherapy. J Nucl Med. 2004;45:1829–1837.
32. Willett CG, Boucher Y, di Tomaso E, et al. Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer. Nat Med. 2004;10:145–147.
33. Willett CG, Duda DG, di Tomaso E, et al. Efficacy, safety, and biomarkers of neoadjuvant bevacizumab, radiation therapy, and fluorouracil in rectal cancer: a multidisciplinary phase II study. J Clin Oncol. 2009;27:3020–3026.

8

Discussion and future perspectives

THE USE OF IMAGING BIOMARKERS TO EVALUATE VASCULAR MODULATION

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^{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

scan. 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 $H_2{}^{15}O$ 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 $H_2{}^{15}O$ 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^{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^{trans} showed that patients with an increase of more than 15% in the standard deviation of tumor K^{trans} (K^{trans} heterogeneity), i.e. an increase in regions with low or high K^{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^{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^{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^{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^{trans} measurements due to signal saturation and the use of a population based AIF. This could well explain why K^{trans} heterogeneity was more discriminative than mean K^{trans}. Δ (mean) K^{trans} requires the absolute value of the week 3 scan to be compared with that of baseline, while K^{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^{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.

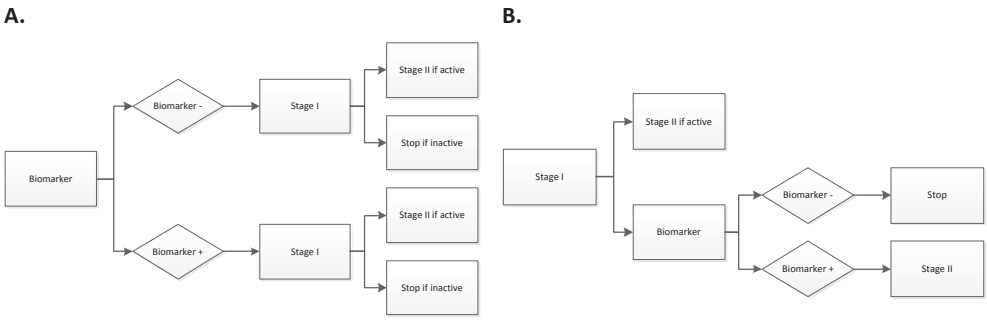
Recently, the Radiological Society of North America (RSNA) initiated QIBA (Quantitative Imaging Biomarkers Alliance) to advance quantitative imaging and the use of imaging biomarkers in clinical trials and clinical practice by engaging researchers, healthcare professionals and industry *(http://qibawiki.rsna.org)*. 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 endpoint (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].



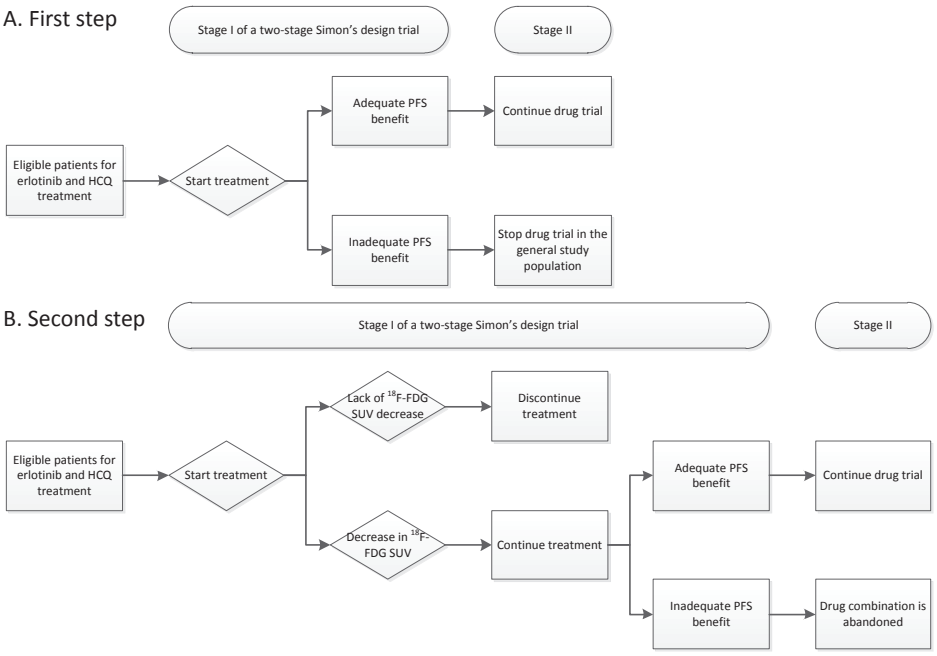
A. Adaptive parallel design [32, 33]. Two two-stage phase II trials are conducted in parallel; one in the biomarker positive group (expected to benefit more from treatment) and one in the biomarker negative group of patients. After the first stage, the trial may continue in all patients or only in the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



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 optimalised 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

1. Jain RK. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science 2005;307(5706):58-62.
2. Raichle ME, Martin WR, Herscovitch P, Mintun MA, Markham J. Brain blood flow measured with intravenous H2(15)O. II. Implementation and validation. J Nucl Med 1983;24(9):790-798.
3. Brooks DJ, Frackowiak RS, Lammertsma AA, Herold S, Leenders KL, Selwyn AP, Turton DR, Brady F, Jones T. A comparison between regional cerebral blood flow measurements obtained in human subjects using 11C-methylalbumin microspheres, the C15O2 steady-state method, and positron emission tomography. Acta Neurol Scand 1986;73(4):415-422.
4. Bergmann SR, Fox KA, Rand AL, McElvany KD, Welch MJ, Markham J, Sobel BE. Quantification of regional myocardial blood flow in vivo with H215O. Circulation 1984;70(4):724-733.
5. Araujo LI, Lammertsma AA, Rhodes CG, McFalls EO, Iida H, Rechavia E, Galassi A, De SR, Jones T, Maseri A. Noninvasive quantification of regional myocardial blood flow in coronary artery disease with oxygen-15-labeled carbon dioxide inhalation and positron emission tomography. Circulation 1991;83(3):875-885.
6. Schafers KP, Spinks TJ, Camici PG, Bloomfield PM, Rhodes CG, Law MP, Baker CS, Rimoldi O. Absolute quantification of myocardial blood flow with H(2)(15)O and 3-dimensional PET: an experimental validation. J Nucl Med 2002;43(8):1031-1040.
7. Rimoldi O, Schafers KP, Boellaard R, Turkheimer F, Stegger L, Law MP, Lammerstma AA, Camici PG. Quantification of subendocardial and subepicardial blood flow using 15O-labeled water and PET: experimental validation. J Nucl Med 2006;47(1):163-172.
8. Koeppe RA, Hutchins GD, Rothley JM, Hichwa RD. Examination of assumptions for local cerebral blood flow studies in PET. J Nucl Med 1987;28(11):1695-1703.
9. Parkka JP, Niemi P, Saraste A, Koskenvuo JW, Komu M, Oikonen V, Toikka JO, Kiviniemi TO, Knuuti J, Sakuma H, Hartiala JJ. Comparison of MRI and positron emission tomography for measuring myocardial perfusion reserve in healthy humans. Magn Reson Med 2006;55(4):772-779.
10. Brasch R, Turetschek K. MRI characterization of tumors and grading angiogenesis using macromolecular contrast media: status report. Eur J Radiol 2000;34(3):148-155.
11. Knopp MV, Weiss E, Sinn HP, Mattern J, Junkermann H, Radeleff J, Magener A, Brix G, Delorme S, Zuna I, van KG. Pathophysiologic basis of contrast enhancement in breast tumors. J Magn Reson Imaging 1999;10(3):260-266.
12. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 2009;50 Suppl 1:122S-150S.
13. Vesselle H, Grierson J, Muzi M, Pugsley JM, Schmidt RA, Rabinowitz P, Peterson LM, Vallieres E, Wood DE. In vivo validation of 3'deoxy-3'-[(18)F]fluorothymidine ([(18)F]FLT) as a proliferation imaging tracer in humans: correlation of [(18)F]FLT uptake by positron emission tomography with Ki-67 immunohistochemistry and flow cytometry in human lung tumors. Clin Cancer Res 2002;8(11):3315-3323.
14. Kenny L, Coombes RC, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3'-deoxy-3'-[18F]fluorothymidine positron emission tomography. Eur J Nucl Med Mol Imaging 2007;34(9):1339-1347.
15. Muzi M, Vesselle H, Grierson JR, Mankoff DA, Schmidt RA, Peterson L, Wells JM, Krohn KA. Kinetic analysis of 3'-deoxy-3'-fluorothymidine PET studies: validation studies in patients with lung cancer. J Nucl Med 2005;46(2):274-282.
16. Zander T, Scheffler M, Nogova L, Kobe C, Engel-Riedel W, Hellmich M, Papachristou I, Toepelt K, Draube A, Heukamp L, Buettner R, Ko YD, Ullrich RT, Smit E, Boellaard R, Lammertsma AA, Hallek M, Jacobs AH, Schlesinger A, Schulte K, Querings S, Stoelben E, Neumaier B, Thomas RK, Dietlein M, Wolf J. Early prediction of nonprogression in advanced non-small-cell lung cancer treated with erlotinib by using [(18)F]fluorodeoxyglucose and [(18)F]fluorothymidine positron emission tomography. J Clin Oncol 2011;29(13):1701-1708.
17. Spigel DR, Ramlau ER, Goldschmidt DJA, Blumenschein GR, Krzakowski MJ, Clement-Duchene C, Barlesi F. Final efficacy results from OAM4558g, a randomized phase II study evaluating MetMAb or placebo in combination with erlotinib in advanced NSCLC. J Clin Oncol 2011;29(15):May 20 Supplement: 7505.
18. Bahce I, Lubberink M, Hendrikse NH, van der Veldt AAM, Yaqub M, Windhorst AD, Schuit RC, Thunnissen E. In vivo selection of non-small cell lung cancer patients with activating mutations in the tumor epidermal growth factor receptor using [11C]erlotinib and positron emission tomography. 2011.
19. Buyse M, Michiels S, Sargent DJ, Grothey A, Matheson A, de GA. Integrating biomarkers in clinical trials. Expert Rev Mol Diagn 2011;11(2):171-182.
20. Sunaga N, Oriuchi N, Kaira K, Yanagitani N, Tomizawa Y, Hisada T, Ishizuka T, Endo K, Mori M. Usefulness of FDG-PET for early prediction of the response to gefitinib in non-small cell lung cancer. Lung Cancer 2008;59(2):203-210.

21. Su H, Bodenstein C, Dumont RA, Seimbille Y, Dubinett S, Phelps ME, Herschman H, Czernin J, Weber W. Monitoring tumor glucose utilization by positron emission tomography for the prediction of treatment response to epidermal growth factor receptor kinase inhibitors. Clin Cancer Res 2006;12(19):5659-5667.
22. Hong X, Choi H, Loyer EM, Benjamin RS, Trent JC, Charnsangavej C. Gastrointestinal stromal tumor: role of CT in diagnosis and in response evaluation and surveillance after treatment with imatinib. Radiographics 2006;26(2):481-495.
23. Beloueche-Babari M, Chung YL, Al-Saffar NM, Falck-Miniotis M, Leach MO. Metabolic assessment of the action of targeted cancer therapeutics using magnetic resonance spectroscopy. Br J Cancer 2010;102(1):1-7.
24. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. AJR Am J Roentgenol 1996;167(4):851-854.
25. Sohns C, Mangelsdorf J, Sossalla S, Konietschke F, Obenauer S. Measurement of response of pulmonal tumors in 64-slice MDCT. Acta Radiol 2010;51(5):512-521.
26. Beer AJ, Wieder HA, Lordick F, Ott K, Fischer M, Becker K, Stollfuss J, Rummeny EJ. Adenocarcinomas of esophagogastric junction: multi-detector row CT to evaluate early response to neoadjuvant chemotherapy. Radiology 2006;239(2):472-480.
27. Lee SM, Kim SH, Lee JM, Im SA, Bang YJ, Kim WH, Kim MA, Yang HK, Lee HJ, Kang WJ, Han JK, Choi BI. Usefulness of CT volumetry for primary gastric lesions in predicting pathologic response to neoadjuvant chemotherapy in advanced gastric cancer. Abdom Imaging 2009;34(4):430-440.
28. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: computerized quantification of tumor response--initial results. Radiology 2006;241(3):892-898.
29. Zhao B, Oxnard GR, Moskowitz CS, Kris MG, Pao W, Guo P, Rusch VM, Ladanyi M, Rizvi NA, Schwartz LH. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. Clin Cancer Res 2010;16(18):4647-4653.
30. Yaghmai V, Miller FH, Rezai P, Benson AB, III, Salem R. Response to treatment series: part 2, tumor response assessment--using new and conventional criteria. AJR Am J Roentgenol 2011;197(1):18-27.
31. Ng CS, Kodama Y, Mullani NA, Barron BJ, Wei W, Herbst RS, Abbruzzese JL, Charnsangavej C. Tumor blood flow measured by perfusion computed tomography and 15O-labeled water positron emission tomography: a comparison study. J Comput Assist Tomogr 2009;33(3):460-465.
32. Jones CL, Holmgren E. An adaptive Simon Two-Stage Design for Phase 2 studies of targeted therapies. Contemp Clin Trials 2007;28(5):654-661.
33. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. Clin Cancer Res 2009;15(6):1898-1905.
34. Pusztai L, Anderson K, Hess KR. Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer. Clin Cancer Res 2007;13(20):6080-6086.

Nederlandse samenvatting

ACHTERGROND

Longkanker is een ziekte die moeilijk is te behandelen. Veelal zijn er uitzaaiingen naar andere organen op het moment van diagnose en kan genezing niet meer worden bereikt. In dat geval wordt chemotherapie gegeven om de ziekte af te remmen met als doel levensverlenging en klachtenvermindering. Omdat de winst beperkt is, wordt met wetenschappelijk onderzoek gezocht naar nieuwe medicijnen die effectiever zijn dan de bestaande.

Tumoren bestaan uit een complex netwerk van kwaadaardige cellen en ondersteunende structuren zoals bloedvaten en bindweefsel waar op microscopisch niveau vele processen tegelijkertijd plaats vinden die de tumor in staat stellen te groeien en zijn destructieve werk uit te voeren. Het profiel van deze processen verschilt tussen typen kanker (bijvoorbeeld tussen borstkanker en longkanker) en de chemotherapeutische behandeling gericht tegen deze processen is dan ook afhankelijk van het type kanker. Recent onderzoek laat zien dat zelfs binnen hetzelfde type kanker dit profiel verschillend kan zijn. Dit betekent dat patiënt A met longkanker het beste behandeld kan worden met chemotherapie X, terwijl patiënt B met longkanker het beste behandeld kan worden met chemotherapie Y. Om erachter te komen welke behandeling het meest effectief is op individueel niveau, dient de behandelend specialist voorafgaand aan en tijdens de behandeling op de hoogte te zijn van het tumorprofiel. Onderzoeken die gericht zijn op het vinden van specifieke kenmerken in het tumorprofiel noemen we “biomarkers”. Door de informatie van deze biomarkers te relateren aan de prognose van een patiënt, de respons op behandeling, of het overlevingsvoordeel na behandeling krijgen zij een voorspellend vermogen en ontstaan respectievelijk prognostische, voorspellende en surrogaat eindpunt biomarkers. Daarnaast kunnen biomarkers inzicht bieden over het werkingsmechanisme van de behandeling. Dit worden farmacodynamische biomarkers genoemd en geven aan of een biologisch proces heeft plaatsgevonden of is veranderd door behandeling. Belangrijk is om biomarker informatie zo vroeg mogelijk te verkrijgen, het liefst voor aanvang van de behandeling, om langdurige therapie met niet of onvoldoende werkzame medicatie te voorkomen. Dit laatste is namelijk geassocieerd met onnodige bijwerkingen en achteruitgang in conditie waardoor patiënten niet meer in aanmerking komen voor een andere behandeling en daardoor korter leven.

De behandelend specialist heeft meerdere onderzoeken tot zijn beschikking die als biomarker kunnen fungeren. Voorbeelden zijn tumorpunctie, bloedonderzoek en beeldvormend onderzoek. De ideale biomarker is een onderzoek dat accuraat, reproduceerbaar, makkelijk uitvoerbaar en toepasbaar, minimaal invasief, veilig en goedkoop is. Beeldvormend onderzoek is hiervoor uitermate geschikt. Dit type onderzoek kan meerdere malen worden herhaald zonder al te grote belasting voor de patiënt, mede omdat het een niet-invasief onderzoek is (in tegenstelling tot bijvoorbeeld herhaalde puncties). Daarnaast maakt beeldvormend onderzoek het mogelijk meerdere tumorlaesies tegelijkertijd te anal-

yseren, waardoor gekeken kan worden naar heterogeniteit van het tumorprofiel binnen de patiënt.
De huidige standaard voor beeldvormend onderzoek is de CT scan, een techniek die de tumor diameter in beeld brengt. Aan dit onderzoek kleven meerdere bezwaren. Afname en toename van tumor diameter is een relatief langzaam proces waardoor het niet als vroege biomarker kan worden ingezet. Daarnaast heeft het een geringe relatie met (progressie-vrije) overleving. Omdat veel nieuwe behandelingen resulteren in stabilisatie van tumor diameter en niet zozeer in afname, is tumor diameter een minder gevoelige maat voor therapie evaluatie.
Een alternatief is positron emissie tomografie (PET). PET scanners detecteren het verval van radioactieve stoffen (tracers). Door een stof te volgen in het lichaam die belangrijk is voor de tumor kan inzicht worden verkregen in het tumorprofiel. Het is een dynamische techniek die kijkt naar het gedrag van de tumor (hoe gaat de tumor met een stof om en hoe verandert dit door behandeling) in tegenstelling tot de statische CT techniek die enkel kijkt naar tumor diameter. Als bekend is dat een bepaalde stof (als voorbeeld nemen we glucose) cruciaal is voor tumorgroei en een behandeling is gevonden die de toevoer of het metabolisme van deze stof (glucose) blokkeert, kan deze stof fungeren als een goede PET biomarker. Voor aanvang van behandeling kan een scan worden gemaakt. Tumoren die veel glucose opnemen zullen waarschijnlijk goed reageren op de behandeling, terwijl tumoren met weinig glucose opname minder afhankelijk zijn van de brandstof voor groei en waarschijnlijk minder baat zullen hebben van het medicament. Een scan tijdens behandeling laat zien of glucose opname en metabolisme ook echt afnemen. Misschien zijn er wel meerdere mechanismen voor glucose opname en metabolisme en blokkeert het medicament net het verkeerde mechanisme bij sommige patiënten. Hier kan dan weer nieuw onderzoek naar verricht worden om te kijken of voor deze patiënten een nieuw medicijn is te ontwikkelen.
Een tweede alternatief voor CT is “dynamic contrast-enhanced magnetic resonance imaging” (DCE-MRI). Dit is een dynamische MRI techniek waarbij de structuur en functie van bloedvaten kunnen worden beoordeeld door de passage van een contrastmiddel met hele kleine deeltjes daarin te volgen door het tumor vaatnetwerk. Met DCE-MRI kunnen de doorbloeding, het bloedvolume en de mate van lekkage van het vaatnetwerk in een tumor worden berekend.
Net als geneesmiddelen moeten ook biomarkers ontwikkeld en getest worden voordat zij toegepast kunnen worden in de behandeling van patiënten. Dit gebeurt door validatie en kwalificatie. Validatie is onderzoek naar de accuraatheid (meet de biomarker wat het beoogt te meten en is deze meting gevoelig en specifiek?) en de reproduceerbaarheid. Kwalificatie omschrijft het onderzoek naar de voorspellende waarde. De eisen waaraan een biomarker moet voldoen dient overeen te komen met de mate van invloed op de behandeling van een patiënt. Wanneer een biomarker directe invloed heeft op de behandelstrategie moet worden voldaan aan de hoogste kwaliteitseisen.

Het doel van dit proefschrift is om PET biomarkers te valideren en te kwalificeren als surrogaat eindpunt biomarkers voor de respons op een specifieke kanker behandeling. Deze behandeling is gericht tegen de bloedvaten van tumoren waardoor de tumor minder bloed en dus minder zuurstof en voedingsstoffen krijgt. Geneesmiddelen in deze klasse worden antiangiogenese middelen genoemd. Ze zijn werkzaam door vaatnieuwvorming tegen te gaan en het bestaande vaatnetwerk te normaliseren. Dit laatste is zinvol omdat bloedvaten in een tumor erg poreus zijn waardoor de druk in een tumor hoog is en chemotherapie moeilijk in de tumor doordringt.

TUMOR DOORBLOEDING METEN MET PET EN MRI

In hoofdstuk twee wordt de meting van tumordoorbloeding met PET en MRI besproken aan de hand van de bestaande literatuur. Hoewel beide scanmethoden informatie geven over tumor doorbloeding, is de aard van het signaal verschillend. PET maakt gebruik van radioactief water wat vrij kan bewegen door het lichaam en de tumor. MRI maakt gebruik van Gadolinium, een contrastmiddel dat niet vrij door de vaatwand kan bewegen en dus afhankelijk is van de mate van poreusheid (permeabiliteit) van bloedvaten. Hierdoor meet PET specifiek de doorbloeding in een tumor, terwijl MRI een combinatie van doorbloeding, vaatwand oppervlak en permeabiliteit meet. Deze resultante wordt de endotheliale transfer constante (K^{trans}) genoemd.
Hoofdstuk twee laat zien dat PET en MRI betrouwbaar de tumordoorbloeding en de endotheliale transfer constante meten, op voorhand dat een zekere mate van standaardisatie in de meetmethode en data analyse wordt toegepast. Beide scanmethoden zijn in staat om therapie effecten van angiogenese remmers te detecteren. De verschillende aard van het MRI en PET signaal maakt hun informatie uniek waardoor deze elkaar kunnen aanvullen. In theorie kan de permeabiliteit over het tumorvaatoppervlak worden geïsoleerd uit het K^{trans} signaal door deze te corrigeren voor de doorbloeding, gemeten met PET.

VALIDATIE VAN PET BIOMARKERS

De behandeling die onderzocht wordt op werkzaamheid in dit proefschrift is deels gericht tegen een groeifactor voor bloedvaten (vasculaire endotheliale groei factor, VEGF) en deels tegen een specifiek groeiproces in de tumor dat geactiveerd wordt door activatie van een receptor op tumorcellen (de epidermale groei factor receptor, EGFR). Om het effect van deze therapie zo vroeg mogelijk te kunnen detecteren kwamen drie PET biomarkers ($H_2{}^{15}O$, ^{18}F-FLT, ^{18}F-FDG) en één MRI biomarker (K^{trans}) in aanmerking.

Reproduceerbaarheid

Hoofdstuk 3 tot en met 5 beschrijven studies die de reproduceerbaarheid van de drie PET biomarkers onderzoeken. Deze studies laten zien dat alle onderzochte PET tracers een goede reproduceerbaarheid hebben.

In hoofdstuk drie wordt de reproduceerbaarheid van ^{18}F-FDG, een marker van glucose metabolisme, besproken. Een meta-analyse (gezamenlijke analyse van data uit meerdere studies) werd verricht om te kijken naar de samengestelde test-retest variabiliteit. De opname werd gekwantificeerd door gebruik te maken van de "standardized uptake value" (SUV). De SUV kan bepaald worden voor een deel van de tumor of voor de hele tumor. Als alleen het meest actieve deel van de tumor wordt bekeken wordt deze maat de SUV_{max} genoemd, terwijl SUV_{mean} het gemiddelde glucose metabolisme voor de hele tumor reflecteert.

Vergeleken met SUV_{max} is SUV_{mean} beter reproduceerbaar. De test-retest variabiliteit was afhankelijk van de mate van ^{18}F-FDG opname. Tumoren met geringe opname hebben een slechtere reproduceerbaarheid. Tumor volume heeft slechts een verwaarloosbare invloed op de reproduceerbaard van SUV. Als twee PET scans met elkaar worden vergeleken geeft een verschil van minimaal 20% en 1.2 SUV_{mean} units een werkelijk verschil aan in glucose metabolisme. Kleinere verschillen kunnen het gevolg zijn van meetfouten.

In hoofdstuk vier wordt de reproduceerbaarheid van $H_2{}^{15}O$, een marker van doorbloeding, besproken. Omdat het contrast van doorbloedingsplaatjes met PET laag is werd met de patiënt in exact dezelfde positie, een aanvullend plaatje gemaakt met ^{18}F-FLT. Op dit laatste plaatje konden de plek en de grenzen van de tumor goed worden beoordeeld. Voor dit volume werd de gemiddelde doorbloeding berekend aan de hand van de $H_2{}^{15}O$ data.

Als twee PET scans met elkaar worden vergeleken geeft een verschil van minimaal 18% een werkelijk verschil aan in tumordoorbloeding. Ook werd gekeken naar het verdelingsvolume van water (V_T), waarbij een toename of afname van 32% of meer de test-retest variabiliteit overstijgt.

In hoofdstuk vijf wordt de reproduceerbaarheid van ^{18}F-FLT, een marker van cel proliferatie, besproken. De mate van tracer opname (en dus cel proliferatie) werd, net als bij ^{18}F-FDG, weergegeven met SUV. Daarnaast werd met een geavanceerde scantechniek ook in detail gekeken naar de verschillende constanten van opname, metabolisme en uitscheiding van de tracer door de tumor.

Als twee PET scans met elkaar worden vergeleken dan geeft een verschil van minimaal 15% in SUV_{mean} en 20-25% in SUV_{max} en K_i (de bloed plasma naar tumor transfer constante) een werkelijk verschil aan in tumorproliferatie.

KWALIFICATIE VAN $H_2{}^{15}O$ EN ^{18}F-FDG PET BIOMARKERS

Na de validatiestudies werd een klinische studie geïnitieerd met als doel de PET tracers te kwalificeren als vroege voorspellende biomarkers voor een langere progressie-vrije overleving (PFS) na behandeling. Hiervoor werden patiënten met niet-kleincellig longkanker behandeld met een combinatie van twee middelen gericht tegen EGFR en VEGF.

Drie tracers kwamen in aanmerking voor biomarker kwalificatie; ^{18}F-FDG, $H_2{}^{15}O$ en ^{18}F-FLT. De korte radioactieve halfwaardetijd van ^{15}O (2 minuten) maakt het mogelijk om aansluitend aan een $H_2{}^{15}O$ scan, een aanvullende scan met een andere tracer te maken om aanvullende metingen te verrichten aan het tumorprofiel. De lange halfwaardetijd van ^{18}F maakt een aanvullende scan met een ^{18}F tracer binnen één sessie onmogelijk. Daarom moest een keuze worden gemaakt tussen ^{18}F-FDG en ^{18}F-FLT. Vanwege de gedetailleerde kennis over de connectie tussen de EGFR cascade en glucose metabolisme werd gekozen voor ^{18}F-FDG.

Hoofdstuk zes en zeven beschrijven de resultaten van een klinische studie waarin een groep patiënten met uitgezaaid longkanker werden behandeld met bevacizumab (anti-VEGF) en erlotinib (anti-EGFR) tot progressie van ziekte (tumorgroei op een CT scan).

De voorspellende waarde van doorbloeding en glucose metabolisme metingen voor PFS zijn geëvalueerd met PET en MRI.

De mediane progressie-vrije en overall overleving in deze studie waren respectievelijk 3 en 6,9 maanden. Bij patiënten met een langer dan gemiddelde PFS nam de tumordoorbloeding met gemiddeld 20% af en K^{trans} en SUV met gemiddeld 17%. Dit betekent dat de behandeling bij patiënten met een progressie-vrij overlevingsvoordeel heeft geresulteerd in afname van tumordoorbloeding en glucose metabolisme. Afname in tumordoorbloeding en SUV was niet zichtbaar bij patiënten zonder progressie-vrij overlevingsvoordeel. Tumor K^{trans} liet een daling zien die onafhankelijk was van PFS.

Om de metingen te kwalificeren als surrogaat eindpunt biomarkers, hebben we de verandering in tumor diameter, doorbloeding, K^{trans} en glucose metabolisme na drie weken therapie gerelateerd aan de PFS. Afname in tumor diameter, doorbloeding, K^{trans} en SUV werd gezien in respectievelijk 14%, 26%, 18% en 43% van de patiënten. Afname van tumor diameter na drie weken was geassocieerd met ziektevrij overlevingsvoordeel (4,6 maanden vs. 2,9 maanden voor patiënten zonder afname van tumor diameter). Echter, 40% van de patiënten die later in het behandeltraject een afname in tumor diameter lieten zien, hadden nog geen afname na drie weken. Deze uitkomst komt overeen met resultaten uit eerdere studies waaruit blijkt dat verandering in tumor diameter een relatief traag proces is en kort na aanvang van behandeling een matige voorspeller is van PFS.

In patiënten met stabiele tumor diameter na drie weken was de tumordoorbloeding wel afgenomen

en de heterogeniteit van K^{trans} niet toegenomen na drie weken, wat laat zien dat deze parameters een beter onderscheidend vermogen hebben. Een toename in tumordoorbloeding van 20% of meer was geassocieerd met een PFS voordeel (12,5 vs. 2,9 maanden voor patiënten zonder afname in tumordoorbloeding).

Hoewel de afname in K^{trans} laat zien dat de biomarker in staat is om een verandering in het tumor vaatnetwerk te meten, kon de parameter niet gekwalificeerd worden als voorspellende biomarker door het ontbreken van een relatie met PFS. Exploratieve analyse van K^{trans} heterogeniteit binnen de tumor liet echter zien dat een toename van 15% of meer in de standaard deviatie van tumor K^{trans} (een toename van gebieden met hoge en lage K^{trans} waarden) na drie weken geassocieerd was met lagere PFS (2,3 vs. 7,0 maanden).

Een toename in K^{trans} heterogeniteit betekent grote verschillen in de endotheliale transfer constante binnen de tumor en dus grote wisselingen in tumor doorbloeding en/of vaatwand permeabiliteit en oppervlakte; kenmerken van een pathologisch vaatbed met gebieden met verhoogde vaatwand permeabiliteit en/of doorbloeding en gebieden met sterk verminderde doorbloeding en/of vaatwand permeabiliteit en oppervlakte.

Hoewel, K^{trans} heterogeniteit veelbelovend lijkt te zijn, is de parameter nog niet afdoende gevalideerd en gekwalificeerd. In deze studie bleek een afkapwaarde van 15% voorspellend te zijn voor PFS voordeel, maar of dit stand houdt in toekomstige studies en of dit de test-retest variabiliteit overstijgt blijft speculatief op dit moment.

Onafhankelijk van de verandering in tumor diameter hadden patiënten met een afname in tumor glucose metabolisme van 20% of meer na drie weken een PFS voordeel (9,7 vs. 2,8 maanden voor patiënten zonder afname in glucose metabolisme).

Concluderend hebben we in deze dissertatie laten zien dat PET biomarkers voor tumor glucose metabolisme, proliferatie en doorbloeding reproduceerbaar zijn. K^{trans} heterogeniteit (gemeten met MRI) en tumordoorbloeding en glucose metabolisme (gemeten met PET) konden worden gekwalificeerd als surrogaat eindpunt biomarkers voor gecombineerde VEGF en EGFR behandeling in patiënten met niet-kleincellig longkanker.

TOEKOMST

Momenteel worden veel studies verricht met nieuwe medicijnen in ongeselecteerde patiëntgroepen waarbij enkel wordt gekeken naar het tumortype en niet naar het tumorprofiel. Deze strategie benadeelt patiënten door hen op individueel niveau de beste therapie te onthouden. Vaak geeft de behandeling gemiddeld voor de hele groep een klein overlevingsvoordeel, maar is de spreiding op individueel gebied groot. Er zijn patiënten die enkel bijwerkingen ondervinden terwijl anderen er juist veel baat van hebben.

Doordat biomarkers in staat zijn een schifting te maken tussen deze groepen kunnen zij van groot belang zijn bij de ontwikkeling van medicijnen en de behandeling van patiënten. Reeds in een vroeg stadium van medicijnontwikkeling (tijdens cel- en dierproeven) kunnen zij aantonen of een medicijn in staat is het tumorprofiel te veranderen. In kleine klinische studies kan het biomarker resultaat worden gerelateerd aan een overlevingsvoordeel. Hierdoor kan de ontwikkeling van niet of onvoldoende werkzame medicijnen al in een vroeg stadium worden gestaakt zodat de ontwikkelingskosten kunnen worden beperkt en zo min mogelijk patiënten worden blootgesteld aan niet of onvoldoende werkzame medicijnen.

In dit proefschrift hebben we PET en MRI parameters gevalideerd als surrogaat eindpunt biomarkers die correleren met de progressie-vrije overleving. Deze biomarkers correleerden beter met de PFS dan de huidige standaard, CT. De PET of MRI uitslag werd echter niet gebruikt om de behandeling wel of niet te continueren. De beslissing hiervoor werd genomen op basis van de CT scan uitslag. Omdat veel patiënten na drie weken nog een stabiele tumor diameter hebben, kan op basis van het CT resultaat hierover geen goede beslissing worden genomen. Daarom zijn nieuwe studies nodig die PET of MRI gebruiken ter evaluatie van het wel of niet continueren van behandeling, zodat duidelijk wordt of het gebruik van deze biomarkers leidt tot een betere behandeling van patiënten, resulterend in een langere overleving en minder bijwerkingen. Daarnaast zijn er veel nieuwe beeldvormende onderzoeken in ontwikkeling die momenteel getest worden op hun voorspellende waarde. Het is de hoop dat dit onderzoek leidt tot een betere behandeling van toekomstige patiënten.

Dankwoord

DANKWOORD

Hoe vreemd is het om “mijn boekje” na al die jaren af te ronden!? Het is toch een beetje mijn kindje geworden en ik heb nu het gevoel er afstand van te moeten nemen. De afgelopen jaren hebben we een haat-liefde verhouding met elkaar gehad. We hebben veel in elkaar geïnvesteerd. Vaak elkaar vervloekt, maar ook mooie tijden met elkaar beleefd. Als ik denk aan de congresbezoeken in Amerika, mijn groeiproces als wetenschapper en arts, het bericht van de editor van een wetenschappelijk tijdschrift dat een artikel was geaccepteerd en de boeiende besprekingen met mijn inspirerende promotoren, dan blijven toch vooral de positieve herinneringen hangen.
Ik heb “mijn boekje” echter niet kunnen maken zonder de steun van velen in mijn omgeving. Misschien kan ik het daarom ook beter “ons boekje” noemen. Een aantal personen wil ik graag in het bijzonder bedanken, waarbij ik mij besef dat ik mensen ga vergeten. If so, let me know, dan maak ik het goed.

Prof. dr. O.S. Hoekstra, beste Otto, jij bent van onschatbare waarde geweest en dat ben je nog steeds. Je hebt mij altijd gesteund, gestimuleerd en alle kansen gegeven. Het promotietraject, maar ook bijvoorbeeld de congresbezoeken en cursussen met als hoogtepunt “FLIMS” waar ik een onderzoeksvoorstel heb geschreven dat nu bij de METC ligt. Daarnaast kon ik altijd bij jou terecht als ik even de weg in het onderzoek kwijt was. Jij zag waar we heen moesten en als ik er aan twijfelde of de studies wel onder één noemer in een boekje kon worden gebracht, sprak jij bemoedigende woorden. Je hebt gelijk gekregen ;). Daarnaast kunnen we het ook op persoonlijk vlak goed met elkaar vinden en spreek je me soms vaderlijk toe. Als ik het me goed herinner heb je zelfs een keer tegen me gezegd: ‘Doe je voorzichtig’. Ik hoop dat we elkaar nog vaak zullen zien, zowel als collega’s en als vrienden.

Prof. dr. E.F. Smit, beste Egbert, jouw visie, durf, schoppen tegen dogma’s, lak aan bureaucratische regels, en je tomeloze energie stimuleren mij enorm. Van jou heb ik geleerd om onderzoek op een efficiënte en gestructureerde manier te doen. Bij elke stap in het onderzoek nadenken of die noodzakelijk is om de onderzoeksvraag te beantwoorden. Ik geniet van je directheid, eerlijkheid en inzicht. Daarnaast maak jij duidelijk dat een succesvolle academische carrière prima valt te combineren met een succesvol sociaal leven. Met jou praten over werk en privé, vaak onder het genot van een biertje, is voor mij waardevol. Ik kijk er naar uit om met jou een plan uit te stippelen voor een mooie (wetenschappelijke) carrière in de longoncologie.

Prof. dr. P.E. Postmus, beste Piet, Jij bent altijd geïnteresseerd geweest in mijn onderzoek en hebt het mij mogelijk gemaakt om dit te combineren met de opleiding tot longarts. Er was en is altijd ruimte om naar congressen en cursussen te gaan en je was altijd bereid om naar knelpunten te kijken als die zich

voordeden. Daarnaast was je altijd nieuwsgierig naar de vorderingen (‘wanneer is het klaar?’). Het is een voorrecht om onder jouw supervisie longarts te mogen worden.

Dr. M. Lubberink, beste Mark, als copromotor heb jij mij geholpen bij de meeste artikelen. Het review stuk hebben wij goeddeels samen geschreven. Bij de water PET analyses met de parametrische plaatjes hebben we leuke en interessante discussies gehad en ook veel samen gewerkt. Daarnaast liep ik vaak even bij je binnen voor een kop koffie of om je op te halen voor lunch. Hier bewaar ik fijne herinneringen aan. Voor je het weet sta ik in Uppsala voor je deur met een lege mok.

De leden van de leescommissie, prof. dr. S. Stroobants, prof. dr. P.E. Postmus, prof. dr. A.A. Lammertsma, prof. dr. G.A.M.S. van Dongen en dr. A.M.C. Dingemans, wil ik bedanken voor hun goedkeuring van het proefschrift en hun aanwezigheid bij de verdediging ervan.

Prof. dr. R. Boellaard, beste Ronald, zoals Otto stond ook jij altijd voor mij klaar en was je buitengewoon geïnteresseerd in wat ik deed. Inhoudelijk met jou discussiëren is een genot. Je SUV tool is een uitvinding, al heb ik ‘m ook wel eens vervloekt omdat je het programma maar bleef verbeteren en ik de oude ROIs niet meer kon openen in nieuwere versies van het programma. Je hebt je hoogleraarschap meer dan verdiend en er zullen in de (nabije) toekomst veel studenten met plezier promoveren onder jouw bezielende leiding.

Prof. dr. A.A. Lammertsma, beste Adriaan, zoals vaak genoemd ben jij een autoriteit op het gebied van PET perfusie. Hoewel ik het meeste met Otto besprak, heb ik bij theoretische en methodologische vragen over perfusie metingen veel aan jouw kennis gehad. Daarnaast ben jij in staat de materie op het denkniveau van een klinische dokter uit te leggen. Best prettig!

Dr. J.M.A. Daniels, beste Hans, bij jou is het onderzoek allemaal begonnen. Ik weet nog dat ik had bedacht om chirurg te worden. Jij had een A4-tje op de geneeskunde faculteit opgehangen met de boodschap: student gezocht voor wetenschappelijke stage bij de afdeling thoraxchirurgie. Ik dacht bingo. Ik kwam bij jou op de kamer te zitten in het MCA en het klikte meteen. Je enthousiasmeerde mij in het onderzoek en gaf me een grote mate van zelfstandigheid. Dit is erg belangrijk geweest en heeft me deels gebracht tot waar ik nu sta. We hebben mooie plannen voor de behandelkamer en ik hoop dat we dat in de nabije toekomst waar kunnen maken.

Dr. P. Raijmakers, beste Pieter, dank voor jouw hulp bij het schrijven van de PET/CT meta-analyse. Het stuk is niet opgenomen in dit boekje, maar ik heb veel van je geleerd in deze periode op het gebied van statistiek en het analytisch ontleden van artikelen.

Dr. A.M.C. Dingemans, beste Anne-Marie, jij hebt de klinische studie getrokken die de basis vormt voor dit boekje. Jouw enthousiasme is aanstekelijk en het is plezierig om met jou samen te werken.

Vivian (van den Boogaart), wij hebben samen de data-analyses verricht op enige afstand van elkaar. Jij het MRI gedeelte in Maastricht en ik het PET gedeelte in Amsterdam. Altijd tijd tekort en te grote afstand waardoor het soms lastig was om samen te werken. Toch hebben we het samen voor elkaar gekregen. Ik hoop dat ook jij snel zal promoveren. Ik heb bewondering hoe je dat combineert met je gezinsleven en de opleiding tot longarts.

Prof. dr. E. Coomans, beste Emile, als ik twijfelde over een scan, waar houdt de tumor op en begint de rib, welk lymfklier station is dit? Je was en bent altijd bereid om mee te denken en een betrokken dokter in hart en nieren. Jouw kennis over de materie overstijgt je vakgebied ruimschoots.

Andrew (Vincent) en Harm (van Tinteren), dank voor jullie hulp bij het analyseren van de data van veel van mijn stukken. De energie die jullie hebben gestoken in het FDG test-retest stuk is niet met deze pen te beschrijven. Dat het stuk maar veel geciteerd mag worden!

Dr. M.A. Paul, beste Rick, dank voor jouw begeleiding in het begin van mijn “onderzoekscarrière”.

Amanda (Kroonenberg-Kalwij), even bijkletsen met jou was altijd verhelderend en ontspannen. Je stond altijd klaar voor mij met raad, daad en koffie. Volgens mij weet jij beter waar Otto is dan hij zelf.

Wiebe (Douma), mede door jouw hulp heb ik überhaupt het VWO af kunnen ronden met een bèta pakket. De avonden dat jij mij op pro-deo basis hebt geholpen bij het vak wiskunde bleken later van onschatbare waarde te zijn. Hoewel ik het nu niet meer kan snappen dat ik het ooit niet heb gesnapt, moest het kwartje even vallen. Jij hebt het naar mij toe gegooid waarvoor ik je dank.

Laboranten van de PET, analisten van het RNC en andere betrokkenen bij de PET, Rob, Suzette, Femke, Remi, Robert, Gert, Nikkie, Cemile: veel dank voor die situaties dat er toch kon worden geschoven als ik weer eens een patiënt door jullie strot probeerde te drukken. Dank voor het scannen van de patiënten en de bloedafnames.

Reina en Virginie, dank voor jullie hulp bij de data-analyses.

Beste Marthe, dank voor de lay-out van dit boekje en het ontwerp van de omslag.

Jasper, Bart, Jacco, Jan-Hein, Cees, Mark (Reisz), Marc (Bijvoet), Diederik en Lebohang. Heren, zonder de ontspanning met jullie zou dit boekje er niet liggen. Het heeft me natuurlijk ook veel tijd gekost door de vele brakke dagen in bed, maar ik heb het er graag voor over gehad. Dat er nog veel brakke dagen mogen volgen. De 28e zal er zeker één van zijn!

Lieve collega AIOS van de longziekten, VUmc, dank voor jullie interesse in mijn onderzoek en het opvangen van klinische taken bij afwezigheid door congressen en studiereizen.

Alle medewerkers van de afdeling, polikliniek en behandelkamer longziekten wil ik bedanken voor hun steun.

Atie (van Wijk), dank voor jouw hulp bij de behandeling en het begeleiden van de studie patiënten.

Gerard, Carla en Maurice, dank voor jullie steun en interesse in mijn onderzoek in de afgelopen jaren.

Linda en Daan, mijn lieve zussen. Ik ben er trots op dat jullie mijn paranimfen zijn. We vormen een mooi trio!

Rutger, Oliver en kids, dank voor jullie interesse de afgelopen jaren.

Lieve Iris, sometimes it lasts in love but sometimes it hurts instead. Jij stond gedurende het hele promotietraject aan mijn zijde en hebt mij altijd gesteund. Ik heb je lief.

Lieve papa en mama, dankzij jullie heb ik een onbezorgde jeugd gehad met alle mogelijkheden. Hiervoor ben ik jullie dankbaar. Ik was niet altijd even makkelijk, maar jullie brachten mij altijd weer op het juiste spoor als ik een zijsprongetje maakte. Jullie onvoorwaardelijke steun voel ik dagelijks. I love you!

Curriculum Vitae

CURRICULUM VITAE

Adrianus Johannes (Joop) de Langen was born on the 7th of April 1978 in Alkmaar. He graduated secondary school in 1997 at the Petrus Canisius College in Alkmaar. That same year he started the study medical biology at the VU University Amsterdam in Amsterdam. In 2000 he was allowed to start with the study medicine. He finished these two studies in 2005 and 2007, respectively. During the last two years of his master, Joop followed two scientific internships. In 2004 he headed to North-Namibia, Africa, for six months to perform a field study on malaria. The results were published in the medical journal 'Tropical Medicine & International Health'. In 2005 he performed research for five months at the departments of Pulmonary Surgery and Nuclear Medicine at the VU University Medical Center. The results of this work were published in the medical journal 'European Journal of Cardio-Thoracic Surgery'. During this scientific internship, his interest was definitively raised for pulmonary oncology. During the last two years of his study medicine he continued his research activities. In 2007 he commenced his specialist training in Pulmonology at the VU University Medical Center in Amsterdam and combined this work with his PhD. With the publication of his thesis, he finishes the latter. He hopes to finish his specialist training in pulmonology in August 2013. In the near future, Joop hopes that he can continue to combine patient care with scientific research.

CURRICULUM VITAE

Adrianus Johannes (roepnaam Joop) de Langen werd op 7 april 1978 geboren in Alkmaar. In 1997 behaalde hij zijn Atheneum diploma aan het Petrus Canisius College te Alkmaar. In datzelfde jaar gaat hij medische biologie studeren aan de Vrije Universiteit in Amsterdam. Na drie keer te zijn uitgeloot voor de studie geneeskunde wordt hij in 2000 alsnog ingeloot. Deze studies rondt hij met succes af in respectievelijk 2005 en 2007. In de afstudeerfase liep Joop de Langen twee wetenschappelijke stages. In 2004 vertrok hij voor een half jaar naar het noorden van Namibië in Afrika om onderzoek te verrichten naar malaria. De resultaten werden gepubliceerd in het tijdschrift 'Tropical Medicine & International Health'. In 2005 liep hij gedurende vijf maanden stage bij de afdelingen Longchirurgie en Nucleaire Geneeskunde. De resultaten van dit onderzoek werden gepubliceerd in het tijdschrift 'European Journal of Cardio-Thoracic Surgery'. Tijdens deze wetenschappelijke stage werd zijn interesse definitief gewekt voor de longoncologie. Tijdens de laatste twee jaar van zijn studie geneeskunde (de coschappen) zette Joop dit onderzoek voort. In 2007 startte hij met de opleiding tot longarts in combinatie met een promotietraject. Met de publicatie van dit proefschrift rondt hij de promotie af. De opleiding tot longarts hoopt hij in augustus 2013 af te ronden. In de toekomst hoopt Joop patiëntenzorg te kunnen blijven combineren met wetenschappelijk onderzoek.

Publications

PUBLICATIONS

de Langen AJ, Raijmakers P, Riphagen I, Paul MA, Hoekstra OS. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. Eur J Cardiothorac Surg. 2006; 29(1): 26-29.

de Langen AJ, van Dillen J, de Witte P, Mucheto S, Nagelkerke N, Kager P. Automated detection of malaria pigment: feasibility for malaria diagnosing in an area with seasonal malaria in northern Namibia. Trop Med Int Health. 2006; 11(6):809-816.

de Langen AJ, Lubberink M, Boellaard R, Spreeuwenberg MD, Smit EF, Hoekstra OS, Lammertsma AA. Reproducibility of tumor perfusion measurements using ^{15}O-labeled water and PET. J Nucl Med. 2008; 49(11): 1763-1768.

de Langen AJ, van den Boogaart VE, Marcus JT, Lubberink M. Use of $H_2{}^{15}O$-PET and DCE-MRI to measure tumor blood flow. Oncologist. 2008; 13(6): 631-644.

de Langen AJ, Klabbers B, Lubberink M, Boellaard R, Spreeuwenberg MD, Slotman BJ, de Bree R, Smit EF, Hoekstra OS, Lammertsma AA. Reproducibility of quantitative ^{18}F-3'-deoxy-3'-fluorothymidine measurements using positron emission tomography. Eur J Nucl Med Mol Imaging. 2009; 36(3): 389-395.

de Langen AJ, van den Boogaart V, Lubberink M, Backes WH, Marcus JT, van Tinteren H, Pruim J, Brans B, Leffers P, Dingemans AM, Smit EF, Groen HJ, Hoekstra OS. Monitoring response to antiangiogenic therapy in non-small cell lung cancer using imaging markers derived from PET and dynamic contrast-enhanced MRI. J Nucl Med. 2011; 52(1): 48-55.

Frings V, de Langen AJ, Smit EF, van Velden FH, Hoekstra OS, van Tinteren H, Boellaard R. Repeatability of metabolically active volume measurements with ^{18}F-FDG and ^{18}F-FLT PET in non-small-cell lung cancer. J Nucl Med. 2010; 51(12): 1870-1877.

Dingemans AM, de Langen AJ, van den Boogaart V, Marcus JT, Backes WH, Scholtens HT, van Tinteren H, Hoekstra OS, Pruim J, Brans B, Thunnissen FB, Smit EF, Groen HJ. First-line erlotinib and bevacizumab in patients with locally advanced and/or metastatic non-small-cell lung cancer: a phase II study including molecular imaging. Ann Oncol. 2011; 22(3): 559-566.

Cheebsumon P, van Velden FH, Yaqub M, Frings V, de Langen AJ, Hoekstra OS, Lammertsma AA, Boellaard R. Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment. J Nucl Med. 2011; 52(10): 1550-1558.