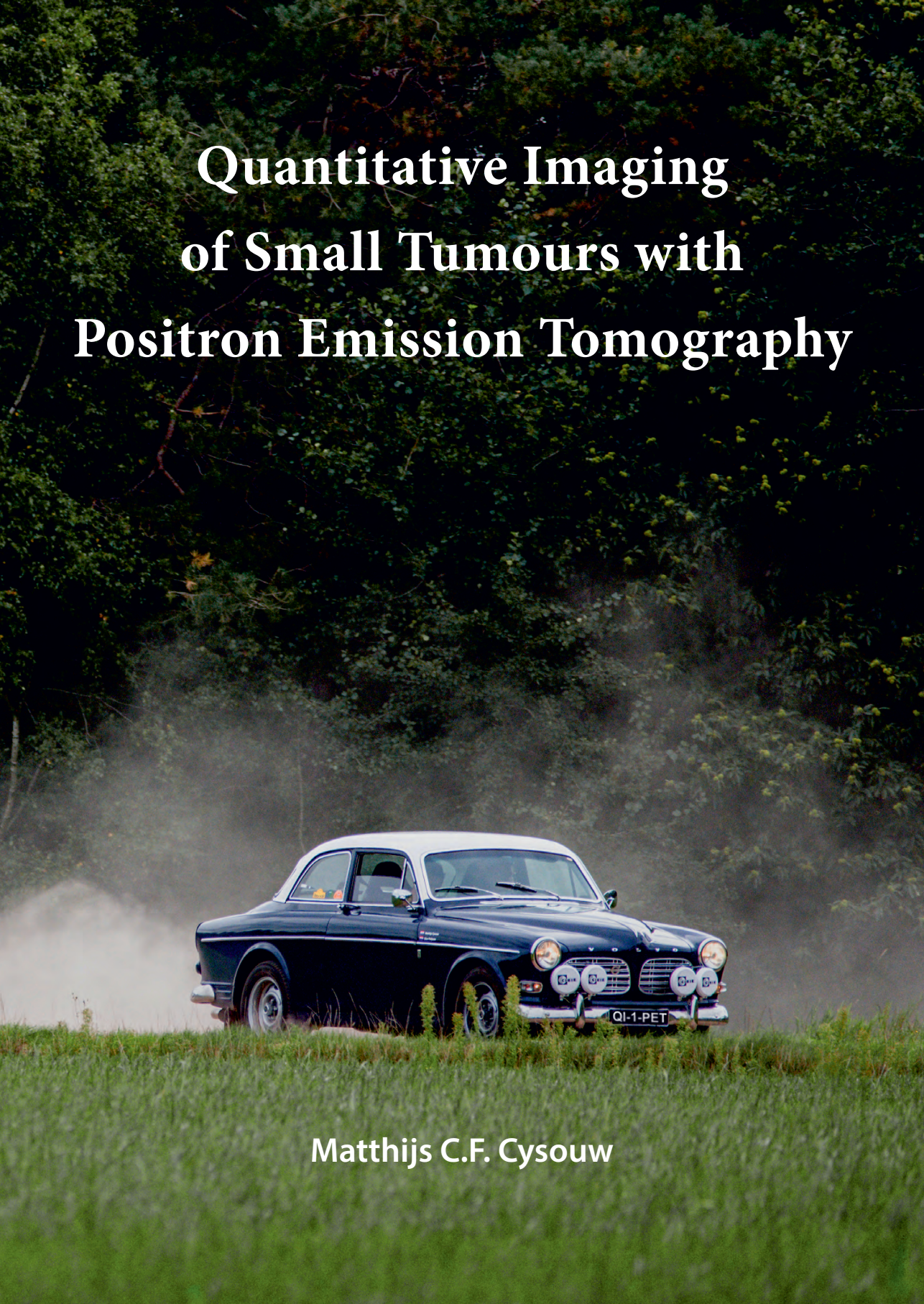
Quantitative Imaging of Small Tumours with Positron Emission Tomography



Matthijs C.F. Cysouw

Quantitative Imaging of Small Tumours with Positron Emission Tomography

Matthijs C.F. Cysouw

2020

ISBN: 978-94-6380-983-2

Photos: Niko Bloemendal Fotografie
Lay-out: Ferdinand van Nispen tot Pannerden, *my*-thesis.nl
Printing: Proefschriftmaken.nl

VRIJE UNIVERSITEIT

Quantitative Imaging of Small Tumours with Positron Emission Tomography

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 19 november 2020 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Matthijs Cornelis François Cysouw

geboren te Terneuzen

promotoren:	prof.dr. O.S. Hoekstra prof.dr. R. Boellaard
copromotoren:	dr. D.E. Oprea-Lager dr. J. Voortman

Aan mijn ouders

Contents


QI-1-PET

Chapter 1

General introduction

M.C.F. Cysouw

Cancer

Cancer is a highly prevalent disease and a major global cause of morbidity and mortality (*1*). While major advancements in screening, diagnosis, and treatments have been made, it is still the leading cause of death in the Netherlands, with nearly a third of all deaths being cancer-related (*2*). It is caused by genetic changes such as DNA mutations in the cell DNA—either inherited or acquired during life—that typically result in unbridled cell proliferation (*3,4*). Cancer cells are complex, highly adaptive and have many cellular characteristics (so called hallmarks) that result in their uncontrolled proliferation (*4*). These hallmarks have been investigated thoroughly and have become the main targets for systemic treatments, but also are the main cause of emerging resistance to these targeted drugs (*5*). The primary distinctions that separate cancer cells from normal cells are their ability to invade normal tissue and to metastasize throughout the body. Metastasis occurs via blood vessels to bone or distant organs, via the lymphatic system to local or distant lymph nodes, or *per continuitatum*. For cancer that has not metastasized beyond local lymph nodes, treatment is usually with curative intent and generally consists of surgery or radiotherapy, with or without systemic treatment. In case distant metastases are found, either at presentation or during follow-up, the disease is most often deemed incurable (with some exeptions, such as testicular cancer or oligometastatic disease). For incurable disease, palliative treatment can be used, the aim of which is to extend life expectancy and improve or maintain quality of life.

In oncology, biomarkers are essential parts of clinical practice and research. A biomarker can be defined as '*a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*' (*6*). Biomarkers can be used for diagnostic purposes, risk assessment, or to determine or predict response to cancer treatment (*7-9*). Hereby, they allow for rational and personalized use of treatments in individual patients based on their 'biomarker-phenotype'. Roughly, biomarkers can be divided into those that are measured in 'biospecimens' (e.g. tissue, blood, or urine), or measured using quantitative imaging techniques (*7,8*). An example of a blood-based biomarker is the prostate specific antigen (PSA), which can be used clinically for diagnosis, prediction, and in follow-up after and during prostate cancer treatment (*10*). A general drawback of (current) blood-based biomarkers is that they do not provide information on inter- or intralesional characteristics and, if at all, only provide indirect information of the total disease burden.

Imaging biomarkers have the advantage that they allow for measurements of individual lesions as well as the disease burden as a whole, which additionally allows for characterization of phenotypical heterogeneity. Commonly used imaging modalities are computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) (*11,12*). CT and MRI typically have high resolutions and provide detailed images of anatomical structures (*11*). However, anatomical data does not directly provide information on the tumour biology and/or functional processes. Also, changes in tumour size detected during treatment may occur only at a late stage, might be difficult to measure, or might partly represent non-viable tissue (*13,14*). PET, on the other hand, allows for non-invasive and objective in-vivo measurements of a multitude of cancer characteristics such as metabolism, functional processes, or drug targeting (*12,15-17*).

Prostate cancer

Prostate cancer (PCa) is the most prevalent cancer type in men in the Western world, with an incidence of 12,600 new cases in the Netherlands (2018) (*18*). It is a heterogeneous cancer type, both biologically and clinically (*19*). Occurrence and progression of this disease is strongly driven by the androgen receptor (AR) (*20*). Clinically, it can be divided into the (early) hormone-sensitive stage and the (late and lethal) castration-resistant phase. In the hormone-sensitive stage, standard androgen deprivation by chemical or surgical castration suffices to achieve disease control. In case of disease progression despite castration levels of testosterone, the disease is deemed to be castration-resistant. In both hormone-sensitive and castrate-resistant disease, targeting the AR axis is the mainstay of treatment (*21-25*). As more AR-targeted drugs are becoming available, and with the advent of metastasis-directed approaches (*26*), novel biomarkers for personalized treatment selection are highly needed. For this purpose, modern imaging modalities such as PET are very promising. Specifically, PET scans using tracers targeting the prostate-specific membrane antigen (PSMA) have recently revolutionized the place of functional imaging in recurrent PCa care (*27*).

Lung cancer

Lung cancer, specifically non-small cell lung cancer (NSCLC), frequently occurs in both men and women. Its incidence is rising (mainly in women), from 6800 new diagnoses in 2000 to 9300 in 2015 in the Netherlands (*28*). At diagnosis, half

of patients have distant metastases rendering curation impossible (*29*). In case of limited disease, surgery or radiotherapy (whether or not combined with adjuvant systemic treatment) are treatments of choice (*30*). For metastasized patients, many effective systemic treatments, including chemotherapy, targeted treatments, and immunotherapy, are available (*31-33*). Rational use of these treatments strongly require novel biomarkers for accurate patient selection and monitoring of treatment response (34).

Positron emission tomography

PET is a medical imaging modality exploiting the physical properties of radioactive decay of unstable positron-emitting isotopes (e.g. ^{18}F, ^{68}Ga, ^{15}O, or ^{11}C) (*35*). Initially, PET was developed as a non-invasive method to measure spatial (3D) or spatiotemporal (4D) functional or physiological processes *in-vivo*, with glucose-analogue [^{18}F]FDG as one of the first tracers used for measuring regional metabolism in healthy brain or cardiac structures (*36*). After the introduction of PET systems allowing for whole body imaging, the focus of PET development in oncology shifted towards high image quality for visual analysis rather than quantitative accuracy, e.g. exploiting the high sensitivity of PET to detect malignant tumours (*36*). However, the quantitative nature of PET uniquely allows for non-invasive objective *in-vivo* measurements of tumour biomarkers.

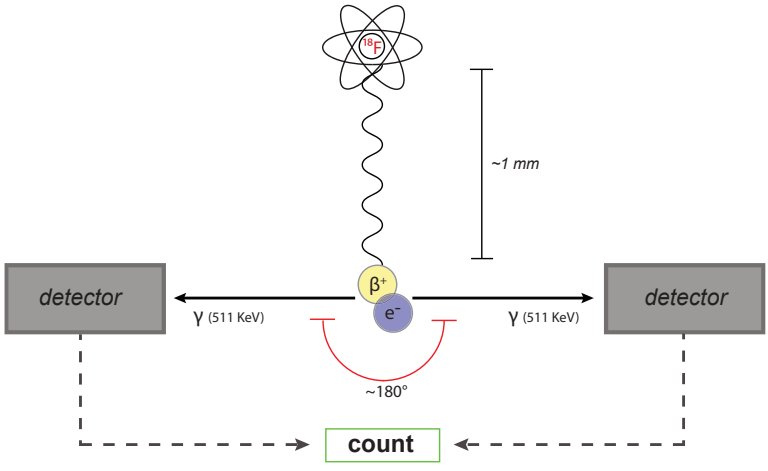


Figure 1.1: Schematic overview of the principle of PET.

Principle and technology

After intravenous administration of a labelled positron-emitting tracer, the tracer will be distributed throughout normal and (if present) malignant tissue and excreted following its specific pharmacokinetic properties. The positrons emitted by this tracer will travel a short distance (~1-3mm) through tumour tissue before thermalizing through interactions with electrons (Figure 1.1). The thermalization event results in a positronium, which annihilates yielding two 511 KeV photons that are emitted in (approximately) opposite directions (180°) (*35*). These photons are then detected by the PET scanner, eventually leading to the acquired PET image.

A PET scanner consists of multiple closely aligned rings of detectors (scintillator crystals) which allow for detection of coinciding photons ('counts') along a line of response (LOR) (*35*). By measuring counts at multiple angles across detector rings, this enables 3-dimensional measurement of radioactivity. The raw PET data result in a sinogram, which can subsequently be reconstructed as a 3-dimensional PET image.

Typically, a large fraction of emitted photons are scattered in tissue before reaching a detector. This may result in registration of counts at an erroneous LOR. Also, when two annihilation events occur approximately simultaneously, a random coincidence may be detected when one photon of each annihilation is detected at approximately the same time. Scatter and random coincidences in PET data acquisition result in substantially increased background activity and image noise (*35*). Another factor hampering detection of true counts is attenuation of the emitted photons due to absorption in tissue (*35*). This will result in most radioactivity being detected at body surfaces, since there attenuation is lowest. As CT basically yields attenuation maps of low energy photons, the rescaled Hounsfield units from low-dose CT can be used to correct the PET image for attenuation effects. Moreover, the CT can be used clinically as an anatomical correlate for the measured regional activity concentrations on PET. Currently used PET image reconstruction algorithms incorporate corrections for scatter, random coincidences, and attenuation, discussion of which is beyond the scope of this thesis.

Lastly, current state-of-the-art PET scanners can take into account the radial distance between the source of the detected photons of an event and each scintillator by measuring the difference in timing between detection of each coinciding photon. This is referred to as time-of-flight (TOF), which improves

contrast-to-noise ratios as the true location of the positron-emitting tracer along the LOR can be estimated more precisely (*37*).

Radiotracers

In PET, many of the hallmarks of cancer have been exploited for tumour imaging using a multitude of radioactively labelled tracers. An ideal radiotracer used for tumour PET imaging is metabolically stable, has high tumour-specific uptake and has low background activity.

[^{18}F]FDG

Cancer cells have an aberrant metabolism compared to non-malignant cells called the 'Warburg effect' (*38*). More specifically, malignant cells not only exhibit higher metabolic rates through uncontrolled growth and proliferation, but also mainly use aerobic glycolysis instead of oxidative phosphorylation as the primary source of energy. Glycolysis is inefficient since it generates less ATP per glucose molecule than oxidative phosphorylation, yielding increased glucose consumption in cancerous cells compared to normal cells. [^{18}F]FDG is a glucose-analogue able to capture this effect for tumour imaging, and has been used in virtually all cancer types (*39*). A drawback of [^{18}F]FDG is that inflammatory processes also exhibit increased uptake due to activation and proliferation of immune cells, limiting its specificity (*40*).

[^{18}F]FLT

^{18}F-fluorothymidine ([^{18}F]FLT) is a radiolabeled nucleoside that targets the elevated DNA synthesis in cancer cells caused by their high proliferation rates (*41*). An advantage of [^{18}F]FLT is that it is presumed not to accumulate in inflammatory processes. Clinical studies have shown that [^{18}F]FLT PET is a predictor of progression-free and disease-free survival in multiple cancer types such as lymphoma and NSCLC, and that it might be suitable for assessing response to chemo- and/or radiotherapy (*42*).

[^{18}F]FCH

Cancer cells are known to exhibit altered choline transport, upregulation of choline kinase, and increased proliferation (*43*). This has been targeted with ^{11}C- and ^{18}F-labelled choline analogues, such as [^{18}F]-fluoromethylcholine ([^{18}F]FCH) (*44*). These tracers have mainly been used for detection of prostate cancer

metastases at biochemical recurrence, but may also be useful in assessing response to chemotherapy (*45*). A limitation of [^{18}F]FCH is that it is less suitable for detection of prostate cancer metastases at low PSA levels (*46*).

[^{18}F]FDHT

Prostate cancer progression is driven by the androgen receptor (AR) through several mechanisms, designating the AR as an attractive target for molecular imaging (*20*). Hence, [^{18}F]-fluorodihydrotestosterone ([^{18}F]FDHT) was developed as a radiotracer allowing for visualization and quantification of tumour AR expression and its heterogeneity in-vivo. [^{18}F]FDHT was successfully used in early phase clinical trials to demonstrate AR-specific drug binding (*47,48*). Also, interlesional [^{18}F]FDHT heterogeneity seems to be predictive for survival after AR-targeted treatment (*15*).

[^{18}F]DCFPyL

PSMA is a type II transmembrane glycoprotein that is overexpressed in prostate cancer cells in all stages of the disease. Using [^{68}Ga] or [^{18}F]-labelled PSMA-ligands, such as [^{18}F]DCFPyL, this characteristic has been exploited with high success for prostate cancer PET imaging (*49*). Mainly in the setting of biochemical recurrence, many reports have shown that PSMA-ligand PET has superior lesion detection rates than conventional modalities, with a resulting high impact on clinical management (*27*). Nonetheless, its place in primary staging of the disease has yet to be established. As performance of visual image assessment seems to be lacking, quantitative approaches of image analysis could be of high benefit in this setting (*50,51*).

Quantification on PET

The gold standard for quantification of radiotracer uptake in tumours is full pharmacokinetic analysis (*52*). Such analysis requires dynamic PET acquisitions over a certain time frame, with blood sampling from an arterial line or venous samples to derive a tracer input function (*17,44,53,54*). Full quantitative analysis on PET yields kinetic rate constants that can be used to calculate macroparameters, such as the volume of distribution (V_T), binding potential (BP), or Ki, which are assumed to represent the biological behavior of radiotracers (*52*). Unfortunately, dynamic PET acquisitions are not feasible in clinical practice due to the limited field of view (FOV), the long duration of imaging, and the need for blood

sampling. Also, many tracers require complex and labor-intensive analyses of blood samples to correct for metabolites and derive parent fractions (*44,53*). Total body PET scanners that may facilitate whole-body dynamic imaging have been developed, but dynamic imaging may be clinically unwanted due to the extended acquisition times and need for blood sampling (*55*).

Routine clinical static PET scans have the advantage that they allow for whole body image acquisition. However, temporal information on tracer kinetics is lost, limiting quantification to single reads of spatial radioactivity. Tumour quantification on static PET typically yields mean, maximum, or peak 'standardized uptake values' (SUV) (*56*). In SUV calculation, tumour uptake is normalized to patient distribution volume (e.g. bodyweight or lean body mass) and the net injected dosage. Hereby, SUVs allow for comparison of tumour uptake between patients. Due to their simplicity SUVs are used worldwide, but they are known to be affected by many factors (*56*). Therefore, the European Association of Nuclear Medicine (EANM) has aimed to standardize imaging protocols across centers to enable reliable multicenter PET studies (*12*).

Apart from basic (first-order) statistics, such as SUV or the metabolically active tumour volume (MATV), static PET images also allow for higher-order features to be extracted. This entails high throughput 'data mining' from tumours on imaging, yielding extensive image-based tumour phenotyping named radiomics (*57*). These radiomic features may characterize the tumour micro-environment and intratumoural heterogeneity, which may reflect tumour aggressiveness or metastatic potential (*58,59*). Radiomics is a rapidly evolving field, and many PET studies have shown promising results for clinical use in oncology (*60*). Due to the high dimensionality of the radiomics data and potential complex non-linear relations with tumour or patient characteristics, analyses based on artificial intelligence algorithms are needed to make optimal use of the extracted data (*61*).

PET quantification: accuracy and precision

Accuracy indicates to what extent a certain measured value will differ from the true value (Figure 1.2). This implicates that, in order to assess accuracy, a ground truth needs to be known. Naturally, this is not the case in tumour PET-CT imaging. Therefore, alternative approaches where a ground truth is available need to be used to assess PET accuracy, such as phantoms or PET simulations. In using phantoms, an actual PET scan is made of an object filled with radioactive solutions in spherical shapes ('tumours'). In simulations, depending on the specific

implementation, a PET image of a 'cancer' patient with a tumour of a specific size, shape, and avidity is simulated, yielding a clinically realistic experimental setting (*62*).

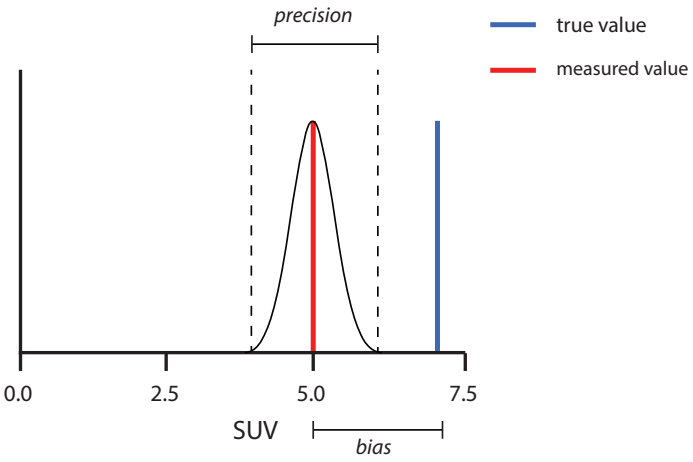


Figure 1.2: Conceptual illustration of accuracy and precision using the SUV measurement as example. The measurement bias is used to assess accuracy. The spread in measurement values represent the measurements' precision.

Precision is a measure of random variability in measurements that are repeated under the same or similar circumstances (Figure 1.2). Precision of tumour quantification on PET-CT can be evaluated in clinical research by performing test-retest studies (also referred to as repeatability studies). To this end, cancer patients not currently receiving systemic anti-cancer treatment are scanned twice on PET-CT with a short interval between these scans. The variability in tumour measurements between these two scans is deemed to be random, e.g. due to both biological variability, technical variability, and Poisson image noise. If quantitative measures of tracer uptake on PET are to be used as biomarkers, high precision is crucial. Similarly, exact knowledge of its precision is crucial to allow for PET response monitoring studies, as only a difference in tracer uptake that exceeds the day-to-day variability can be regarded as an actual treatment response or progression of disease.

Challenges of small tumour quantification on PET

PET is characterized by a limited spatial resolution, especially in comparison with anatomical imaging modalities such as CT or MRI. The limited resolution results in 'blurry' images, which hampers the detection limit of PET for small lesions, and is known to negatively affect quantification of tracer uptake in small or heterogeneous structures.

The result of the limited spatial resolution is referred to as the 'partial-volume effect' (PVE) (*63*). This term is somewhat misleading, as it does not refer to a partial detection of the tumour volume on PET. In fact, it refers to the volume-dependency of the effect, mainly pertaining to small lesions sized <2-3 times the spatial resolution. The consequence of the PVE is an apparent spill-out of the regional radioactivity concentrations. A clear illustration of the PVE is given in Figure 1.3, which shows that smaller lesions appear to have much lower uptake on PET than larger lesions due to a spill-out of activity, even though the true uptake is identical between these lesions.

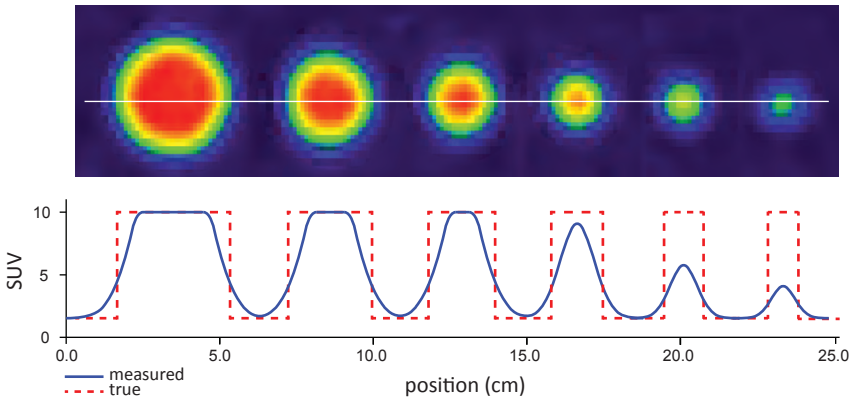


Figure 1.3: Illustration of the partial-volume effect. Differently sized spheres were filled with identical activity concentrations. Spill-out of observed activity at lesion edges into background and volume-dependent underestimations of activity are noted.

The PVE is caused by a Gaussian-shaped uncertainty in localization of the positron-emitting isotope due to several physical properties of PET, which are all portrayed in Figure 1.1: *i)* positron-range, *ii)* non-collinearity, and *iii)* detector size (*63*). The positron-range depends on the isotope-specific positron energy, meaning a higher energy level will result in a larger travelled distance before

annihilation. Non-collinearity results from the momentum of the annihilating particles that gives some variability in the angle between two emitted photons.

Another factor resulting in underestimations of tumour uptake (mainly at lesions edges) is the so-called 'tissue-fraction effect' (*63*). This is due to the relatively large voxel sizes in PET (varying 2 to 4mm), where voxels can contain both malignant and normal background tissue. Decreasing voxel sizes may mitigate this effect, but propagates image noise if tracer dosages and/or acquisition times are not adjusted accordingly.

Clinical consequences

The PVE can have several clinical consequences for use of both visual image analysis and quantitative reads (*63,64*).

Visual analysis

Activity spill-out can result in small tumours remaining below detection limits. As this yields false-negative interpretations, it may negatively affect diagnosis and staging. Also, small lesions could be falsely interpreted as being benign, e.g. a small lymph node metastasis might appear as a reactive lymph node due to the spill-out of activity. Several studies have shown that the number of lesions detected on PET increases when the PVE is taken into consideration within PET image reconstruction (*65-68*).

Quantitative analysis

PVE could negatively impact the use of quantitative PET analyses for diagnosis, staging, determining prognosis, and predicting or monitoring treatment response (*64*).

Based on the hypothesis that malignancies exhibit higher uptake on PET than benign lesions, quantification of tracer uptake can be used to discern benign versus malignant tumours. To this end, often a certain parameter threshold is defined, e.g. SUV=2.5 for solitary pulmonary nodules, to characterize lesions as being either benign or malignant (*69*). As small malignant lesions are more prone to remain below these thresholds due to the PVE, it may yield false negative reads that hamper diagnosis and (nodal) staging.

Secondly, in many cancer types a relationship between patient prognosis and tumour tracer uptake on PET has been observed (mainly for [^{18}F]FDG), probably since tumour grade and aggressiveness are proportional to glycolytic activity and/or proliferation rate (*70-73*). Similarly, tumour tracer uptake on PET can be

predictive for treatment outcomes. The PVE could result in small tumours with a poor prognosis to have low SUVs, hence hampering accurate prognostication. An illustrative study was performed by Vesselle et al. These authors observed that SUV of the primary tumour in NSCLC correlated with tumour stage and was prognostic for patient outcome, but this association disappeared after PVC was applied. This exemplifies that PVE implicitly correlates tumour size and tracer uptake, and therefore any observed prognostic or predictive value of tumour tracer uptake can be confounded by lesion size (*74*).

Third, PET can be used to assess treatment response by repeated imaging before and after (or during) treatment (*13,17,75*). Here, changes in tumour volume or size could result in changes in measured tracer uptake due to the PVE instead of a true treatment effect on the cancer cells. Simply stated, tumour shrinkage might result in a false detection of partial metabolic response on PET, and tumour growth might result in false detection of progressive metabolic disease on PET (*76*).

Finally, due to spill-over between voxels, the PVE might have a large impact on the characterization of intratumoural heterogeneity using radiomics. To date, few studies have taken into account the potential impact of the PVE and/or tissue fraction effect on the clinical performance of classification or regression models using radiomics (*77*). Applying PVC in radiomics studies may improve measurements of textural tumour features that depict the relationship between voxel values within tumours, and its impact on radiomics-based clinical predictions should be analyzed.

Partial-volume correction

Since decades, many studies have sought to find methods able to correct for the PVE (*63,64,78,79*). These partial-volume correction (PVC) methods can roughly be divided into i) reconstruction-based or post-reconstruction methods, and into ii) region-based or voxel-based (parametric) methods. Also, the different PVC methods differ in their inherent assumptions regarding tumour shape, size, homogeneity of activity distributions, knowledge on exact spatial resolution, and definition of tumour boundaries. Evaluation of the accuracy of PVC methods is challenging, since a ground truth cannot be derived from clinical images. Therefore, such studies commonly rely on phantom scans to determine an algorithm's accuracy under rather ideal circumstances. Clinically realistic simulations might be more suitable for this purpose.

Most PVC methods developed to date are post-reconstruction methods, which can be applied to PET images that are acquired clinically. The region-based methods require exact definitions of anatomical or PET-based regions, which often requires co-registered high-definition images (*80*). In oncological PET, this is problematic since low-dose CT does not allow for exact tumour delineation. The region-based PVC methods are mainly used in functional neuroimaging, where grey and white matter regions are more easily segmented (*80,81*). For tumour imaging, most studies have used a simple recovery coefficient correction using phantom data (*82-84*). More advanced or parametric PVC methods might be more suitable for oncological PET since these do not rely on segmentation of lesion boundaries. However, it is yet unclear how the net accuracy is affected by tumour delineation accuracy (*85,86*).

In recent years, use of reconstruction-based methods has become more common, especially since vendors started to provide such algorithms with novel PET systems (*87-89*). These reconstructions have the advantage of higher detection rates through improved spatial resolution, but could have lower quantitative accuracy due to increased image noise characteristics and Gibbs ringing artefacts (*90,91*). As these image reconstructions with corrections for PVE are being implemented worldwide, the impact of PVC needs to be taken into account in quantitative PET studies. This might be of particular importance for test-retest and response monitoring studies, as the lower precision caused by PVC methods might affect the minimal detectable change in tumour uptake.

Aim and outline of the thesis

The clinical use of quantitative assessment of PET images in oncology is maturing and merging into clinical practice. Therefore, it is crucial to know if and how small tumours can be accurately and reliably assessed using quantitative PET analysis. Moreover, the clinical value of these analyses needs to be defined. In the present thesis we aim to perform technical validation of small tumour quantitative PET imaging and investigate its clinical application. In **Part 1** (chapter 2-4) of this thesis we focus on methodological aspects of quantitative PET imaging of small tumours, specifically investigating the performance of PVC. In **Part 2** (chapter 5-8) we explore the clinical benefit of quantitative assessment of small tumours on PET and the impact of applying PVC.

In **Chapter 2** we investigate the accuracy and precision of several PVC methods, in combination with several different methods for tumour delineation on PET in [^{18}F]FDG and [^{18}F]FCH PET-CT. In **Chapter 3** we assess how parametric PVC affects pharmacokinetic modelling on dynamic [^{18}F]FLT PET-CT and validation of simplified parameters in NSCLC patients undergoing systemic treatment. **Chapter 4** focusses on the influence and interplay of image noise and PVC on the repeatability of quantitative tumour assessment on [^{18}F]FDHT PET-CT in metastatic PCa patients. In **Chapter 5** we assess the repeatability of [^{18}F] DCFPyL PET-CT in metastatic PCa patients and evaluate the impact of PVC. In **Chapter 6** we perform machine learning-based analysis of [^{18}F]DCFPyL PET-CT radiomics for risk-stratification of primary PCa patients, and assess the impact of PVC and tumour delineation methods.

In **Chapter 7** we systematically review and perform meta-analysis on the clinical application of PVC in PET studies in oncology. In **Chapter 8** we evaluate whether PVC can improve prediction of outcome after stereotactic body radiotherapy (SBRT) for oligometastatic prostate cancer using [^{18}F]FCH PET-CT. In **Chapter 9** we discuss methodological aspects of quantification of [^{18}F]DCFPyL and [^{18}F]FDHT in prostate cancer, and provide clinical illustration of its use in response assessment. In **Chapter 10** we benchmark the predictive value of the [^{18}F]DCFPyL radiomics analysis from Chapter 6 against methods used in clinical practice. Lastly, a summarizing discussion of findings is presented in **Chapter 11.**

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. ***CA Cancer J Clin.*** 2019;69:7-34.
2. Centraal Bureau voor de Statistiek, Doodsoorzakenstatistiek. 2017.
3. Hanahan D, Weinberg RA. The hallmarks of cancer. ***Cell.*** 2000;100:57-70.
4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. ***Cell.*** 2011;144:646-674.
5. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. ***Nat Rev Cancer.*** 2013;13:714-726.
6. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. ***Clin Pharmacol Ther.*** 2001;69:89-95.
7. Mishra A, Verma M. Cancer biomarkers: are we ready for the prime time? ***Cancers (Basel).*** 2010;2:190-208.
8. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. ***Nat Rev Clin Oncol.*** 2017;14:169-186.
9. O'Connor JP, Jackson A, Asselin MC, Buckley DL, Parker GJ, Jayson GC. Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives. ***Lancet Oncol.*** 2008;9:766-776.
10. Schalken J, Dijkstra S, Baskin-Bey E, van Oort I. Potential utility of cancer-specific biomarkers for assessing response to hormonal treatments in metastatic prostate cancer. ***Ther Adv Urol.*** 2014;6:245-252.
11. Gonzalez Hernando C, Esteban L, Canas T, Van den Brule E, Pastrana M. The role of magnetic resonance imaging in oncology. ***Clin Transl Oncol.*** 2010;12:606-613.
12. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
13. 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.
14. Mac Manus MP, Hicks RJ, Matthews JP, et al. Positron emission tomography is superior to computed tomography scanning for response-assessment after radical radiotherapy or chemoradiotherapy in patients with non-small-cell lung cancer. ***J Clin Oncol.*** 2003;21:1285-1292.
15. Fox JJ, Gavane SC, Blanc-Autran E, et al. Positron Emission Tomography/Computed Tomography-Based Assessments of Androgen Receptor Expression and Glycolytic Activity as a Prognostic Biomarker for Metastatic Castration-Resistant Prostate Cancer. ***JAMA Oncol.*** 2018;4:217-224.
16. Niemeijer AN, Leung D, Huisman MC, et al. Whole body PD-1 and PD-L1 positron emission tomography in patients with non-small-cell lung cancer. ***Nat Commun.*** 2018;9:4664.
17. Frings V, Yaqub M, Hoyng LL, et al. Assessment of simplified methods to measure 18F-FLT uptake changes in EGFR-mutated non-small cell lung cancer patients undergoing EGFR tyrosine kinase inhibitor treatment. ***J Nucl Med.*** 2014;55:1417-1423.
18. Cysouw MCF, Kramer GM, Frings V, et al. Baseline and longitudinal variability of normal tissue uptake values of [(18)F]-fluorothymidine-PET images. ***Nucl Med Biol.*** 2017;51:18-24.
19. Scher HI, Morris MJ, Stadler WM, et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. ***J Clin Oncol.*** 2016;34:1402-1418.
20. Heinlein CA, Chang C. Androgen receptor in prostate cancer. ***Endocr Rev.*** 2004;25:276-308.
21. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. ***N Engl J Med.*** 2011;364:1995-2005.
22. Rathkopf DE, Antonarakis ES, Shore ND, et al. Safety and Antitumor Activity of Apalutamide (ARN-509) in Metastatic Castration-Resistant Prostate Cancer with and without Prior Abiraterone Acetate and Prednisone. ***Clin Cancer Res.*** 2017;23:3544-3551.
23. Ryan CJ, Molina A, Griffin T. Abiraterone in metastatic prostate cancer. ***N Engl J Med.*** 2013;368:1458-1459.
24. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. ***N Engl J Med.*** 2012;367:1187-1197.

25. James ND, de Bono JS, Spears MR, et al. Abiraterone for Prostate Cancer Not Previously Treated with Hormone Therapy. ***N Engl J Med.*** 2017;377:338-351.

26. Ost P, Reynders D, Decaestecker K, et al. Surveillance or Metastasis-Directed Therapy for Oligometastatic Prostate Cancer Recurrence: A Prospective, Randomized, Multicenter Phase II Trial. ***J Clin Oncol.*** 2018;36:446-453.

27. Donin NM, Reiter RE. Why Targeting PSMA Is a Game Changer in the Management of Prostate Cancer. ***J Nucl Med.*** 2018;59:177-182.

28. Lai JKC, Teo E, Lichtenstein M, Toh HB, Sivaratnam D. Cardiac FDG/PET imaging mimicking sarcoidosis. ***J Nucl Cardiol.*** 2019;26:1025-1027.

29. Walters S, Maringe C, Coleman MP, et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007. ***Thorax.*** 2013;68:551-564.

30. Chang JY, Senan S, Paul MA, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. ***Lancet Oncol.*** 2015;16:630-637.

31. Baxevanos P, Mountzios G. Novel chemotherapy regimens for advanced lung cancer: have we reached a plateau? ***Ann Transl Med.*** 2018;6:139.

32. Sim EH, Yang IA, Wood-Baker R, Bowman RV, Fong KM. Gefitinib for advanced non-small cell lung cancer. ***Cochrane Database Syst Rev.*** 2018;1:CD006847.

33. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. ***Lancet.*** 2019.

34. Smit EF, de Langen AJ. Pembrolizumab for all PD-L1-positive NSCLC. ***Lancet.*** 2019.

35. Zanzonico P. Positron emission tomography: a review of basic principles, scanner design and performance, and current systems. ***Semin Nucl Med.*** 2004;34:87-111.

36. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. ***J Nucl Med.*** 2017;58:1019-1024.

37. Vandenberghe S, Mikhaylova E, D'Hoe E, Mollet P, Karp JS. Recent developments in time-of-flight PET. ***EJNMMI Phys.*** 2016;3:3.

38. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? ***Trends Biochem Sci.*** 2016;41:211-218.

39. Som P, Atkins HL, Bandoypadhyay D, et al. A fluorinated glucose analog, 2-fluoro-2-deoxy-D-glucose (F-18): nontoxic tracer for rapid tumor detection. ***J Nucl Med.*** 1980;21:670-675.

40. Love C, Tomas MB, Tronco GG, Palestro CJ. FDG PET of infection and inflammation. ***Radiographics.*** 2005;25:1357-1368.

41. Been LB, Suurmeijer AJ, Cobben DC, Jager PL, Hoekstra HJ, Elsinga PH. [18F]FLT-PET in oncology: current status and opportunities. ***Eur J Nucl Med Mol Imaging.*** 2004;31:1659-1672.

42. Bollineni VR, Kramer GM, Jansma EP, Liu Y, Oyen WJ. A systematic review on [(18)F]FLT-PET uptake as a measure of treatment response in cancer patients. ***Eur J Cancer.*** 2016;55:81-97.

43. Kwee SA, DeGrado TR, Talbot JN, Gutman F, Coel MN. Cancer imaging with fluorine-18-labeled choline derivatives. ***Semin Nucl Med.*** 2007;37:420-428.

44. Verwer EE, Oprea-Lager DE, van den Eertwegh AJ, et al. Quantification of 18F-fluorocholine kinetics in patients with prostate cancer. ***J Nucl Med.*** 2015;56:365-371.

45. Oprea-Lager DE, van Kanten MP, van Moorselaar RJ, et al. [18F]fluoromethylcholine as a chemotherapy response read-out in prostate cancer cells. ***Mol Imaging Biol.*** 2015;17:319-327.

46. Afshar-Oromieh A, Zechmann CM, Malcher A, et al. Comparison of PET imaging with a (68)Ga-labelled PSMA ligand and (18)F-choline-based PET/CT for the diagnosis of recurrent prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2014;41:11-20.

47. Scher HI, Beer TM, Higano CS, et al. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. ***Lancet.*** 2010;375:1437-1446.

48. Rathkopf DE, Morris MJ, Fox JJ, et al. Phase I study of ARN-509, a novel antiandrogen, in the treatment of castration-resistant prostate cancer. ***J Clin Oncol.*** 2013;31:3525-3530.

49. Jansen BHE, Yaqub M, Voortman J, et al. Simplified Methods for Quantification of (18)F-DCFPyL Uptake in Patients with Prostate Cancer. ***J Nucl Med.*** 2019.

50. Uprimny C, Kroiss AS, Decristoforo C, et al. (68)Ga-PSMA-11 PET/CT in primary staging of prostate cancer: PSA and Gleason score predict the intensity of tracer accumulation in the primary tumour. ***Eur J Nucl Med Mol Imaging.*** 2017;44:941-949.

51. Gorin MA, Rowe SP, Patel HD, et al. Prostate Specific Membrane Antigen Targeted (18)F-DCFPyL Positron Emission Tomography/Computerized Tomography for the Preoperative Staging of High Risk Prostate Cancer: Results of a Prospective, Phase II, Single Center Study. ***J Urol.*** 2018;199:126-132.

52. Morris ED, Endres CJ, Schmidt KC, Christian BT, Muzic RF, Fisher RE. Kinetic modeling in positron emission tomography. ***Emission Tomography: The Fundamentals of PET and SPECT Academic, San Diego.*** 2004.

53. Kramer GM, Yaqub M, Vargas HA, et al. Assessment of simplified methods for quantification of (18) F-FDHT uptake in patients with metastatic castration-resistant prostate cancer. ***J Nucl Med.*** 2019.

54. Lubberink M, Direcks W, Emmering J, et al. Validity of simplified 3'-deoxy-3'-[18F]fluorothymidine uptake measures for monitoring response to chemotherapy in locally advanced breast cancer. ***Mol Imaging Biol.*** 2012;14:777-782.

55. Badawi RD, Shi H, Hu P, et al. First Human Imaging Studies with the EXPLORER Total-Body PET Scanner. ***J Nucl Med.*** 2019;60:299-303.

56. 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.

57. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. ***Eur J Cancer.*** 2012;48:441-446.

58. De Bernardi E, Buda A, Guerra L, et al. Radiomics of the primary tumour as a tool to improve (18) F-FDG-PET sensitivity in detecting nodal metastases in endometrial cancer. ***EJNMMI Res.*** 2018;8:86.

59. Li K, Sun H, Lu Z, et al. Value of [(18)F]FDG PET radiomic features and VEGF expression in predicting pelvic lymphatic metastasis and their potential relationship in early-stage cervical squamous cell carcinoma. ***Eur J Radiol.*** 2018;106:160-166.

60. Cook GJR, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and Promises of PET Radiomics. ***Int J Radiat Oncol Biol Phys.*** 2018;102:1083-1089.

61. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning methods for Quantitative Radiomic Biomarkers. ***Sci Rep.*** 2015;5:13087.

62. Pfaehler E, De Jong JR, Dierckx R, van Velden FHP, Boellaard R. SMART (SiMulAtion and ReconsTruction) PET: an efficient PET simulation-reconstruction tool. ***EJNMMI Phys.*** 2018;5:16.

63. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. ***J Nucl Med.*** 2007;48:932-945.

64. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. ***Phys Med Biol.*** 2012;57:R119-159.

65. Hotta M, Minamimoto R, Yano H, Gohda Y, Shuno Y. Diagnostic performance of (18)F-FDG PET/CT using point spread function reconstruction on initial staging of rectal cancer: a comparison study with conventional PET/CT and pelvic MRI. ***Cancer Imaging.*** 2018;18:4.

66. Lasnon C, Hicks RJ, Beauregard JM, et al. Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non-small cell lung cancer. ***Clin Nucl Med.*** 2012;37:971-976.

67. Bellevre D, Blanc Fournier C, Switsers O, et al. Staging the axilla in breast cancer patients with (1) (8)F-FDG PET: how small are the metastases that we can detect with new generation clinical PET systems? ***Eur J Nucl Med Mol Imaging.*** 2014;41:1103-1112.

68. Kawashima K, Kato K, Tomabechi M, et al. Clinical evaluation of (18)F-fludeoxyglucose positron emission tomography/CT using point spread function reconstruction for nodal staging of colorectal cancer. ***Br J Radiol.*** 2016;89:20150938.

69. Groheux D, Quere G, Blanc E, et al. FDG PET-CT for solitary pulmonary nodule and lung cancer: Literature review. ***Diagn Interv Imaging.*** 2016;97:1003-1017.

70. Spaepen K, Stroobants S, Dupont P, et al. Prognostic value of positron emission tomography (PET) with fluorine-18 fluorodeoxyglucose ([18F]FDG) after first-line chemotherapy in non-Hodgkin's lymphoma: is [18F]FDG-PET a valid alternative to conventional diagnostic methods? ***J Clin Oncol.*** 2001;19:414-419.

71. Bonomo P, Merlotti A, Olmetto E, et al. What is the prognostic impact of FDG PET in locally advanced head and neck squamous cell carcinoma treated with concomitant chemo-radiotherapy? A systematic review and meta-analysis. ***European Journal of Nuclear Medicine and Molecular Imaging.*** 2018;45:2122-2138.

72. Berghmans T, Dusart M, Paesmans M, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project. ***J Thorac Oncol.*** 2008;3:6-12.

73. Geus-Oei LF, Oyen WJ. Predictive and prognostic value of FDG-PET. ***Cancer Imaging.*** 2008;8:70-80.

74. Vesselle H, Turcotte E, Wiens L, et al. Relationship between non-small cell lung cancer fluorodeoxyglucose uptake at positron emission tomography and surgical stage with relevance to patient prognosis. ***Clin Cancer Res.*** 2004;10:4709-4716.

75. 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.

76. Anouan KJ, Lelandais B, Edet-Sanson A, et al. 18F-FDG-PET partial volume effect correction using a modified recovery coefficient approach based on functional volume and local contrast: physical validation and clinical feasibility in oncology. ***Q J Nucl Med Mol Imaging.*** 2017;61:301-313.

77. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. ***Sci Rep.*** 2017;7:358.

78. Rousset O, Rahmim A, Alavi A, Zaidi H. Partial Volume Correction Strategies in PET. ***PET Clin.*** 2007;2:235-249.

79. Hoetjes NJ, van Velden FH, Hoekstra OS, et al. Partial volume correction strategies for quantitative FDG PET in oncology. ***Eur J Nucl Med Mol Imaging.*** 2010;37:1679-1687.

80. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. ***J Nucl Med.*** 1998;39:904-911.

81. Muller-Gartner HW, Links JM, Prince JL, et al. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. ***J Cereb Blood Flow Metab.*** 1992;12:571-583.

82. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. ***J Comput Assist Tomogr.*** 1979;3:299-308.

83. Geworski L, Knoop BO, de Cabrejas ML, Knapp WH, Munz DL. Recovery correction for quantitation in emission tomography: a feasibility study. ***Eur J Nucl Med.*** 2000;27:161-169.

84. Avril N, Dose J, Janicke F, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. ***J Clin Oncol.*** 1996;14:1848-1857.

85. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. ***J Nucl Med.*** 2007;48:802-810.

86. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. ***Neuroimage.*** 2008;39:1570-1584.

87. Rahmim A, Qi J, Sossi V. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. ***Med Phys.*** 2013;40:064301.

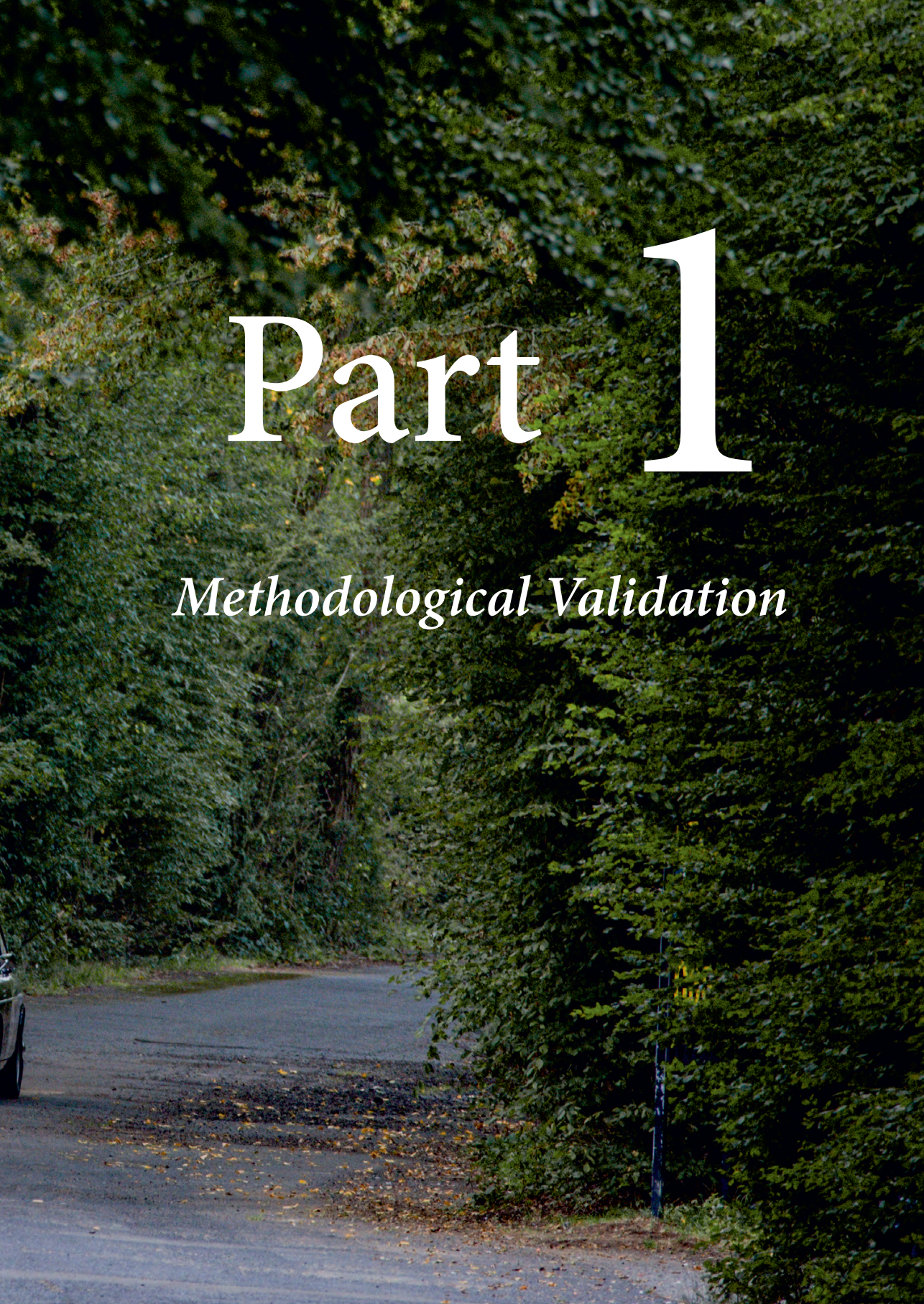
88. Kuhnert G, Boellaard R, Sterzer S, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. ***Eur J Nucl Med Mol Imaging.*** 2016;43:249-258.

89. Munk OL, Tolbod LP, Hansen SB, Bogsrud TV. Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative. ***EJNMMI Phys.*** 2017;4:5.

90. Kidera D, Kihara K, Akamatsu G, et al. The edge artifact in the point-spread function-based PET reconstruction at different sphere-to-background ratios of radioactivity. ***Ann Nucl Med.*** 2016;30:97-103.

91. Rahmim A, Tang J. Noise propagation in resolution modeled PET imaging and its impact on detectability. ***Phys Med Biol.*** 2013;58:6945-6968.


QI-1-PET



Part 1

Methodological Validation


QI-1-PET

Chapter 2

Accuracy and Precision of Partial-Volume Correction in Oncological PET/CT studies

M.C.F. Cysouw, G.M. Kramer, O.S. Hoekstra, V. Frings, A.J. de Langen, E.F. Smit, A.J.M. van den Eertwegh, D.E. Oprea-Lager, R. Boellaard

J Nucl Med. 2016 Oct;57(10):1642-1649.

Abstract

Accurate quantification of tracer uptake in small tumours using PET is hampered by the partial-volume effect as well as by the method of volume-of-interest (VOI) delineation. This study aimed to investigate the effect of partial-volume correction (PVC) combined with several VOI methods on the accuracy and precision of quantitative PET.

Methods: Four image-based PVC methods and resolution modeling (applied as PVC) were used in combination with several common VOI methods. Performance was evaluated using simulations, phantom experiments, and clinical repeatability studies. Simulations were based on a whole-body ^{18}F-FDG PET scan in which differently sized spheres were placed in lung and mediastinum. A National Electrical Manufacturers Association NU2 quality phantom was used for the experiments. Repeatability data consisted of an ^{18}F-FDG PET/CT study on 11 patients with advanced non–small cell lung cancer and an ^{18}F-fluoromethylcholine PET/CT study on 12 patients with metastatic prostate cancer.

Results: Phantom data demonstrated that most PVC methods were strongly affected by the applied resolution kernel, with accuracy differing by about 20%–50% between full-width-at-half-maximum settings of 5.0 and 7.5 mm. For all PVC methods, large differences in accuracy were seen among all VOI methods. Additionally, the image-based PVC methods were observed to have variable sensitivity to the accuracy of the VOI methods. For most PVC methods, accuracy was strongly affected by more than a 2.5-mm misalignment of true (simulated) VOI. When the optimal VOI method for each PVC method was used, high accuracy could be achieved. For example, resolution modeling for mediastinal lesions and iterative deconvolution for lung lesions were 99% ± 1.5% and 99% ± 0.9% accurate, respectively, for spheres 15–40 mm in diameter. Precision worsened slightly for resolution modeling and to a larger extent for some image-based PVC methods. Uncertainties in delineation propagated into uncertainties in PVC performance, as confirmed by the clinical data.

Conclusion: The accuracy and precision of the tested PVC methods depended strongly on VOI method, resolution settings, contrast, and spatial alignment of the VOI. PVC has the potential to substantially improve the accuracy of tracer uptake assessment, provided that robust and accurate VOI methods become available. Commonly used delineation methods may not be adequate for this purpose.

Introduction

Quantitative PET provides clinical oncology with a powerful tool for diagnosis, staging, restaging, and response monitoring (*1,2*). To allow for appropriate quantification of radioactive tracer uptake, PET data need to be corrected for several physical effects, including decay, scatter, random coincidences, and attenuation. An effect not regularly corrected for, but having a major impact on PET accuracy in small tumours, is the partial-volume effect (PVE) (*3*).

PVE originates from the finite spatial resolution of the PET scanner, described by the point spread function (PSF), and the tissue fraction effect (*4*). In hot lesions, PVE causes a net spill-out of activity into the background, leading to considerable underestimation of the measured activity concentration (*3-6*). Although clinical application of partial-volume correction (PVC) has led to contradictory results to date (*7*), both accurate and precise PVC methods may have a significant clinical impact and substantially change quantitative reads (*8*).

Many PVC methods have been developed (*4,7,9*), such as the recovery coefficient method (*5,6,10*), the geometric transfer matrix (*11*), the Müller-Gärtner method (*12*), and iterative deconvolution (*13,14*). However, each method has its limitations, and new methodology is still being developed. Some are adaptations of the recovery coefficient method (*15,16*), but others are more refined, such as resolution modeling (*17,18*), adaptations of iterative deconvolution (*19-21*), adaptations of the geometric transfer matrix (*22,23*), and background-adapted PVC algorithms (*24*).

Besides being affected by PVE, PET accuracy is strongly affected by the applied volume-of-interest (VOI) method, noise level, and tumour-to-background ratio (*25*). In addition, several PVC methods use predefined VOI boundaries to correct for PVE. Hoetjes et al. argued that the performance of PVC methods may benefit from exact (e.g., CT-based) VOI definition (*3*). We therefore hypothesized that PVC performance, and hence PET accuracy, is strongly affected by VOI definition methodology.

Because PVC performance is a function of not only the PVC method and settings but also the VOI method and settings, their interplay may affect the accuracy and precision of PVE-corrected quantitative PET metrics. In the present study, we investigated the effect of several combinations of PVC methods and VOI methods on the accuracy and precision of PET using phantoms and simulations. We also investigated the impact of PVC on the repeatability of ^{18}F-FDG and

^{18}F-fluoromethylcholine PET in patients with advanced non–small cell lung cancer and metastatic prostate cancer, respectively.

Materials and Methods

We used phantom experiments, simulations, and clinical data to evaluate PVC performance as a function of PVC method, VOI method, spatial kernel settings, noise-level, and alignment of VOI. The analyses are summarized in Supplemental Table 1, available at http://jnm.snmjournals.org.

PVC Methods

Four image-based PVC methods were applied: iterative deconvolution Lucy–Richardson PVC (*3,14*), background-adapted PVC (*24*) using local and global background regions, and mask-based spillover PVC (*3,4*). We optimized the spatial kernel settings using phantom data, setting the Gaussian kernel at 5.0–7.5 mm (0.5-mm intervals).

Reconstruction-Based PVC

We applied the resolution modeling (*17*) approach (PSF reconstruction) as part of the reconstruction process provided by the vendor (Philips Healthcare). The default settings were used with noise regularization (1 PSF iteration, 6-mm regularization), implemented within binary, large-object ordered-subset time-of-flight iterative reconstruction.

VOI Methods

The following threshold-based VOI methods (in-house–developed software (*26*) (26)) were applied to all data: 42% and 50% of the maximal voxel value, 42% and 50% of the maximal voxel value adapted for local background uptake, 50% and 70% of the peak value (i.e., average value of a 12-mm sphere positioned to yield the highest value) adapted for local background uptake, and iteratively defined background-adapted relative threshold level using the system PSF (*27*). In simulations, we also used the true sphere volume as VOI.

Phantom Experiments

We used a National Electrical Manufacturers Association NU2 quality phantom to calibrate the spatial resolution kernel for image-based PVC methods. The phantom contained 6 spheres with diameters ranging from 10 to 37 mm. Spheres and background were filled with ^{18}F-FDG solutions of 12.38 and 1.46 kBq/mL, respectively. A 30-min scan was obtained on an Ingenuity TF PET/CT scanner (Philips Healthcare). Reconstruction was performed using ordered-subset time-of-flight iterative reconstruction with and without resolution modeling.

Simulations (*25*)

A mathematic phantom was derived from an ^{18}F-FDG whole-body scan. Next, 10- to 40-mm-diameter spheres (5-mm intervals) were placed within mediastinum and lung. The voxel values within the spheres were set to 10 kBq/mL, providing local tumour-to-background ratios of about 6.7 and 3.3 for lung and mediastinum, respectively.

Using forward projection, we generated noise-free sinograms. In addition, we added noise to the sinograms using Poisson statistics simulating 3 noise levels, corresponding to data collected for 4, 3, and 2 min per bed position, as is typical for clinical practice. Noise-free images and images corresponding to data collected for 4, 3, and 2 min had liver uptake coefficients of variation of 6.2%, 13.2%, 13.6%, and 18.2%, respectively (as determined by a 3-cm spheric VOI placed in the right liver lobe). For each combination of sphere size and noise level, 10 sinograms were generated (except for noise-free sinograms).

Images were reconstructed using ordered-subset expectation maximization, with and without resolution modeling, and were post-smoothed with a 5-mm Gaussian filter. The number of iterations (6) and subsets (16) was set such as to ensure a minimal level of convergence and to avoid limited contrast recovery. In this way, PVE was affected mainly by the spatial resolution and voxel size.

Clinical Data

Clinical repeatability data consisted of an ^{18}F-FDG PET/CT study (*28*) on 11 patients with advanced non–small cell lung cancer and an ^{18}F-fluoromethylcholine PET/CT study (*29*) on 12 patients with metastatic prostate cancer. At the time the patients underwent PET, they received no treatment. Both studies were approved by the Medical Ethical Committee of the VU Medical Centre, and the patients gave informed consent to participate.

The patients were scanned using a Gemini TF-64 PET/CT scanner (Philips Healthcare). They fasted for 6 and 4 h before undergoing ^{18}F-FDG and ^{18}F-fluoromethylcholine PET, respectively. PET/CT scans were acquired at 60 and 40 min after injection of 185 MBq of ^{18}F-FDG and 200 MBq of ^{18}F-fluoromethylcholine, respectively. Images were reconstructed using ordered-subset time-of-flight iterative reconstruction with 3 iterations and 33 subsets, with and without resolution modeling. All data were corrected for decay, scatter, random coincidences, and attenuation.

PVC Performance Metrics

For the phantom experiment and simulations, accuracy was calculated using the recovery coefficient, defined as follows:

$$\text{Recovery coefficient} = \frac{\text{ACmeasured}}{\text{ACtrue}} \qquad \text{Eq. 1}$$

where ACmeasured is measured mean activity concentration (Bq/mL) and ACtrue is true (simulated) activity concentration (Bq/mL). Bias was calculated as follows:

$$\text{Bias} = \frac{(\text{ACmeasured} - \text{ACtrue})}{\text{ACtrue}} \qquad \text{Eq. 2}$$

For volumetric accuracy, recovery coefficient and bias were calculated in the same manner (volumes [mL] instead of activity concentrations).
Activity concentration ratios were defined as follows:

$$\text{Ratio} = \frac{\text{ACpvc}}{\text{ACuncorrected}} \qquad \text{Eq. 3}$$

where ACpvc is mean activity concentration with PVC and ACuncorrected is mean activity concentration without PVC.

SUVmean, normalized to body weight, was calculated for clinical data. Total lesion glycolysis (TLG) was calculated as SUVmean × metabolically active lesion volume (mL). All metrics were derived with and without PVC.

Statistical Analysis

The normality of SUVs and TLGs was assessed with the Shapiro–Wilks test. The intraclass correlation coefficient (ICC; 2-way mixed model with an absolute agreement definition) was calculated for each combination of VOI and PVC

method. For nonnormal distributions, log-transformed SUVmean and TLG were used to calculate ICC. Analyses were performed using SPSS Statistics, version 22.0 (IBM).

Results

Phantom Experiments

Image-based PVC methods required that the applied spatial kernel be optimized for each VOI method. For all VOI methods, Lucy-Richardson PVC and spillover PVC demonstrated differences in recovery coefficients ranging from 0.2 to 0.5 between full width at half maximum (FWHM) settings of 5.0–7.5 mm. The accuracy of global-background–adapted PVC was not affected by FWHM setting and for local-background–adapted PVC the recovery coefficients demonstrated differences of only 0.03 to 0.3 between FWHM settings of 5.0–5.5, 6.0, and 6.5–7.5 mm. Large differences in accuracy among the various VOI methods were seen for all image-based PVC methods, especially for the 13- and 17-mm spheres (typically yielding overcorrection). Even for the optimal FWHMs, the PVC methods still failed for the 10-mm sphere.

Volumetric accuracy was better in non-PSF reconstruction for the 17- to 37-mm spheres, excepting the 37-mm sphere delineated with 42% maximal (Figs. 2.1A and 2.1B). Notably, the 10- and 13-mm spheres were delineated more accurately using 42% maximal, 50% maximal, and relative threshold level in PSF reconstruction. The smallest differences in volumetric accuracy were seen for background-adapted VOIs. PET-based VOIs generated on PSF reconstructed images were smaller than those generated on non–PSF reconstructed images (Fig. 2.1C). No difference in volume was found for the 10-mm sphere delineated with background-adapted 42% maximal or background-adapted 50% maximal, whereas delineation with background-adapted 50% peak and background-adapted 70% peak provided negligibly larger volumes (0.064 mL larger).

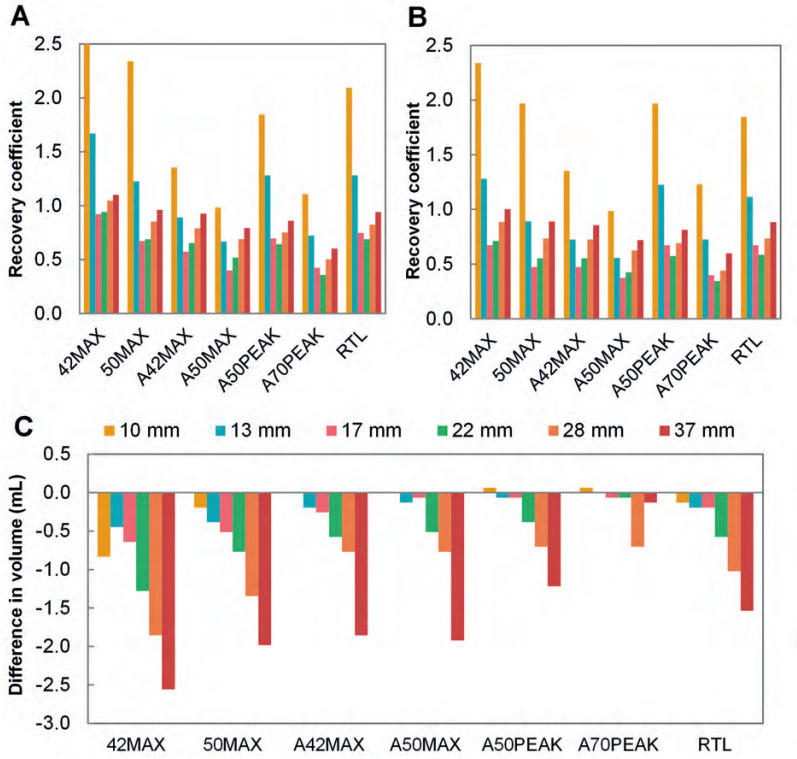


Figure 2.1: Volume recovery coefficients for non–PSF PVC images (A) and PSF PVC images (B) per VOI method and sphere size, and differences in PET-based volumes between non–PSF PVC images and PSF PVC images (C). Negative volume differences indicate smaller volumes for PSF reconstructed images than for non–PSF reconstructed images. Key indicates sphere diameters. Ten-millimeter sphere delineated with 42% maximal had recovery coefficient of 3.9 in non–PSF reconstructed images. 42MAX = 42% of maximal voxel value; 50MAX = 50% of maximal voxel value; A42MAX = 42% of maximal voxel value adapted for local background uptake; A50MAX = 50% of maximal voxel value adapted for local background uptake; A50PEAK = 50% of peak voxel value adapted for local background uptake; A70PEAK = 70% of peak voxel value adapted for local background uptake; RTL = relative threshold level.

Simulations

Large differences in PVC performance were seen among all VOI methods (Supplemental Figs. 1 and 2). The optimal combinations of PVC method and VOI method are shown in Figure 2.2 and Table 2.1. Generally, recovery coefficients were lower in lung than in mediastinum. For spheres 15 mm or larger, PSF reconstruction with adapted 70% peak yielded the highest accuracy in mediastinum (99% ± 1.5%), whereas Lucy–Richardson PVC with adapted 42% maximal yielded the highest accuracy in lung (99% ± 0.9%). Global- and local-background–adapted PVCs considerably overcorrected true activity concentration when using background-adapted 42% maximal, background-adapted 50%

maximal, background-adapted 50% peak, or background-adapted 70% peak. Both local-background–adapted PVC and spillover PVC performed excellently (100% accuracy overall) when using true (simulated) VOIs and were within 10% accurate when using relative-threshold-level VOIs (≥15 mm). Figure 2.3 demonstrates the percentage bias in sphere volumes in lung. We found a strong relationship between underestimation of true volume and overcorrection of activity concentration recovery coefficients for global- and local-background–adapted PVCs (recovery coefficients up to 3 and 2.25, respectively). Spillover PVC was moderately affected (recovery coefficients ≤ 1.33), and the recovery coefficients for Lucy–Richardson PVC and PSF reconstruction did not significantly correlate with a negative bias in volume (recovery coefficients, 0.6–1.05). There was a moderate inverse correlation between recovery coefficients and a positive bias in volume for all methods other than PSF reconstruction. Similar correlations were observed for mediastinal spheres, but bias in volume, and thus in activity concentration, was larger.

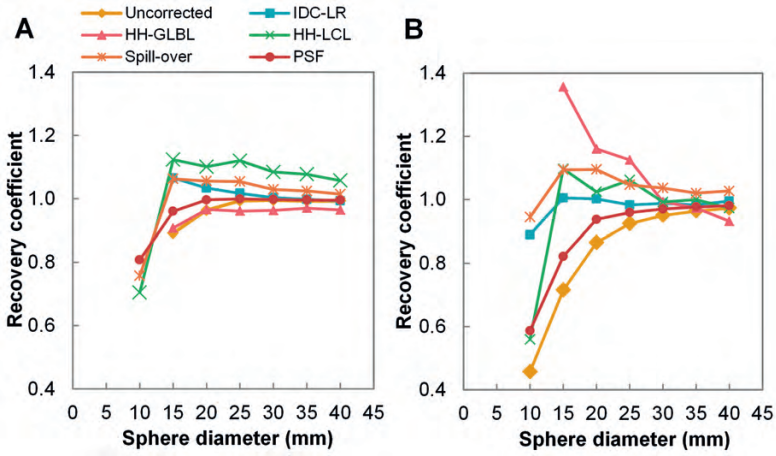


Figure 2.2: Activity concentration recovery coefficients as function of sphere diameter for all PVC methods, and uncorrected data, with their optimal PET-based VOI method (Table 1) for spheres in mediastinum (A) and lung (B). Missing values are due to delineation failure. HH-GLBL = global-background–adapted PVC; IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-LCL = local-background–adapted PVC.

Table 2.1: Optimal PET-based VOI method for each PVC method, in lung and mediastinum.

	Uncorrected	IDC-LR	HH-GLBL	HH-LCL	Spill-over	PSF
VOI method: (Mediastinum)	A70PEAK (97±4.1)	A50MAX (102±2.7)	50MAX (96±2.4)	RTL (109±2.6)	RTL (104±2.0)	A70PEAK (99±1.5)
VOI method: (Lung)	A70PEAK (90±9.8)	A42MAX (99±0.9)	42MAX (109±15.8)	50MAX (103±4.7)	A42MAX (105±3.3)	A70PEAK (94±6.0)

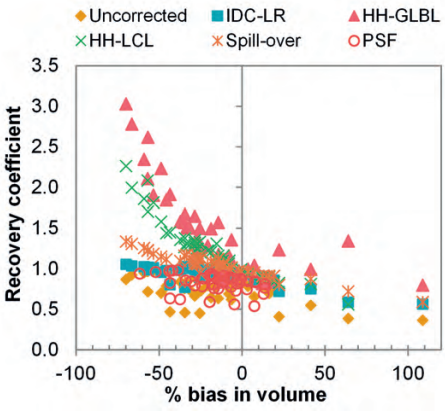


Figure 2.3: Activity concentration recovery coefficients as function of volumetric bias. Shown are results for all VOI methods for spheres in lung (noise-free images). HH-LCL = local-background–adapted PVC; IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-GLBL = global-background–adapted PVC.

Figure 2.4 illustrates activity concentration recovery coefficients as a function of misalignment of true VOI. Global-background–adapted PVC demonstrated only a slight decrease in recovery coefficient for a misalignment of 10 mm or more and was more than 94% accurate in lung but overcorrected by up to 20% in mediastinum. Local-background–adapted PVC was 98%–100% accurate when misalignment was less than 5 mm in lung and mediastinum. Spillover PVC performed slightly worse than local-background–adapted PVC. The performance of Lucy–Richardson PVC and PSF reconstruction was poorest when true VOI was used, but their sensitivity to misalignment was similar to that of local-background–adapted PVC and spillover PVC. Similar trends were obtained for all sphere sizes, but sensitivity to misalignment increased with decreasing sphere size.

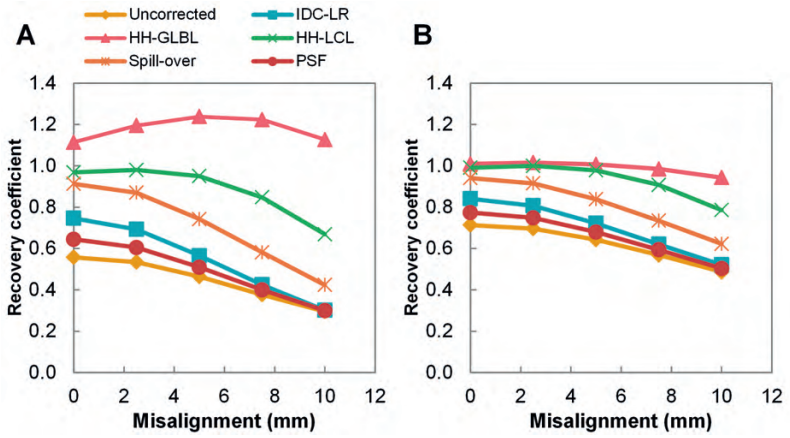


Figure 2.4: Activity concentration recovery coefficients as function of misalignment of true VOI. Shown are results from 15-mm (A) and 25-mm (B) spheres in lung (noise-free images). HH-LCL = local-background–adapted PVC; IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-GLBL = global-background–adapted PVC.

There was a positive association between noise level and recovery coefficient, with recovery coefficients becoming larger as VOI thresholds increased. The activity concentration ratios of mediastinal spheres increased with noise level for spillover PVC, global-background–adapted PVC, and local-background–adapted PVC when background-adapted VOIs were used, whereas in lung these ratios were equal for all noise levels (Fig. 2.5; similar but inverse trends were observed for volumes). In contrast, the results for noise-free images were similar to those for the highest noise level. With true (simulated) VOI, recovery coefficients were similar at all noise levels, both in mediastinum and in lung.

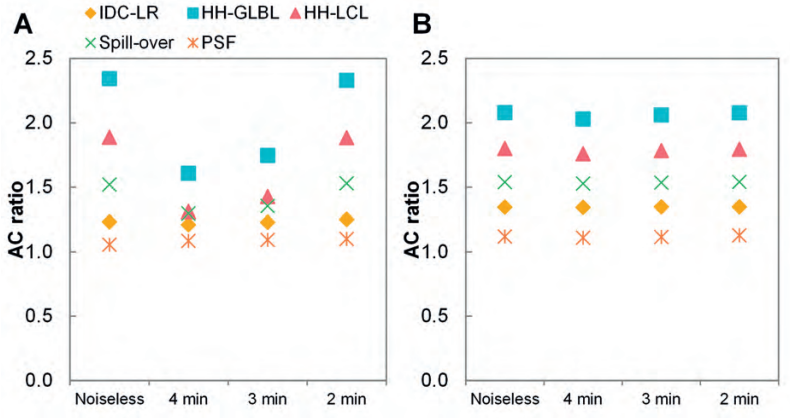


Figure 2.5: Activity concentration ratios as function of simulated acquisition time (thus, noise level). Shown are results from background-adapted 50% peak for 20-mm sphere (corresponding to median volumes of 18F-FDG and 18F-fluoromethylcholine PET cohorts delineated with background-adapted 50% peak) in mediastinum (A) and lung (B), respectively. AC = activity concentration; IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-GLBL = global-background–adapted PVC; HH-LCL = local-background–adapted PVC.

The impact of PVC on precision for spheres in lung is illustrated in Figure 2.6. In general, PVC increased SDs—an effect that was most pronounced for global- and local-background–adapted PVCs. Precision depended on the applied combination of VOI method and PVC method. When the true volume was used, the SDs were smallest, suggesting that uncertainties in PET-based VOI performance propagate into uncertainties in PVC performance. PET-based VOIs generally resulted in larger SDs in mediastinum than in lung.

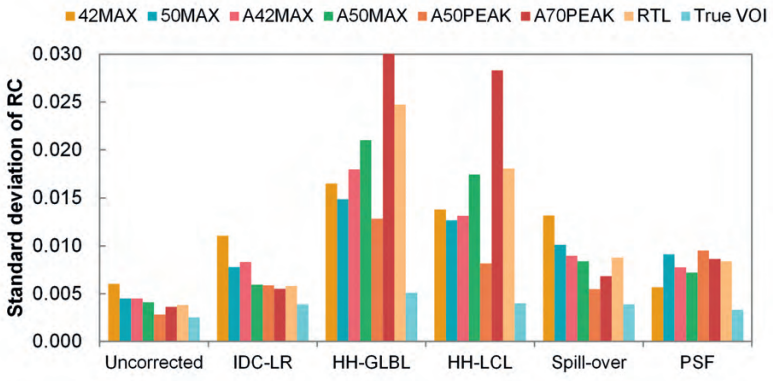


Figure 2.6: SDs of recovery coefficients for all combinations of PVC method. Shown are results for 20-mm spheres in lung (corresponding to median volumes of 18F-FDG and 18F-fluoromethylcholine PET cohorts delineated with background-adapted 50% peak). y-axis is scaled for visual interpretation; SD of global-background–adapted PVC using background-adapted 70% peak was 0.049. IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-GLBL = global-background–adapted PVC; HH-LCL = local-background–adapted PVC; 42MAX = 42% of maximal voxel value; 50MAX = 50% of maximal voxel value; A42MAX = 42% of maximal voxel value adapted for local background uptake; A50MAX = 50% of maximal voxel value adapted for local background uptake; A50PEAK = 50% of peak voxel value adapted for local background uptake; A70PEAK = 70% of peak voxel value adapted for local background uptake; RTL = relative threshold level.

Clinical Data

Table 2.2 describes the clinical cohorts. The feasibility (i.e., percentage of lesions successfully delineated) of the VOI methods was better in PSF reconstructed images than non–PSF reconstructed images (Supplemental Tables 2.2 and 2.3). Global-background–adapted PVC failed (providing negative activity concentrations) in 2.4% and 2.8% of lesions in the ^{18}F-FDG and ^{18}F-fluoromethylcholine PET cohorts, respectively.

Table 2.2: Patient characteristics. Median volumes determined with A50PEAK, the most accurate VOI method as determined in phantom experiment, on baseline.

			^{18}F-FDG-PET (28)	^{18}F-FCH-PET (29)
Type			NSCLC	mPC (n=4 castration-resistant)
No. of patients			11	12
No. of lesions			70	67
Age (mean±SD, years)			60±7	64±8
Gender			7 male, 4 female	12 male
Lesion localization			16 intrapulmonary, 54 extrapulmonary	44 bone metastases 23 lymph node metastases
Median volume (mL)		*Non-PSF*	3.94 (IQR 10.85)	5.76 (IQR 8.64)
		PSF	3.90 (IQR 20.10)	5.28 (IQR 7.92)

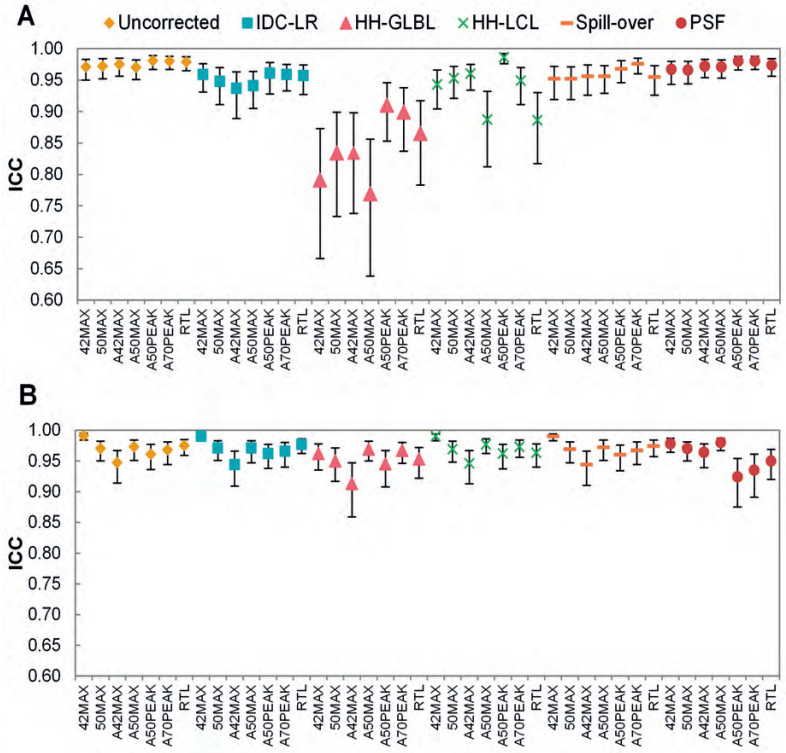


Figure 2.7: ICCs of SUVmean (A) and TLG (B) for all combinations of PVC method. Shown are results for ^{18}F-FDG PET cohort. Error bars represent 95% confidence intervals. Similar results were obtained for ^{18}F-fluoromethylcholine PET cohort (Supplemental Fig. 3). IDC-LR = iterative deconvolution Lucy–Richardson PVC; HH-GLBL = global-background–adapted PVC; HH-LCL = local-background–adapted PVC; 42MAX = 42% of maximal voxel value; 50MAX = 50% of maximal voxel value; A42MAX = 42% of maximal voxel value adapted for local background uptake; A50MAX = 50% of maximal voxel value adapted for local background uptake; A50PEAK = 50% of peak voxel value adapted for local background uptake; A70PEAK = 70% of peak voxel value adapted for local background uptake; RTL = relative threshold level.

ICCs were calculated to quantify, and facilitate comparison between, the repeatability of SUVmean and TLG (Fig. 2.7). Repeatability was best for uncorrected SUVmean (ICC, ~0.97–0.98), with comparable SUVmean ICCs for Lucy–Richardson PVC, spillover PVC, and PSF reconstruction. The ICCs for local-background–adapted PVC were slightly lower, depending on the VOI method. For all VOI methods, global-background–adapted PVC demonstrated the worst SUVmean repeatability (ICC, ~0.77–0.83). The SUVmean ICCs were comparable between VOI methods, except for global- and local-background–adapted PVCs. All PVE-corrected TLGs had ICCs almost equal to uncorrected TLG, except for PSF reconstruction. Similar trends in ICCs were seen among the volumes delineated with the various VOI methods (Supplemental Table 4) and their respective TLGs. Overall, ICCs were lower for the ^{18}F-fluoromethylcholine PET cohort than for the ^{18}F-FDG PET cohort.

Discussion

PVE introduced substantial error to the quantification of tracer uptake in mediastinal lesions smaller than 25 mm in diameter and lung lesions smaller than 30 mm. The current guidelines for response evaluation with PET do not include PVC (*2,30,31*). PERCIST (*2*) advises assessment of only tumours larger than 2 cm at baseline, to avoid overestimation of metabolic response with shrinkage during therapy, whereas the European Organization for Research and Treatment of Cancer (*31*) merely recommends documentation of tumour size in relation to scanner resolution. However, it is unclear how lesion selection strategies in metastasized disease affect the clinical performance of imaging biomarkers of response, especially in the case of targeted therapy with potentially heterogeneous inter- or intralesional target expression. Of note, the median volumes of lesions in the ^{18}F-FDG and ^{18}F-fluoromethylcholine cohorts corresponded to 20- to 22-mm equal-volume spheres—well within the range of lesions affected by PVE. PVE may also compromise diagnosis or prognosis when SUV-based thresholds are used in small tumours (*7*), even when guidelines for scanner calibration, image acquisition, and reconstruction are implemented (*32*). Taken together, these factors lead us to estimate that appropriate PVC may prove to be of greater clinical importance than considered so far. Our results demonstrated that PVC methods have the potential to be accurate and precise. However, the performance

of PVC depends heavily on the applied VOI method and factors influencing VOI method performance, such as lesion size, tumour-to-background ratio, noise, and spatial alignment. We recommend that the focus of future research into PVC be to develop robust and standardized PVC–VOI combinations and to assess their clinical impact using valid clinical reference standards.

Phantom and Simulation Studies

Adjustment of FWHM for the image-based PVC methods had a major effect on the performance of most methods. Lucy–Richardson PVC and spillover PVC substantially differed in accuracy between different FWHM settings, with the recovery coefficient increasing with the FWHM setting. The reason for this difference was most likely the fact that both methods directly use the applied FWHM for PVC, warranting accurate calibration. The performance of global-background–adapted PVC was equal for all FWHM settings, whereas local-background–adapted PVC differed for only some of the settings. For that PVC method, it is advisable that the FWHM not be underestimated, ensuring that the entire spill-out of signal is contained within the spill-out region, in accordance with the results of Hofheinz et al. (*24*).

VOIs were generated on both PSF reconstructed images and non–PSF reconstructed images. Therefore, differences in volume and volumetric accuracy between the two were assessed. In general, PSF reconstruction resulted in smaller VOIs, most likely because of improved tumour-to-background ratios and enhanced edges. However, volumetric accuracy was worse, apart from some VOIs generated on the smallest spheres.

In simulations, the performance of PVC differed between VOI methods. Recovery coefficients tended to be lower for spheres in lung than in mediastinum, which in the case of the simulated uniform spheres can be explained by a larger PVE in lung due to a higher tumour-to-background ratio. Without PVC, activity concentrations obtained with adapted 70% peak proved most accurate. This VOI method results in very small volumes, including only the core of spheres and thereby bypassing the PVE, which occurs mainly at lesion edges. PSF reconstruction increased accuracy by 2%–16%, with the increase being most pronounced for the smallest spheres. Even though accuracy was only moderately improved in lung, VOI methods tended to be more feasible on PSF reconstructed images (Supplemental Tables 2 and 3). Whereas Teo et al. found iterative deconvolution to perform optimally in a phantom study when an 80%

maximal VOI was applied (*13*), our simulation study suggested that Lucy–Richardson PVC performs excellently using background-adapted VOIs with a fixed threshold. Global- and local-background–adapted PVCs were sensitive to underestimation of volume, probably because of inclusion of the sphere activity concentration within the spill-out region, thus substantially overestimating the true activity concentration (Fig. 3). Overall, local-background–adapted PVC performed better than global-background–adapted PVC, most likely because the former can account for heterogeneity of activity within the background. Spillover PVC had excellent performance using relative threshold level and adapted 42% maximal, with accuracies of 104% ± 2.0% and 105% ± 3.3% for spheres 15 mm or larger in mediastinum and lung, respectively. Notably, when the true VOI was used, spillover PVC, local-background–adapted PVC, and global-background–adapted PVC performed excellently (accuracy, ~100%). This is understandable since theoretically, with homogeneous uptake, accuracy should be 100% when these methods are applied using perfect tumour boundaries and true FWHM. In addition, the true VOI demonstrated the highest precision. However, in the clinical setting, perfect alignment between CT and PET images is not realistic because of patient movement and breathing, and the CT-based anatomic volume may include nonviable tumour tissue. Thus, application of CT-based VOIs when using local-background–adapted PVC or spillover PVC may result in less accurate results because of sensitivity to misalignment of the VOI (Fig. 2.4) and inclusion of nonviable tumour tissue. Global-background–adapted PVC was unaffected by misalignment for spheres 20 mm or larger, most likely because of the large background region. However, some dependency on tumour-to-background ratio was seen using the true VOI (20% overcorrection in mediastinum).

PVC methods directly using VOI boundaries (i.e., spillover PVC, global-background–adapted PVC, and local-background–adapted PVC) differed considerably between noise levels for mediastinal spheres when background-adapted VOIs were applied. In contrast, similar performance was observed for all PVC methods at each noise level in lung, where high contrast resulted in very similar VOI delineations between noise levels. Thus, at low contrast, background-adapted VOIs become unreliable, propagating into an unreliable performance for PVC methods sensitive to volumetric accuracy.

PVC negatively affected precision to only a small extent (Fig. 2.6), except for global- and local-background–adapted PVCs, for which SDs increased considerably. Overall, background-adapted 50% peak seemed most precise when

image-based PVC was applied, most likely because peak values are less sensitive to noise than maximal values.

Clinical Studies

A previous study showed PVC to have no significant effect on tracer uptake repeatability, but only one PET-based VOI method (adapted 50% maximal) was used in that study (*3*). In the present study, the repeatability of SUVmean using the various VOI delineation methods was consistent—with comparable ICCs—after all types of PVC except for local-background–adapted and global-background–adapted PVC. Those two demonstrated large differences in ICC among various VOI methods and broader confidence intervals overall, illustrating worsened precision in accordance with the precisions observed in the simulations.

Erlandsson et al. proposed using PVE-corrected TLG in clinical settings, since uncorrected SUV might retain important volumetric information that is lost when SUV is corrected (*7*). Our results demonstrated that corrected TLG and uncorrected TLG had similar repeatability characteristics (Fig. 2.6). ICCs were similar among all PVC methods except for PSF reconstruction. The difference in the latter was most likely caused by the similarity between trends in the ICCs of TLG between VOI methods and trends in the ICCs of VOI volumes (Supplemental Table 4), emphasizing the importance of volumetric information to precision in PVC. For global- and local-background–adapted PVCs applied with their optimal VOI methods, PVE-corrected TLG might be suitable for acquiring data with optimal accuracy and precision.

Limitations

In phantoms and simulations, lesions are spheric and have homogeneous uptake. In reality, however, tumours rarely have spheric dimensions, let alone homogeneous uptake. Yet, we observed PVC-performance trends similar to those for clinical data, and the simulations allowed us to gain insight into the performance of PVC methods with the advantage of known (simulated) truth. In addition, the quantitative accuracy of PET can be negatively affected not only by PVE but by motion blurring due to breathing and peristaltic movement. To mitigate the effects of breathing, for example, respiration-gated PET/CT studies may be performed (*33*). Respiration-gated PET/CT is, however, not yet routinely applied in all centers.

Conclusion

We investigated the performance of PVC as a function of VOI delineation, resolution settings, tumour-to-background ratio, and noise. We conclude that the investigated PVC methods may greatly improve the quantitative accuracy of oncologic PET studies while maintaining good precision. However, the performance of PVC depends heavily on the VOI method and differs considerably between lung and mediastinum. For most image-based PVC methods, it is critical that there be no more than a 2.5-mm error in the spatial alignment of the VOI and tumour. Some methods that directly use predefined VOIs to correct PVE are less dependent on correct alignment but more sensitive to volumetric accuracy. Furthermore, uncertainties in PET-based VOIs propagate into uncertainties in PVC performance. PVC can substantially improve the accuracy with which tumours 15–25 mm in diameter are quantified in oncologic PET studies. However, without highly accurate and precise VOI methods, PVC may actually worsen accuracy and precision. Even with contemporary scanners and modern reconstruction methods, quantification of tracer uptake in tumours smaller than 15 mm in diameter is still not recommended.

References

1. Hoekstra CJ, Paglianiti I, Hoekstra OS, et al. Monitoring response to therapy in cancer using [^{18}F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. ***Eur J Nucl Med.*** 2000;27:731-743.
2. 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.
3. Hoetjes NJ, van Velden FH, Hoekstra OS, et al. Partial volume correction strategies for quantitative FDG PET in oncology. ***Eur J Nucl Med Mol Imaging.*** 2010;37:1679-1687.
4. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. ***J Nucl Med.*** 2007;48:932-945.
5. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. ***J Comput Assist Tomogr.*** 1979;3:299-308.
6. Geworski L, Knoop BO, de Cabrejas ML, Knapp WH, Munz DL. Recovery correction for quantitation in emission tomography: a feasibility study. ***Eur J Nucl Med.*** 2000;27:161-169.
7. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. ***Phys Med Biol.*** 2012;57:R119-159.
8. Kuhnert G, Boellaard R, Sterzer S, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. ***Eur J Nucl Med Mol Imaging.*** 2016;43:249-258.
9. Rousset O, Rahmim A, Alavi A, Zaidi H. Partial Volume Correction Strategies in PET. ***PET Clin.*** 2007;2:235-249.
10. Avril N, Dose J, Janicke F, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. ***J Clin Oncol.*** 1996;14:1848-1857.
11. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. ***J Nucl Med.*** 1998;39:904-911.
12. Muller-Gartner HW, Links JM, Prince JL, et al. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. ***J Cereb Blood Flow Metab.*** 1992;12:571-583.
13. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. ***J Nucl Med.*** 2007;48:802-810.
14. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. ***Neuroimage.*** 2008;39:1570-1584.
15. Gallivanone F, Canevari C, Gianolli L, et al. A partial volume effect correction tailored for ^{18}F-FDG-PET oncological studies. ***Biomed Res Int.*** 2013;2013:780458.
16. Krempser AR, Ichinose RM, Miranda de Sa AM, Velasques de Oliveira SM, Carneiro MP. Recovery coefficients determination for partial volume effect correction in oncological PET/CT images considering the effect of activity outside the field of view. ***Ann Nucl Med.*** 2013;27:924-930.
17. Rahmim A, Qi J, Sossi V. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. ***Med Phys.*** 2013;40:064301.
18. Wallsten E, Axelsson J, Sundstrom T, Riklund K, Larsson A. Subcentimeter tumor lesion delineation for high-resolution ^{18}F-FDG PET images: optimizing correction for partial-volume effects. ***J Nucl Med Technol.*** 2013;41:85-91.
19. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. ***Eur J Nucl Med Mol Imaging.*** 2009;36:1064-1075.
20. Bhatt R, Adjouadi M, Goryawala M, Gulec SA, McGoron AJ. An algorithm for PET tumor volume and activity quantification: without specifying camera's point spread function (PSF). ***Med Phys.*** 2012;39:4187-4202.
21. Merlin T, Visvikis D, Fernandez P, Lamare F. A novel partial volume effects correction technique integrating deconvolution associated with denoising within an iterative PET image reconstruction. ***Med Phys.*** 2015;42:804-819.

22. Sattarivand M, Kusano M, Poon I, Caldwell C. Symmetric geometric transfer matrix partial volume correction for PET imaging: principle, validation and robustness. ***Phys Med Biol.*** 2012;57:7101-7116.

23. Sattarivand M, Armstrong J, Szilagyi GM, Kusano M, Poon I, Caldwell C. Region-Based Partial Volume Correction Techniques for PET Imaging: Sinogram Implementation and Robustness. ***Int J Mol Imaging.*** 2013;2013:435959.

24. Hofheinz F, Langner J, Petr J, et al. A method for model-free partial volume correction in oncological PET. ***EJNMMI Res.*** 2012;2:16.

25. 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.

26. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. ***Radiology.*** 2014;273:539-548.

27. van Dalen JA, Hoffmann AL, Dicken V, et al. A novel iterative method for lesion delineation and volumetric quantification with FDG PET. ***Nucl Med Commun.*** 2007;28:485-493.

28. Kramer GM, Frings V, Hoetjes N, et al. Repeatability of quantitative uptake measures of whole body [^{18}F]FDG PET/CT in NSCLC patients. ***J Nucl Med.*** 2015;56(suppl 3):1379.

29. Oprea-Lager DE, Kramer G, van de Ven PM, et al. Repeatability of Quantitative ^{18}F-Fluoromethylcholine PET/CT Studies in Prostate Cancer. ***J Nucl Med.*** 2016;57:721-727.

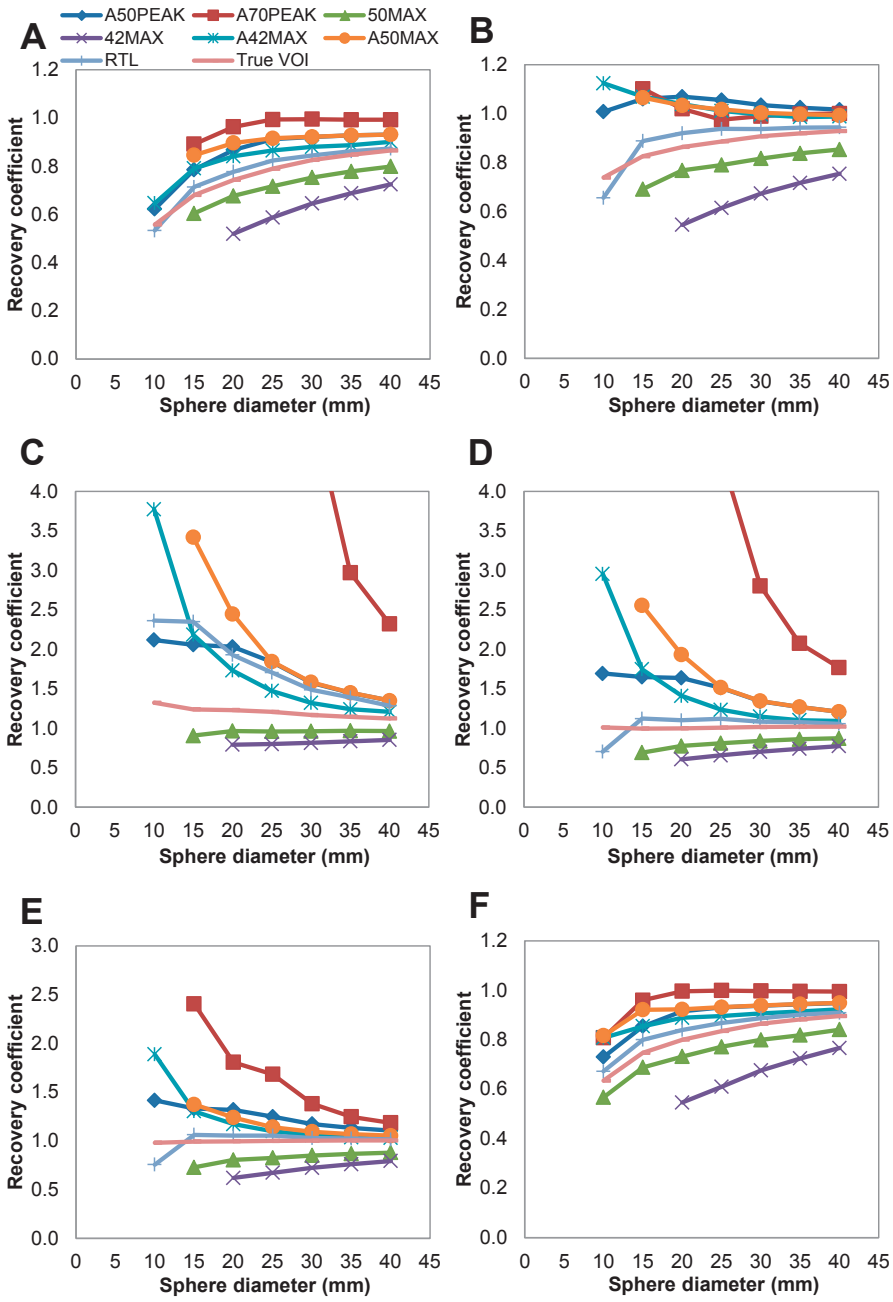
30. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [^{18}F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. ***Eur J Cancer.*** 1999;35:1773-1782.

31. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of ^{18}F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. ***J Nucl Med.*** 2006;47:1059-1066.

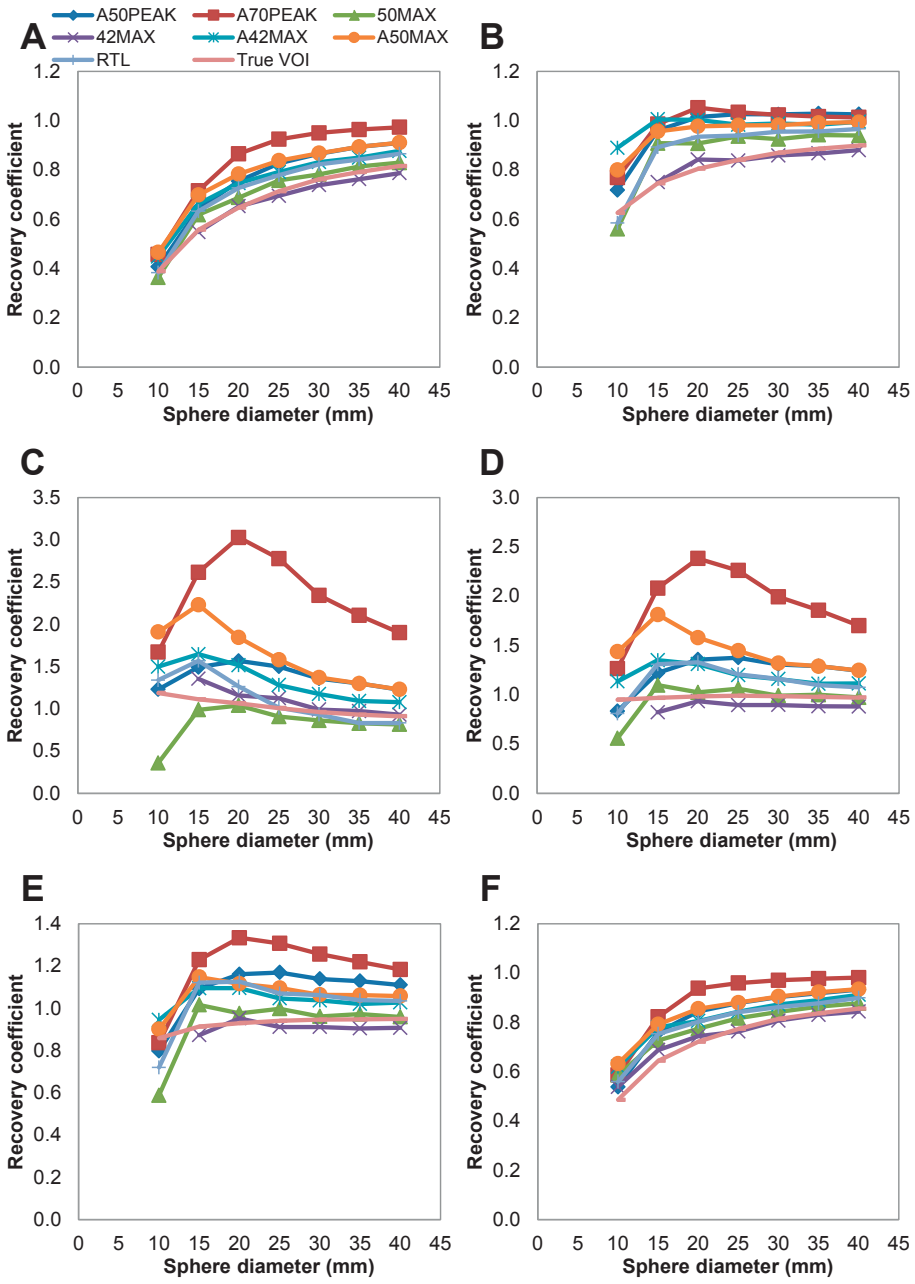
32. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.

33. Daouk J, Fin L, Bailly P, Meyer ME. Respiratory-gated positron emission tomography and breath-hold computed tomography coupling to reduce the influence of respiratory motion: methodology and feasibility. ***Acta Radiol.*** 2009;50:144-155.

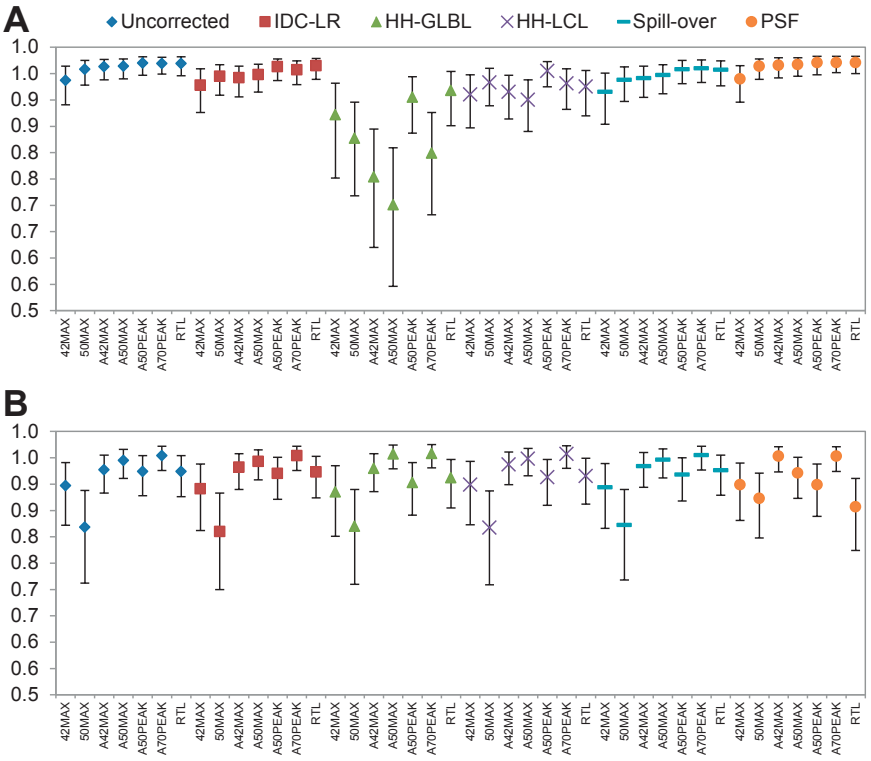
Supplemental files



Supplemental Figure 2.1: Recovery coefficients as a function of sphere diameter: uncorrected (A), IDC-LR (B), HH-GLBL (C), HH-LCL (D), spill-over (E), and PSF-reconstruction (F) of spheres in mediastinum, on noise-less images. Please note: RCs of A70PEAK exceed the y-axis maximum, up to 11.8 and 8.2 in graphs C and D, respectively. Missing values are due to failure of VOI delineation.



Supplemental Figure 2.2: Recovery coefficients as a function of sphere diameter: uncorrected (A), IDC-LR (B), HH-GLBL (C), HH-LCL (D), spill-over (E), and PSF reconstruction (F) of spheres in lung, on noiseless images. Missing values are due to failure of VOI delineation.



Supplemental Figure 2.3: ICCs of SUVmean (A) and TLG (B) of all combinations of VOI and PVC method. Shown are the results of the 18F-FCH-PET cohort. Error-bars represent 95% confidence intervals.

Supplemental Table 2.1: Summary of performed analyses and the respective datasets used.

Analysis	Sub-analysis	Data
Resolution kernel calibration	-	Phantom experiment
Volumetric accuracy non-PSF- vs. PSF-reconstruction	-	Phantom experiment
Accuracy of PVC methods	Effect of: – VOI method – Volumetric accuracy – True VOI misalignment – Noise-level	Simulations
Precision of PVC and VOI methods	-	Simulation experiment and clinical data

Supplemental Table 2.2: VOI method feasibility (% of 70 lesions successfully delineated) and median SUVmean and TLG for all combinations of VOI and PVC method of the ^{18}F-FDG-PET cohort. *numbers in parentheses represent IQR.

		Scan 1			Scan 2		
VOI	PVC	feasibility (% of 70)	Median SUVmean*	Median TLG*	feasibility (% of 70)	Median SUVmean*	Median TLG*
42MAX	Uncorrected	81.4	6.2 (3.5)	41.8 (211.2)	81.4	5.9 (3.9)	42.1 (234.2)
	IDC-LR		8.2 (6.2)	57.6 (245.8)		7.7 (5.9)	54.6 (264.9)
	HH-GLBL		9.1 (8)	65.3 (304.5)		9.4 (7.2)	73.6 (310.6)
	HH-LCL		8.8 (6)	63.8 (264)		8.5 (6.6)	61.6 (277.6)
	Spill-over		9.2 (7)	67.1 (269.2)		9.3 (7.2)	64.8 (282.8)
	PSF	92.9	7.4 (3.9)	36.4 (225.1)	88.6	7.3 (4.6)	37.9 (215)
50MAX	Uncorrected	94.3	6.8 (3.5)	28.4 (165.8)	84.3	6.3 (4.4)	27.8 (156.4)
	IDC-LR		9.5 (7)	44 (194.7)		8.7 (6.7)	38.6 (193.6)
	HH-GLBL		11.2 (9.6)	60.5 (242.6)		12.2 (11.4)	60.8 (254.7)
	HH-LCL		10.5 (7.8)	50.5 (224.8)		10.5 (8.2)	47 (217.2)
	Spill-over		10.1 (7.1)	46.9 (210.8)		9.9 (7.5)	42.1 (214.7)
	PSF	97.1	8.1 (4.6)	21.4 (145.6)	91.4	7.9 (5.4)	18.6 (112.7)
A42MAX	Uncorrected	94.3	6.7 (3.5)	18.9 (131.2)	91.4	6.4 (4.1)	18.9 (77.5)
	IDC-LR		9.9 (6.4)	30.4 (161.8)		9.1 (5.4)	29.5 (93.7)
	HH-GLBL		13.4 (8.1)	54.5 (206.5)		13.1 (6.7)	43.4 (185.8)
	HH-LCL		11 (6.9)	35.9 (181.9)		11.3 (6.2)	34 (118.1)
	Spill-over		10.1 (6.3)	29.9 (175.1)		9.5 (5.6)	29.1 (99.6)
	PSF	97.1	8.1 (4.3)	17.7 (109)	95.7	7.8 (5)	19.5 (54.3)
A50MAX	Uncorrected	97.1	7.4 (3.8)	13.4 (59.8)	91.4	7 (4.4)	13.9 (37.8)
	IDC-LR		10.3 (6.1)	19.9 (71)		9.4 (5.3)	19 (43.7)
	HH-GLBL		17.6 (10.7)	41 (164.5)		19.5 (12.7)	40.5 (124.5)
	HH-LCL		14 (8.9)	30.1 (100.6)		15.2 (8.1)	30.2 (71.9)
	Spill-over		10.1 (5.7)	19.3 (75.1)		10.3 (5.7)	19.7 (47.9)
	PSF	100.0	9 (4.7)	12.4 (34.2)	94.3	8.8 (5.3)	13.7 (31.6)

A50PEAK	Uncorrected	90.0	6.3 (3.8)	24.8 (131.3)	88.6	6 (4.1)	24.8 (116.7)
	IDC-LR		9 (5.3)	36.8 (162)		8.1 (5)	37.4 (129.5)
	HH-GLBL		11.1 (6.4)	50.1 (202.6)		11.1 (6.5)	50.5 (201.4)
	HH-LCL		9.4 (6.6)	43.1 (188.3)		9.3 (6)	38.9 (173.8)
	Spill-over		8.9 (5.6)	38.7 (181.8)		8.8 (5.5)	38.9 (161.6)
	PSF	97.1	7.3 (4.1)	28.3 (166.1)	95.7	7.1 (4.7)	27.3 (120)
A70PEAK	Uncorrected	97.1	7.3 (4.2)	14.8 (34.9)	91.4	7.1 (4.6)	13.2 (25.4)
	IDC-LR		9.9 (5.8)	21.5 (42.4)		9.5 (5.4)	17.8 (29.3)
	HH-GLBL		20.5 (12.9)	37.5 (113.6)		21.3 (13.2)	40.6 (84.8)
	HH-LCL		14.6 (8.8)	30.8 (78.7)		14.7 (8.7)	26.8 (51)
	Spill-over		10 (5.6)	21.3 (48.8)		9.8 (5.8)	18.1 (33.6)
	PSF	100.0	8.2 (5.2)	17.2 (38.7)	95.7	8.2 (5.6)	13.1 (25.4)
RTL	Uncorrected	88.6	6 (3.6)	36.1 (180.8)	90.0	5.7 (4.2)	35.4 (189.2)
	IDC-LR		8.4 (6)	53 (206.3)		7.8 (5.3)	50.8 (206.8)
	HH-GLBL		8.9 (6.3)	63.7 (248)		8.8 (6.9)	64.7 (265.9)
	HH-LCL		8.9 (6.3)	61.5 (243.7)		8.1 (7.2)	61.2 (240.2)
	Spill-over		9.6 (6.4)	60.8 (234.1)		8.6 (6.2)	60.3 (250.8)
	PSF	97.1	7.7 (4.1)	31.2 (160.7)	94.3	7.4 (4.7)	25.3 (142.4)

Supplemental Table 2.3: VOI method feasibility (% of 67 lesions successfully delineated and median SUVmean and TLG for all combinations of VOI and PVC method of the ^{18}F-FCH-PET cohort. *numbers in parentheses represent IQR.

		Scan 1			Scan 2		
VOI	PVC	feasibility (% of 67)	Median SUVmean*	Median TLG*	feasibility (% of 67)	Median SUVmean*	Median TLG*
42MAX	Uncorrected	74.6	6.4 (3.4)	52.6 (92.7)	79.1	6.6 (3.9)	54.2 (90)
	IDC-LR		8.5 (4.8)	70.9 (109.5)		9.2 (5.5)	73.2 (108.7)
	HH-GLBL		10.4 (4.6)	71.6 (127.5)		11.2 (5.9)	92.5 (150.9)
	HH-LCL		9 (5.4)	76.8 (124.9)		9.7 (5.6)	86.7 (131.5)
	Spill-over		9.3 (5.1)	78 (120.8)		9.8 (6)	83.2 (126.8)
	PSF	84.3	7.7 (4.6)	55.7 (87.7)	80.6	7.4 (4.5)	50.1 (70.7)
50MAX	Uncorrected	92.5	6.4 (4)	43.6 (73.6)	89.6	6.4 (3.7)	45.2 (56.6)
	IDC-LR		9.1 (5.9)	62.5 (95.8)		9.3 (6)	60.1 (71.7)
	HH-GLBL		13.8 (8.3)	77.3 (135.3)		14.2 (10.3)	82 (134.1)
	HH-LCL		10.7 (7)	73.6 (113.8)		11.1 (8.3)	77.1 (95.3)
	Spill-over		9.8 (6)	66.3 (106.7)		9.8 (7)	65.4 (82.4)
	PSF	94.3	7.7 (4.6)	40 (66.6)	92.5	7.6 (5.9)	39.6 (51.1)
A42MAX	Uncorrected	95.5	6.5 (4)	25.8 (51.4)	95.5	6.3 (4.2)	27.5 (51.4)
	IDC-LR		9.3 (5.5)	33.6 (66)		9.2 (6.2)	35.4 (66.9)
	HH-GLBL		13.5 (7.1)	60.7 (104.6)		14.3 (9.4)	67.4 (107)
	HH-LCL		11.2 (6.8)	43 (83.5)		11.2 (7.2)	47.9 (81.2)
	Spill-over		9.3 (5.9)	35.6 (70.6)		9.2 (6)	38.2 (68.3)
	PSF	95.7	7.9 (4.8)	27.1 (59.9)	95.5	7.6 (5.5)	28.2 (48.5)
A50MAX	Uncorrected	95.5	7.2 (4.4)	17.7 (35.5)	97.0	6.8 (4.6)	19.3 (31.3)
	IDC-LR		9.2 (5.5)	21.4 (42.8)		9.3 (6.4)	25.5 (39.5)
	HH-GLBL		18.6 (10.4)	49.6 (74.7)		20.6 (11.1)	45.5 (71.2)
	HH-LCL		14.2 (7.2)	32.2 (63.4)		14.4 (8.4)	37.2 (62.9)
	Spill-over		9.3 (5.4)	22.5 (46.5)		9.3 (6.4)	25.4 (40.9)
	PSF	95.7	8.9 (5.1)	19 (30.4)	97.0	8.2 (6)	19.5 (33.1)

A50PEAK	Uncorrected	94.0	6 (3.8)	35.4 (61.4)	94.0	5.9 (3.8)	35.5 (58.7)
	IDC-LR		8.4 (5.1)	47.9 (83.6)		8.3 (5.3)	48 (76.7)
	HH-GLBL		11.3 (6.3)	68 (116.5)		12.7 (6)	73.3 (110.2)
	HH-LCL		9.7 (5.4)	59.7 (92.9)		9.6 (6.3)	60.1 (85.2)
	Spill-over		8.8 (5.3)	52.3 (85.1)		8.8 (5.5)	52.2 (80.7)
	PSF	95.7	7.2 (4.7)	41.6 (69.9)	95.5	7.1 (4.6)	33 (60)
A70PEAK	Uncorrected	97.0	7.4 (4.2)	15.1 (26)	98.5	7.3 (4.7)	14.4 (26.5)
	IDC-LR		9.6 (5.1)	20 (33.8)		9.3 (5.7)	19.8 (35.8)
	HH-GLBL		21.9 (10.9)	47 (75)		24.5 (11.1)	41.2 (62.9)
	HH-LCL		14.9 (8.6)	33.4 (51.6)		14.7 (9.4)	32.3 (58.9)
	Spill-over		10.3 (5.8)	20.9 (34.6)		9.5 (6.6)	19.9 (37.4)
	PSF	97.1	8.9 (5.1)	5.4 (30.8)	98.5	8.5 (5.4)	17.4 (28.5)
RTL	Uncorrected	91.0	6.1 (3.5)	43.6 (67.4)	94.0	6 (3.6)	44.8 (64.2)
	IDC-LR		8.4 (5)	58.1 (84.7)		8.3 (5.4)	60.9 (81.1)
	HH-GLBL		11.1 (7.4)	74.4 (123.3)		12 (7.8)	79.1 (110.4)
	HH-LCL		11.3 (6.4)	72.7 (127.9)		11 (7.5)	79.6 (110.4)
	Spill-over		9.4 (5.1)	68.3 (94)		9.5 (5.9)	70.4 (93.2)
	PSF	94.3	7.3 (4.5)	46.3 (73.9)	95.5	7.1 (5)	45.3 (62.6)

Supplemental Table 2.4: ICCs of VOI volumes generated on non-PSF- and PSF-reconstructed images in the 18F-FDG-PET cohort. 95%-CI in parentheses.

	42MAX	50MAX	A42MAX	A50MAX	A50PEAK	A70PEAK	RTL
Non-PSF	0.98 (0.97-0.99)	0.96 (0.93-0.97)	0.92 (0.87-0.95)	0.96 (0.94-0.98)	0.94 (0.90-0.96)	0.95 (0.92-0.97)	0.96 (0.94-0.98)
PSF	0.97 (0.94-0.98)	0.95 (0.92-0.97)	0.95 (0.91-0.97)	0.97 (0.95-0.98)	0.88 (0.81-0.93)	0.89 (0.82-0.93)	0.93 (0.88-0.95)


QI-1-PET

Chapter 3

Partial-Volume Correction in Dynamic PET-CT: Effect on Tumor Kinetic Parameter Estimation and Validation of Simplified Metrics

M.C.F. Cysouw, S.V.S. Golla, V. Frings, E.F. Smit, O.S. Hoekstra, G.M. Kramer, R. Boellaard, and on behalf of the QuIC-ConCePT Consortium

EJNMMI Res. 2019 Feb 4;9(1):12.

Abstract

Background: Partial-volume effects generally result in an underestimation of tumour tracer uptake on PET-CT for small lesions, necessitating partial-volume correction (PVC) for accurate quantification. However, investigation of PVC in dynamic oncological PET studies to date is scarce. The aim of this study was to investigate PVC's impact on tumour kinetic parameter estimation from dynamic PET-CT acquisitions and subsequent validation of simplified semi-quantitative metrics. Ten patients with EGFR-mutated non-small cell lung cancer underwent dynamic ^{18}F-fluorothymidine PET-CT before, 7 days after, and 28 days after commencing treatment with a tyrosine kinase inhibitor. Parametric PVC was applied using iterative deconvolution without and with highly constrained backprojection (HYPR) denoising, respectively. Using an image-derived input function with venous parent plasma calibration, we estimated full kinetic parameters VT, K1, and k3/k4 (BPND) using a reversible two-tissue compartment model, and simplified metrics (SUV and tumour-to-blood ratio) at 50–60 min post-injection.

Results: PVC had a non-linear effect on measured activity concentrations per timeframe. PVC significantly changed each kinetic parameter, with a median increase in VT of 11.8% (up to 25.1%) and 10.8% (up to 21.7%) without and with HYPR, respectively. Relative changes in kinetic parameter estimates vs. simplified metrics after applying PVC were poorly correlated (correlations 0.36–0.62; $p < 0.01$). PVC increased correlations between simplified metrics and VT from 0.82 and 0.81 ($p < 0.01$) to 0.90 and 0.88 ($p < 0.01$) for SUV and TBR, respectively, albeit non-significantly. PVC also increased correlations between treatment-induced changes in simplified metrics vs. VT at 7 (SUV) and 28 (SUV and TBR) days after treatment start non-significantly. Delineation on partial-volume corrected PET images resulted in a median decrease in metabolic tumour volume of 14.3% (IQR −22.1 to −7.5%), and increased the effect of PVC on kinetic parameter estimates.

Conclusion: PVC has a significant impact on tumour kinetic parameter estimation from dynamic PET-CT data, which differs from its effect on simplified metrics. However, it affected validation of these simplified metrics both as single measurements and as biomarkers of treatment response only to a small extent. Future dynamic PET studies should preferably incorporate PVC.

Background

In clinical oncology positron-emission tomography (PET) is a valuable tool allowing guidance of treatment on a per-patient basis (*1*). Clinical decision-making using PET-CT is commonly limited to visual analysis, where local disease and the presence of nodal or distant metastases is evaluated (*2,3*). However, since PET is an inherently quantitative technique, it may also be used for quantitative assessment of tumour metabolic, proliferative, or drug targeting characteristics (*1,4,5*).

For quantitative PET-CT to be of practical clinical utility, metrics need to be easily extracted from static whole-body PET-CT images as performed in routine clinical practice. To this end, standardized uptake values (SUV) are typically used as simplified semi-quantitative measures of tracer uptake (*6*). However, pharmacokinetic modeling using dynamic PET-CT acquisitions with arterial or venous blood sampling is an essential first step to technically validate the clinical use of these simplified metrics as biomarkers of, e.g., response to treatment (*4,5,7,8*).

As is well known, quantification of tracer distribution on PET-CT scans is hampered by several sources of error. Amongst these are attenuation, Compton scatter, random coincidences, and decay, all accounted for by contemporary image reconstruction algorithms. However, due to the inherently limited spatial resolution of PET-CT, acquired images still suffer from partial-volume effects (*9*). Partial-volume effects lead to spill-in and spill-out of measured activity distributions, generally resulting in net underestimations of tracer uptake, the extent of which depend on tumour size, shape, and contrast (*9*). Hence, partial-volume correction (PVC) is needed for accurate quantification, especially for small and/or heterogeneous lesions (*9-12*).

In oncological studies, PVC has been predominantly applied to static PET-CT images (in contrast with brain (*13-22*) or cardiac (*23,24*) PET imaging). However, in dynamic acquisitions the activity spill-over in and from tumours due to partial-volume effects may vary over time. The impact of PVC on tumour kinetic parameter estimates could therefore differ from its impact on simplified measures of uptake. Consequently, it may not only affect absolute quantitative reads, but also validation of simplified parameters for clinical implementation.

The present study aims to evaluate the impact of frame-wise parametric PVC in dynamic PET-CT studies on tumour kinetic micro- and macroparameter

estimations, and evaluate the correlation between its effect on kinetic parameters and simplified metrics. Secondly, PVC's effect on technical validation of simplified ^{18}F-fluorothymidine (^{18}F-FLT) PET-CT metrics as biomarkers of response to treatment of non-small cell lung cancer (NSCLC) with tyrosine kinase inhibitors (TKI) will be investigated.

Methods and Materials

Patients

The present study is a retrospective analysis of a prospective cohort study (*5*). Patients with metastatic epidermal growth factor receptor (EGFR) mutated NSCLC scheduled for treatment with an EGFR-TKI were included. All patients were scanned with ^{18}F-FLT PET-CT on three occasions: at baseline, 7 days after, and 28 days after commencing treatment with a TKI (gefitinib or erlotinib), respectively. The Amsterdam UMC (location VUmc) institutional review board approved this study (Dutch Trial Register, NTR3557) and all included patients provided informed consent for study participation.

PET-CT image acquisition and reconstruction

The EARL-compliant imaging protocol was described previously (*5*). All scans were acquired on a Philips Gemini TF-64 PET-CT scanner (Philips Healthcare). Patients were instructed not to eat 4 hours prior to each scan. A thoracic field of view was placed such that it contained the primary tumour, using a transmission scan for positioning. A 60min dynamic PET acquisition started directly after injection of 370MBq ^{18}F-FLT in 5mL saline (flushed with 20mL saline). Afterwards, a low-dose CT was acquired for attenuation correction (120kV, 50mAs). The PET emission scan was binned into 36 frames with varying durations (1x10, 8x5, 4x10, 3x20, 5x30, 5x60, 4x150, 4x300 and 2x600 seconds). Images were reconstructed with a time-of-flight 3D row action maximum likelihood algorithm (3 iterations, 33 subsets), as provided by the vendor, with corrections for Compton scatter, random coincidences, attenuation, and decay. PET image dimensions were 144x144x45 voxels with voxel dimensions of 4x4x4 mm. Venous blood samples were drawn at 5, 10, 20, 30, 40, and 60min post-injection of ^{18}F-FLT. From each sample, the whole blood and plasma activity concentrations and parent fractions were measured.

Image processing

For PVC we applied a post-reconstruction iterative deconvolution algorithm (Lucy-Richardson [LR]) (*25*). This parametric (voxel-wise) method aims to deblur images by iteratively correcting the activity spill-over, only assuming approximate knowledge of the PET-CT scanner's spatial resolution. We set the full-width at half-maximum (FWHM) of a spatially invariant Gaussian point spread function at 7.5mm, as previously calibrated in phantom experiment for the used scanner (*11*), with 10 iterations allowing for sufficient convergence. PVC was applied to each image frame. As iterative deconvolution is known result in lower signal-to-noise ratios (SNR), in order to evaluate effect of image noise we additionally applied a highly constrained backprojection (HYPR) algorithm shown to improve SNR for dynamic PET studies (*26,27*). Iterative deconvolution was applied without and with HYPR denoising (denotated as LR and LR+HYPR, respectively). HYPR settings were optimized, comparing a single composite image ($HYPR_{single}$) and several moving frame composite images ($HYPR_{moving}$), using a Gaussian 7.5mm FWHM low-pass filter (F). The HYPR implementation can be described as follows (*21,26*):

$$I_H = I_c \times I_w \quad \text{Eq.1}$$

$$I_c = \sum I_i \times \Delta t_i \quad \text{Eq.2}$$

$$I_w = \frac{F \otimes I_o}{F \otimes I_c} \quad \text{Eq.3}$$

where I_H is the HYPR image; I_c is the composite image, which is a duration weighted summed average of either all frames in the dynamic image ($HYPR_{single}$) or a set of frames around the to be denoised frame ($HYPR_{moving}$), with Δt_i as the individual frame duration; I_o is the original dynamic frame being denoised; and I_w is the weighting image computed as the ratio between the spatially filtered original frame and spatially filtered composite image.

Kinetic modeling and semi-quantitative analysis

Lesions were delineated using in-house developed software (VU University Medical Center) on a volume-of-interest (VOI) basis (*28*). Tumour delineation was performed on a summation of the last 3 PET frames of the original (non-PVC) image. In short, a rough manual delineation was performed, warranting all peak ^{18}F-FLT-avid tumour activity was contained in the VOI and no non-

tumour structures with high uptake were included. Second, this VOI was shrunk to an isocontour based on 50% of the peak value (mean activity in a 12mm sphere positioned to provide the highest uptake value), with correction for local background activity. VOIs were then projected onto each frame of both the original and partial-volume corrected PET images to acquire time activity curves from both the datasets (without and with PVC). To explore the effect of PVC on tumour delineation, tumours were also delineated on the LR+HYPR images using the same approach. Metabolically active tumour volume (MATV) was defined as the sum of voxel volumes within a VOI.

A 2x2 voxel (8x8 mm) region was placed centrally in ascending aorta on 5 adjacent slices to acquire an image-derived input function (IDIF), aiming to avoid partial-volume effects. Parent plasma input functions were generated by calibrating IDIFs using the activity concentrations measured in the venous blood samples, and correcting for metabolites and plasma-to-blood ratio. Full quantitative parameters derived from kinetic modeling and simplified measures were extracted using in-house developed software in MATLAB. We used a reversible two-tissue model with blood volume parameter, which has been identified as the optimal compartment model for ^{18}F-FLT by Frings et al. (*5*). Pharmacokinetic parameters rate of influx of the tracer from blood to tissue (K_1), volume of distribution (V_T), and binding potential (BP_{ND}) of each lesion were derived using non-linear regression, where:

$$V_T = \frac{K1}{k2}\left(1 + \frac{k3}{k4}\right) \quad \text{Eq.4}$$

$$BP = \frac{k3}{k4} \quad \text{Eq.5}$$

V_T served as the preferred reference parameter for validation of simplified metrics for ^{18}F-FLT (*5*). The simplified metrics, mean SUV and tumour-to-blood ratio (TBR; parent plasma), were derived at a 50-60min post-injection scan interval, where:

$$SUV = \frac{activity\ concentration\ \left[\frac{Bq}{mL}\right]}{\left(\frac{injected\ activity\ [Bq]}{lean\ body\ mass}\right)} \quad \text{Eq.6}$$

$$TBR = \frac{tumor\ activity\ concentration\ \left[\frac{Bq}{mL}\right]}{blood\ activity\ concentration\ \left[\frac{Bq}{mL}\right]} \quad \text{Eq.7}$$

Statistical analysis

Data were described as mean with standard deviation (SD), median with interquartile range (IQR), minimum and maximum. Correlations between pairwise data were investigated using Spearman correlation. To assess technical validation of simplified metrics, we assessed correlations between both single measurements of kinetic parameter estimations and simplified metrics as well as correlations between relative changes in these parameters during treatment. Differences were tested using the Wilcoxon signed rank test (two related) or the Friedman test (multiple related), with significance level $p<0.05$. SPSS Statistics v22 (IBM) was used for statistical analyses.

Results

Patients

Ten patients with EGFR-mutated NSCLC were included, consisting of 4 men and 6 women with a mean age of 64±8 years. Treatment consisted of gefitinib and erlotinib in 7 and 3 patients, respectively. In one patient, the baseline scan was not evaluable due to scanner failure (scan at 7 and 28 days could still be used for lesion-based analyses). Another patient had no visible lesions at PET-CT. Twenty-four suspected lesions were detected on ^{18}F-FLT PET-CT (*5*).

HYPR optimization

A single composite ($HYPR_{single}$) provided most SNR improvement (Supplemental Figure 1, available at https://ejnmmires.springeropen.com). However, it eliminated the temporal dynamics of PVC (Figure 1). A $HYPR_{moving}$ setting with a composite image consisting of ±3 frames relative to the denoised frame provided an adequate trade-off between SNR improvement and partial-volume correction and was hence used in further analyses.

Image-derived input functions

We verified the assumption that partial-volume effects do not affect ascending aorta-derived IDIFs (based on the 2x2 voxel VOI approach used that minimized or avoided partial volume effects). First, PVC introduced only small relative differences in IDIF area-under-the-curve (AUC; Table 3.1), which were mitigated by $HYPR_{moving}$ and reduced to 0% by $HYPR_{single}$ (the latter providing most noise

mitigation). As a consequence, IDIF AUCs of uncorrected and PVC images were highly correlated (Supplemental Table 1). Similar results were observed for parent plasma calibrated input curves. Also, kinetic parameter estimates derived from uncorrected images using uncorrected vs. PVC input functions were very similar (Supplemental Table 2); small but significant differences in V_T and K_1 were observed for LR and LR+$HYPR_{moving}$ IDIFs, but not when $HYPR_{single}$ was applied. Therefore, we continued our analyses using the parent plasma calibrated input functions derived from uncorrected PET images.

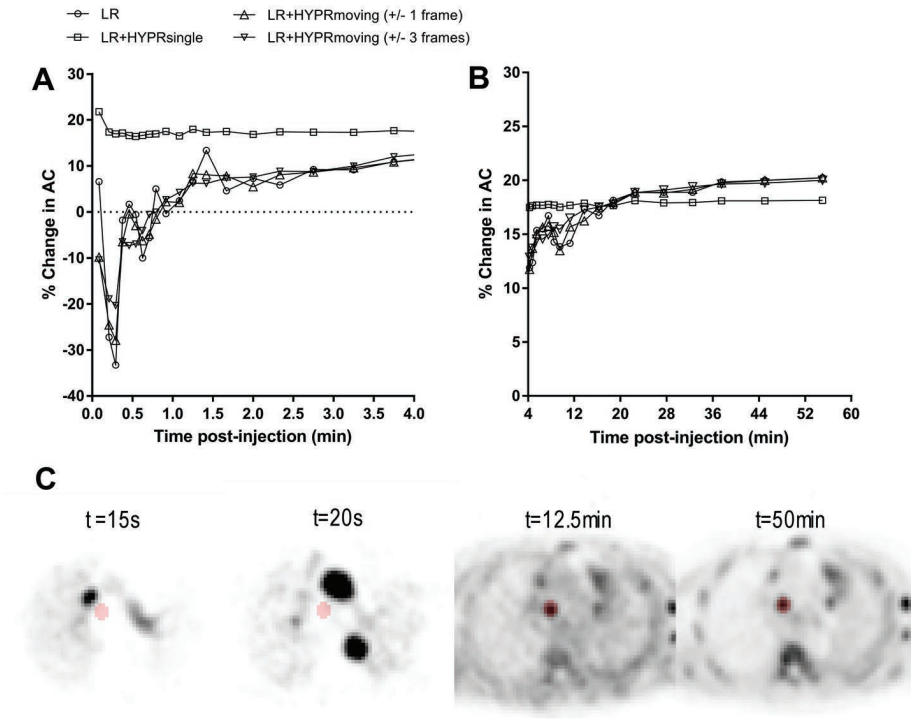


Figure 3.1: Time-activity curves of relative change in activity concentrations (AC) after PVC using several HYPR settings. Frames of 0-4 minutes (A) and 4-60 minutes (B) post-injection. Results of a typical mediastinal lymph node metastasis are shown. Note the temporality of PVE with a spill in at early timeframes. Corresponding original PET images (C) with the lesion volume-of-interest in red demonstrate blood pool activity near the VOI and increasing tumor-to-background contrast over time.

Table 3.1: Median relative differences (% with IQR) in IDIF AUC of PVC-images compared to uncorrected images.

	Entire curve		Peak only (2.5 min)	
	Image-derived	PP calibrated	Image-derived	PP calibrated
LR	-0.8 (-1.2 to 0.6)	-0.7 (-1.3 to -0.2)*	-2.0 (-3.4 to -0.9)*	-1.8 (-3.7 to -0.9)*
$LR+HYPR_{moving}$	-0.7 (-1.2 to 0.6)	-0.6 (-1.1 to -0.1)*	-2.2 (-3.2 to -0.5)*	-2.0 (-3.3 to -1.1)*
$LR+HYPR_{single}$	-0.8 (-1.2 to 0.6)	0.0 (0.0 to 0.0)	-0.9 (-1.2 to 0.7)	0.0 (-0.1 to 0.1)

*p<0.05. PP= parent plasma

Table 3.2: Relative changes (%) in kinetic parameter estimates and simplified metrics after PVC.

	Mean	Median	SD	IQR	Min	Max	p-value
				LR			
V_T	11.8	13.2	7.1	6.0-16.4	-15.2	25.1	<0.001
K_1	6.6	6.8	7.5	2.6-11.1	-16.7	32.3	<0.001
BP	6.1	6.0	8.8	2.1-10.7	-21.9	34.6	<0.001
SUV	13.1	13.2	6.1	7.3-17.1	3.3	28.4	<0.001
TBR	13.1	13.2	6.1	7.3-17.1	3.3	28.3	<0.001
				LR+HYPR			
V_T	10.8	11.7	6.1	6.1-15.5	-13.6	21.7	<0.001
K_1	5.7	4.3	6.9	2.3-10.0	-14.9	25.1	<0.001
BP	3.7	4.4	6.4	0.1-7.1	-20.6	19.8	<0.001
SUV	12.6	12.9	5.8	7.0-16.7	2.1	24.7	<0.001
TBR	12.8	12.9	6.0	7.0-17.0	3.1	27.3	<0.001

Table 3.3: Correlation (Spearman, with 95% confidence intervals) between PVC-induced relative changes in kinetic parameter estimates and simplified metrics.

	V_T	K_1	BP
		LR	
SUV	0.58* (0.38-0.73)	0.61* (0.42-0.75)	0.51* (0.30-0.68)
TBR	0.58* (0.38-0.73)	0.61* (0.42-0.75)	0.51* (0.30-0.68)
		LR+HYPR	
SUV	0.62* (0.43-0.75)	0.47* (0.24-0.65)	0.36* (0.11-0.56)
TBR	0.62* (0.43-0.75)	0.48* (0.26-0.66)	0.36* (0.12-0.57)

*p<0.01

Kinetic parameter estimates and simplified metrics

Relative differences between uncorrected and PVC data for K_1, V_T, BP_{ND}, SUV, and TBR are presented in Table 3.2. Both LR and $LR+HYPR_{moving}$ significantly ($p<0.001$) increased each parameter. Overall, LR provided larger changes in parameters than $LR+HYPR_{moving}$ for both kinetic parameters and simplified

metrics. Regarding kinetic parameters, largest changes were seen for V_T, which was increased by median 13.2% up to 25.1% using LR. Changes in K_1 and BP_{ND} were very similar (median 6.8% and 6.0%, respectively, using LR). Changes in SUV and TBR after PVC were almost identical, as expected, and were comparable to changes in V_T. LR and LR+HYPR$_{moving}$ decreased V_T, K_1 and BP_{ND} in some lesions, but only provided increases for SUV and TBR. Changes in V_T, K_1 and BP_{ND} after PVC had low but significant correlations with changes in SUV and TBR after PVC (Table 3.3); highest correlations were seen between relative changes in V_T and changes in SUV and TBR (up to 0.62).

We plotted relative changes in V_T, K_1, BP_{ND}, and SUV after PVC as a function of lesion (original) MATV to provide insight into effect of lesion size on PVC performance (Figure 3.2). For LR, the correlations between MATV and relative change in V_T, K_1, BP_{ND}, SUV, and TBR were -0.39, -0.47, -0.36, -0.80, and -0.80, respectively ($p<0.01$). For LR+HYPR, these correlations were -0.43, -0.34, -0.24, -0.81, -0.80, respectively ($p<0.01$, except for BP_{ND}; $p=0.07$). Compared to tumour delineation on uncorrected images, delineation on partial-volume corrected images (LR+HYPR$_{moving}$) provided a median relative decrease in MATV of 14.3%, (IQR -22.1 to -7.5, minimum -69.2, maximum 5.3; Figure 3.3). Also, the effect of PVC on kinetic parameters and simplified metrics was higher when using VOIs generated on PVC images compared to when using original VOIs (Supplemental Table 3). Here, largest increases after PVC were seen for V_T, SUV, and TBR with median increases of 13.9% (IQR 7.6-18.7; max 37.8%), 15.8% (IQR 8.4-20.4; max 31.5), and 15.8% (IQR 8.4-20.7; max 34%), respectively.

Technical validation of simplified metrics

PVC increased the correlations between SUV and V_T and K_1, but not for BP_{ND} (Table 3.4). PVC increased the correlations between TBR and V_T, K_1, and BP_{ND} (Table 3.4). Largest increases in these correlations were seen between V_T and SUV (0.82 to 0.90; Figure 3.4). However, confidence intervals of these correlations overlapped and therefore were not statistically significant.

During treatment, V_T, BP_{ND}, SUV, and TBR significantly decreased, while K_1 did not change (as was also observed in Frings et al. (*5*)), regardless of PVC (p-values in Supplemental Table 4). At 7 and 28 days after starting treatment, original MATV demonstrated a median decrease of 16.1% (IQR -38.9 to -0.6), and 17.6% (IQR -58.3 to 4.3). We correlated treatment-induced relative changes in kinetic parameters to treatment-induced relative changes in simplified metrics

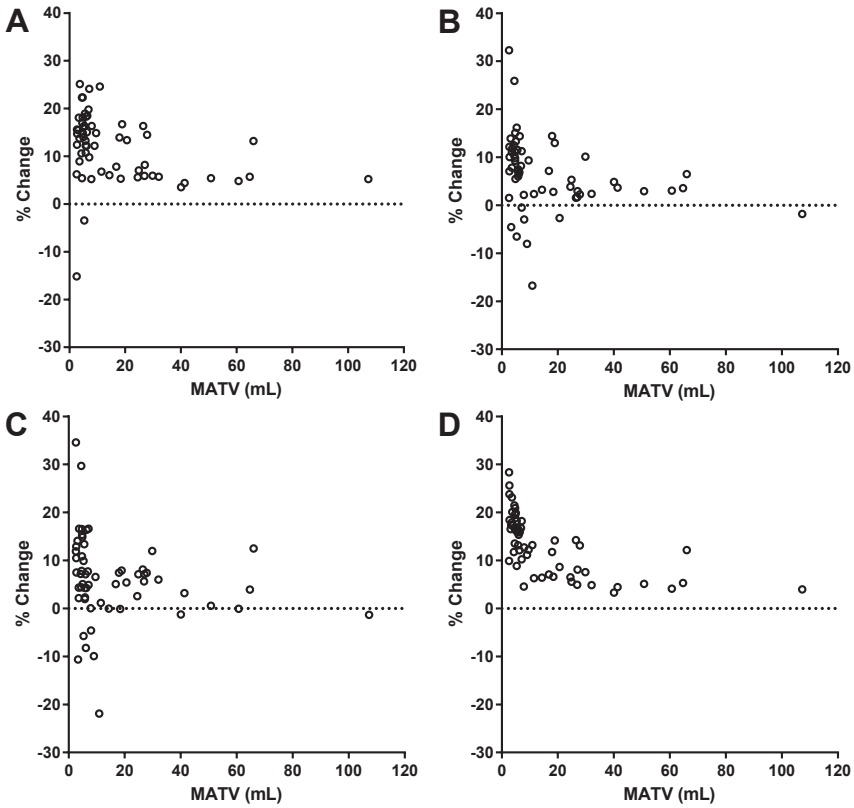


Figure 3.2: Relative change (%) in quantitative parameters after PVC (LR) as a function of lesion MATV (mL) for VT (A), K1 (B), BP (C), and SUV (D). TBR is not displayed since it was virtually identical to SUV.

Table 3.4: Correlation (Spearman, with 95% confidence intervals) between kinetic parameter estimates and simplified metrics, with and without PVC.

	V_T	K_1	BP
Uncorrected			
SUV	0.82* (0.72-0.89)	0.43* (0.19-0.62)	0.89* (0.82-0.93)
TBR	0.81* (0.69-0.88)	0.47* (0.24-0.65)	0.82* (0.72-0.89)
LR			
SUV	0.90* (0.83-0.94)	0.45* (0.22-0.63)	0.89* (0.82-0.93)
TBR	0.88* (0.81-0.93)	0.48* (0.26-0.65)	0.84* (0.74-0.90)
LR+HYPR			
SUV	0.90* (0.83-0.94)	0.48* (0.26-0.65)	0.89* (0.81-0.93)
TBR	0.88* (0.81-0.93)	0.51* (0.30-0.68)	0.83* (0.73-0.90)

*p<0.01

during treatment with TKIs for the uncorrected data as well as those with PVC (Figure 3.5). At both 7 and 28 days after treatment start, changes in V_T and BP_{ND} were significantly correlated (0.79-0.98 and 0.44-0.91, respectively) with changes in SUV and TBR (with the exception of correlation between changes in BP_{ND} vs. TBR on LR images at 7 days; 0.45, p>0.05), regardless of PVC. PVC (both LR and LR+HYPR) did not improve correlations between treatment induced changes in BP and changes in SUV or TBR. PVC increased the correlation between treatment-induced changes in SUV and V_T at 7 days and 28 days (increases in correlation ranging 0.05-0.09, with overlapping confidence intervals). Also, PVC increased the correlation between treatment-induced changes in TBR with changes in V_T at 28 days, but not at 7 days, after treatment start by 0.06 for both LR and LR+HYPR, with overlapping confidence intervals.

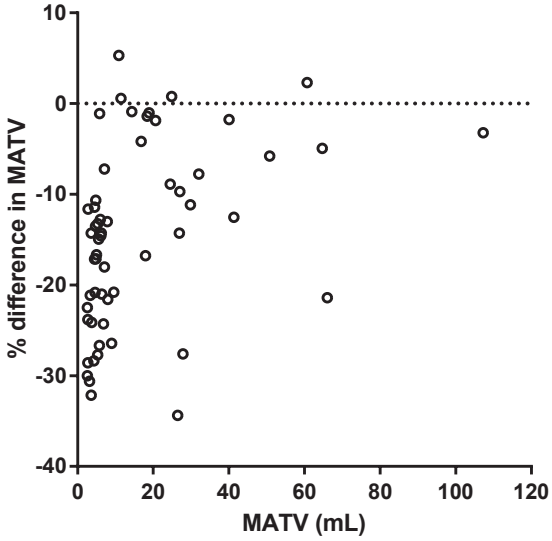


Figure 3.3: Relative difference (%) in lesion MATV (mL) between uncorrected and PVC images (LR+HYPR) as function of MATV on uncorrected images. Y-axis was scaled to -40%; for one lesion of 5.8ml MATV was 69% smaller on PVC image.

Discussion

In the present study we evaluated the impact of frame-wise parametric PVC on tumour kinetic parameter estimation derived from dynamic PET-CT scans and the resulting effect on validation of simplified metrics. PVC significantly increased

both tumour micro- and macrokinetic parameters, and we observed that partial-volume effects varied over time due to blood pool activity and changing tumour contrast. Hence, the effect of PVC on kinetic parameter estimates was not in full concordance with its effect on simplified metrics (SUV and TBR), and as a consequence PVC was found to affect the validation of SUV using V_T both for single measurements and as biomarker of treatment response to a small extent (albeit non-significantly).

Application of PVC in oncologic dynamic PET-CT studies is scarce. Mankoff et al. (2003) applied PVC in dynamic FDG-PET of breast cancer patients using a simple method with recovery coefficients, assuming lesions are spherical with homogenous tracer distributions (*29*). They observed that applying PVC in response measurements reduced changes in metabolic rate of FDG and blood flow of responding patients, reducing significance of parameter changes (albeit still statistically significant). By using this method, however, kinetic parameters were solely corrected for (changes in) tumour size, and no correction for spill-in from blood pool structures and/or heterogeneous tumour background was applied. In 2007, Teo et al. validated the use of iterative deconvolution as an image-based PVC method not requiring anatomical segmentation or knowledge of lesion size, and suggested its potential application in kinetic modeling, which to the best of our knowledge has not been performed to date for oncologic PET-CT (*30*).

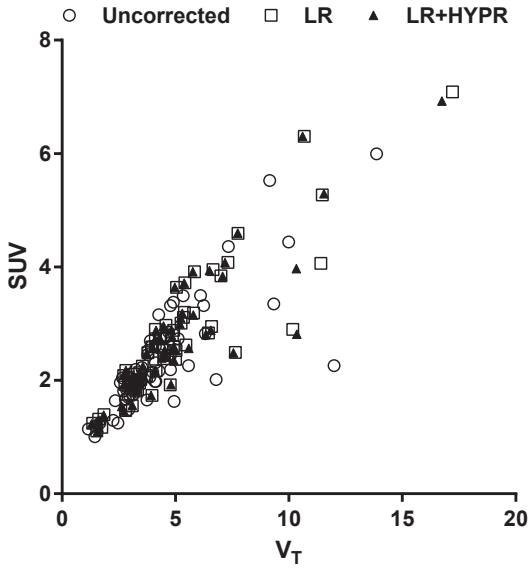


Figure 3.4: Scatter plot of VT versus SUV, without and with PVC. For both LR and LR+HYPR, the Spearman correlation between VT and SUV increased from 0.82 to 0.90 after PVC.

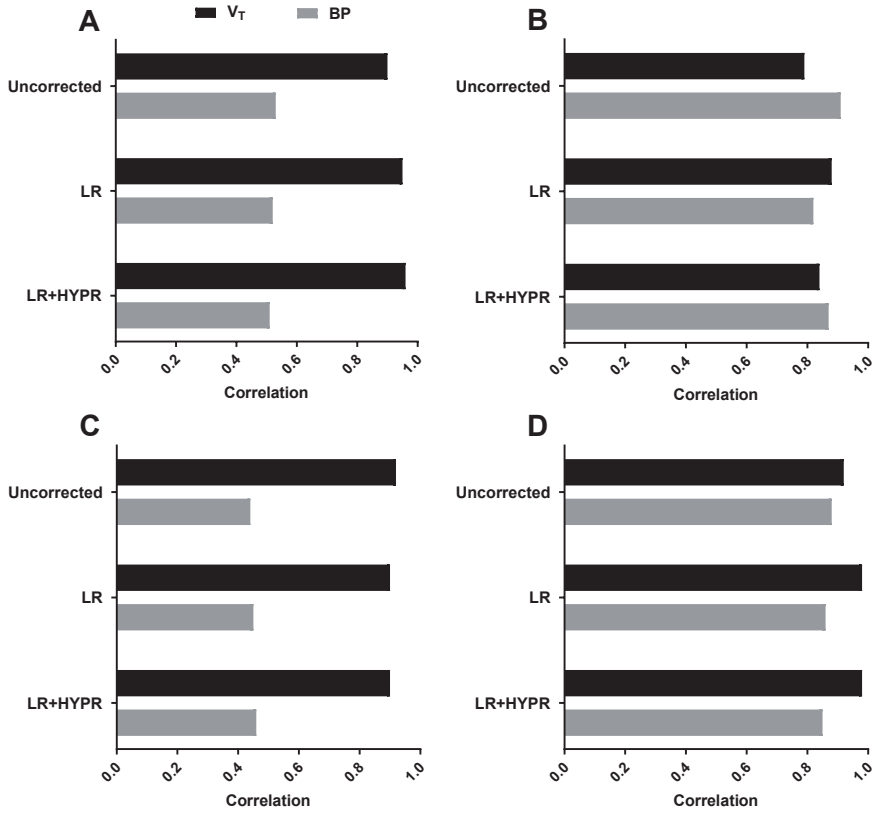


Figure 3.5: Correlation (Spearman) between changes in kinetic parameter estimates vs. simplified metrics during treatment with TKI, with and without PVC. Results shown are for SUV at 7 (A) and 28 (B) days, and for TBR at 7 (C) and 28 (D) days after treatment start.

Both tumour macroparameters V_T and BP_{ND}, and microparameter K_1 significantly changed after application of PVC. This corresponds with results from applications of PVC in brain dynamic PET studies, where similar increases in kinetic parameter estimations have been observed when applying PVC in case of activity spill-out (*19-21,31*). Interestingly, the effect of PVC on kinetic parameter estimates was poorly (albeit significantly) correlated with its effect on simplified measures. As previously described(*9*), the effect of PVC on SUV of (hotspot) lesions on static PET-CT scans is straightforward: an expected net increase in activity, mainly dependent on lesion size (and, in lesser extent, shape and local contrast). This can be seen in Figure 3.2, where change in SUV after PVC is highly (inversely) correlated to tumour volume, whilst the kinetic parameter estimations are not. This illustrates that impact of PVC on tumour kinetic

parameter estimation is more complex, as seen in Figure 3.1 which displays the non-linear temporality of partial-volume effects for a typical mediastinal lymph node metastasis. Here, an early spill-in of activity due to blood pool proximity is noted, with increasing activity spill-out afterwards as tumour uptake increases and background activity decreases. Hence, across lesions the effect of PVC on kinetic parameters may differ depending not only on size, but as well on presence of proximate high activity structures, rate of tracer uptake during the scan, and background activity.

For quantification of functional tumour characteristic on PET-CT in clinical practice, a simplified quantitative method is necessary, obviating the need for complex and extended dynamic image acquisitions, need for blood sampling, and facilitating the possibility of whole-body acquisitions. To this end, per radiotracer and cancer type simplified metrics needs to be technically validated by pharmacokinetic modeling using dynamic PET-CT (*4*). In the current study the effect of PVC on kinetic parameter estimates was different from its effect on simplified metrics, which explains why it might affect validation of these simplified metrics (using V_T). We observed a trend that PVC increased correspondence of SUV with V_T in single measurements (correlations improving from 0.82 to 0.90) and as a biomarker of treatment response (correlations improving from 0.90 to 0.95 at 7 days and from 0.79 to 0.88 at 28 days after treatment start). However, confidence intervals of these correlations overlapped, which might at least partly be due to the sample size (inherent to this type of study), and therefore these differences are not statistically significant. Therefore, while PVC is mandated to acquire accurate quantitative reads, it only increases correspondence of kinetic parameters with simplified metrics to a small extent on a cohort level. This indicates that the impact of image resolution on technical validation of simplified metrics of ^{18}F-FLT as biomarkers of response to TKI might be small, and that PET images without PVC seem non-inferior for this purpose. It should be noted that for response assessment to treatments that affect tracer kinetics and blood pool activity to a larger extent than TKIs and for other cancer types more affected by spill-in (eg. prostate cancer lesions with urinary tract proximity), PVC may have a larger impact on validation of simplified metrics.

Spill-out due to PVE will result in overestimation of metabolic tumour volumes, which increases the underestimation of true tracer uptake since background activity is included (*11*). A parametric PVC method may therefore theoretically reduce inaccuracies in delineation. However, iterative deconvolution

has been proposed with use of VOIs defined on uncorrected images, due to the expected propagation of image noise after PVC (*30*). We evaluated the impact of delineation on deconvoluted images with HYPR denoising, and found not only substantial decreases in MATVs (Figure 3) but also an increase in PVCs effect on kinetic parameter estimates (Supplemental Table 3). Nonetheless, our previous study demonstrated that the reduction in MATV after PVC may not necessarily lead to more accurate definition of tumour volumes (*11*).

In brain PET studies, frequently a small vessel such as the carotid artery needs to be utilized for IDIF generation. This mandates PVC due to the small artery diameter (*32,33*). In this study on thoracic oncological PET-CTs, the ascending aorta, a large vessel, was used for IDIF generation. We noted that PVC introduced negligible differences in IDIF area-under-the-curves, and that without denoising this introduced small but significant differences in kinetic parameter estimates (Supplemental Table 2). However, since HYPR denoising using a single composite image (providing maximum noise reduction) appeared to completely mitigate this effect, the effect of PVC on these input functions seems to be based on PVC-induced noise-propagation. Therefore, when input functions derived from large blood pool structures are used, PVC is preferably avoided to evade noise-induced inaccuracies in kinetic parameter estimates (assuming no spillover from nearby high activity structures).

Iterative deconvolution algorithms are known to propagate image noise, which may necessitate denoising methods to be applied to preserve image quality. Several approaches have been proposed, such as wavelet-based denoising for static PET-CT and HYPR denoising for dynamic acquisitions, respectively (*26,34*). We observed that HYPR needs to be optimized for tracer kinetics using a moving composite image, since when applied using a single composite image (maximal denoising) it seems to lose the temporal dynamic course of the PVC (Figure 3.1). Including $HYPR_{moving}$ resulted in very similar outcomes compared to PVC alone, and slightly mitigated the increase in kinetic parameter estimates after PVC. The latter may not only be attributed to reduced statistical noise, but also to some smoothing effects inherent to the algorithm. Also, at late time frames it had no effect on intratumoural COV% (Supplemental Figure 1). This might be explained by the high tumour contrast and high count number (due to the long frame duration), as Golla et al. previously demonstrated (*21*). The increase in COV% at late time frames thus seems to be a resultant of increased intratumoural heterogeneity by PVC itself. Therefore, in region-based non-linear

regression analyses the impact of PVC-induced increased image noise on kinetic parameter estimation seems negligible. However, it may have significant impact when tumours are analyzed on a parametric level.

While the presence of PVE and the consequent need for PVC are well recognized, to date PVC has rarely been applied in oncological PET studies. This may be because to date there is no consensus on the optimal correction strategy and data yielded from application of PVC does not seem to have triggered routine clinical application (*12,35*). Our study now demonstrates that PVC should not only be performed in future regular static PET-CT studies, but in dynamic PET-CT studies as well, also when simplified quantitative metrics are validated for clinical applications. If not applied, small lesions should preferably be excluded from analyses, as recommended and performed in previous studies using a 2-3cm diameter cut-off to avoid PVE (*36,37*). Still, our data demonstrate that lesions above these size thresholds are also affected by PVE (Figure 3.2).

Only data from ^{18}F-FLT PET-CT was used. However, the current dataset from a widely used whole body TOF PET-CT scanner allowed for both kinetic modeling and extraction of simplified parameters per lesion, at time points used in clinical practice due to the long acquisition time (0-60min post-injection). Also, the dataset included both large and small lesions, both nearby and remote from large blood pool structures. Additionally, it facilitated evaluation of PVCs effect on validation of simplified parameters both in single measurements and during systemic treatment. Since we have demonstrated the significant effect of PVC in kinetic parameter estimation, future dynamic PET studies focusing on other PET-tracers in small tumours (e.g. PSMA-ligand tracers in prostate cancer metastases) should apply PVC as a similar (or larger) impact of PVC may be expected. In the current study no correction was made for potential motion blurring effects, which is another factor possibly affecting accuracy of kinetic parameter estimations (*38*). Efforts should be made to incorporate both PVC and motion correction methodologies simultaneously for dynamic PET studies. Also, the impact of PVC on parametric kinetic analyses of oncologic dynamic PET warrants further investigation, which will require HYPR denoising to be optimized for this purpose.

Conclusion

Parametric PVC using iterative deconvolution had a significant impact on tumour kinetic macro- and microparameter estimations from dynamic PET-CT. The relative effects of PVC on kinetic parameter estimations and simplified metrics were poorly correlated. This resulted in a non-significant trend in higher correlation between V_T and SUV in single reads and affected its technical validation as a biomarker of treatment response to a small extent. Therefore, the impact of image resolution on technical validation of simplified metrics for clinical use seems to be small. When optimized according to tracer kinetics, HYPR denoising may adequately reduce PVC-induced image noise for low count and low contrast timeframes. However, it has only limited effect on kinetic parameter estimations and thus may be obviated for region-based non-linear regression analysis. Future oncologic dynamic PET-CT studies should preferably incorporate partial-volume correction to acquire accurate quantitative reads.

References

1. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
2. Gallamini A, Zwarthoed C, Borra A. Positron Emission Tomography (PET) in Oncology. ***Cancers (Basel).*** 2014;6:1821-1889.
3. Han S, Woo S, Kim YJ, Suh CH. Impact of (68)Ga-PSMA PET on the Management of Patients with Prostate Cancer: A Systematic Review and Meta-analysis. ***Eur Urol.*** 2018;74:179-190.
4. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. ***J Nucl Med.*** 2017;58:1019-1024.
5. Frings V, Yaqub M, Hoyng LL, et al. Assessment of simplified methods to measure 18F-FLT uptake changes in EGFR-mutated non-small cell lung cancer patients undergoing EGFR tyrosine kinase inhibitor treatment. ***J Nucl Med.*** 2014;55:1417-1423.
6. 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.
7. Castell F, Cook GJ. Quantitative techniques in 18FDG PET scanning in oncology. ***Br J Cancer.*** 2008;98:1597-1601.
8. Verwer EE, Oprea-Lager DE, van den Eertwegh AJ, et al. Quantification of ^{18}F-fluorocholine kinetics in patients with prostate cancer. ***J Nucl Med.*** 2015;56:365-371.
9. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. ***J Nucl Med.*** 2007;48:932-945.
10. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. ***Phys Med Biol.*** 2012;57:R119-159.
11. Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial-Volume Correction in Oncological PET/CT Studies. ***J Nucl Med.*** 2016;57:1642-1649.
12. Cysouw MCF, Kramer GM, Schoonmade LJ, Boellaard R, de Vet HCW, Hoekstra OS. Impact of partial-volume correction in oncological PET studies: a systematic review and meta-analysis. ***Eur J Nucl Med Mol Imaging.*** 2017;44:2105-2116.
13. Wong DF, Harris JC, Naidu S, et al. Dopamine transporters are markedly reduced in Lesch-Nyhan disease in vivo. ***Proc Natl Acad Sci U S A.*** 1996;93:5539-5543.
14. Giovacchini G, Lerner A, Toczek MT, et al. Brain incorporation of 11C-arachidonic acid, blood volume, and blood flow in healthy aging: a study with partial-volume correction. ***J Nucl Med.*** 2004;45:1471-1479.
15. Price JC, Klunk WE, Lopresti BJ, et al. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B. ***J Cereb Blood Flow Metab.*** 2005;25:1528-1547.
16. Ahmad R, Postnov A, Bormans G, Versijpt J, Vandenbulcke M, Van Laere K. Decreased in vivo availability of the cannabinoid type 2 receptor in Alzheimer's disease. ***Eur J Nucl Med Mol Imaging.*** 2016;43:2219-2227.
17. Greve DN, Svarer C, Fisher PM, et al. Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data. ***Neuroimage.*** 2014;92:225-236.
18. Bauer M, Karch R, Neumann F, et al. Age dependency of cerebral P-gp function measured with (R)-[11C]verapamil and PET. ***Eur J Clin Pharmacol.*** 2009;65:941-946.
19. Rousset OG, Deep P, Kuwabara H, Evans AC, Gjedde AH, Cumming P. Effect of partial volume correction on estimates of the influx and cerebral metabolism of 6-[(18)F]fluoro-L-dopa studied with PET in normal control and Parkinson's disease subjects. ***Synapse.*** 2000;37:81-89.
20. Bowen SL, Byars LG, Michel CJ, Chonde DB, Catana C. Influence of the partial volume correction method on (18)F-fluorodeoxyglucose brain kinetic modelling from dynamic PET images reconstructed with resolution model based OSEM. ***Phys Med Biol.*** 2013;58:7081-7106.
21. Golla SSV, Lubberink M, van Berckel BNM, Lammertsma AA, Boellaard R. Partial volume correction of brain PET studies using iterative deconvolution in combination with HYPR denoising. ***EJNMMI Res.*** 2017;7:36.

22. Wolters EE, Golla SSV, Timmers T, et al. A novel partial volume correction method for accurate quantification of [(18)F] flortaucipir in the hippocampus. ***EJNMMI Res.*** 2018;8:79.

23. Hunter CR, Klein R, Beanlands RS, deKemp RA. Patient motion effects on the quantification of regional myocardial blood flow with dynamic PET imaging. ***Med Phys.*** 2016;43:1829.

24. Wang T, Wu KY, Miner RC, Renaud JM, Beanlands RSB, deKemp RA. Reproducible quantification of cardiac sympathetic innervation using graphical modeling of carbon-11-meta-hydroxyephedrine kinetics with dynamic PET-CT imaging. ***EJNMMI Res.*** 2018;8:63.

25. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. ***Neuroimage.*** 2008;39:1570-1584.

26. Christian BT, Vandehey NT, Floberg JM, Mistretta CA. Dynamic PET denoising with HYPR processing. ***J Nucl Med.*** 2010;51:1147-1154.

27. Floberg JM, Mistretta CA, Weichert JP, Hall LT, Holden JE, Christian BT. Improved kinetic analysis of dynamic PET data with optimized HYPR-LR. ***Med Phys.*** 2012;39:3319-3331.

28. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. ***Radiology.*** 2014;273:539-548.

29. 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.

30. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. ***J Nucl Med.*** 2007;48:802-810.

31. Kim E, Shidahara M, Tsoumpas C, et al. Partial volume correction using structural-functional synergistic resolution recovery: comparison with geometric transfer matrix method. ***J Cereb Blood Flow Metab.*** 2013;33:914-920.

32. Mourik JE, Lubberink M, Klumpers UM, Comans EF, Lammertsma AA, Boellaard R. Partial volume corrected image derived input functions for dynamic PET brain studies: methodology and validation for [11C]flumazenil. ***Neuroimage.*** 2008;39:1041-1050.

33. Sari H, Erlandsson K, Law I, et al. Estimation of an image derived input function with MR-defined carotid arteries in FDG-PET human studies using a novel partial volume correction method. ***J Cereb Blood Flow Metab.*** 2017;37:1398-1409.

34. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. ***Eur J Nucl Med Mol Imaging.*** 2009;36:1064-1075.

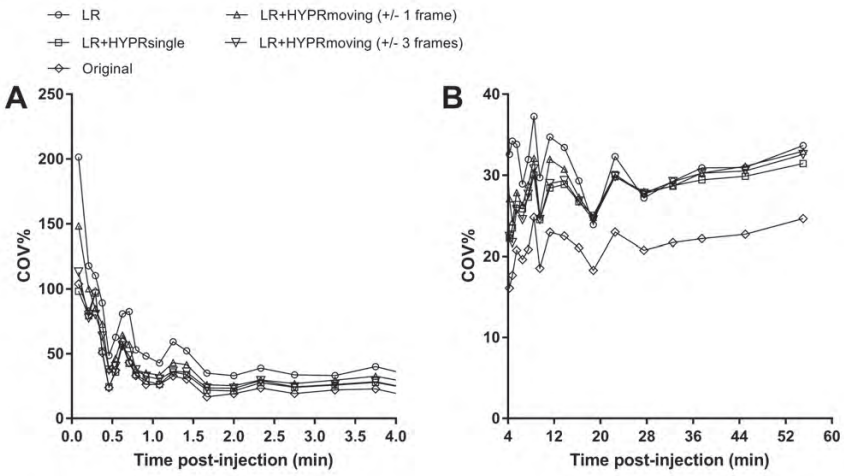
35. Alavi A, Werner TJ, Hoilund-Carlsen PF, Zaidi H. Correction for Partial Volume Effect Is a Must, Not a Luxury, to Fully Exploit the Potential of Quantitative PET Imaging in Clinical Oncology. ***Mol Imaging Biol.*** 2018;20:1-3.

36. 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.

37. Frings V, de Langen AJ, Smit EF, et al. 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;51:1870-1877.

38. Merlin T, Visvikis D, Fernandez P, Lamare F. Dynamic PET image reconstruction integrating temporal regularization associated with respiratory motion correction for applications in oncology. ***Phys Med Biol.*** 2018;63:045012.

Supplemental files



Supplemental Figure 3.1: Time-activity curves of intralesional image noise (COV%) without and with PVC using several HYPR settings. Frames of 0-4 minutes (A) and 4-60 minutes (B) post-injection. Results of a typical mediastinal lymph node metastasis are shown.

Supplemental Table 3.1: Spearman correlations between IDIF AUCs of PVC-images and uncorrected images. All correlations were significant with p<0.001.

	Entire curve		Peak only (2.5 min)	
	Image-derived	**Calibrated**	**Image-derived**	**Calibrated**
LR	0.994	0.991	0.993	0.988
LR+HYPR (-/+ 3 frames)	0.994	0.990	0.992	0.987
LR+HYPR (single composite)	0.994	0.992	0.997	0.996

Supplemental Table 3.2: Median relative differences (% with IQR) in K1, Vt, and k3/k4 of uncorrected images using uncorrected versus corrected IDIFs (PVC without and with HYPR denoising). *p<0.05 Wilcoxon-signed-rank test.

IDIF:	K1	Vt	k3/k4
LR	2.9 (0.2 to 7.4)*	0.9 (-1.1 to 2.6)*	-0.2 (-3.7 to 2.4)
LR+HYPR (-/+ 3 frames)	3.1 (0.9 to 5.5)*	1.1 (-0.4 to 4.5)*	-0.8 (-3.2 to 1.6)
LR+HYPR (single composite)	0.0 (-0.2 to 0.1)	0.0 (0.0 to 0.1)	0.1 (-0.1 to 0.1)

Supplemental Table 3.3: Relative changes (%) in kinetic parameter estimates and simplified metrics after PVC using VOIs delineated on PVC images (LR+HYPR).

	Mean	Median	SD	IQR	Min	Max	p-value
LR+HYPR:							
Vt	13.6	13.9	7.9	7.6 – 18.7	5.5	37.8	<0.001
K1	6.6	5.3	6.7	2.5 – 10.1	4.4	30.3	<0.001
BP	5.7	5.4	7.6	1.2 – 8.9	3.7	38	<0.001
SUV	15.4	15.8	7.1	8.4 – 20.4	3.6	31.5	<0.001
TBR	15.6	15.8	7.2	8.4 – 20.7	5.7	34.0	<0.001

Supplemental Table 3.4: P-values of testing (Friedman's test) between changes in kinetic parameter estimates and simplified metrics (with and without PVC) during treatment with TKI at 7 and 28 days after treatment start.

	Vt	K1	BP	SUV	TBR
Uncorrected	<0.001	0.45	0.038	0.001	0.002
LR	<0.001	0.819	0.005	0.001	0.002
LR+HYPR	<0.001	0.819	0.031	0.001	0.002


QI-1-PET

Chapter 4

Sensitivity of [^{18}F]Fluorodihydrotestosterone PET-CT to Count Statistics and Reconstruction Protocol in Metastatic Castration-Resistant Prostate Cancer

M.C.F. Cysouw, G.M. Kramer, D. Heijtel, R.C. Schuit, M.J. Morris, A.J.M. van den Eertwegh, J. Voortman, O.S. Hoekstra, D.E. Oprea-Lager, R. Boellaard

EJNMMI Res. 2019 Jul 30;9(1):70.

Abstract

Objectives: Whole body [^{18}F]-fluorodihydrotestosterone positron emission tomography ([^{18}F]FDHT PET) imaging directly targets the androgen receptor and is a promising prognostic and predictive biomarker in metastatic castration-resistant cancer (mCRPC). To optimize of [^{18}F]FDHT PET-CT for diagnostic and response assessment purposes, we assessed how count statistics and reconstruction protocol affect its accuracy, repeatability, and lesion detectability.

Methods: Whole body [^{18}F]FDHT PET-CT scans were acquired on an analogue PET-CT on two consecutive days in 14 mCRPC patients harboring a total of 336 FDHT-avid lesions. Images were acquired at 45 min post-injection of 200 MBq [^{18}F]FDHT at 3 min per bed position. List-mode PET data were split on a count-wise basis, yielding two statistically independent scans with each 50% of counts. Images were reconstructed according to current EANM Research Ltd. (EARL1: 4mm voxel) and novel EARL2 guidelines (4mm voxel + PSF). Per lesion we measured SUVpeak, SUVmax, SUVmean, and contrast-to-noise ratio (CNR). SUV was normalized to dose per bodyweight as well as to the parent plasma input curve integral. Variability was assessed with repeatability coefficients (RC).

Results: Count reduction increased liver coefficient of variation from 9.0 to 12.5% and from 10.8 to 13.2% for EARL1 and EARL2, respectively. SUVs of EARL2 images were 12.0-21.7% higher than EARL1. SUVs of 100% and 50% count data were highly correlated ($R2>0.98$; slope=0.97-1.01; ICC=0.99-1.00). Intrascan variability was volume-dependent and count reduction resulted in higher intrascan variability for EARL2 than EARL1 images. Intrascan RCs were lowest for SUVmean (8.5-10.6%), intermediate for SUVpeak (12.0-16.0%) and highest for SUVmax (17.8-22.2%). Count reduction increased test-retest variance non-significantly ($p>0.05$) for all SUV types and normalizations. For SUVpeak at 50% of counts, RCs remained <30% when small lesions were excluded. Splitting data reduced CNR by median 4.6% (interquartile range 1.2-8.7%) and 4.6% (interquartile range 1.2-8.7%) for EARL1 and EARL2 images, respectively.

Conclusions: Reducing [^{18}F]FDHT PET acquisition time from 3 min to 1.5 per bed position resulted in a repeatability of SUVpeak (bodyweight) remaining ≤30%, which is generally acceptable for response monitoring purposes. However, EARL2 reconstruction was more affected, especially for SUVmax whose repeatability tended to exceed 30%. Lesion detectability was only slightly impaired by reducing acquisition time, which might not be clinically relevant in mCRPC.

Introduction

The androgen receptor (AR) axis plays a central role in hormone sensitive as well as castrate resistant prostate cancer (CRPC) (*1*). In the last decades several AR signalling inhibitor (ARSi) therapies have been developed and approved for treatment of metastatic (m)CRPC patients (*2-5*). Results of initial treatment with ARSi therapies (e.g. enzalutamide and abiraterone) are excellent, with mild toxicity profiles. Unfortunately, initial treatment response and response durability are variable, and response to second-line ARSi therapies is often short (*3,5*). Therefore, a predictive biomarker for response to these ARSi drugs is urgently needed. Currently used imaging modalities (e.g. CT and bone scintigraphy) for restaging and detection of disease progression in CRPC are not suited for this purpose (*6*).

^{18}F-fluorodihydrotestosterone ([^{18}F]FDHT) positron emission tomography-computed tomography (PET-CT) directly targets the AR in whole body imaging (*7,8*). Hereby it can assess AR-status on a lesion-by-lesion level allowing for characterization of AR expression and its intra-patient heterogeneity in-vivo(*9*). This may not only enable prognostication for ARSi therapies, but also facilitate novel AR-targeted drug development (*10,11*). Recently, technical validation studies on the optimal simplified metrics and their repeatability have been performed and clinical studies evaluating the value of ^{18}F-FDHT PET/CT as imaging biomarker in clinical setting are ongoing (*8,12*).

Crucial elements of validation and clinical implementation of novel oncologic PET tracers and their imaging protocols are patient burden and cost of imaging. The latter two should be as low as possible, whilst maintaining high quantitative and qualitative accuracy for clinical purposes such as prediction or monitoring of treatment response. Until now, whole-body [^{18}F]FDHT PET-CT studies have been acquired at 3-4 minutes per bed position, resulting in a typical in-scanner time of about 30 min for a single scan session (*7,12*). As mCPRC patients often have extensive (painful) metastatic disease, frequently involving the spine, reducing acquisition time could diminish patient burden, reduce cost of imaging, and improve department efficiency. This requires that the effect of count statistics on the performance of [^{18}F]FDHT PET-CT is known. For [^{18}F]FDG PET-CT it has been shown that reducing acquisition times may reduce image quality, but does not necessarily affect lesion detection rates (*13,14*).

The finite spatial resolution of current PET scanners lead to blurring of images and cause partial-volume effects. Therefore, the EANM Research Ltd.

(EARL) group has incorporated the point spread function (PSF) in reconstruction algorithms in the novel EARL2 (2019) guideline to improve image resolution (*15,16*). These novel standards could negatively affect quantitative precision and thereby hamper both prediction or monitoring of treatment response (*17,18*). Also, they could affect comparability of data between centres in multicentre trials. Therefore, it is important to know how these reconstruction protocols and their sensitivity to count statistics affect accuracy and precision of [^{18}F]FDHT PET-CT studies.

To technically validate [^{18}F]FDHT PET-CT for trials and future clinical use, i.e. for drug development, prognostication, and prediction or monitoring of response, it is crucial to know whether and how accuracy and precision of [^{18}F]FDHT PET-CT are a function of image count statistics and reconstruction protocol. Therefore, the aim of this study was to assess how count statistics and reconstruction protocol (EARL1 [2015] vs EARL2 [2019] guidelines) affect accuracy, repeatability, and lesion detectability of analogue whole body [^{18}F] FDHT PET-CT.

Methods

Patients

Fourteen histologically proven mCRPC patients were prospectively included at the Amsterdam UMC (location VUmc), the Netherlands, between February 2015 and April 2016, as part of a multicentre cohort study (*12*). Patient eligibility criteria were: castrate levels of serum testosterone (<1.7 nmol/L [50 ng/dL]); ≥1 month since their last anti-cancer pharmacologic therapy; no concurrent malignancies; and progressive disease based on any of the following: (a) a rise in PSA through 3 consecutive measurements; (b) RECIST 1.1 imaging evidence of progressive disease and/or (c) bone scan showing at least two new metastatic lesions not attributable to flare phenomenon. Patients without orchiectomy remained on androgen depletion therapy with a gonadotropin-releasing hormone analogue or inhibitor during the study. The Amsterdam UMC (location VUmc) institutional review board approved this prospective study and each subject gave written informed consent prior to study enrolment.

PET imaging protocol

Patients were scanned on two consecutive days on a whole body time-of-flight Gemini TF64 PET-CT scanner (Philips Healthcare, Netherlands) with EARL accreditation (*16*). A 4h fasting period was included to minimize intra-intestinal bile activity. Intravenous injection of ±200 MBq [18F]FDHT was followed by a 30 min dynamic scan of the chest (with aorta in field of view) to acquire an image derived input function. Venous blood samples were drawn at 5, 10 and 30 min. Analysis of venous samples included measurements of whole blood and plasma activity concentrations, parent fraction, and metabolites (details in (*19*)). A whole body scan (3min/bed position) was made from mid-thigh to skull vertex at 45 min post-injection. Complying with the EARL1 guideline (*16*), whole body PET images were reconstructed with standard iterative time-of-flight reconstruction algorithm (BLOB-OS-TF) with 3 iterations and 33 subsets, with a matrix size of 144x144 and voxels 4x4x4mm. Images were corrected for scatter, random coincidences, decay, and attenuation (low-dose CT; 80 mA at 120-140 kV, 5 mm slice thickness).

We additionally reconstructed images with the PSF algorithm as provided by the vendor (Philips Healthcare) to conform with EARL2 guidelines. This comprises post-reconstruction image processing using the Richardson-Lucy iterative deconvolution algorithm with sieve noise regularization (PSF option: 1 iteration, regularization full-width-at-half-max at 6 mm) as resolution recovery method (*20*). This algorithm uses a scanner-specific spatially variant PSF to improve image resolution, and is described as follows (*20*):

$$\boldsymbol{I_{i+1} = \frac{I_i}{f*s}\left(f * s * \frac{I_o}{I_i \otimes s \otimes f}\right)} \qquad \text{Eq. 1}$$

where I_{i+1} is the current image estimate; I_i is the image estimate from the i^{th} iteration; f is the system Gaussian PSF; s is the sieve kernel; and I_o is the original measured image.

To evaluate the impact of count statistics (i.e., acquisition time), we split the original list-mode data of each whole body PET scan on an alternating count-wise basis into two new datasets, which were subsequently reconstructed into whole body images (as proposed in (*21,22*)) using both EARL1 (4 mm) and EARL2 (4 mm+PSF) reconstructions. This generated two statistically equivalent but count-independent PET-images each containing 50% of the original counts (referred to as split 1 and split 2). Due to the linear relationship between (decay corrected)

number of counts and acquisition time, the whole body images reconstructed from split 50% count data served as surrogates for images acquired at 1.5min per bed position (Figure 4.1).

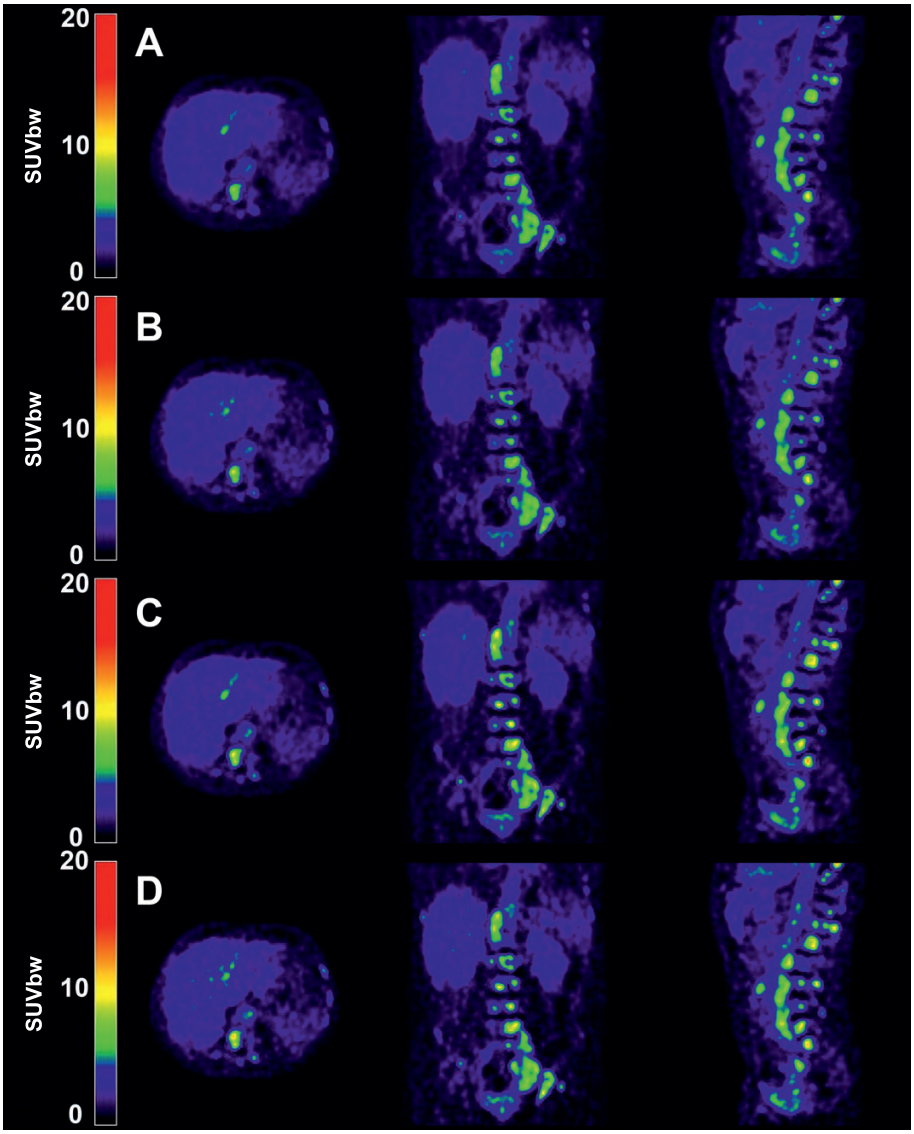


Figure 4.1: Illustration of a PET image of a typical mCRPC patient with extensive [18F]FDHT-avid bone metastases reconstructed with (A) 100% counts EARL1, (B) 50% counts EARL1, (C) 100% counts EARL2, and (D) 50% counts EARL2. Axial (left column), coronal (middle column), and sagittal views (right column) are shown.

Image analysis

All suspicious FDHT-avid lesions with uptake exceeding background were included. Volumes of interest (VOI) were delineated on the original PET images using a semi-automatic algorithm using a threshold of 50% of the peak value within the area of interest with correction for local background uptake(*23*). From each VOI, we derived the average, peak, and maximum activity concentrations (AC; Bq/cc). Next, SUV was derived by dividing the tumour AC to a normalization factor. Two normalizations were used (*12*): a) injected dose per kg bodyweight (bw), and b) area-under-the-curve of the parent plasma calibrated image-derived input function (AUC-PP). The AR-positive tumour volume (ARTV; mL) was defined as the sum of all voxel volumes within a VOI. To assess lesion detectability, we generated a single voxel thick shell around the tumour VOI to determine local background activity, yielding the contrast-to-noise ratio (CNR) as follows:

$$CNR = \frac{(AC_{avg} - AC_{bgr})}{SD_{bgr}} \quad \text{Eq. 2}$$

where ACavg is the average tumour AC; ACbgr is the average background AC, and SDbgr is the standard deviation of AC in voxels included in the background shell.

To compare image noise levels of the 100% and 50% count images, a 3cm diameter spherical VOI was placed in the liver, from which the coefficient of variation (COV%) was calculated as follows:

$$COV\% = \frac{SD_{liver}}{AC_{liver}} \times 100 \quad \text{Eq. 3}$$

where SDliver is the standard deviation of the ACs of voxels within the liver VOI, and ACliver is the mean AC of voxels within the liver VOI.

Statistical Analysis

Analyses were performed using SPSS statistics (v22, IBM) and Excel datasheets. Intrascan variability was defined as the difference in SUVs between the split scans of each original scan (eg. split 1 versus split 2). Interscan variability was defined as the test-retest variability (repeatability) of SUVs. Both intra- and interscan variability were assessed on a per-lesion basis. Repeatability coefficients (RC%) were calculated from the standard deviations of the relative differences of measured SUVs between test-retest scans (day 1 vs day 2) and between split

scans (split 1 vs split 2). To evaluate test-retest variability, 50% count scans were compared to the mean since split scans could not be directly compared as this would yield 4 comparisons (Figure 4.2). Calculation of test-retest RCs was as follows (Equation 4-8):

$$d = \frac{SUV_2 - SUV_1}{\overline{SUV_{orig}}} \times 100 \quad \text{Eq. 4}$$

$$d_{i,j} = \frac{SUV_{i,j} - \overline{SUV_{i,j}}}{\overline{SUV_{i,j}}} \times 100 \quad \text{Eq. 5}$$

$$SD = \sqrt{\frac{\sum(d-\bar{d})^2}{n-1}} \quad \text{Eq. 6}$$

$$SD_{split} = \sqrt{\frac{\sum(d_{i,j} - \overline{d_{i,j}})^2}{n-1}} \quad \text{Eq. 7}$$

$$RC = SD * 1.96 \quad \text{Eq. 8}$$

where d is the relative difference between day 1 (SUV_1) and day 2 (SUV_2) for original data; $d_{i,j}$ is the relative difference of each split *i* (split 1 and split 2) on each day *j* (days 1 and 2) compared to the average $SUV_{i,j}$; SD is the standard deviation of relative test-retest differences (SD_{split} was scaled by factor 2 since SUVs were compared to the mean $SUV_{i,j}$). RCs of intrascan variability were calculated using eq. 4, 6, and 8.

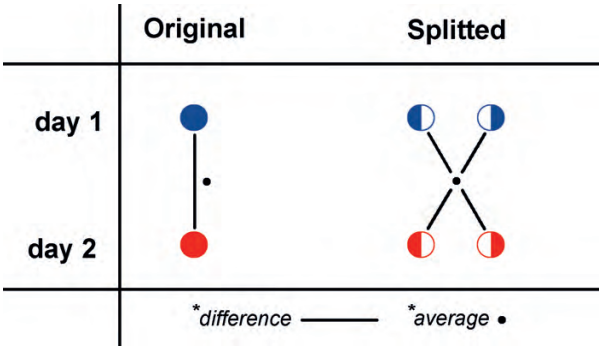


Figure 4.2: Schematic representation of assessment of test-retest variability of original 100% count scans and split 50% count scans, respectively. Note that, in contrast with original scans, split scans cannot be directly compared, as this would yield 4 individual comparisons underestimating true test-retest variability.

Bland-Altman plots with 95% limits of agreement, R^2 and intraclass correlation coefficients (ICC) were calculated were calculated to assess inter- and intrascan variability (*24*). ICCs represent the fraction of the total variability attributable to between lesion variability, and were calculated using a two-way mixed model with absolute agreement definition (*25*). To test for differences in repeatability between 100% and 50% count scans we used a Wilcoxon signed-rank test designed to compare variances of dependent data ($p<0.05$) (*26*).

Results

Fourteen patients with a median age of 65 (IQR 47-75) years were included. Median Gleason score was 8 (IQR 5-10) and median PSA at imaging was 103 (IQR 11-1602) ng/ml. Median injected dosages of [18F]FDHT on day 1 and day 2 were 194 MBq (range 152-216) and 193 MBq (range 186-215) with residual activity in syringes/tubes of 37 MBq (range 26-63) and 36 MBq (18-54), respectively (*12*). In two patients, no FDHT-avid lesions were detected. In the remaining 12 patients, 336 FDHT-avid lesions were visually detectable on both test and retest PET-CT scans.

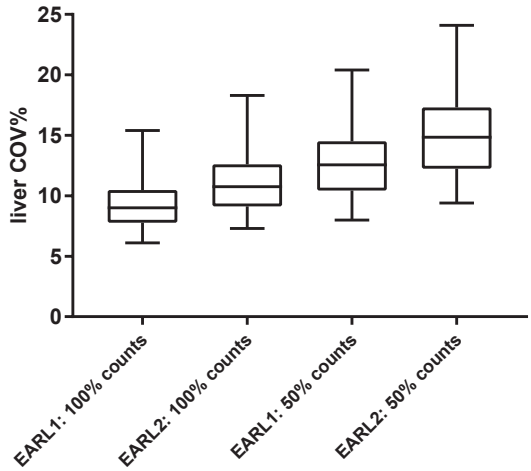


Figure 4.3: Liver COV% for 100% count and 50% count EARL1 and EARL2 images.

Image noise

Liver COV% of EARL1 images increased from a median 9.0 (IQR 7.9-10.4) to 12.5 (IQR 10.5-14.5) after count reduction (Figure 4.3). For EARL2 images, liver COV% increased from a median 10.8 (IQR 9.2-12.4) to 13.2 (IQR 12.4-17.3) after count reduction.

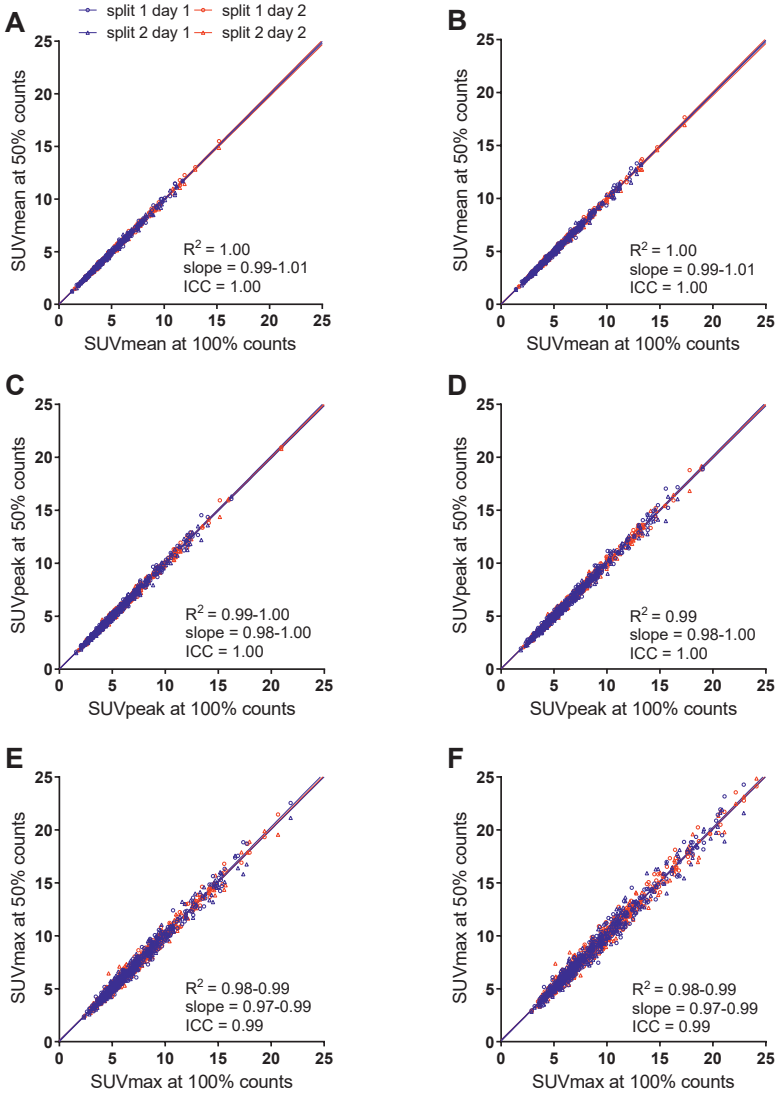


Figure 4.4: Correlations between SUVbw of original 100% count scans and split 50% count scans for SUVmean (A and B), SUVpeak (C and D), and SUVmax (E and F). Results from both EARL1 images (A, C, E) and EARL2 images (B, D, F) are shown.

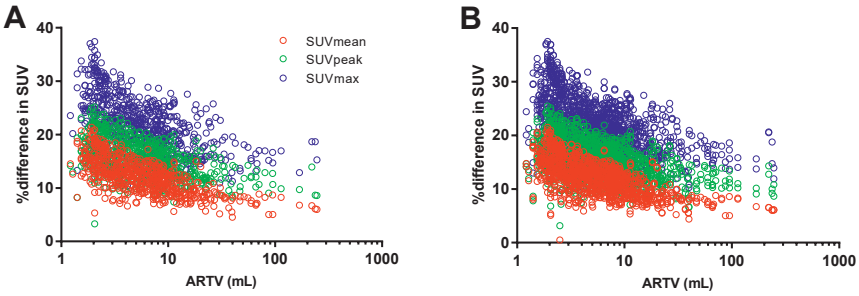


Figure 4.5: Relative difference (%) between SUVs derived from EARL1 images compared to SUVs derived from EARL2 images as a function of lesion ARTV. (A) Results from original (100% of counts) scans and (B) split (50% of counts) images.

Semi-quantitative measurements

On EARL1 images, SUVmean, SUVpeak, and SUVmax of 100% count scans were highly correlated with those of 50% count scans (Figure 4.4), with SUVmax being most affected by count reduction (albeit still with R^2>0.98 and ICC=0.99-1.00). Similar results were observed for EARL2 images (Figure 4.4), with again SUVmax being most affected by count reduction (R^2>0.98 and ICC=0.99-1.00). On 100% count EARL2 images, SUVmean, SUVpeak and SUVmax were median 12.1% (IQR 9.8-14.2%), 15.6% (IQR 13.1-17.9%), and 21.7% (IQR 18.4-25.1%) higher compared to EARL1 images. Similarly, on 50% count EARL2 images, SUVmean, SUVpeak and SUVmax were a median 12.0% (9.6-14.2%), 15.5% (IQR 13.2-17.9%), and 21.6% (IQR 18.5-25.1%) higher compared to EARL1 images. These relative differences were inversely related to lesion ARTV (Figure 4.5).

Intrascan variability

RCs between 50% count scans on day 1 and 2, respectively, were 9.9% and 8.5% for SUVmean, 14.3% and 12.0% for SUVpeak, and 19.6% and 17.8% for SUVmax (Figure 4.6) on EARL1 images. ICCs and R^2 values between SUVs of 50% count EARL1 scans were high (ICC=0.97-1.00; R^2=0.95-0.99). On EARL2 images, RCs between 50% count scans on day 1 and 2, respectively, were 10.6% and 9.1% for SUVmean, 16.0% and 13.3% for SUVpeak, and 22.2% and 19.8% for SUVmax (Figure 4.6). For EARL2 images, ICCs and R^2 between SUVs of 50% count scans were almost identical to EARL1 (ICC=0.97-0.99.;R^2=0.94-0.99.). SUV intrascan variability was volume-dependent for both EARL1 and EARL2 images (Supplemental Figure 1, available at https://ejnmmires.springeropen.com).

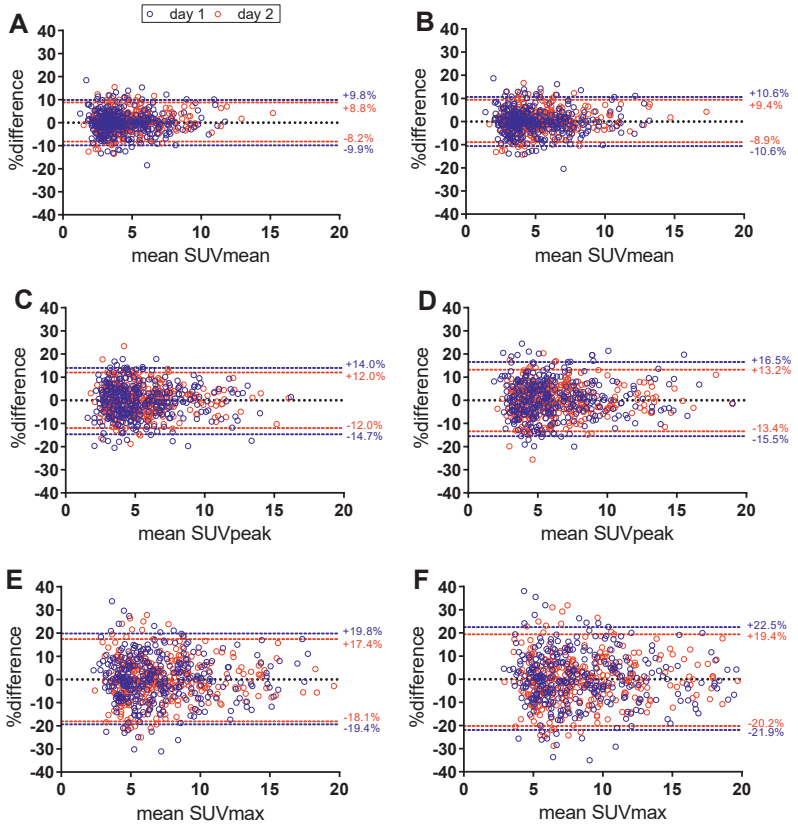


Figure 4.6: Bland-Altman graph of intrascan variability due to 50% count reduction for SUVmean (A and B), SUVpeak (C and D), and SUVmax (E and F). Results from both EARL1 images (A, C, E) and EARL2 images (B, D, F) are shown. Variability was derived from the relative difference in SUV between split 1 and 2 of each scan on each day. Note that bw or AUC-PP normalization are not reported separately since normalization factors are identical for split 1 and split 2.

Interscan variability

For EARL1 images, RCs of 50% count scans were higher than RCs of 100% count scans, but differences in variances were not significant (Figure 4.7 and Table 4.1; ICC 0.94-0.97). A similar effect of count reduction on RCs was observed for selected lesions with ARTV >4.2 mL, but in general RCs for lesions >4.2 mL were lower (Table 4.1). Repeatability of EARL2 was worse than EARL1 at both 100% and 50% count data (Figure 4.7 and Table 4.1). Repeatability of EARL2 images was more affected by count reduction than EARL1 images, yet differences between variances of 100% and 50% count data were not significant (p=0.53-1.00; Table 4.1). Normalizing SUVs to bw (Figure 4.7) resulted in lower RCs than normalizing to AUC-PP (Supplemental Figure 2).

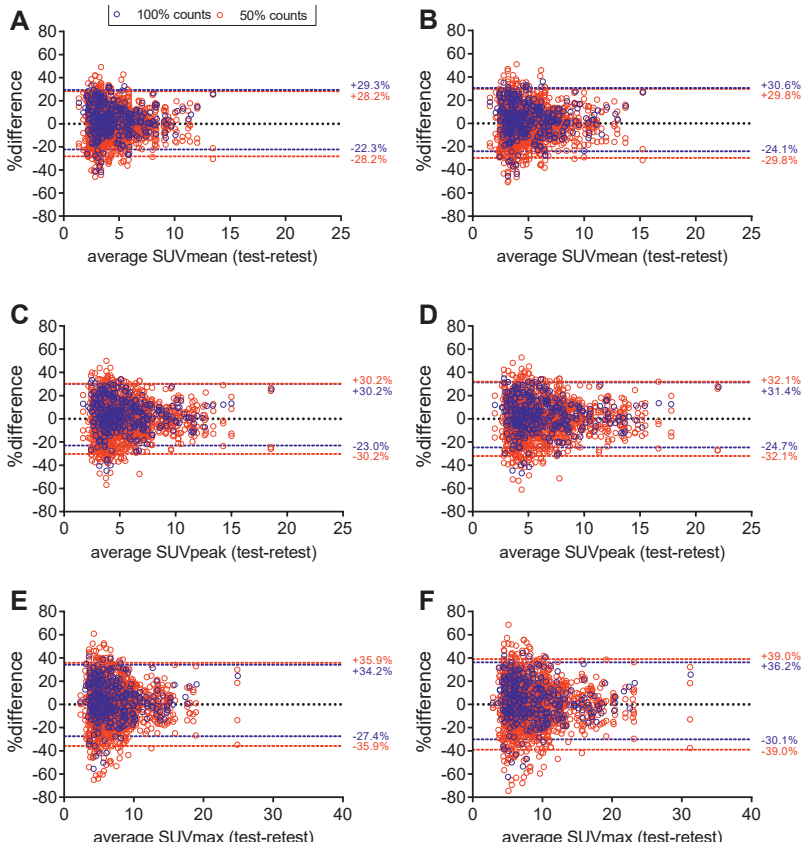


Figure 4.7: Bland-Altman graph of interscan (test-retest) variability of SUVmean (A and B), SUVpeak (C and D), and SUVmax (E and F) normalized to bodyweight at 100% and 50% of counts. Results from both EARL1 images (A, C, E) and EARL2 images (B, D, F) are shown.

Lesion detectability

In general, the impact of count reduction on lesion detectability was small, with a median 4.6% (IQR 1.2-8.7%) reduction in CNR from median 3.7 (IQR 3.1-4.3) to 3.5 (IQR 2.9-4.1) on EARL1 images after count reduction. For EARL2 images, there was a median 4.6% (IQR 1.2-8.7%) reduction in CNR from median 3.9 (IQR 3.1-4.7) to 3.7 (IQR 2.9-4.5) after count reduction.

Table 4.1: Interscan variability (test-retest). Agreement between SUVs from test-retest scans (day 1 vs day 2) for original (100% count) scans and split (50% count) scans of both EARL1 and EARL2 images.

Image	SUV normalization	SUV type	100% of counts		50% of counts			
			RC% (all lesions)	RC% (>4.2 ml)	RC% (all lesions)	RC% (>4.2 ml)	p-value (all lesions)	p-value (>4.2 ml)
EARL1	Bodyweight	SUVmean	25.8	23.23	28.2	25.0	0.18	0.80
		SUVpeak	26.6	24.70	30.2	27.9	0.13	0.97
		SUVmax	30.8	26.92	35.9	31.5	0.50	0.41
	AUC-PP	SUVmean	27.6	27.53	29.7	30.0	0.52	0.95
		SUVpeak	27.7	28.17	31.3	32.2	1.00	1.00
		SUVmax	31.7	30.39	36.8	35.8	0.42	0.31
EARL2	Bodyweight	SUVmean	27.3	24.4	29.8	26.2	0.58	0.91
		SUVpeak	28.1	26.0	32.1	29.6	0.70	0.99
		SUVmax	33.1	28.6	39.0	34.0	0.37	0.74
	AUC-PP	SUVmean	28.5	28.0	30.6	30.3	0.37	0.99
		SUVpeak	28.8	29.2	32.8	33.4	0.69	1.00
		SUVmax	33.6	31.6	39.5	37.7	0.56	0.53

RC% = repeatability coefficient; p-value for differences between variances of test-retest differences of 100% vs 50% count images.

Discussion

We investigated how accuracy, precision, and lesion detectability of analogue whole body [18F]FDHT PET-CT are affected by image count statistics and reconstruction protocol, to optimize imaging protocols for research and clinical use. Reducing counts by 50% introduced <20% SUV intrascan variability for EARL1 images, which only increased test-retest variability to a small extent. Improving image spatial resolution by adhering to EARL2 guidelines might reduce the size-dependent bias in SUV, but it hampers repeatability and increases sensitivity to count statistics. Lesion detectability is only slightly affected by reduced counts and only marginally increased by resolution modelling.

SUVs of 50% count scans correlated highly with SUVs of 100% count scans, indicating accuracy is preserved at lower count statistics. However, when comparing split scans directly a variability in SUV ranging 8.5% (SUVmean EARL1) to 22.2% (SUVmax EARL2) was observed. Hence, while SUV accuracy is maintained at low counts, its precision might be hampered. Still, test-retest variability only increased to a small and non-significant extent, which indicates that the statistical Poisson image noise is a minor determinant of SUV repeatability for [18F]FDHT.

SUV repeatability of oncological 18F-tracers (ie. [18F]FDG, [18F]-fluorothymidine, [18F]-fluoromethylcholine, [18F]FDHT) ranges between 10-30%, yielding 30% as the preferred upper threshold for SUV variability for use in e.g. response monitoring studies (*27-30*). As expected, repeatability of SUVmax was most affected by count reduction and EARL2 reconstruction, yielding RCs >30%. In contrast, SUVpeak seemed to be robust to both count statistics and reconstruction protocol, yielding an RC of approximately 30% after count reduction, which was even lower (27.9%) when only lesions >4.2mL were considered. The improved repeatability of SUV when excluding small lesions seems a direct consequence of the size-dependency of intrascan variability at reduced counts (Supplemental Figure 1). Note that test-retest variability of [18F]FDHT can be even lower when evaluating only selected target lesions, or analysing on a patient- instead of lesion-basis (*12*). In the current study, all avid lesions were primarily included to avoid selection bias and also evaluate the effect of count reduction on smaller and less avid lesions.

Between SUV normalizations, differences in test-retest variability were observed, with larger variability in SUVauc-pp (>30%) compared SUVbw. While

SUV normalized to AUC-PP correlates better with reference pharmacokinetic parameters than SUV normalized to bodyweight (*19*), deriving it is more technically demanding and less precise compared to more simple factors such as dose per bodyweight, making it less suitable for multicentre studies. Hence, a trade-off between accuracy, precision, and ease of use has to be made when selecting the preferred SUV normalization. For example, while SUVpeak normalized to bodyweight had a RC of 30% at half of counts, it exceeded 30% when normalizing to AUC-PP rendering it unfit for response assessment.

Partial-volume effects generally result in volume-dependent underestimations of tumour SUV and possibly hamper lesion detectability (*31*). Correcting for PVE in the reconstruction algorithms might be particularly important in [^{18}F]FDHT due to the high frequency of small (e.g. <4.2 ml) detected lesions. Novel reconstruction algorithms incorporating the PSF either within or after reconstruction have been proposed to improve image resolution (*17*). The EARL2 standards have adopted these algorithms as a step forward in scanner calibration harmonization between centres (*15*). However, PSF reconstructions are known to suffer from noise propagation and image artefacts (e.g. Gibbs phenomenon resulting in edge overshoot), which might lead to misinterpretation regarding treatment effects (*17,18,32*). Indeed, we observed that repeatability was worse for the EARL2 reconstruction with higher sensitivity to count statistics, resulting in a higher minimal detectable change for response assessment.

Previous reports argued that PSF reconstructions should be used for qualitative purposes (i.e. lesion detection), and that non-PSF images (such as EARL1) should be used for tumour quantification (*18,33*). However, Quak et al. found that with additional image filtering the higher lesion detection and image resolution of PSF images do not need to be impaired in order to meet the EARL criteria (*34*). In the present study we observed a very small increase in lesion CNR when PSF was applied. This will not likely result in clinically relevant different conclusions regarding the extent of disease or intrapatient heterogeneity (Figure 1) due to the vast amount of detected lesions (336 lesions in 12 patients). The small reduction in CNR by <5% after count reduction is also not likely to have clinical consequences (Figure 1). This corresponds to [^{18}F]FDG PET-CT data in several cancer types, where reducing acquisition time from 3 to 1.5min per bed position reduced image quality, but did not impair lesion detection rates (*13*).

Another factor affecting image count statistics is the injected tracer dosage. In the present cohort, patients received a relatively low dosage compared to other

cohorts from the recent multicentre study (*12*). However, while SUV test-retest variability varied between centres, the authors did not observe a direct relationship between injected dosage and repeatability (*12*). This might be explained by differences in other factors determining repeatability, such as the observer variability in tumour delineation, PET system specifics, adherence to imaging protocols (i.e. uptake interval), and methods for acquiring the SUV normalization factors. Hence, count statistics did not appear to be the main determinant of [18F]FDHT repeatability, which we confirm in the current study where non-significant increases in test-retest RCs were observed after count reduction. Therefore, a potentially modifiable and important determinant of SUV variability in [18F]FDHT imaging seems to be the choice of normalization factors, which, again, need some trade-off between accuracy and precision to be made.

The present study contains several limitations. First, while splitting data on a count-wise basis enables evaluation of Poisson noise induced by count reduction, the 50% count scans do not fully represent a 50% shorter image acquisition. However, [18F]FDHT kinetics commonly reach a plateau after 20-30min, yielding stable SUV during the whole body acquisition (*8*). Second, the present study contains data acquired on a PET system of a single vendor. As between vendors the overlap between bed positions differs, count reduction might have a different impact on measurement variability for these PET systems. Also, for novel PET systems, which may have higher sensitivities and better time of flight performance, in particular for the new digital systems, the impact of reducing acquisition times on measurement variability will be even smaller. Hence, for these systems acquisition times may be reduced even further, but this remains to be investigated for each type of system. As investigated in the present study for analogue PET, a reduction up to 50% compared with current standard practice seems to be feasible for diagnostic and response assessment purposes, warranted that use of SUVmax is avoided.

The current approach for evaluating the sensitivity of whole body PET-CT acquisition to scan statistics can be extended to other tracers currently being investigated and/or implemented in clinical practice, such as PSMA-ligand PET-CT. For adequate evaluation of these tracers, however, test-retest data should be available.

Conclusion

In [^{18}F]FDHT PET-CT studies, noise-induced SUV variability leads to small increases in test-retest variability, which improves when excluding small lesions. Novel EARL2-compliant reconstruction increases lesion SUVs and marginally increases CNRs. However, it requires higher count statistics to preserve adequate precision. For SUVpeak normalized to bodyweight, test-retest variability remained below 30% when lesions <4.2 ml were excluded, which is generally acceptable for oncological [^{18}F]-tracers. In contrast, SUVmax was substantially affected by count reduction and EARL2 reconstruction, hence its use should be avoided. Lesion detectability is only slightly impaired by reducing counts by 50%, which might not be clinically relevant in the mCRPC population.

Taken together, with the current imaging procedure on an analogue PET-CT system count statistics are more than sufficient and could even be reduced by 50% without affecting diagnostic performance and at a small expense of reduced precision of response assessments. Acquisition time reduction is feasible for staging and response assessment purposes.

References

1. Heinlein CA, Chang C. Androgen receptor in prostate cancer. ***Endocr Rev.*** 2004;25:276-308.
2. Ryan CJ, Smith MR, Fizazi K, et al. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naive men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. ***Lancet Oncol.*** 2015;16:152-160.
3. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. ***N Engl J Med.*** 2011;364:1995-2005.
4. Beer TM, Armstrong AJ, Rathkopf DE, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. ***N Engl J Med.*** 2014;371:424-433.
5. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. ***N Engl J Med.*** 2012;367:1187-1197.
6. Scher HI, Morris MJ, Stadler WM, et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. ***J Clin Oncol.*** 2016;34:1402-1418.
7. Larson SM, Morris M, Gunther I, et al. Tumor localization of 16beta-18F-fluoro-5alpha-dihydrotestosterone versus 18F-FDG in patients with progressive, metastatic prostate cancer. ***J Nucl Med.*** 2004;45:366-373.
8. Beattie BJ, Smith-Jones PM, Jhanwar YS, et al. Pharmacokinetic assessment of the uptake of 16beta-^{18}F-fluoro-5alpha-dihydrotestosterone (FDHT) in prostate tumors as measured by PET. ***J Nucl Med.*** 2010;51:183-192.
9. Fox JJ, Gavane SC, Blanc-Autran E, et al. Positron Emission Tomography/Computed Tomography-Based Assessments of Androgen Receptor Expression and Glycolytic Activity as a Prognostic Biomarker for Metastatic Castration-Resistant Prostate Cancer. ***JAMA Oncol.*** 2018;4:217-224.
10. Scher HI, Beer TM, Higano CS, et al. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. ***Lancet.*** 2010;375:1437-1446.
11. Rathkopf DE, Morris MJ, Fox JJ, et al. Phase I study of ARN-509, a novel antiandrogen, in the treatment of castration-resistant prostate cancer. ***J Clin Oncol.*** 2013;31:3525-3530.
12. Vargas HA, Kramer GM, Scott AM, et al. Reproducibility and Repeatability of Semiquantitative (18) F-Fluorodihydrotestosterone Uptake Metrics in Castration-Resistant Prostate Cancer Metastases: A Prospective Multicenter Study. ***J Nucl Med.*** 2018;59:1516-1523.
13. Hausmann D, Dinter DJ, Sadick M, Brade J, Schoenberg SO, Busing K. The impact of acquisition time on image quality in whole-body ^{18}F-FDG PET/CT for cancer staging. ***J Nucl Med Technol.*** 2012;40:255-258.
14. Kadrmas DJ, Oktay MB, Casey ME, Hamill JJ. Effect of Scan Time on Oncologic Lesion Detection in Whole-Body PET. ***IEEE Trans Nucl Sci.*** 2012;59:1940-1947.
15. Kaalep A, Sera T, Oyen W, et al. EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems. ***Eur J Nucl Med Mol Imaging.*** 2018;45:412-422.
16. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
17. Rahmim A, Qi JY, Sossi V. Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls. ***Medical Physics.*** 2013;40.
18. Munk OL, Tolbod LP, Hansen SB, Bogsrud TV. Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative. ***EJNMMI Phys.*** 2017;4:5.
19. Kramer GM, Yaqub M, Vargas HA, et al. Assessment of simplified methods for quantification of (18) F-FDHT uptake in patients with metastatic castration-resistant prostate cancer. ***J Nucl Med.*** 2019.
20. Narayanan M, Perkins A. ***Resolution recovery in the Ingenuity TF PET/CT*** [White Paper]. Retrieved January 23, 2019 from Philips Healthcare: http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/PetCT/Resolution-recovery-in-the-Ingenuity-TF-PET-CT.
21. Kramer GM, Frings V, Heijtel D, et al. Parametric Method Performance for Dynamic 3'-Deoxy-3'-(18) F-Fluorothymidine PET/CT in Epidermal Growth Factor Receptor-Mutated Non-Small Cell Lung Carcinoma Patients Before and During Therapy. ***J Nucl Med.*** 2017;58:920-925.

22. Jauw YWS, Heijtel DF, Zijlstra JM, et al. Noise-Induced Variability of Immuno-PET with Zirconium-89-Labeled Antibodies: an Analysis Based on Count-Reduced Clinical Images. ***Mol Imaging Biol.*** 2018;20:1025-1034.

23. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. ***Radiology.*** 2014;273:539-548.

24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. ***Lancet.*** 1986;1:307-310.

25. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. ***J Chiropr Med.*** 2016;15:155-163.

26. Wilcox RR. Comparing the Variances of Dependent Groups. ***Psychometrika.*** 1989;54:305-315.

27. Oprea-Lager DE, Kramer G, van de Ven PM, et al. Repeatability of Quantitative ^{18}F-Fluoromethylcholine PET/CT Studies in Prostate Cancer. ***J Nucl Med.*** 2016;57:721-727.

28. Lodge MA. Repeatability of SUV in Oncologic (^{18})F-FDG PET. ***J Nucl Med.*** 2017;58:523-532.

29. Kramer GM, Liu Y, de Langen AJ, et al. Repeatability of quantitative (18)F-FLT uptake measurements in solid tumors: an individual patient data multi-center meta-analysis. ***Eur J Nucl Med Mol Imaging.*** 2018;45:951-961.

30. 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.

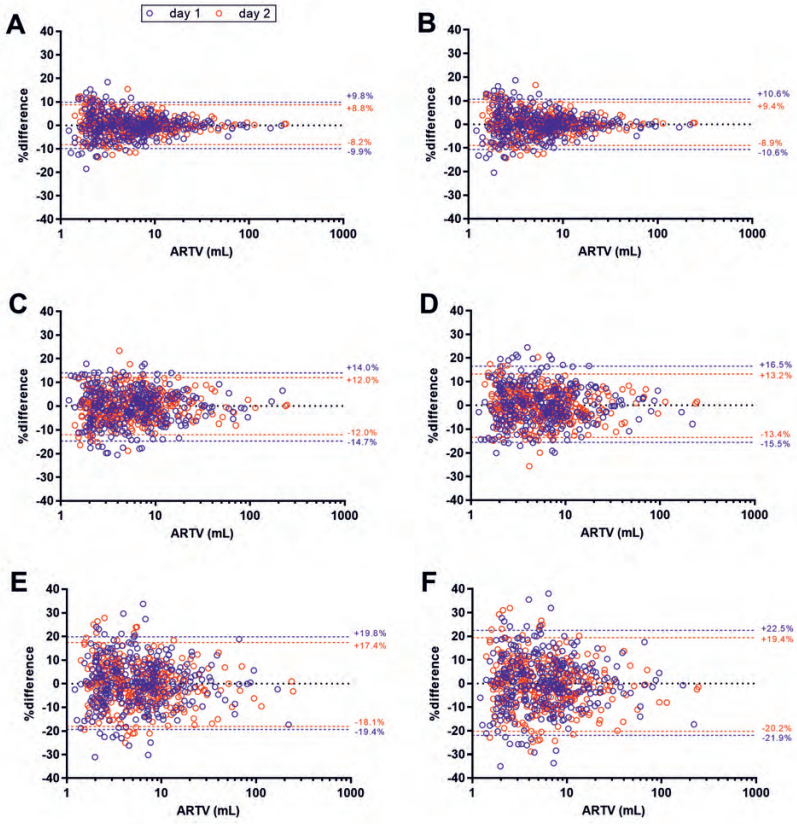
31. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. ***J Nucl Med.*** 2007;48:932-945.

32. Rahmim A, Tang J. Noise propagation in resolution modeled PET imaging and its impact on detectability. ***Physics in Medicine and Biology.*** 2013;58:6945-6968.

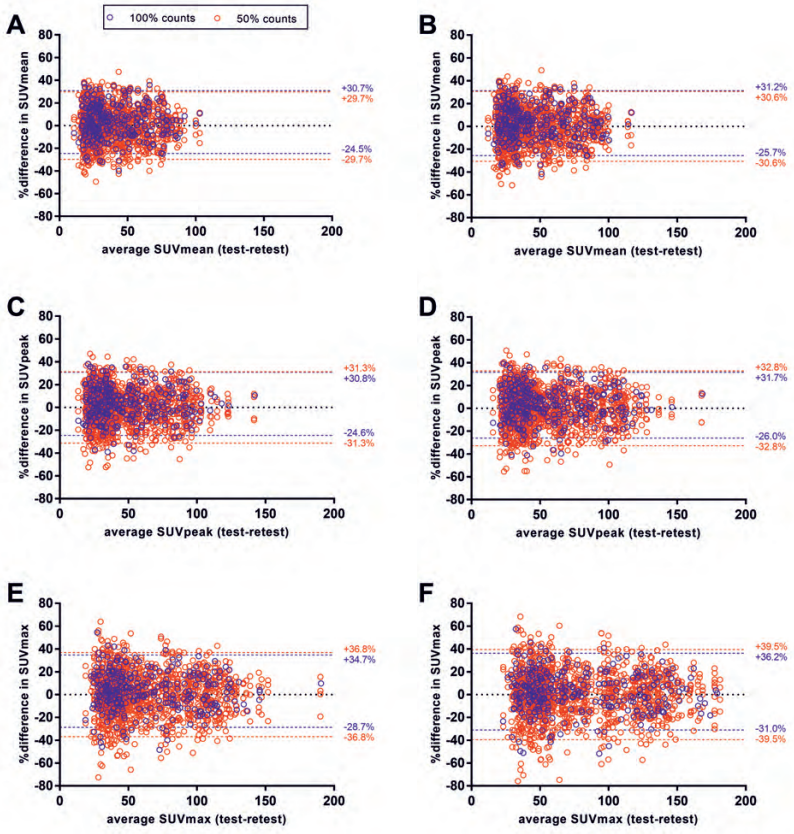
33. Devriese J, Beels L, Maes A, Van de Wiele C, Pottel H. Impact of PET reconstruction protocols on quantification of lesions that fulfil the PERCIST lesion inclusion criteria. ***EJNMMI Phys.*** 2018;5:35.

34. Quak E, Le Roux PY, Hofman MS, et al. Harmonizing FDG PET quantification while maintaining optimal lesion detection: prospective multicentre validation in 517 oncology patients. ***Eur J Nucl Med Mol Imaging.*** 2015;42:2072-2082.

Supplemental files



Supplemental Figure 4.1: Intrascan variability due to 50% count reduction as function of lesion ARTV for SUVmean (A and B), SUVpeak (C and D), and SUVmax (E and F). Results from both EARL1 images (A, C, E) and EARL2 images (B, D, F) are shown, with limits of agreement from Bland-Altman analyses.



Supplemental Figure 4.2: Bland-Altman graph of interscan (test-retest) variability of SUVmean (A and B), SUVpeak (C and D), and SUVmax (E and F) normalized to AUC-PP at 100% and 50% of counts. Results from both EARL1 images (left column) and EARL2 images (right column) are shown.


QI-1-PET

Chapter 5

Repeatability of Quantitative [^{18}F]DCFPyL PET-CT Measurements in Metastatic Prostate Cancer

B.H.E. Jansen, M.C.F. Cysouw, A.N. Vis, R.J.A. van Moorselaar, J. Voortman, Y.J.L. Bodar, P.R. Schober, N.H. Hendrikse, O.S. Hoekstra, R. Boellaard, D.E. Oprea-Lager

J Nucl Med. 2020 Jan 10.
doi: 10.2967/jnumed.119.236075.

Abstract

Quantitative evaluation of radiolabelled Prostate-Specific Membrane Antigen (PSMA) PET scans may be used to monitor treatment response in patients with prostate cancer (PCa). To interpret longitudinal differences in PSMA uptake, the intrinsic variability of tracer uptake in PCa lesions needs to be defined. The aim of this study was to investigate the repeatability of quantitative [^{18}F]DCFPyL (a second generation [^{18}F]PSMA-ligand) PET/CT measurements in patients with PCa.

Methods: Twelve patients with metastatic PCa were prospectively included, of which 2 were excluded from final analyses. Patients received two whole-body [^{18}F]DCFPyL PET/CT scans (median dose 317 MBq; uptake time 120 min), within median 4 days (range 1-11 days). After semi-automatic (isocontour-based) tumour delineation, the following lesion-based metrics were derived: Tumour-to-Blood ratio (TBRmean, TBRpeak, and TBRmax), Standardized Uptake Value (SUVmean, SUVpeak, SUVmax, normalized to bodyweight), tumour volume, and total lesion tracer uptake (TLU). Additionally, patient-based Total Tumour Volume (sum of PSMA-positive tumour volumes; TTV) and Total Tumour Burden (sum of all lesion TLUs; TTB) were derived. Repeatability was analysed using repeatability coefficients (RC) and intra-class correlations (ICC). Additionally, the effect of point spread function (PSF) image reconstruction on the repeatability of uptake metrics was evaluated.

Results: In total, 36 [^{18}F]DCFPyL PET positive lesions were analysed (up to 5 lesions per patient). RCs of TBRmean, TBRpeak, and TBRmax were 31.8%, 31.7%, and 37.3%, respectively. For SUVmean, SUVpeak, SUVmax the RCs were 24.4%, 25.3% and 31.0%, respectively. All ICC were ≥0.97. Tumour volume delineations were well repeatable, with RC 28.1% for individual lesion volumes and RC 17.0% for TTV. TTB had a RC of 23.2% and 33.4%, when based on SUVmean and TBRmean, respectively. Small lesions (<4.2mL) had worse repeatability for volume measurements. The repeatability of SUVpeak, TLU, and all patient-level metrics were not affected by PSF-reconstruction.

Conclusion: [^{18}F]DCFPyL uptake measurements are well repeatable and can be used for clinical validation in future treatment response assessment studies. Patient-based TTV may be preferred for multicenter studies since its repeatability was both high and robust to different image reconstructions.

Introduction

Prostate cancer (PCa) is the second most common cancer in men worldwide, with an estimated annual number of deaths over 350.000 (*1*). Prostate-Specific Membrane Antigen (PSMA) Positron Emission Tomography (PET) is increasingly used for PCa diagnostics (*2*). PSMA is a class II trans-membrane glycoprotein that provides a valuable target for radiolabelled imaging, as its expression is upregulated in malignant prostate cells and associated with aggressive disease characteristics (*3*). Due to larger availability, ^{68}Gallium-labeled PSMA tracers have been studied most frequently to date, demonstrating high detection rates for metastatic disease (*2,4*). Alternatively, ^{18}Fluorine-labeled tracers have been developed, including [^{18}F]DCFPyL (2-(3-(1-carboxy-5-[(6-^{18}F-fluoro-pyridine-3-carbonyl)-amino]-pentyl)-ureido)-pentanedioic acid), a second-generation small-molecule ligand that strongly binds to PSMA (*5,6*). The ^{18}F-radionuclide provides for PET-images with a higher resolution compared to ^{68}Ga, due to a shorter positron range and higher positron yield (*2*).

Quantitative analysis of PSMA uptake may be used to predict or evaluate treatment response, as changes in PSMA uptake over time may indicate response to treatment or progression of disease (*7-9*). Recently, we performed a full pharmacokinetic analysis of [^{18}F]DCFPyL to validate simplified methods for tumour uptake quantification. Tumour-to-Blood Ratios (TBR; tracer activity concentration in the tumour normalized to the whole blood activity concentration on PET) were found to best describe the tumour tracer uptake (*10*). For reliable use of quantitative PSMA PET metrics in clinical practice, it is important to determine their repeatability. Only changes that exceed random variability should be interpreted as treatment response or disease progression. To the best of our knowledge, this is the first study reporting on the day-to-day variability of [^{18}F]DCFPyL uptake in PCa lesions. The aim of this study was to evaluate the repeatability of quantitative [^{18}F]DCFPyL PET/CT measurements in patients with metastatic PCa.

Materials and methods

Patients

Twelve patients were prospectively included in the Amsterdam UMC between January and May 2019. Inclusion criteria were: (1) histologically proven PCa; (2) at least two metastases, detected by any imaging modality; and (3) at least one metastasis of ≥1.5cm in size (to minimize partial-volume effects). Patients with multiple malignancies and claustrophobia were excluded.

The study was approved by the ethical review board of the Amsterdam UMC and all subjects signed informed consent. This trial was registered under EudraCT number 2017-000344-18 and the Netherlands Trial Registry number 6477. Personal and demographic data regarding age, length, body weight, Gleason score, prostate-specific antigen level (ng/mL) at the time of PET/CT, and information on prior therapy were collected.

Data Acquisition

All patients underwent two [^{18}F]DCFPyL PET/CT scans within 4 days (median, range 1-11 days). PET/CT imaging adhered to routine clinically used protocols. [^{18}F]DCFPyL was synthesized under Good Manufacturing Practices conditions at the Amsterdam UMC (Radionuclide Center), using the precursor of ABX GmbH® (Germany) (Supplemental Text 1, available at http://jnm.snmjournals.org (*11*)). No fasting was required and no diuretics were administered prior to imaging. PET was performed using an European Association of Nuclear Medicine Research Ltd (EARL) calibrated hybrid Philips Ingenuity TF scanner (Philips Healthcare®, the Netherlands/USA) (*12,13*). Our imaging protocol included a target injected [^{18}F] DCFPyL dose of 300 MBq, with an uptake interval after injection of 120 minutes. First, a CT scan was made for attenuation correction (30-120 mAs; 120 kV). Next, whole-body PET acquisitions were then acquired from mid-thigh to skull base (4 minutes per bed position).

Images were corrected for decay, scatter, random coincidences, and photon attenuation. Images were reconstructed using the (EARL1-compliant (*13*)) Ordered Subsets Expectation Maximization with Time-of-Flight algorithm (3 iterations; 33 subsets). Additionally, images were post-processed using Lucy-Richardson iterative deconvolution (henceforth referred to as Point Spread Function [PSF] reconstruction) (*14*). Using a NEMA-NU-2 Image Quality Phantom the full-width-at-half-maximum of this reconstruction was calibrated at 7.0mm for adequate signal recovery complying with novel EARL2 guidelines (*15*).

Data Analysis

All scans were controlled for image-quality (*16*) and visually interpreted by an experienced nuclear medicine physician (DO), who identified suspicious PCa metastases in bone and/or lymph nodes. Lesions were semi-automatically delineated using in-house developed software (*ACCURATE-tool*, previously benchmarked against commercially available image-analysis tools (*17*)) using a 50% isocontour of SUV_{peak} (sphere of 1.2 cm diameter, positioned to maximize its mean value) with correction for local background uptake to obtain volumes-of-interest (VOIs) (*18*). Blood activity concentrations (for TBR calculation) were measured in the ascending aorta using: i) a single image slice 3x3 voxel (12x12 mm) VOI, and ii) a 3x3 voxel VOI in 5 consecutive slices (*10*). VOIs were created on both original (EARL1) and PSF-reconstructions (EARL2).

From each VOI, the following metrics were recorded on a lesion-level: tumour volume (mL), TBR (tumour VOI activity concentration/blood activity concentration), SUV, and total lesion uptake (TLU). TBR was calculated using both the mean, peak, and max activity within the VOI, yielding TBR_{mean}, TBR_{peak}, and TBR_{max}, respectively. SUV variants included SUV_{max} (maximum SUV within the VOI), SUV_{peak} (mean SUV within a 12mm diameter sphere positioned within the VOI to yield the highest value), and SUV_{mean} (mean SUV within the VOI). SUV was normalized to body weight. TLU was defined as lesion SUV_{mean} or TBR_{mean} multiplied by lesion volume, yielding TLU_{SUV} and TLU_{TBR}, respectively. Additionally, two patient-level metrics were derived: Total PSMA-positive Tumour Volume (TTV) and PSMA Total Tumour Burden (TTB). TTV was defined as the sum of the delineated tumour volumes within a patient. TTB was defined as the sum of the TLU_{SUV} and TLU_{TBR} within a patient, yielding TTB_{SUV} and TTB_{TBR}, respectively. As recommended by PERCIST guidelines for PET response assessment(*19*), and to balance the number of analysed lesions between patients, for lesion-based analyses we selected the 5 hottest lesions in case of >5 PSMA-avid lesions. For patient-level analysis, all suspicious PSMA-avid lesions were included.

Statistical Analysis

To assess difference in uptake intervals and injected dosages between test and retest scans we used the Wilcoxon singed-rank test for paired data. Repeatability of quantitative PET metrics was quantified using repeatability coefficients (RC; in percentages). The RCs were calculated as 1.96 times the standard deviation of relative test-retest differences *d,* that were calculated as follows:

$$d = \frac{X2\text{-}X1}{\overline{X}} * 100 \qquad \text{Eq. 1}$$

Where *X1* and *X2* are the lesion- or patient-level metrics on day 1 (test) and day 2 (retest), respectively. $\overline{X}$ is the average between *X1* and *X2*. Bland-Altman plots were used for visual inspection of test-retest differences. Also, intra-class correlations (ICC; two-way mixed model with an absolute agreement definition) were calculated between test and retest data. The Pitman-Morgan test was used to test for differences in repeatability between paired data (correlated variances) (*20*), with α set at 0.05. P-values were corrected for multiple comparisons using Holms-Bonferroni method (*21*). The Levene's test was used to compare variances of independent groups in subgroup analysis (bone vs. lymph node metastases, >4.2mL lesions vs <4.2mL lesions). SPSS version 22.0 (IBM) and Excel (Microsoft) worksheets were used for statistical analyses.

Results

Patients

Twelve patients were enrolled, of which two patients could not be analysed. Patient characteristics and disease stage of the ten finally evaluated patients are presented in Table 5.1. Seven (70%) patients were using androgen-deprivation therapy (luteinizing hormone–releasing hormone agonist), all of which had been treated for at least 3 months at time of PET. In one excluded patient reliable comparison of the [^{18}F]DCFPyL scans was impeded due to significant radiolysis of the tracer (evident from a visually altered biodistribution as well as highly abnormal bone uptake (*16*)). The radiolysis was likely caused by a relatively high radioactivity concentration in the production batch (268 MBq/mL), combined by a long interval between delivery of the tracer and injection (>3 hours). Tracer logistics and storage were improved after this incidental finding and no other radiolysis problems occurred during this study. Another patient was excluded because of unconfirmed malignancy upon performing post-hoc CT-guided histological biopsy during clinical follow-up for two highly suspicious bone lesions on [^{18}F]DCFPyL PET. There were no significant differences between uptake intervals, injected dosages, and injected masses between test and retest scans ($P = 0.799$, $P = 0.499$, and $P = 0.878$ respectively).

Table 5.1: Patient and scan characteristics of the patients included in the repeatability analysis (*n*=10).

Patient Characteristics	**Median**	**Range**
Age (years)	74	(61-79)
Initial Gleason score	8	(6-9)
PSA at PET/CT (ng/mL)	9	(1-2796)
Length (cm)	178	(168-192)
Weight (kg)	88	(68-94)
PCa stage:	***n***	%
Primary metastastic	2	20.0
Biochemically recurrent	3	30.0
Castration-resistant	5	50.0
Analyzed lesion type:	***n***	%
Bone	21	58.3
Lymph node	12	33.3
Intraprostatic	3[a]	8.3
	n	%
Androgen Deprivation at PET/CT	7	70.0
prior docetaxel	3	30.0
Injected activity: Test (MBq)	317	(280-331)
Injected activity: Retest (MBq)	313	(254-341)
Uptake time: Test (min)	120	(118-153)
Uptake time: Retest (min)	122	(111-149)
Test-Retest diff. injected activity (MBq)[b]	28	(8-63)
Test-Retest diff. uptake time (min)[b]	3	(0-22)

[a]*two intraprostatic foci in one patient;* [b] *differences were not significant (p>0.05).*

Repeatability of Lesion-Level Metrics

In total, 36 [18F]DCFPyL PET-avid lesions were analysed, including 21 bone lesions (58.3%), 12 lymph node metastases (33.3%), and 3 intraprostatic foci (8.3%). Descriptive values of the analysed PET parameters are shown in Table 5.2 (PSF-reconstruction data in Supplemental Table 1). The best repeatability was observed for SUV_{mean} (RC 24.4%) and SUV_{peak} (RC 25.3%). SUV_{max} had poorer repeatability (RC 31.0%; Table 5.3), but the differences between repeatability of SUVs were not significant (p=0.06-0.60). Blood activity derived from a 1-slice and 5-slice VOI had a repeatability of 23.1% and 17.3%, respectively. Consequently, calculating TBR using 5-slice blood measurements had better repeatability compared to single-slice measurements (RC 31.7-37.3% versus 34.1-40.1%) and

was used henceforth. Overall, TBRs had worse repeatability than SUVs, but only repeatability of TBR_{mean} was significantly lower than that of SUV_{mean} (RC 31.8% versus 24.4%, p=0.03; Figure 5.1).

Table 5.2: Descriptive data of lesion and patient-based uptake metrics on test and retest scans.

	Test		Retest	
	Median	***IQR***	***Median***	***IQR***
Lesion-level				
Volume	4.6	2.8 – 8.7	4.6	2.5 – 8.6
SUV_{mean}	16.6	9.5 – 24.4	17.1	9.7 – 28.0
SUV_{peak}	21.7	10.5 - 28.3	21.6	11.4 – 32.2
SUV_{max}	28.1	16.0 – 41.0	29.8	17.2 – 51.2
TBR_{mean}	13.4	7.1 – 24.1	14.6	8.6 – 22.7
TBR_{peak}	17.7	7.7 – 28.4	18.8	9.8 – 26.7
TBR_{max}	25.0	11.7 – 40.9	23.6	14.1 – 38.8
TLU_{SUV}	85.6	32.3 – 192.7	80.1	30.1 – 194.0
TLU_{TBR}	67.6	24.6 – 189.4	66.7	23.5 – 152.6
Patient-level				
TTV	21.4	10.6 – 63.2	21.8	10.3 – 69.7
TTB_{SUV}	317.8	70.4 – 1920.7	285.5	70.6 – 1846.4
TTB_{TBR}	236.6	63.2 – 1920.1	224.9	68.2 – 1720.0

IQR = interquartile range; SUV = Standardized Uptake Value; TBR = Tumor-to-Blood Ratio; TLU = Total Lesion Uptake; TTV = Total Tumor Volume; TTB = Total Tumor Burden

Table 5.3: Repeatability of lesion- and- patient-based ^{18}F-DCFPyL uptake metrics.

Parameter	Mean test-retest difference %	RC%	ICC (95% CI)
Lesion-level			
Volume	-1.1	28.1	1.00 (0.99-1.00)
SUV_{mean}	1.0	24.4	0.99 (0.98-0.99)
SUV_{peak}	1.8	25.3	0.99 (0.97-0.99)
SUV_{max}	1.9	31.0	0.97 (0.94-0.99)
TBR_{mean}	1.9	31.8	0.98 (0.96-0.99)
TBR_{peak}	2.6	31.7	0.98 (0.96-0.99)
TBR_{max}	2.7	37.3	0.97 (0.94-0.98)
TLU_{SUV}	-0.1	32.1	0.99 (0.98-1.00)
TLU_{TBR}	-3.5	39.3	0.98 (0.96-0.99)
Patient-level			
TTV	-2.2	17.0	1.00 (0.99-1.00)
TTB_{SUV}	-0.2	23.2	0.99 (0.97-1.00)
TTB_{TBR}	-2.1	33.4	0.98 (0.91-0.99)

IQR = interquartile range; SUV = Standardized Uptake Value; TBR = Tumor-to-Blood Ratio; TLU = Total Lesion Uptake; TTV = Total Tumor Volume; TTB = Total Tumor Burden

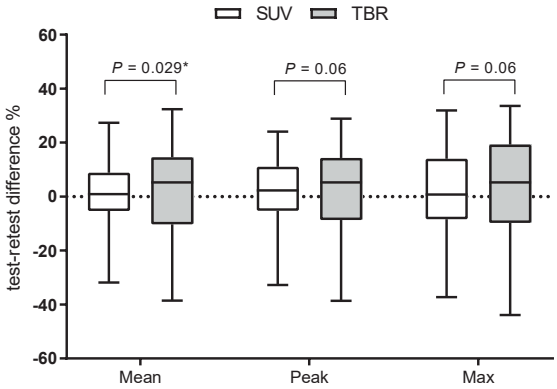


Figure 5.1: Test-retest variability of SUV and TBR variants. Significant differences have been indicated with an asterix (Holms-Bonferroni corrected p-values). Differences in repeatability between SUVs and between TBRs were not significant.

Repeatability of semi-automatic tumour volume measurement was 28.1%. Repeatability of TLU_{TBR} (RC 39.3%) was non-significantly lower than that of TLU_{SUV} (RC 32.1%, p=0.08). Bland-Altman plots did not demonstrate a skewed variability, but variability of SUV and TBR tended to be less for higher values (Figure 5.2). In subgroup analysis, no significant differences between repeatability of metrics derived from bone versus lymph node metastases were observed (p=0.06-0.98). Only volume measurements had a significantly different repeatability for lesions >4.2mL versus <4.2mL (RC 17.6% and 36.8%, respectively; *p*=0.015).

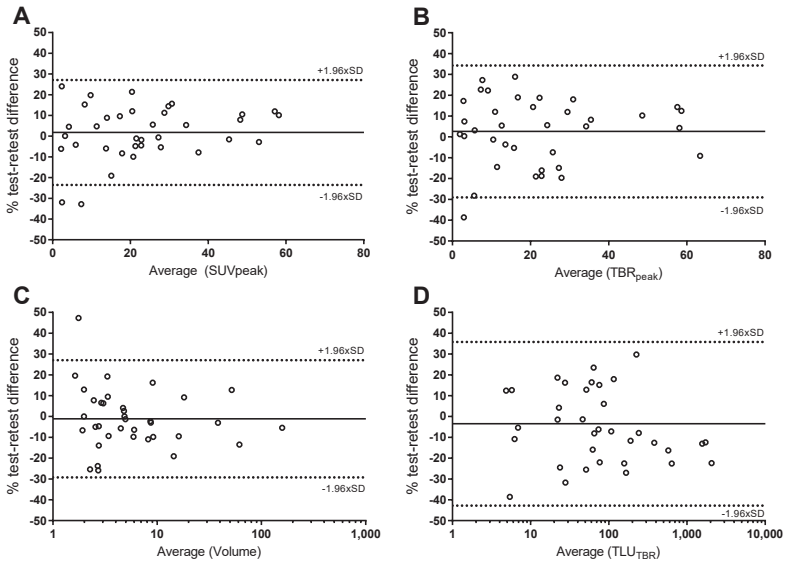


Figure 5.2: Bland-Altman plots of lesion-level metrics (A) SUV_{peak}, (B) TBR_{peak}, (C) volume, and (D) TLU_{TBR}. Y-axis in (C) and (D) were log-scaled for visual interpretation.

Repeatability of Patient-Level Metrics

The highest repeatability was observed for TTV (RC 17%). TTB_{SUV} had better repeatability than TTB_{TBR}, albeit non-significantly (RC 23.2% versus 33.4%, p=0.19). Bland-Altman plots demonstrated no skewed variability (Figure 5.3).

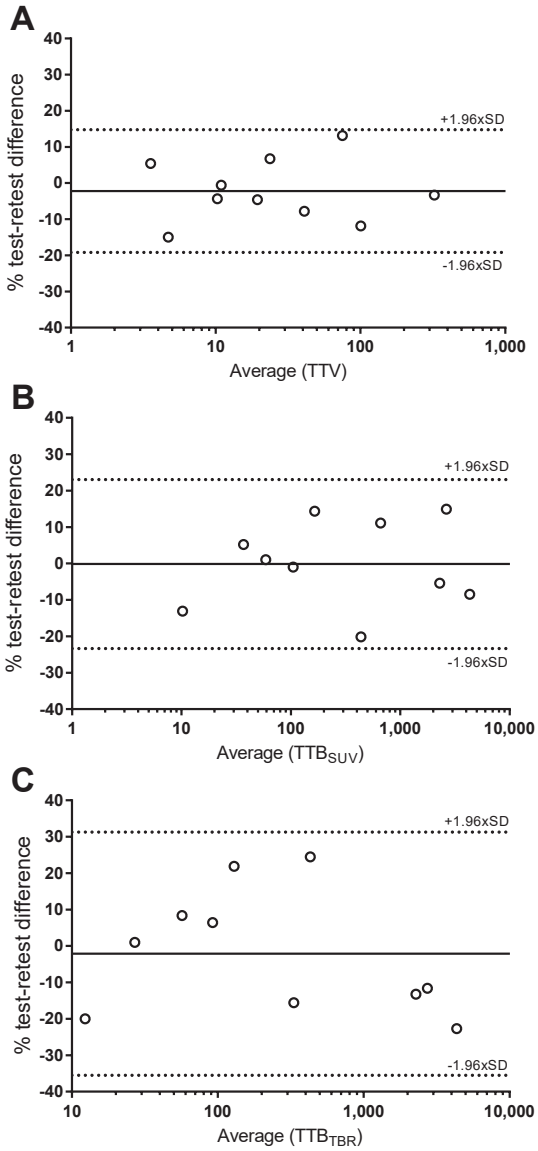


Figure 5.3: Bland-Altman plots of patient-level metrics TTV, TTB_{SUV}, and TTB_{TBR}. Y-axes were log-scaled for visual interpretation.

Effect of PSF-Reconstruction on Repeatability

PSF-reconstruction worsened repeatability significantly for the TBRs, SUV_{mean}, and SUV_{max} ($p \leq 0.005$; Supplemental Table 2). However, the repeatability of tumour volume (RC 32.0%, p=0.43), SUV_{peak} (RC% 27.8%, p=0.15), TLU_{SUV} (RC 30.3%, p=0.62), and TLU_{TBR} (RC 41.3%, p=0.70) was not affected. Notably, repeatability of all patient-level metrics was not significantly affected by the PFS-reconstruction (p=0.15-0.59; Supplemental Table 2).

Discussion

In this study we investigated the repeatability of [18F]DCFPyL uptake and volume measurements in metastatic PCa patients. Knowledge of the day-to-day variation in these metrics is indispensable for use of [18F]DCFPyL metrics as novel biomarkers for assessment of response to systemic treatments. We conclude that [18F]DCFPyL uptake metrics are highly repeatable (ICC ≥0.97) and are thus suited for response monitoring purposes. SUV metrics tend to have higher repeatability than TBRs. The best repeatability was observed for patient-based TTV measurements.

In routine static PET acquisitions, [18F]DCFPyL pharmacokinetics are most accurately quantified using the TBR (*10*), which demonstrated a repeatability of 31.8% in this study. Hence, a change in TBR exceeding 32% may indicate a change in tumoural [18F]DCFPyL -uptake that exceeds the physiological variability, due to (e.g.) disease progression, treatment response, a true flare phenomenon, or an imaging protocol deviation. Repeatability of tumour SUV_{mean} was superior to TBR (Figure 5.1), which can be explained by added variability of blood pool activity measurements used in TBR calculation (blood pool RC 17.3%). Still, in our pharmacokinetic analysis we concluded that SUV measurements do not universally correlate with the underlying [18F]DCFPyL pharmacokinetics (K_i), as intrapatient tumour volumes appear to affect the bioavailability of the tracer (a so-called sink-effect) (*10,22*). At higher tumour loads, SUV tends to underestimate [18F]DCFPyL uptake in lesions, while TBR (partly) corrects for this and thus better reflects changes in [18F]DCFPyL during response monitoring. These findings are in line with other prostate cancer radiotracers (18F-fluoromethylcholine, 18F-fluordihydrotestosterone) where tumour uptake measurements normalized to blood pool activity are more accurate metrics for tracer quantification than SUV, but have worse repeatability (*23,24*). All taken together, TBR may be preferred

over SUV metrics despite its lower repeatability, which we recently illustrated in a clinical case (*22*). The higher variability of TBR compared to SUV will only have a negative impact on response assessment in patients with small tumour volumes with small treatment effect sizes.

Interestingly, the semi-automatically delineated total intrapatient tumour volume (TTV) demonstrated the highest overall repeatability (RC 17%). These favourable outcomes are likely explained by the high tumour to background ratio that [^{18}F]DCFPyL provides (and PSMA tracers in general), permitting reliable (semi)automatic identification of tumour extent. On a lesion-basis, however, variability of volume measurements was larger (RC 28.1%), which can at least partly be explained by the volume-dependency of its variability. The high repeatability of TTV may be of benefit for longitudinal assessments of total PSMA burden in patients receiving systemic treatments. Especially for PSMA-targeted radioligand therapies (e.g. ^{177}Lutetium-PSMA), assessment of changes in the total tumour volume as a whole, instead of individual lesion responses, may be clinically useful.

In multicenter studies, use of different PET/CT systems with varying image-reconstruction protocols require quantitative metrics that are robust to such factors. Advanced reconstruction methods may improve lesion detection (*25*), but repeatability may be hampered by the inherent image noise propagation. In line with previous observations for ^{18}F-fluorodihydrotestosterone (*26*), we observed lower repeatability of several metrics when using an image reconstruction with improved signal recovery, adhering to novel imaging guidelines (EARL2). However, repeatability of SUV_{peak}, TLU and patient-level measurements were not affected by the PSF-reconstruction, rendering them fit for use in multicenter studies were PET imaging protocols differ between centers. As blood activity measurements are susceptible to noise, repeatability of TBR was negatively affected by PSF-reconstruction. Overall, non-PSF reconstruction images (EARL1-compliant) may therefore be preferred for quantitative assessment.

Our study has limitations, most notably the small patient sample. Still, results were in line with findings on other ^{18}F-labeled PCa radiotracers, as well as ^{18}F-FDG (*19,23,24*). Factors contributing to the total variability in quantitative PET-metrics include biological variation in tracer uptake, image noise, scan protocol deviations between scans, and the analysis software used. We acknowledge the patients' heterogeneity in terms of disease stages (primary metastatic disease, biochemical recurrence, castration-resistance), but subgroup

analysis per disease stage was not feasible at the current sample size. We have no reason to assume that tracer uptake variability attributable to tumour biology will differ between disease stages, however. In our single-center evaluation, only a single type of PET-scanner and analysis software package was used – multicenter variability may be higher. We welcome other investigators using [18F]DCFPyL to repeat our study in their own center, or even in a multicenter setting, to validate our current findings in a larger cohort. In the present study, the tracer uptake time and injected dosages of both test and retest scan were similar (Table 5.1). As our pharmacokinetic data indicated that tumour [18F]DCFPyL uptake continues to rise at 120 min after injection (*10*), test-retest variability might be higher in clinical practice, where uptake times between scans may vary more. Clinical imaging protocols for [18F]DCFPyL regarding uptake time intervals, total scan duration, and patient positioning (*i.e.* feet first or head first) should be stringently adhered to, especially in response assessment studies.

Conclusions

In this study we assessed the repeatability of quantitative [18F]DCFPyL PET/CT measurements in patients with metastatic PCa, concluding that [18F]DCFPyL uptake metrics are well repeatable. The variability limits proposed in this study should be validated in future clinical studies. To this end, any change in TBR exceeding 32% can be considered a change in tracer uptake beyond physiologic day-to-day variability (in case of comparable image-acquisition parameters). Additionally, as TTV measurements are highly repeatable (RC 17%) they may be specifically suitable for longitudinal assessment of PSMA-targeted radioligand therapy effects. The repeatability of SUV_{peak}, Total Lesion Uptake, and patient-level metrics (TTV and TTB) of [18F]DCFPyL uptake is robust to differences in image reconstructions.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. ***CA Cancer J Clin.*** 2018;68:394-424.
2. Rowe SP, Gorin MA, Allaf ME, et al. PET imaging of prostate-specific membrane antigen in prostate cancer: current state of the art and future challenges. ***Prostate Cancer Prostatic Dis.*** 2016;19:223-230.
3. Perner S, Hofer MD, Kim R, et al. Prostate-specific membrane antigen expression as a predictor of prostate cancer progression. ***Hum Pathol.*** 2007;38:696-701.
4. Perera M, Papa N, Christidis D, et al. Sensitivity, specificity, and predictors of positive 68Ga-prostate-specific membrane antigen positron emission tomography in advanced prostate cancer: a systematic review and meta-analysis. ***Eur Urol.*** 2016;70:926-937.
5. Robu S, Schmidt A, Eiber M, et al. Synthesis and preclinical evaluation of novel (18)F-labeled Glu-urea-Glu-based PSMA inhibitors for prostate cancer imaging: a comparison with (18)F-DCFPyl and (18)F-PSMA-1007. ***EJNMMI Res.*** 2018;8:30.
6. Chen Y, Pullambhatla M, Foss CA, et al. 2-(3-{1-Carboxy-5-[(6-[18F]fluoro-pyridine-3-carbonyl)-amino]-pentyl}-ureido)-pen tanedioic acid, [18F]DCFPyL, a PSMA-based PET imaging agent for prostate cancer. ***Clin Cancer Res.*** 2011;17:7645-7653.
7. Fendler WP, Eiber M, Beheshti M, et al. (68)Ga-PSMA PET/CT: Joint EANM and SNMMI procedure guideline for prostate cancer imaging: version 1.0. ***Eur J Nucl Med Mol Imaging.*** 2017;44:1014-1024.
8. Gupta M, Choudhury PS, Rawal S, Goel HC, Rao SA. Evaluation of RECIST, PERCIST, EORTC, and MDA criteria for assessing treatment response with Ga68-PSMA PET-CT in metastatic prostate cancer patient with biochemical progression: a comparative study. ***Nucl Med Mol Imaging.*** 2018;52:420-429.
9. Seitz AK, Rauscher I, Haller B, et al. Preliminary results on response assessment using (68)Ga-HBED-CC-PSMA PET/CT in patients with metastatic prostate cancer undergoing docetaxel chemotherapy. ***Eur J Nucl Med Mol Imaging.*** 2018;45:602-612.
10. Jansen BHE, Yaqub M, Voortman J, et al. Simplified methods for quantification of (18)F-DCFPyL uptake in patients with prostate cancer. ***J Nucl Med.*** 2019; doi: 10.2967/jnumed.119.227520.
11. Ravert HT, Holt DP, Chen Y, et al. An improved synthesis of the radiolabeled prostate-specific membrane antigen inhibitor, [(18) F]DCFPyL. ***J Labelled Comp Radiopharm.*** 2016;59:439-450.
12. Kolthammer JA, Su KH, Grover A, Narayanan M, Jordan DW, Muzic RF. Performance evaluation of the Ingenuity TF PET/CT scanner with a focus on high count-rate conditions. ***Phys Med Biol.*** 2014;59:3843-3859.
13. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
14. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. ***Neuroimage.*** 2008;39:1570-1584.
15. Kaalep A, Sera T, Rijnsdorp S, et al. Feasibility of state of the art PET/CT systems performance harmonisation. ***Eur J Nucl Med Mol Imaging.*** 2018;45:1344-1361.
16. Jansen BHE, Kramer GM, Cysouw MCF, et al. Healthy Tissue Uptake of (68)Ga-Prostate Specific Membrane Antigen (PSMA), (18)F-DCFPyL, (18)F-Fluoromethylcholine (FCH) and (18) F-Dihydrotestosterone (FDHT). ***J Nucl Med.*** 2019;60(8):1111-1117.
17. Boellaard R. Quantitative oncology molecularanalysis suite: ACCURATE May 1st, 2018 ed: J Nucl Med; 2018;59 Suppl 1:1753.
18. Frings V, Yaqub M, Hoyng LL, et al. Assessment of simplified methods to measure 18F-FLT uptake changes in EGFR-mutated non-small cell lung cancer patients undergoing EGFR tyrosine kinase inhibitor treatment. ***J Nucl Med.*** 2014;55:1417-1423.
19. 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.
20. Pitman E. A note on normal correlation. ***Biometrika.*** 1939;31:9-12.
21. Holm S. A simple sequentially rejective multiple test procedure. ***Scandinavian journal of statistics.*** 1979:65-70.

22. Cysouw MCF, Jansen BHE, Yaqub M, et al. Letter to the Editor re: Semiquantitative parameters in PSMA-targeted PET imaging with [18F]DCFPyL: impact of tumor burden on normal organ uptake. ***Mol Imaging Biol.*** 2019. doi: 10.1007/s11307-019-01438-y.

23. Oprea-Lager DE, Kramer G, van de Ven PM, et al. Repeatability of Quantitative 18F-Fluoromethylcholine PET/CT Studies in Prostate Cancer. ***J Nucl Med.*** 2016;57:721-727.

24. Vargas HA, Kramer GM, Scott AM, et al. Reproducibility and repeatability of semi-quantitative (18) F-fluorodihydrotestosterone (FDHT) uptake metrics in castration-resistant prostate cancer metastases: a prospective multi-center study. ***J Nucl Med.*** 2018; 59(10):1516-1523.

25. Jansen BHE, Jansen RW, Wondergem M, et al. Lesion detection and interobserver agreement with advanced image-reconstructions for (18)F-DCFPyL PET/CT in patients with biochemically recurrent prostate cancer. ***J Nucl Med.*** 2019. doi: 10.2967/jnumed.118.222513.

26. Cysouw MCF, Kramer GM, Heijtel D, et al. Sensitivity of (18)F-fluorodihydrotestosterone PET-CT to count statistics and reconstruction protocol in metastatic castration-resistant prostate cancer. ***EJNMMI Res.*** 2019;9:70.

Supplemental files

Supplemental Table 5.1: Descriptive data of lesion- and- patient-based uptake metrics on test and retest scans for PSF-reconstruction images.

	Test		Retest	
	Median	*Range*	*Median*	*Range*
Lesion-level				
Volume	3.4	2.0 – 7.8	3.6	1.9 – 7.5
SUV_{mean}	25.0	15.0 – 40.3	24.2	14.9 – 47.2
SUV_{peak}	27.2	15.3 – 38.8	29.5	16.3 – 48.8
SUV_{max}	52.9	29.2 – 93.9	47.1	29.5 – 98.6
TBR_{mean}	19.7	11.6 – 33.7	19.4	13.1 – 34.1
TBR_{peak}	23.8	10.4 – 37.7	23.0	13.2 – 36.0
TBR_{max}	49.3	22.0 – 83.3	36.8	26.9 – 76.1
TLU_{SUV}	100.3	38.5 – 223.0	99.9	35.6 – 240.9
TLU_{TBR}	84.0	29.2 – 213.1	82.2	27.1 – 255.7
Patient-level				
TTV	17.2	7.5 – 46.0	17.9	7.8 – 47.0
TTB_{SUV}	375.3	74.6 – 1909.1	340.9	77.4 – 2094.6
TTB_{TBR}	276.9	66.3 – 1933.5	263.5	74.8 – 2215.6

IQR = interquartile range; SUV = Standardized Uptake Value; TBR = Tumor-to-Blood Ratio; TLU = Total Lesion Uptake; TTV = Total Tumor Volume; TTB = Total Tumor Burden

Supplemental Table 5.2: Repeatability of lesion and patient-based uptake metrics for PSF-reconstruction images. P-value indicates a significant differences in repeatability between original (non-PSF) and PSF-reconstructions.

Parameter	Mean test-retest difference %	RC%	ICC (95% CI)	*p*-value *non-PSF vs. PSF*
Lesion-level				
Volume	0.1	32.0	1.00 (1.00-1.00)	0.427
SUV_{mean}	0.6	32.3	0.97 (0.95-0.99)	**0.005**
SUV_{peak}	2.2	27.8	0.98 (0.97-0.99)	0.147
SUV_{max}	-0.9	52.1	0.87 (0.77-0.93)	**<0.001**
TBR_{mean}	2.2	41.7	0.96 (0.92-0.98)	**<0.001**
TBR_{peak}	3.7	38.1	0.97 (0.94-0.98)	**<0.001**
TBR_{max}	0.7	58.9	0.88 (0.78-0.94)	**<0.001**
TLU_{SUV}	0.7	30.3	0.99 (0.98-0.99)	0.616
TLU_{TBR}	2.2	41.3	0.97 (0.95-0.99)	0.699
Patient-level				
TTV	1.7	13.9	1.00 (1.00-1.00)	0.152
TTB_{SUV}	1.9	23.4	0.99 (0.98-1.00)	0.952
TTB_{TBR}	3.9	37.6	0.98 (0.92-1.00)	0.588


QI-1-PET

Chapter 6

Machine Learning-based Analysis of [^{18}F]PSMA PET Radiomics for Risk Stratification in Primary Prostate Cancer.

M.C.F. Cysouw, B.H.E. Jansen, T. van de Brug, D.E. Oprea-Lager, E. Pfaehler, B.M. de Vries, R.J.A. van Moorselaar, O.S. Hoekstra, A.N. Vis, R. Boellaard

Published in revised form in
Eur J Nucl Med Mol Imaging. 2020 Jul 31.
https://doi.org/10.1007/s00259-020-04971-z

Abstract

Quantitative [^{18}F]-Prostate Specific Membrane Antigen (PSMA) analysis may provide for non-invasive and objective risk stratification of primary prostate cancer (PCa) patients. We determined the ability of machine learning-based analysis of [^{18}F] PSMA PET radiomic features to predict metastatic disease or high-risk pathological tumor features. Also, we evaluated the performance of these models with respect to tumor delineation method and use of partial-volume correction (PVC).

Methods: 76 consecutive patients with intermediate- to high-risk PCa scheduled for robot-assisted radical prostatectomy with extended pelvic lymph node dissection prospectively underwent pre-operative [^{18}F]-PSMA PET-CT using [^{18}F]-DCFPyL. Primary tumors were delineated using 50 to 70% peak isocontour thresholds on PET images with and without PVC. 480 standardized radiomic features were extracted per tumor. Random Forest models were trained to predict lymph node involvement (LNI), presence of any metastasis, Gleason score ≥8, and presence of extracapsular extension (ECE). The impact of several dimension reduction methods and use of oversampling on model predictions was evaluated. Model performance was validated using 50-times repeated 5-fold cross-validation yielding the mean receiver-operator characteristic curve AUC. Random permutations were performed to assess cross-validated AUC significance. Differences between the predictive value of radiomics and standard PET metrics were compared using the median p-value from deLong testing within cross-validation.

Results: The radiomics-based models were significantly able to predict LNI (AUC 0.86±0.15, $p<0.01$), any metastasis (AUC 0.86±0.14, $p<0.01$), Gleason score (0.81±0.16, $p<0.01$), and ECE (0.76±0.12, $p<0.01$). Model performance and stability for LNI and any metastasis prediction seemed to improve using PVC and a higher (70%) delineation threshold. PVC affected all feature types, while delineation thresholds mainly affected morphological features. Machine learning pre-processing methods had a minor impact on model performance. AUCs using standard PET metrics were non-significantly lower than those of radiomics-based models.

Conclusion: Machine learning-based analysis of [^{18}F]PSMA PET-CT radiomics can significantly predict metastatic disease and high-risk pathological tumor features in primary PCa patients. Future multicenter external validation is needed to determine the benefits of using radiomics versus standard PET metrics in clinical practice.

Introduction

In primary prostate cancer (PCa), risk stratification is crucial to determine prognosis and treatment strategies. Extended pelvic lymph node dissection (ePLND) is the current standard for identification of lymph node metastases (*1-3*). This procedure, however, is invasive and associated with complications such as lymphocele, venous thrombosis, and extended hospital stays (*4,5*). Hence, patients at risk of lymph node involvement (LNI) are selected using clinical nomograms, but these lack adequate performance(*3*). Also, histopathology data (e.g. Gleason score: GS) used as input for these nomograms are based on error-prone prostate biopsies (*6*). Taken together, a novel biomarker able to pre-operatively stratify high and low-risk patients is highly needed.

Prostate-Specific Membrane Antigen (PSMA) is a type-II transmembrane protein known to be highly overexpressed on PCa cells (*7*). Kaittanis et al. recently demonstrated that PSMA is a stimulator of oncogenic signaling, clarifying the role of PSMA in PCa progression(*8*). Moreover, primary tumor PSMA expression on immunohistochemistry was a prognostic marker for metastasis-free, recurrence-free, disease-free, and overall survival (*8-11*). Therefore, quantitative measures of PSMA-expression are promising biomarkers for risk stratification of primary PCa patients.

PSMA-expression may be quantified and characterized on PSMA-ligand Positron Emission Tomography Computed-Tomography (PET-CT) through radiomics analysis (*12*). Radiomics entails high-throughput image data mining aiming to capture a tumor's phenotype - potentially reflecting its metastatic tendency (*13,14*). In contrast with biopsies, radiomics may characterize local tumor phenotype based on the entire lesion instead of through tumor subsamples. Due to the high dimensionality of radiomics data, machine learning (ML) is necessary to translate radiomics data into clinically actionable predictions (*15*). In radiomics validation, tumor segmentation and image processing methods must be considered as these may affect radiomics data and their predictive value (*16,17*).

In this study we investigated whether ML-based analysis of pre-operative [^{18}F]PSMA PET-CT radiomics could predict metastatic disease and high-risk local tumor features in patients with intermediate- and high-risk primary PCa scheduled to undergo robot-assisted radical prostatectomy and ePLND. Secondly, we assessed the influence of tumor delineation method and image processing on model predictions.

Materials and methods

Patients

Between November 2017 and August 2019, 76 consecutive patients prospectively underwent pre-operative ^{18}F-DCFPyL PET-CT for lymph node and distant metastasis detection (NTR7623). Inclusion criteria were: 1) biopsy-proven prostate adenocarcinoma and 2) clinical indication for robot-assisted radical prostatectomy with ePLND based on either an ≥8% risk score of LNI based on the Memorial Sloan Kettering Cancer (MSKCC) nomogram or any high-risk feature (≥T3, Gleason >7, PSA>20 ng/mL). Patients with distant metastases on PET for whom surgery was omitted were only included in case of histopathological confirmation. Only patients who underwent [^{18}F]DCFPyL PET-CT at the Amsterdam UMC were included. Surgical tissue specimens (prostate and lymph nodes) were reviewed according to international guidelines by uropathologists (*3*). Patients provided written informed consent for collection and analysis of imaging, clinical, and pathology data. The Amsterdam UMC medical ethical committee provided formal approval (2017.543).

Outcomes

All references outcomes were pathology-proven, and dichotomized for ML-based classification: post-operative GS (<8 versus ≥8), presence of extracapsular tumor extension (ECE; ≤pT2b versus ≥pT3a), pathology-proven LNI (N0 versus N1), and presence of any metastasis (pN0 and cM0 versus pN1 and/or pM1).

PET-CT Imaging

Patients were scanned on a time-of-flight PET-CT system (Ingenuity, Philips Healthcare) with European Association of Nuclear Medicine Research Ltd. (EARL) accreditation (*18*). A CT scan was acquired at 120kV and 30-110 mAs. Next, whole-body PET was performed at 122.5±11.1 min post-injection of 310.1±16.2 MBq [^{18}F]DCFPyL, from mid-thighs to skull base, at 4min per bed position. Images were reconstructed using iterative ordered subset expectation maximization reconstruction (3 iterations, 33 subsets) with 4mm voxel dimensions, with corrections for decay, scatter, random coincidences, and attenuation correction. Lucy-Richardson iterative deconvolution (10 iterations) was applied for partial-volume correction (PVC) (*19*). The FWHM for PVC was calibrated at 7.0mm using a NEMA NU2 Quality Phantom, such that signal recovery was in line with EARL2 guidelines (*20*). Original and PVC images were analyzed separately.

Tumor Delineation

An experienced nuclear medicine physician (DO) reviewed all [^{18}F]PSMA PET-CT scans for intra-prostatic tumor localization. A mask was manually drawn around PET-avid intraprostatic tumor volumes to constrain region-growing and prevent inclusion of bladder activity. All masks were reviewed by a second observer. If needed, consensus was reached through joint revision. Next, tumors were delineated using a region-growing algorithm with a background-adapted peak threshold (*19*). The thresholds were varied incrementally from 50% to 70% (5% intervals). Delineation was performed on original and PVC scans separately to mimic clinical reality.

Radiomics Extraction

Radiomic features were extracted from the delineated tumors following descriptions of the Image Biomarker Standardization Initiative, as presented by Zwanenburg et al., using the RaCaT software (*21,22*). Voxel values were scaled to the net injected tracer dosage per kilogram bodyweight (Standardized Uptake Value, SUV). Image voxels and volumes of interest were resampled to 2x2x2 mm isotropic voxels using tri-linear interpolation as recommended (*23,24*). Per tumor we extracted 480 radiomic features on intensity (n=50), morphology (n=22), and texture (n=408). Intensity features encompassed peak intensity, intensity-based statistics, intensity-volume histograms, and intensity histograms. 2D and 3D textural features based on grey-level co-occurrence matrices (GLCM), grey-level run length matrices (GLRLM), grey-level size zone matrices (GLSZM), grey-level distance zone matrices (GLDZM), neighborhood grey-tone difference matrices (NGTDM), and neighboring grey-level dependence matrices (NGLDM) were extracted. Before textural feature calculation, images were discretized using a fixed bin width of 0.25 SUV starting at SUVmin (*23*). To compare with radiomics, from the original and PVC PET images we also extracted standard PET features SUVmean, SUVpeak, SUVmax, PSMA-positive tumor volume, and PSMA-total lesion uptake (the product of SUVmean and volume) and used these data as input for the machine learning pipeline.

Machine Learning

ML algorithms may handle high-dimensional data and/or data with complex non-linear relations with clinical outcomes. We constructed a ML framework in Python 3.6 using *Scikit-learn* library 0.21 (pipeline in Fig. 6.1) (*15,25*). As ML model we used a *Random Forest* classifier (1000 decision trees), a commonly used non-parametric

ensemble algorithm (*26*). To assess model generalizability (i.e. its prediction performance on unseen data), we used a stratified 5-fold cross-validation approach. In each cross-validation fold the Random Forest was trained on 80% of samples and validated on an unseen subset of 20% of samples. This was repeated until each fold had served as the test set. Finally, this 5-fold cross-validation was repeated 50 times to further limit chance findings. Features were scaled using a z-score normalization. Model hyperparameters (tree depth, splitting criterion) were optimized within each training set in nested cross-validation using a randomized search algorithm. All pre-processing and optimization steps were performed within each training fold to prevent leakage of test data into the trained model (Fig. 6.1).

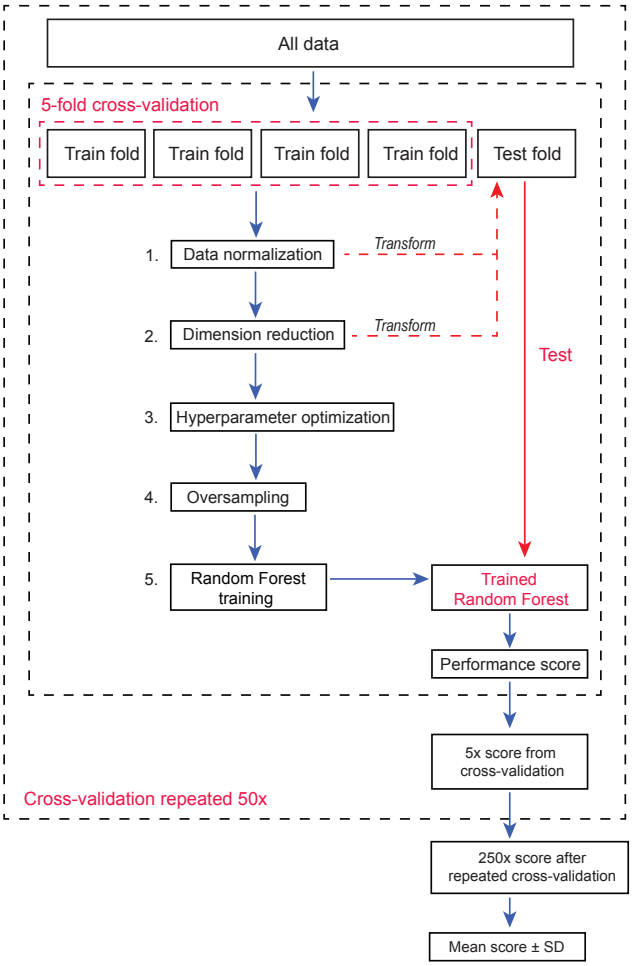


Figure 6.1: Schematic overview of the implemented machine learning pipeline. Data pre-processing and model tuning are performed on the training dataset in repeated cross-validation to prevent leakage of information between training and testing data.

Dimensionality Reduction. To mitigate model overfitting and potentially improve generalizability, we applied three different strategies for dimension reduction that reduced the number of features used as input for the Random Forests: i) a principal component analysis (PCA) retaining 95% of the observed variance, ii) a recursive feature elimination approach using a Random Forest in nested cross-validation, and iii) a univariate selection method based on ANOVA testing that retained the top 10 percentile features. Models were also trained without any dimensionality reduction. When using standard PET metrics as model input, no dimension reduction was applied because of the small number of metrics.

Oversampling. In case of strong class imbalance, a trained ML model may have high accuracy in classifying the majority class, but perform poorly for classifying the minority class. Therefore, oversampling was applied in each training set by generation of 'synthetic' samples with interpolated feature values (SMOTE) (*27*). Models were also trained without oversampling.

Feature Importance. To explore feature importance, coefficients representing the relative importance of each feature within a trained Random Forest model can be derived (the sum of coefficients being equal to 1.0). Per outcome we visualized the top 10% coefficients (n=48) from a Random Forest trained on the entire dataset using the feature selection method that yielded the highest predictions per outcome (excluding PCA as this does not yield interpretable features).

Statistical Analysis

To evaluate model performance, we generated the Receiver-Operator Characteristic curve and calculated the area-under-the-curve (AUC). The Brier score was used to assess model calibration and refinement (0.0 being optimal) (*28*). For each score, we calculated the mean with standard deviation over the repeated cross-validation folds. Random permutations were used to test whether the models performed significantly better than random guessing. To this end, labels were randomly shuffled before performing 10-times repeated 5-fold cross-validation, resulting in a 'random guessing' cross-validated AUC. This was repeated 100 times, yielding a p-value defined as the fraction of repeated cross-validation iterations in which the permutated mean AUC was equal or higher than the actual mean AUC (*29*).

Comparing the cross-validated AUCs of two machine learning models is a known difficulty due to the complex relations between the trained models and the

inherent dependency of train-test iterations. Still, to be able to compare the mean AUCs of radiomics versus standard PET metrics, we used a framework developed by Van De Wiel et al. (*30*): in each fold the AUCs of two models were compared statistically using DeLong test (*31*), and the median of the p-values over the different folds was reported as the final p-value. A disadvantage of this method is that each p-value is based on the test set of a single fold only, resulting in a rather low power to detect true differences.

Intraclass correlation coefficients (ICC: 2-way mixed model, absolute agreement) were calculated for each radiomic feature between original versus PVC images (per delineation threshold), and between delineation thresholds. ICCs were categorized as poor (ICC<0.5), moderate (0.5<ICC<0.75), good (0.75<ICC<0.9), or excellent (ICC>0.9) (*32*).

Results

Patients

We included 76 patients (Table 6.1), of which 71 ultimately underwent surgery. Six patients had uptake suspicious for distant metastases on PET (n=2 nodal, n=1 bone, n=3 both), all of which were biopsied. In 4 of these patients biopsies confirmed malignancy and surgery was omitted; in 2 patients (n=1 bone, n=1 nodal lesion) biopsy did not confirm malignancy and surgery was performed as planned. Additionally, 1 patient had biopsy-proven LNI within the ePLND template, but surgery was omitted due to additional PSMA-positive nodal metastases outside the ePLND template. The final pathology findings are listed in Table 6.2.

Impact Of PVC And Delineation Threshold

Delineated tumor volumes for each delineation threshold with and without PVC are shown in Supplemental Fig. Most radiomic features had a moderate agreement between original and PVC data (Fig. 6.2A). Delineation thresholds mainly affected morphological features, while intensity and textural features were less affected (Fig. 6.2B). For LNI and any metastasis prediction, PVC and a higher delineation threshold tended to improve model stability, reducing the width of the cross-validation AUC distributions (Figs. 6.3A-B). For GS and ECE predictions, there was no optimal delineation threshold and PVC had no apparent benefit (Figs. 6.3C-D).

Table 6.1: Patient characteristics.

Number of patients	n=76
Age (mean ± SD)	66 ± 6 years
PSA at PET (median, [range])	11 (4-70) ng/ml
ISUP Gleason grade (biopsy)	**n (%)**
Group 1	4 (5.3%)
Group 2	21 (27.6%)
Group 3	19 (25.0%)
Group 4	21 (27.6%)
Group 5	11 (14.5%)
Positive biopsies % (mean ± SD)	54.7% ± 27.3%
Clinical T-stage	**n (%)**
T1c	26 (34.2%)
T2a	24 (31.6%)
T2b	12 (15.8%)
T2c	11 (14.5%)
T3a	3 (3.9%)

Table 6.2: Final pathological findings. 71 patients underwent robot-assisted radical prostatectomy with ePLND; 1 patient had biopsy-proven LNI but did not undergo surgery; 4 patients did not undergo surgery due to proven distant metastases. Pathological N-stage refers to LNI, ISUP grade 4-5 represent Gleason ≥8, and ≥T3 to ECE.

	n (%)
ISUP Gleason grade	
Group 1	1 (1.4%)
Group 2	27 (38.0%)
Group 3	24 (33.8%)
Group 4	5 (7.0%)
Group 5	14 (19.7%)
Pathological T-stage	
T2a-c	35 (49.3%)
T3a-b	35 (49.3%)
T4	1 (1.4%)
Pathological N-stage	
N0	62 (86.1%)
N1	10 (13.9%)
Biopsy-proven M-stage	
M0	72 (94.7%)
M1	4 (5.3%)
Resection margin status	
R0	43 (60.6%)
R1	28 (39.4%)

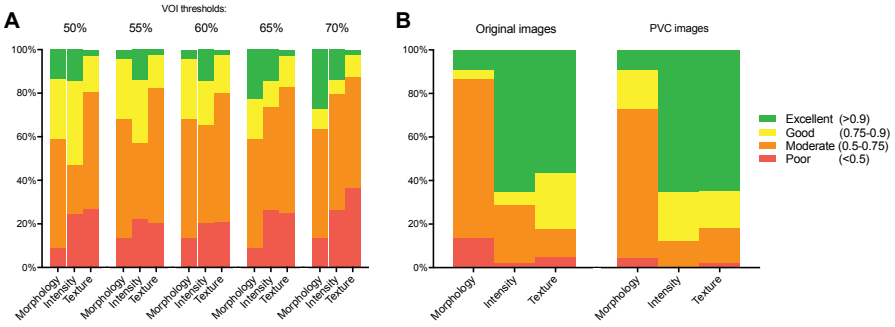


Figure 6.2: Agreement of radiomic features (A) between original versus PVC images at each delineation threshold, and (B) between the applied delineation thresholds for original and PVC images. Shown are the relative distributions of the radiomics ICC values per ICC category (poor, moderate, good or excellent).

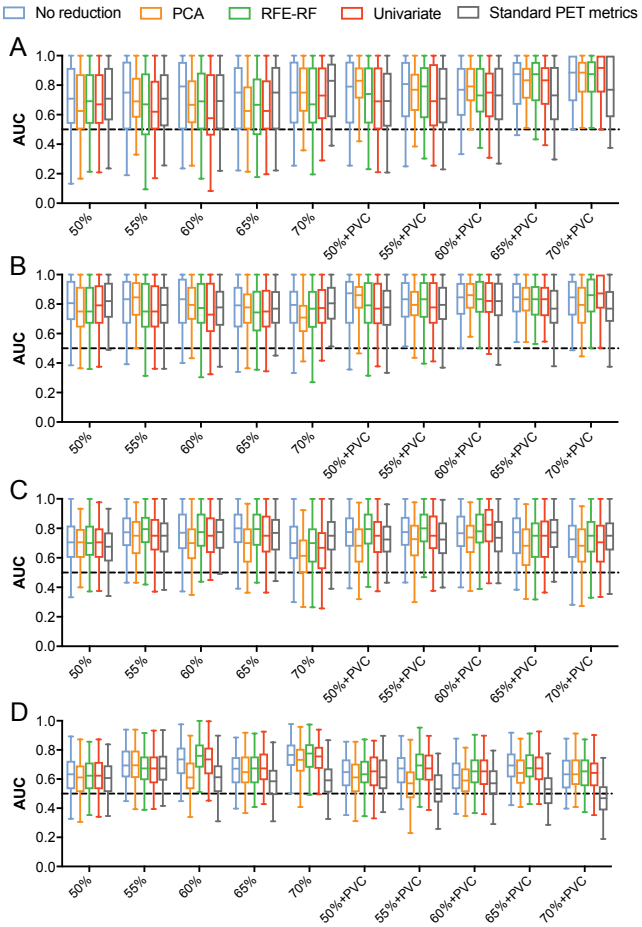


Figure 6.3: Boxplots of cross-validation AUCs for (A) LNI, (B) any metastasis, (C) Gleason score ≥8, and (D) ECE prediction. Results are shown for each delineation threshold on both original and PVC PET images, for all dimension reduction methods. Also, AUCs of standard PET metrics are shown in grey. Radiomics results in (A) and (B) with use of oversampling, results in (C) and (D) without use of oversampling. Standard PET results in (A) and (D) with use of oversampling, results in (B) and (C) without use of oversampling. Boxplots are outlier-trimmed (±2.5 percentile).

Impact Of Data Pre-Processing

Dimension reduction had a limited effect on mean AUCs, with median differences of -0.02 (range -0.11 to 0.07), -0.02 (range -0.07 to 0.04), -0.02 (range -0.11 to 0.04), and 0.00 (range -0.11 to 0.04) for LNI, metastasis, GS, and ECE prediction, respectively. Overall, oversampling tended to slightly increase AUCs for LNI and metastasis prediction, with a median difference in AUCs of +0.02 (range -0.06 to 0.07) and +0.02 (range -0.01 to 0 0.06), respectively. Generally, GS and ECE prediction did not benefit from oversampling, with a median difference in AUCs of 0.0 (ranging -0.02 to 0.05) and 0.0 (no range), respectively.

Final Predictions

Overall, the highest AUC of LNI prediction was 0.86±0.15 ($p<0.01$; Fig. 6.4: Youden sensitivity and specificity of 65% and 91%, respectively). A similar AUC of 0.86±0.14 ($p<0.01$) was found for prediction of any metastasis (Fig. 6.4: Youden sensitivity and specificity of 62% and 99%, respectively). The AUC of GS ≥8 prediction was 0.81±0.16 ($p<0.01$; Fig. 6.4: Youden sensitivity and specificity of 69% and 90%, respectively). The AUC for ECE prediction was 0.76±0.12 ($p<0.01$; Fig. 6.4: Youden sensitivity and specificity of 47% and 99%, respectively).

All LNI, metastasis, and GS AUCs were highest using univariate feature selection and minority class oversampling, on PVC images with a 70% (LNI and metastasis) or a 60% (GS) delineation threshold. ECE prediction AUC was highest using RFE-RF, no oversampling, and a 60% threshold on original images. Using these settings, Brier scores for LNI (0.09±0.05), any metastasis (0.10±0.04), and GS prediction (0.15±0.06) were low, indicating adequate model calibration. For ECE prediction, Brier scores were higher (0.21±0.05).

Feature Importance

For both LNI and metastasis prediction, intensity-based features *difference volume at intensity fraction* (importance coefficient 0.14 and 0.11, respectively) and *volume at intensity fraction 10* (importance coefficient 0.11 and 0.11, respectively) were most important, followed by multiple textural features and in a lesser extent several morphological features (Fig. 6.5). For GS prediction, textural features were evidently most important, specifically *zone size non uniformity* (importance coefficient 0.07), *zone distance non uniformity* (importance coefficient 0.06), and *grey level variance* (importance coefficient 0.05), with minor contributions from intensity and morphological features. For ECE prediction, again the *difference*

volume at intensity fraction (importance coefficient 0.03) and *volume at intensity fraction 10* (importance coefficient 0.02) features were among the most important features, along with *grey level non uniformity* (GLSZM; importance coefficient 0.02).

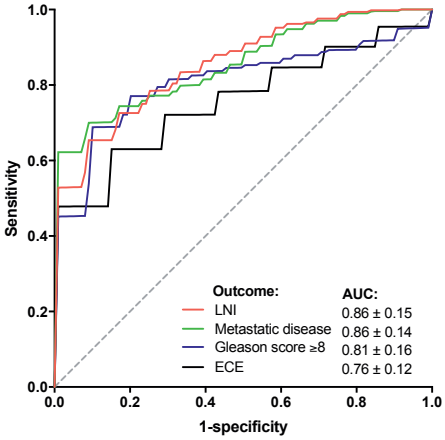


Figure 6.4: Mean cross-validated ROC curves of radiomics-based ML models. Random forest with univariate feature selection and minority class oversampling for LNI, metastasis, and GS prediction. Random forest RFE-RF feature selection without oversampling for ECE prediction.

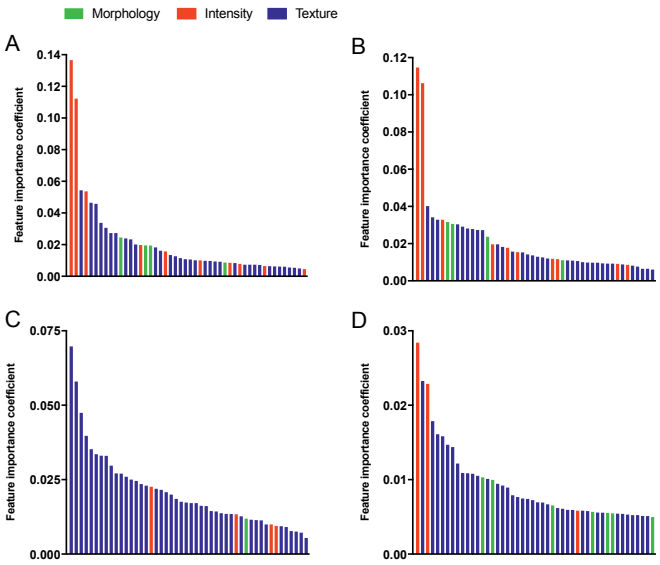


Figure 6.5: Feature importance coefficients from Random Forests trained to predict (A) LNI, (B) any metastasis, (C) Gleason score ≥8, and (D) ECE. Each bar represents the relative feature importance coefficient from a single radiomic feature. Shown are the top 10 percentile feature coefficients.

Standard PET Metrics

For the models using standard PET metrics as input, the highest AUCs for LNI, any metastasis, GS, and ECE prediction were 0.77±0.21 (p=0.03), 0.81±0.16 (p<0.01), 0.76±0.14 (p<0.01), and 0.67±0.14 (p=0.03), respectively. Differences between the highest mean AUCs of the standard PET metrics and radiomics models were -0.09 for LNI, -0.05 for any metastasis, -0.05 for GS, and -0.09 for ECE prediction. While these AUCs are consistently lower than those of the radiomics-based models, differences were not statistically significant (p=0.25-0.29). Still, relative standard deviations were higher (except for GS). Also, the average Brier scores of these models for these outcomes were higher than those of radiomics-based models (0.14±0.06, 0.11±0.04, 0.17±0.05, and 0.24±0.06, respectively).

Discussion

The present study demonstrates that [^{18}F]PSMA PET-CT radiomics can predict disease risk in primary PCa patients before treatment, rendering this approach clinically attractive to identify low-risk patients for whom ePLND will be unnecessary (Fig. 6.6). Our findings indicate that PSMA-expression detected on PET is related to both local tumor histopathology and metastatic tendency. For metastasis prediction, a higher tumor delineation threshold and PVC benefitted model stability. The use of different ML pre-processing methods (dimension reduction and oversampling) did not substantially affect model prediction performance. Standard PET metrics yielded non-significantly lower AUCs than radiomics-based models, a finding that will warrant confirmation in future external validation studies.

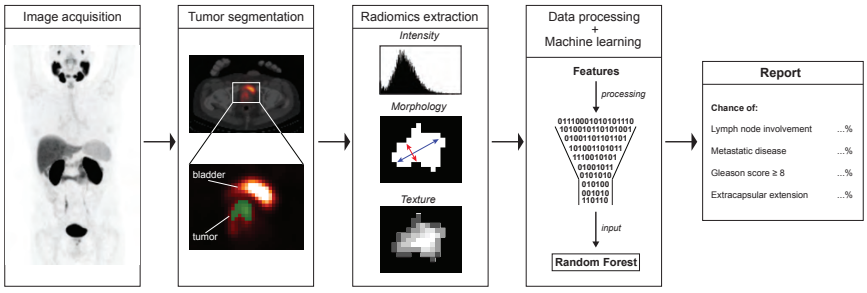


Figure 6.6: Illustration of a potential workflow for using ^{18}F-PSMA radiomics and machine learning in a clinical setting.

Kaittanis et al. observed that PSMA-expression on [^{68}Ga]PSMA PET/MR correlated with phosphorylation of Akt, a kinase involved in oncogenic signaling that drives PCa progression, but less so with GS and PSA (*8*). This might explain why intensity-based features were most important in prediction of metastatic disease (Fig. 6.5). Moreover, a recent study observed that PSMA-expression on [^{68}Ga]PSMA PET correlated with genomic index lesions (*33*). While PSMA-expression correlated with GS on immunohistochemistry, the association between PSMA uptake on PET (expressed in SUVmax) and GS is not fully evident (*34-36*). This may indicate that information on the spatial distribution of PSMA-expression is needed. Indeed, textural features appeared to be most important within the Random Forest models for GS prediction (Fig. 6.5). As texture on PET may be partly related to tumor volume, some caution regarding interpretation of these data is warranted. Still, morphological features appeared to be of limited value for GS prediction. Taken together, PSMA PET radiomics may capture tumor aggressiveness by carrying genomic as well as histopathological information. A full head-to-head comparison of radiomics with genomic, molecular (e.g. PSMA- and androgen receptor expression (*37*)), and histopathological features will be necessary to establish the biological basis of PSMA PET radiomics.

Zamboglou et al. similarly investigated [^{68}Ga]PSMA PET radiomics for prediction of GS ≥8 and LNI, observing similar validation AUCs for GS (AUC 0.84) and LNI prediction (AUC 0.85) (*38*). However, no cross-validation was applied to mitigate bias induced by a limited sample size. Also, the authors selected a single radiomic feature for LNI prediction based on its correlation with GS, which might explain why the AUCs of LNI and GS prediction were similar. Recently, Ferraro et al. evaluated whether standard PET metrics from [^{68}Ga]PSMA could predict LNI, and observed AUCs of 0.70-0.76, similar to the AUCs we observed for standard PET metrics (*39*). While PSMA PET radiomics seem to outperform standard PET features for prediction of LNI, in our study the differences were not significant, which is at least partly due to the limited power of statistical testing in cross-validation.

Validation of radiomics for predictive modelling warrants that methodological PET factors are taken into account (*17*). To date, use of PVC is not often considered in PET radiomics studies (*40*). PVC could improve accuracy of intensity feature measurements in small and heterogeneous lesions, and improve textural features calculation by reducing spill-over between voxels. Also, as PVC increases tumor-to-background contrast it may improve tumor delineation, which

may be of particular benefit for low-grade prostate cancer lesions that tend to be less avid on PSMA PET. We observed that PVC had a substantial impact on most radiomics features (Fig. 6.2A). Also, use of PVC tended to increase the predictive value of radiomics for LNI and any metastasis by improving model stability.

Between prediction of metastatic disease and histopathological features there was no single delineation threshold optimal for both, which may be due to the different importance of features between these outcomes. For metastasis prediction, a higher delineation threshold seemed to benefit model performance, while for GS and ECE prediction no clear trend was observed between the different thresholds and use of PVC. Specifically for multicenter settings, harmonization of tumor delineation method and use of PVC data will be crucial. In order to facilitate radiomics analysis, it may be an option to extract radiomics features using a 70%peak threshold on PVC-images for all predictions as this approach tended to improve LNI and metastasis prediction AUCs and model stability (Fig. 6.3), and had minimal effect on the other outcome predictions.

Some studies have observed that in radiomics analyses, calculation of textural features might be biased in small tumors or provide little added value above lesion volume itself (*41,42*), suggesting small lesions might need to be excluded from such studies. Still, the redundancy of those features will depend on a complex relationship between lesion size distributions, level of correlation between the individual features, and the relative importance of those features within the prediction models. Perhaps, a better approach to determine the clinical added value of small tumor PET radiomics might be to determine its predictive value and benchmark this against that of basic PET features. Also, a potential benefit of PVC needs to be considered. Despite analyzing predominantly small lesions (see Supplemental Fig.), we did find significant predictive value in the radiomics data, with (non-significantly) higher AUCs than based on standard PET-metrics. Also, use of PVC seemed to benefit LNI and metastasis prediction. Hence, small tumor radiomics and use of PVC may indeed allow for worthwhile radiomics studies in cancers with predominantly small lesions. Still, future multicenter external validation is needed to demonstrate true benefits of PSMA-radiomics over standard PET metrics in these small prostate cancer lesions, especially since using different PET systems with potentially different imaging protocols might negatively affect radiomics-based predictions more than those based on standard PET features.

Our study has several limitations. First, the data set was relatively small. Still, the significant high cross-validated prediction scores indicate that even for such a training dataset size the ML models were able to identify high-risk patients in independent data. Enlargement of the current dataset will likely improve model stability and potentially model calibration. Finally, external model validation was not yet performed. In such analysis, harmonization of image processing and tumor delineation method is recommended (*23*).

Conclusions

[^{18}F]PSMA PET radiomic features analyzed with ML are significantly predictive for LNI, presence of any metastasis, and high-risk pathological tumor features in primary PCa patients. These data demonstrate that the spatial distribution and levels of PSMA expression quantified on [^{18}F]PSMA PET are related to both tumor histopathological grade and metastatic tendency. For prediction of nodal and/or distant metastatic disease, PVC and a higher segmentation threshold seemed to improve model stability. Future multicenter external validation is needed to determine the benefits of using radiomics versus standard PET metrics.

References

1. Bader P, Burkhard FC, Markwalder R, Studer UE. Disease progression and survival of patients with positive lymph nodes after radical prostatectomy. Is there a chance of cure? *J Urol.* 2003;169:849-854.
2. Heidenreich A, Varga Z, Von Knobloch R. Extended pelvic lymphadenectomy in patients undergoing radical prostatectomy: high incidence of lymph node metastasis. *J Urol.* 2002;167:1681-1686.
3. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: Screening, diagnosis, and local treatment with curative intent. *Eur Urol.* 2017;71:618-629.
4. Musch M, Klevecka V, Roggenbuck U, Kroepfl D. Complications of pelvic lymphadenectomy in 1,380 patients undergoing radical retropubic prostatectomy between 1993 and 2006. *J Urol.* 2008;179:923-928; discussion 928-929.
5. Briganti A, Chun FK, Salonia A, et al. Complications and other surgical outcomes associated with extended pelvic lymphadenectomy in men with localized prostate cancer. *Eur Urol.* 2006;50:1006-1013.
6. Danneman D, Drevin L, Delahunt B, et al. Accuracy of prostate biopsies for predicting Gleason score in radical prostatectomy specimens: nationwide trends 2000-2012. *BJU Int.* 2017;119:50-56.
7. Israeli RS, Powell CT, Corr JG, Fair WR, Heston WD. Expression of the prostate-specific membrane antigen. *Cancer Res.* 1994;54:1807-1811.
8. Kaittanis C, Andreou C, Hieronymus H, et al. Prostate-specific membrane antigen cleavage of vitamin B9 stimulates oncogenic signaling through metabotropic glutamate receptors. *J Exp Med.* 2018;215:159-175.
9. Ross JS, Sheehan CE, Fisher HA, et al. Correlation of primary tumor prostate-specific membrane antigen expression with disease recurrence in prostate cancer. *Clin Cancer Res.* 2003;9:6357-6362.
10. Hupe MC, Philippi C, Roth D, et al. Expression of prostate-specific membrane antigen (PSMA) on biopsies is an independent risk stratifier of prostate cancer patients at time of initial diagnosis. *Front Oncol.* 2018;8:623.
11. Paschalis A, Sheehan B, Riisnaes R, et al. Prostate-specific membrane antigen heterogeneity and DNA repair defects in prostate cancer. *Eur Urol.* 2019;76(4):469-478.
12. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441-446.
13. De Bernardi E, Buda A, Guerra L, et al. Radiomics of the primary tumour as a tool to improve (18) F-FDG-PET sensitivity in detecting nodal metastases in endometrial cancer. *EJNMMI Res.* 2018;8:86.
14. Li K, Sun H, Lu Z, et al. Value of [(18)F]FDG PET radiomic features and VEGF expression in predicting pelvic lymphatic metastasis and their potential relationship in early-stage cervical squamous cell carcinoma. *Eur J Radiol.* 2018;106:160-166.
15. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging.* 2019;46(13):2638-2655..
16. van Velden FH, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imaging Biol.* 2016;18:788-795.
17. Cook GJR, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and promises of PET radiomics. *Int J Radiat Oncol Biol Phys.* 2018;102:1083-1089.
18. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328-354.
19. Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and precision of partial-volume correction in oncological PET/CT studies. *J Nucl Med.* 2016;57:1642-1649.
20. Kaalep A, Sera T, Rijnsdorp S, et al. Feasibility of state of the art PET/CT systems performance harmonisation. *Eur J Nucl Med Mol Imaging.* 2018;45:1344-1361.
21. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:161207003.*
22. Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. RaCaT: An open source and easy to use radiomics calculator tool. *PLoS One.* 2019;14:e0212223.

23. Pfaehler E, van Sluis J, Merema BBJ, et al. Experimental multicenter and multivendor evaluation of PET radiomic features performance using 3D printed phantom inserts. *J Nucl Med.* 2019.

24. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging.* 2017;44:151-165.

25. Predregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.

26. Breiman L. Random Forests. *Machine Learning.* 2001;45:5-32.

27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357.

28. Murphy AH. A new vector partition of the probability score. *J Appl Meteorol.* 1973;12:595-600.

29. Ojala M, Garriga GC. Permutation tests for studying classifier performance. *J Mach Learn Res.* 2010;11:1833-1863.

30. van de Wiel MA, Berkhof J, van Wieringen WN. Testing the prediction error difference between 2 predictors. *Biostatistics.* 2009;10:550-560.

31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837-845.

32. Koo TK, Li MY. A Guideline of Selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155-163.

33. Kesch C, Radtke JP, Wintsche A, et al. Correlation between genomic index lesions and mpMRI and (68)Ga-PSMA-PET/CT imaging features in primary prostate cancer. *Sci Rep.* 2018;8:16708.

34. Woythal N, Arsenic R, Kempkensteffen C, et al. Immunohistochemical Validation of PSMA Expression Measured by (68)Ga-PSMA PET/CT in Primary Prostate Cancer. *J Nucl Med.* 2018;59:238-243.

35. Prasad V, Steffen IG, Diederichs G, Makowski MR, Wust P, Brenner W. Biodistribution of [(68)Ga] PSMA-HBED-CC in patients with prostate cancer: characterization of uptake in normal organs and tumour Lesions. *Mol Imaging Biol.* 2016;18:428-436.

36. Bravaccini S, Puccetti M, Bocchini M, et al. PSMA expression: a potential ally for the pathologist in prostate cancer diagnosis. *Sci Rep.* 2018;8:4254.

37. Palamiuc L, Emerling BM. PSMA brings new flavors to PI3K signaling: A role for glutamate in prostate cancer. *J Exp Med.* 2018;215:17-19.

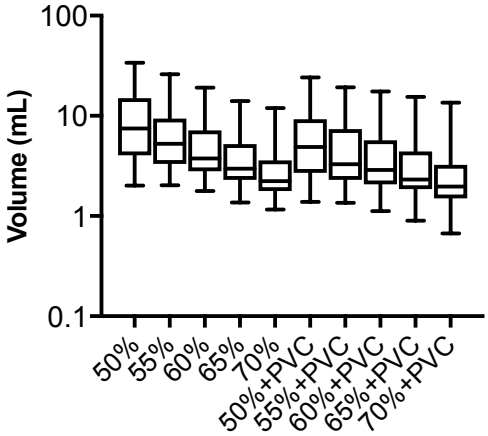
38. Zamboglou C, Carles M, Fechter T, et al. Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate- and high-risk prostate cancer - a comparison study with histology reference. *Theranostics.* 2019;9:2595-2605.

39. Ferraro DA, Muehlematter UJ, Garcia Schuler HI, et al. (68)Ga-PSMA-11 PET has the potential to improve patient selection for extended pelvic lymph node dissection in intermediate to high-risk prostate cancer. *Eur J Nucl Med Mol Imaging. 2020;47(1):147-159.*

40. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour (1)(8)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging.* 2013;40:1662-1671.

41. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med.* 2014;55:37-42.

42. Hatt M, Majdoub M, Vallieres M, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med.* 2015;56:38-44.



Supplemental Figure: Delineated PSMA-positive tumor volumes (mL). Data shown for each delineation threshold (peak reference) with and without partial-volume correction (PVC).


QI-1-PET

Part 2

Clinical Application


QI-1-PET

Chapter 7

Impact of Partial-Volume Correction in Oncological PET Studies: A Systematic Review and Meta-Analysis

M.C.F. Cysouw, G.M. Kramer, L.J. Schoonmade, R. Boellaard, H.C.W. de Vet, O.S. Hoekstra

Eur J Nucl Med Mol Imaging. 2017 Nov;44(12):2105-2116.

Abstract

Positron-emission tomography can be useful in oncology for diagnosis, (re) staging, determining prognosis, and response assessment. However, partial-volume effects hamper accurate quantification of lesions <2-3x the PET system's spatial resolution, the clinical impact of which is not evident. This systematic review provides an up-to-date overview of studies investigating impact of partial-volume correction (PVC) in oncological PET studies.

Methods: We searched in PubMed and Embase databases according to the PRISMA statement, including studies from inception till May 9th 2016. Two reviewers independently screened all abstracts and eligible full-text articles, and performed quality assessment according to QUADAS-2 and QUIPS criteria. For a set of similar diagnostic studies, we statistically pooled the results using bivariate meta-regression.

Results: Thirty-one studies were eligible for inclusion. Overall, study quality was good. For diagnosis and nodal staging, PVC yielded a strong trend of increased sensitivity at expense of specificity. Meta-analysis of six studies investigating diagnosis of pulmonary nodules (679 lesions) showed no significant change in diagnostic accuracy after PVC (p=0.222). Prognostication was not improved for non-small cell lung cancer and esophageal cancer, whereas it did improve for head-and-neck cancer. Response assessment was not improved by PVC for (locally advanced) breast cancer and rectal cancer, and was worsened in metastatic colorectal cancer.

Conclusions: The accumulated evidence to date does not support routine application of PVC in standard clinical PET practice. Consensus on the preferred PVC methodology in oncological PET should be reached. Partial-volume corrected data should be used as adjuncts, but not yet replace, uncorrected data.

Introduction

Positron-emission tomography (PET) enables in-vivo assessment of metabolic and intracellular processes. Whereas in clinical practice PET is predominantly used to qualitatively assess tracer uptake, PET(/CT) may also serve as a surrogate quantitative biomarker of, for example, tumour metabolism and proliferation. Interest into quantitative approaches of tumour assessment has grown considerably, for discriminating between benign and malignant lesions, staging, prognostication, and determining or predicting therapy response (*1-4*).

Accurate quantification of metabolic volumes <2-3x the spatial resolution of PET is hampered by partial-volume effects, leading to underestimations of standardized uptake values (SUV), and possibly compromising lesion detection (*5,6*). Many methods for partial-volume correction (PVC) have been advocated (*7*). The simplest way is to use recovery coefficients obtained from phantom experiments, which assumes that true metabolic volume is known and that lesions are spherically shaped with homogeneous uptake. More sophisticated methods have been developed, but all suffer from limitations (*7,8*). Voxel-wise resolution recovery methods, incorporating the point spread function (PSF) within iterative reconstruction (*9*) (PSF reconstruction) or performing post-reconstruction iterative deconvolution (*10*), could improve both qualitative and quantitative reads. To date, consensus on standardized application of PVC in oncological PET/CT studies is lacking, and perhaps as a consequence PVC is not yet routinely applied. In fact, most current clinical quantitative PET studies merely exclude small lesions (e.g. <2cm in diameter), as recommended in the PERCIST criteria (*3*).

The clinical impact of PVC in oncological setting, and thus the need for standardized application, is not yet fully elucidated (*7*). We performed a systematic review and meta-analysis to assess the impact of PVC in clinical PET studies, focusing on diagnosis, staging, prognostication, and response assessment.

Materials and methods

Search strategy

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement. A comprehensive search (Supplemental Tables 1 and 2, available at https://www.

springer.com/journal/259) was performed in PubMed and Embase.com from inception to May 9th 2016, in collaboration with a medical librarian (LJS). Search terms included controlled terms (MesH in PubMed, EMtree in Embase) as well as free text terms. The following terms were used (including synonyms and closely related words) as index terms or free-text words: 'positron-emission tomography or 'PET' and 'partial volume correction' or 'point spread function reconstruction' and 'neoplasms' or 'cancer'.

Selection process

Abstracts and titles of all studies retrieved from the search were independently screened by two researchers (MCFC and GMK). Afterwards, eligible articles were studied in full-text. In case of differences in judgment, consensus was reached through discussion. Cross-referencing was performed to further identify relevant articles.

Inclusion criteria were: studies applying PVC in clinical PET studies, using oncological patients, reporting PET data with and without PVC, and investigating clinical impact of PVC on either diagnosis, staging, prognostication (reporting survival data), or response assessment.

Exclusion criteria were: reviews, letters, editorials, conference abstracts, case reports, full-text not available or not in English, no adequate reference data, no description of, or reference to, PVC method, combined PVC and motion blur correction method, or overlapping patient cohort with other included study.

Quality assessment

Quality of included articles was assessed (independently by MCFC and GMK) according to the QUADAS-2 (*11*) (n=25) or QUIPS (*12*) (n=12) tools. QUADAS-2 assesses bias and applicability of diagnostic studies, whereas QUIPS assesses bias of studies investigating prognostic factors. Staging and response assessment studies were assigned to either of the quality assessment tools. Consensus was reached through discussion.

Data-extraction and meta-analysis

Both researchers independently extracted results regarding impact of PVC on diagnostic accuracy (for diagnosis and staging), prediction of survival (for prognostication), and response assessment. Measures of diagnostic accuracy were derived with and without PVC. If test characteristics were described for

subgroups, overall measures of accuracy were calculated when possible. When p-values of differences in accuracy between uncorrected and PVC data were not reported, these differences were deemed not statistically significant. Descriptive data regarding cancer type, number of patients, lesion sizes, scanner type, and PVC method were also extracted. Unless stated otherwise, we presented data on SUV quantification.

Diagnostic studies on the same topic were pooled using bivariate random effects meta-regression analysis, which is the recommended method for meta-analysis of diagnostic studies (*13*). This method provides summary estimates of sensitivity and specificity with 95% confidence intervals, taking into account the correlation between sensitivity and specificity and heterogeneity in results between studies. We tested for differences in overall diagnostic accuracy between different diagnostic tests using a likelihood ratio test, comparing models that included and excluded a covariate for the diagnostic test. For illustrative purposes, summary receiver-operator-characteristics (ROC) curves were calculated according to the Moses-Littenberg method (*14*). We used Stata software (StataCorp. 2015. *Stata Statistical Software: Release 14.* College Station, TX: StataCorp LP.) for statistical analyses.

Results

Study selection

Pubmed and EMBASE searches yielded 371 potentially eligible studies (Figure 7.1). Three studies were additionally found through reference screening. Two-hundred and ninety-three abstracts were excluded based on eligibility criteria, leaving 81 for full-text screening. For 19 (5.1%) abstracts, judgments were conflicting, which were resolved through discussion. After full-text reviewing, 31 studies met eligibility criteria (Figure 7.1). Studies on diagnosis (n=10), staging (n=10), prognostication (n=6), and response assessment (n=5) are presented in Tables 7.1-7.4, respectively. Supplemental Table 3 contains the PVC and tumour delineation methodologies, reconstruction settings, full-width at half maximums, and voxel-sizes of each included study. Thirty studies used ^{18}F-FDG as radiopharmaceutical, one study used ^{18}F-Choline.

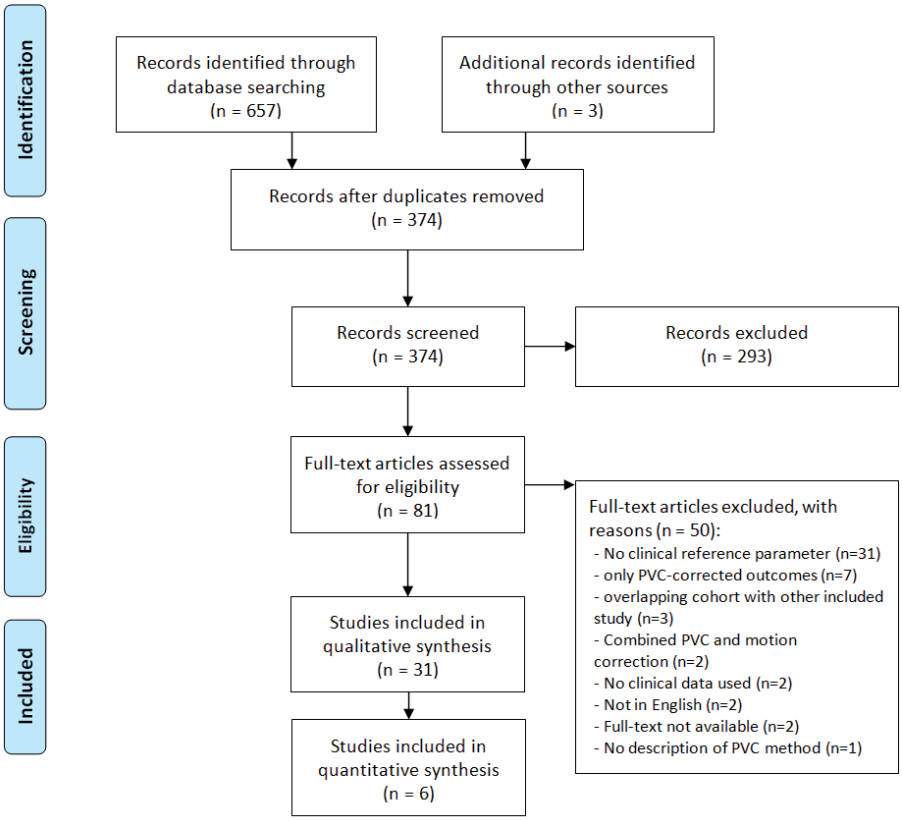


Figure 7.1: PRISMA flowchart.

Quality assessment

For extensive descriptions of the QUADAS-2 and QUIPS scoring criteria, we refer to their respective primary publications (*11,12*).

Considering QUADAS-2 (Figure 7.2a), the 'reference standard' and 'patient selection' items resulted in low risk of bias (high risk of bias in 14% of studies for either item). Elevated risk of bias for the 'reference standard' item was caused by use of multiple reference tests within the same study. Risk of bias in index test was high in 24% of studies due to use of data-driven, instead of pre-defined, SUV cut-offs. Applicability concerns regarding patient selection were mainly caused by large tumour size spectra and unspecified tumour sizes.

Using QUIPS (Figure 7.2b), low risk of bias scores were found in the majority of the studies for the items measurement of outcome and prognostics factors, study attrition, and statistical analysis and reporting. Several studies did

not adequately investigate potential factors of study confounding, which resulted in a moderate risk of bias in 40% of studies and high risk of bias in 40% of studies. Unclear descriptions of included patient cohorts ('study participation' item) resulted in moderate risk of bias in 40% of included studies.

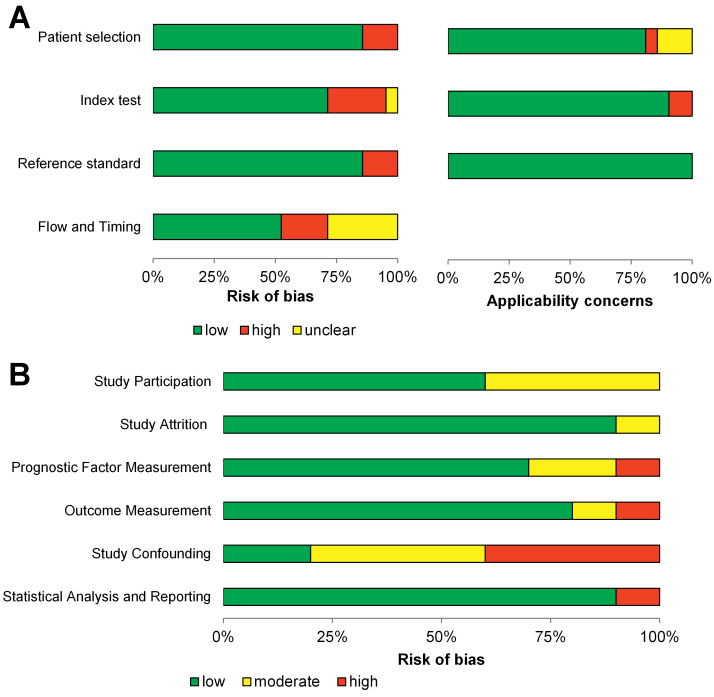


Figure 7.2: Results of quality assessment according to QUADAS-2 (a) and QUIPS (b) tools.

Table 7.1: Eligible diagnostic studies, in chronological order.

Ref.	No. of patients	Target lesions	No. and type lesions	Lesion sizes (mm)*	Cut-off:	Non-PVC PVC	Reference test(s)	Effect on test performance?
(***21***)	73	breast tumors	51 M, 46 B	25±9 (B), 27±17 (M)	data-driven	2.1 2.1	histology	sens ↑ 69 to 81% spec = 90%
(***23***)	27	malignant lymphoma	n.s.	median 18 (range 8-53)	n.a.		follow-up / biopsy	n.a.
(***19***)	127	pulmonary nodules	86 M, 41 B	33±23	pre-defined	2.5 2.5	histology	sens = 94% spec ↓ 76 to 67%
(***15***)	47	pulmonary nodules	36 M, 11 B	21.6±9.7	pre-defined	2.5 2.5	follow-up / biopsy	sens ↑ 72 to 97% spec ↓ 82 to 73%
(***16***)	60	pulmonary nodules	46 M, 14 B	26.3±15.8 (M), 20.4±10.4 (B)	pre-defined	2.5 2.5	histology	sens ↑ 87 to 98% spec ↓ 21 to 14%
(***17***)	265	pulmonary nodules	72 M, 193 B	<10 (n=32), 10-15 (n=57), 16-30 (n=176)	pre-defined	2.5 2.5	follow-up / biopsy	sens ↑ 65 to 90% spec ↓ 92 to 80%
(***18***)	46	pulmonary nodules	26 M, 23 B	20±7 (M), 13±5 (B)	data-driven	2.4 2.9	follow-up / biopsy	sens ↑62 to 73% spec = 80%
(***22***)	42	NHL	26 aggressive 16 indolent	32.4±18.3 (aggressive), 21.9±10.3 (indolent)	data-driven	9.5 11.2	histology	sens = 81% spec ↓ 81 to 63%
(***20***)	131	pulmonary nodules	86 M, 45 B	29.1±18.1	pre-defined	2.5 2.5	histology	sens ↓ 89 to 88% spec ↓ 51 to 42%
(***24***)	22	lymph nodes	8 KFD, 14 NHL	13.8±5.4 (KFD), 25.4±11.8 (indolent), 29.7±18.8 (aggressive)	n.a.		histology	n.a.

*Sizes are presented in mean±SD, unless stated otherwise. M = malignant; B = benign; NHL = non-Hodgkin lymphoma; KFD = Kikuchi-Fujimoto disease; n.a. = not applicable; n.s. = not specified; sens = sensitivity, spec = specificity.

Table 7.2: Eligible studies evaluating staging, in chronological order.

Ref.	No. of patients	Cancer type	No. and type lesions	Lesion sizes (mm)*	T/N/M	Cut-off:	Non-PVC PVC	Method of staging	Effect on test performance?
(***25***)	178	NSCLC	n.s.	range 18±5 - 44±20	TNM	n.a.		imaging / surgery / pathology	n.a.
(***28***)	7	thyroid (mLN)	15 M, 24 B	n.s.	N	data-driven	4.0 10.0	imaging / pathology	sens = 100% spec ↑ 92 to 100%
(***26***)	52	breast	n.s.	n.s.	N	n.s.	n.s.	imaging	sens ↑ 75 to 86% spec ↓ 87 to 83%
(***31***)	58	NSCLC	201	7.2±1.7 (<10mm), 19.2±1.05 (≥10mm)	N	n.a.		pathology / imaging / clinical	sens ↑ 78 to 97% spec ↓ 71 to 58%
(***34***)	35	lung	n.s.	30 (range 8-79)	TNM	n.a.		imaging / pathology	n.a.
(***32***)	50	breast (mLN)	n.s.	8.2±4.3	N	n.a.		pathology	sens ↑ 76 to 85% spec ↓ 75 to 69%
(***27***)	32	HNSCC (mLN)	18 M, 39 B	1.14±1.38mL (M), 0.64±0.93mL (B)	N	data-driven	n.s.	pathology	sens ↑ 57 to 64% spec ↑ 71 to 76%
(***29***)	71	nasopharyngeal (mLN)	35 M, 53 B	<6 (n=55) 6-6.9 (n=7) ≥7 (n=26	N	pre-defined	2.5 n.s	imaging	sens ↑ 77 to 94% spec ↓ 89 to 59%
(***30***)	39	prostate	49 prostatic, 43 nodal	n.s.	TN	data-driven	2.4 5.0	pathology / imaging / PSA	sens ↓ 90 to 84% spec = 73%
(***33***)	38	colorectal	32 M, 115 B	n.s.	N	n.a.		surgery / pathology	sens ↑ 53 to 66% spec = 99.1%

*Sizes are presented in mean±SD, unless stated otherwise. M = malignant; B = benign; NSCLC = non-small cell lung cancer; HNSCC = head-and-neck squamous cell carcinoma; mLN = lymph node metastases; n.a. = not applicable; n.s. = not specified; PSA = prostate specific antigen; sens = sensitivity, spec = specificity.

Table 7.3: Eligible studies evaluating prognostication, in chronological order.

Ref.	No. of patients	Cancer type	No. of lesions	Spectrum of tumor sizes (mm)*	Effect on prognostication?
(***35***)	145	NSCLC	n.s.	median 30 (range 10-110)	not improved
(***37***)	52	esophageal	n.s.	n.s.	not improved
(***38***)	50	esophageal	n.s.	39.9±36.1mL	not improved
(***36***)	191	NSCLC	n.s.	median 23 (range 10-36)	not improved
(***39***)	19	HNC	19	15.2±5.0	improved
(***40***)	19	HNC	19	15±5	improved for subgroup

*Sizes are presented in mean±SD, unless stated otherwise. NSCLC = non-small cell lung cancer; mLN = lymph node metastases; HNC = head-and-neck cancer; n.s. = not specified.

Table 7.4: Eligible studies evaluating response assessment, in chronological order.

Ref.	No. of patients	Cancer type	No. of lesions	Spectrum of tumor sizes (mL)*	Reference test	Effect on response assessment?
(***41***)	35	LABC	n.s.	n.s.	clinical + pathologic	not improved
(***42***)	51	breast	n.s.	median 14 (range 2-227)	pathologic	not improved
(***43***)	28	LARC	n.s.	median 23 (range 2-397)	pathologic	not improved
(***44***)	40	mCRC	101	34.4±66.4	RECIST	worsened
(***45***)	19	NSCLC	24	median 6.95 (range 2.2-46)	clinical	PERCIST classification improved in 5 lesions, confirmed in follow-up

*Sizes are presented in mean±SD, unless stated otherwise, at baseline. LABC = locally-advanced breast cancer; LARC = locally-advanced rectal cancer; NSCLC = non-small cell lung cancer; mCRC = metastatic colorectal cancer; n.s. = not specified.

Diagnosis

Impact of PVC on diagnosis (Table 7.1, n=10) was investigated for pulmonary nodules (n=6), breast lesions (n=1), and lymphoma (n=3). PVC included recovery coefficient-method (n=9) and CT-volume-based PVC (n=1). All studies reported lesion sizes. One study stratified both uncorrected and PVC data for lesion size in secondary analysis.

The six studies evaluating diagnostic accuracy of PET for pulmonary nodules were pooled (Table 7.1, Figure 7.3-7.4), including a total of 352 malignant and 327 benign lesions (*15-20*). Prevalence of malignancy ranged from 27-77% (mean 57%). Five studies applied an RC-method for PVC, one study applied a CT-volume-based correction. Thresholds of PET positivity were predefined in 5/6 studies and data-driven in 1/6 studies. Predefined thresholds were similar for uncorrected and PVC data. Three studies used SUV 2.5 as predefined threshold (*16,17,20*). One study used SUV 2.0 and 2.5 as thresholds (*19*). One study used SUV 1.5, 2.0, 2.5, and 3.0 as thresholds (*15*). In case of multiple predefined thresholds,

results of the SUVmax 2.5 threshold were used in meta-analysis (SUVmean for PVC data in Hickeson et al.) since this was reported in all 5 studies with predefined SUV thresholds. One study used data-driven thresholds specifically for uncorrected (SUV 2.4) and PVC data (SUV 2.9) (*18*). Pooled sensitivity and specificity of uncorrected data were 81% (95%CI 70-89) and 70% (95%CI 48-86), respectively (Figure 7.5). Pooled sensitivity and specificity of partial-volume corrected data were 91% (95%CI 83-95) and 60% (95%CI 37-79), respectively (Figure 7.4). No significant change in diagnostic accuracy after PVC was found (p=0.222), using the SUV thresholds as described above. One of the pulmonary studies (by Hickeson et al.) stratified both uncorrected and corrected data for lesion size (*15*). The authors observed that for lesions <2cm accuracy increased from 59% to 85% using SUV cut-off 2.5, while for lesions >2cm accuracy changed from 95% to 100%.

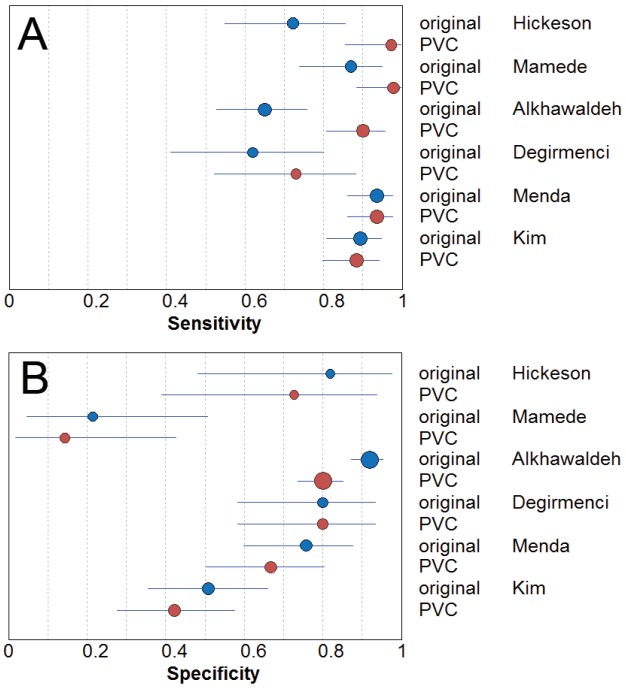


Figure 7.3: Forest plots presenting sensitivity (a) and specificity (b) with 95% CI of discrimination between benign and malignant pulmonary nodules with ^{18}F-FDG-PET.

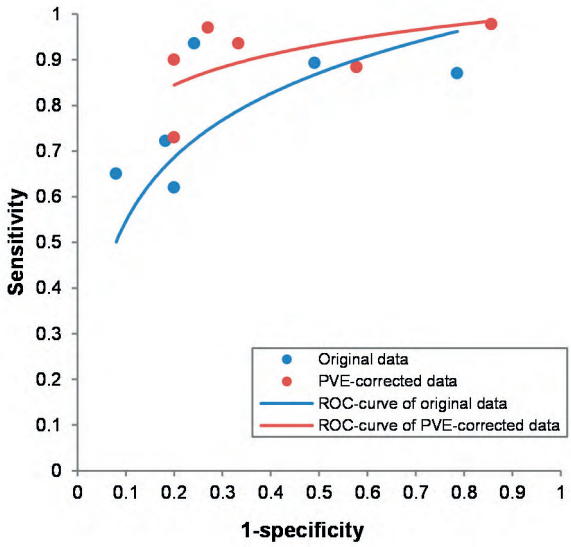


Figure 7.4: Summary ROC curves of discrimination between benign and malignant pulmonary nodules with ^{18}F-FDG-PET.

With diagnosis of breast lesions, using data-driven SUVmean thresholds of 2.1 for PVC and non-PVC, at a fixed specificity of 90%, PVC increased sensitivity from 69 to 81%, but the impact on accuracy was not statistically significant (*21*). In discriminating between aggressive and indolent non-Hodgkin lymphoma (NHL), PVC decreased specificity without affecting sensitivity (*22*). Similarly, PVC did not improve differentiating between high- and low-grade NHL (*23*). PVC to enabled differentiation between indolent NHL and Kikuchi-Fujimoto disease (*24*).

Staging

Studies evaluating the effect of PVC on staging (Table 7.2, n=10) included lung (n=3), breast (n=2), thyroid (n=1), head-and-neck squamous cell (n=1), nasopharyngeal (n=1), prostate (n=1), and colorectal cancer (n=1). Applied PVC methods were the recovery coefficient-method (n=4), PSF reconstruction (n=4), iterative deconvolution (n=1) and geometric transfer matrix (n=1). Most of these studies did not specify SUV thresholds of test positivity for uncorrected and PVC data. Four studies did not specify lesions sizes. One study stratified both uncorrected and PVC data for lesion size in secondary analysis.

In non-small cell lung cancer (NSCLC) patients the association between primary tumour SUVmax and overall TNM stage disappeared after PVC (*25*).

For nodal staging using SUV, non-significant trends of increased accuracy for breast, head-and-neck squamous cell, and thyroid cancer (from 80%, 66% and 95% to 84%, 71% and 100%, respectively) (*26-28*), and decreased accuracy for nasopharyngeal and prostate cancer (from 84% and 85% to 73% and 80%, respectively) were observed (*29,30*). The study investigating accuracy of nodal staging of nasopharyngeal cancer did observe a large increase in accuracy, from 14% to 71%, when stratifying for lesion size (6-7mm diameter) (*29*).

With visual image interpretation, PSF reconstruction tended to increase accuracy of nodal staging in NSCLC, breast, and colorectal cancer (not statistically significant) compared to non-PSF reconstruction (from 76%, 76%, and 89% to 84%, 80%, and 92%, respectively) (*31-33*). Another study found no significant difference in lung cancer (several types) overall staging accuracy between non-PSF and PSF reconstruction (*34*).

Prognosis

Impact of PVC on prognostication (Table 3, n=6) was investigated for NSCLC (n=2), esophageal (n=2), and head-and-neck cancer (n=2). Applied PVC methods were the recovery coefficient-method (n=4), iterative deconvolution (n=1), and mask-based PVC (n=1). Only prognostic studies providing survival data were included. One study did not specify lesion sizes. None of the studies stratified results on PVC for lesion size in secondary analysis.

PVC did not alter the association of SUVmax with disease-free survival of NSCLC (various histological types) patients in multivariate analysis (*35,36*). Similarly, in NSCLC patients (various histologic types) PVC did not alter the ROC area-under-the-curve of primary tumour SUVmax to differentiate between groups of patients in terms of disease-free and overall survival (*36*). Primary tumour SUVs, regardless of PVC, were insufficient as prognostic markers in esophageal (adeno- and squamous cell) cancer in univariate and ROC analysis (*37,38*). In head-and-neck cancer patients, partial-volume-corrected SUV was significantly different between patient groups stratified according to disease-free survival, whereas uncorrected SUV was not (*39*). In univariate analysis, PVC did not affect predictive value of head-and-neck cancer primary tumour SUV on local recurrence-free survival, distant metastasis-free survival, and disease-free survival, but did allow for prediction of distant metastasis-free survival in a subgroup of patients with PET-positive lymph nodes (*40*).

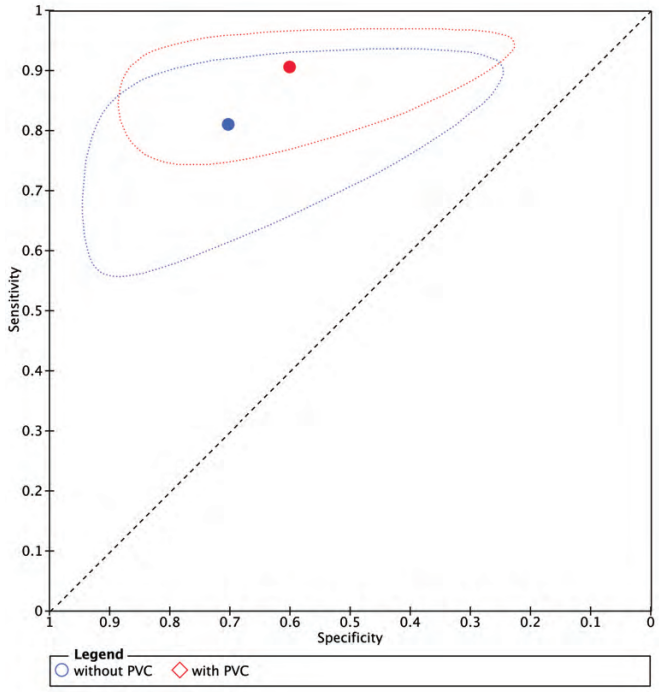


Figure 7.5: Summary sensitivity and specificity with 95% confidence region of discrimination between benign and malignant pulmonary nodules with ^{18}F-FDG-PET.

Response assessment

Impact of PVC on response assessment (Table 4, n=5) was investigated for breast (n=2), rectal (n=1), colorectal (n=1), and NSCLC (n=1). Applied PVC methods were the recovery coefficient-method (n=2), iterative deconvolution (n=2), and both RC-method and iterative deconvolution (n=1). One study did not specify lesion sizes. None of the studies stratified results on PVC for lesion size in secondary analysis.

For locally-advanced breast cancer (*41*), regardless of PVC primary tumour FDG metabolic rate was not able to differentiate between clinical and pathologic responders and non-responders during neoadjuvant chemotherapy (after 2 months). In another study in breast cancer patients PVC did not significantly change prediction of pathologic response with primary tumour SUV during neoadjuvant therapy (after 2 cycles) (*42*)]. In locally-advanced rectal cancer patients treated with (preoperative) chemoradiotherapy, PVC had no impact on histopathological response prediction, at baseline or after 1 or 2 weeks of therapy

(*43*). In patients with metastatic colorectal cancer PVC significantly reduced the ROC area-under-the-curve of SUV in discriminating between responders and non-responders after 2 weeks of chemotherapy, as defined with RECIST (*44*). In NSCLC patients treated with radio- or radiochemotherapy PVC changed PERCIST (*3*) classification of response in 5/24 lesions, which were verified as correct alterations in clinical follow-up (*45*).

Discussion

Quantification of functional tumour characteristics with PET is considered to be useful in clinical oncology, often using semi-quantitative analyses resulting in SUVs. Unfortunately, partial-volume effects are known to cause underestimations of tumour activity, and hence the necessity of PVC for accurate semi-quantitative reads for small lesions is well recognized (*5*)]. However, many factors affect its accuracy and potentially hamper its optimal usage. Perhaps as a consequence, its resulting advantage in oncological PET studies is not yet evident. Additionally, the lack of consensus on the preferred PVC and delineation method may result in suboptimal results and could hamper comparisons between studies. This review discusses the clinical impact of PVC and gives recommendations for specific research questions and analyses in future studies applying PVC.

When applied to diagnosis of primary lesions and (mainly nodal) staging PVC often yielded higher sensitivity at the expense of specificity (Tables 7.1-7.2 and Figures 7.3-7.4), which is an obvious consequence when using the same test positivity SUV thresholds for uncorrected and PVC data. In the subset of studies which allowed statistical pooling (679 lesions), meta-analysis showed that PVC did not significantly alter the overall diagnostic accuracy of characterizing pulmonary lesions with PET (Figure 7.5). When estimating the effect of PVC, the optimal trade-off between sensitivity and specificity (the SUV threshold of test positivity) may be different for PVC and uncorrected data. At an exploratory level, one should define this cut-off for either method. Of note, Degirmenci et al. (on pulmonary nodules) used data-driven SUV cut-offs of 2.4 and 2.9 for uncorrected and PVC data, respectively, which yielded a specificity fixed at 80% with sensitivities of 62 and 73% for uncorrected and PVC data, respectively (*18*). We performed a similar analysis using the (individual patient) data from Hickeson et al. (*15*). At a predefined SUV cutoff of 2.5, PVC decreased specificity

and increased sensitivity (Table 1). However, when applying cut-offs of 2.55 and 2.8 (as derived from ROC analysis) for uncorrected and PVC data, respectively, PVC increased sensitivity from 72% to 94% while specificity remained 91%. This provides further demonstration that PVC may indeed increase diagnostic accuracy when SUV cutoffs are adequately adapted for this correction. Obviously, each proposed threshold requires external validation.

Another explanation of the limited impact of PVC on diagnostic accuracy as published in the literature may relate to the size spectra of included lesions, with the distribution of benign and malignant lesions therein. When performing PVC analysis on all lesions, both large and small simultaneously, the overall impact of PVC on diagnostic accuracy will be diluted. Indeed, several studies demonstrated high impact of PVC on accuracy for small lesions (when stratifying for lesion size), but less so when including all lesions regardless of size (*15,29*). Therefore, we suggest that investigators stratify diagnostic performance results for lesion size in secondary analyses. However, since partial-volume effects are not merely size-dependent, but are also affected by lesion contrast and shape, reliable classification of lesions that are (most) affected by partial-volume effects will be difficult. In our previous simulation study we observed that for high contrast spherical lesions partial-volume effects started to occur below 3cm diameter (*8*). A practical approach for stratification would thus be to stratify results using a 3cm lesion diameter or a 14mL metabolic volume cutoff (corresponding to a 3cm diameter sphere). Even though larger lesions may also be somewhat affected by partial-volume effects depending on their shape and contrast, such a size cutoff will ensure that lesions that are most affected by partial-volume effects are separated. Another approach would be to plot the percentage increases in SUV after PVC as function of metabolic tumour volume to determine an appropriate size cutoff for stratification of results within studies (not possible when applying the RC method).

Regarding visual nodal staging, PSF reconstruction did not significantly alter accuracy, but tended to increase sensitivity in lung, breast, and colorectal cancer (Table 7.2) (*31-33*). This may be attributed to improved qualitative reads, improved (small) lesion detection, and higher diagnostic confidence (*31-33*). Therefore, it may be worthwhile to validate these higher-resolution reconstruction algorithms for use in clinical practice, especially for detection of small lymph node metastases and lesions embedded in high background activity such as in liver or mediastinum. However, PFS reconstructions may suffer from Gibbs artefacts

(overshoot in activity) and, moreover, they are also known not to guarantee full signal recovery (*9*). Also, further research into their impact on compliance with EANM standards is needed to ensure equal scanner calibration in multicenter quantitative PET/CT studies, which may require a SUV harmonization procedure (*46*).

We found that PVC might improve prognostication in head-and-neck cancer (*39,40*), but these studies did not stratify for the human papilloma virus-status, a prognostic marker associated with lower tumour SUV and smaller MATV (*47*). For future studies, please note that appropriate PVC may not necessarily improve prognostication with SUV, but rather may enable it to reflect its true prognostic value. For example, Vesselle et al. found that PVC mitigated the correlation between primary tumour SUV and overall survival in NSCLC patients, and also observed that the correlation between SUV and overall TNM-stage, which in essence is based on patient prognosis, disappeared after PVC, suggesting that the 'prognostic value' of uncorrected SUV was based on tumour volume rather than metabolic activity (*25,48*).

For response assessment no conclusions regarding PVC's effect can be made at this point, due to the small number of heterogeneous studies. One included study demonstrated that after PVC PERCIST classification of response was altered for 5/24 NSCLC lesions during radio- or radiochemotherapy (*45*). This is an important observation, since conceptually PVC may correct changes in SUV during treatment for changes in tumour volume and contrast, allowing for more appropriate PET-based classification of tumour response. Interestingly, two studies (excluded since no clinical verification was performed) demonstrated PVC to alter response classifications according to EORTC or PERCIST criteria in patients with bone metastases and NSCLC (*39,49*). Concluding, future PET response assessment studies should include PVC to allow for metabolic response assessment irrespective of tumour shrinkage or growth and quantify its clinical impact.

To improve comparison of PVC's impact between studies, consensus on the preferred combination of PVC and lesion delineation methodologies should be reached. Many PVC methods have been advocated, some specific for oncological application (*5,7,50,51*). Still, most studies in this review applied a recovery coefficient-method, a quite simple method assuming spherically shaped lesions, homogeneous activity distributions, and known tumour sizes. Using this method, even small errors in tumour size measurements may result

in over- or underestimations of true SUVs. Also, size measurements are often CT-based, whereas partial-volume effects affect metabolic volumes, which may be different from anatomical tumour volume (*52*) (e.g. due to necrosis and treatment effects). In a previous phantom and simulation study we found that voxel-wise PVC methods such as iterative deconvolution may be preferred, since this only assumes approximate knowledge of PET/CT systems' resolution kernel size, has low dependency on accurate delineation, and has only limited effect on precision (*8*). Additionally, such a voxel-wise PVC method could allow for more accurate delineation of tumours (*53*) and, theoretically, heterogeneous tumour background. However, iterative deconvolution is known to increase image noise levels, which may require some form of a denoising algorithm to be applied (*9,54*). Iterative deconvolution may be relatively easy to implement, and has demonstrated to perform well using commonly applied background-adapted threshold-based delineation methods (*8*). Since to date iterative deconvolution has been predominantly applied by the same research group (Supplemental Table 3), more extensive clinical evaluation is warranted. Our previous phantom and simulation study showed that for lesions ≤10mm in diameter even with PVC no fully accurate results can yet be acquired (*8*), which may contribute to the relatively low impact of PVC. Due to heterogeneity between studies the impact of chosen PVC methods on outcomes cannot be established in this review.

A limitation of this systematic review and the meta-analysis was the small number of studies included (only six diagnostic studies could be pooled; which is the maximum number of studies in any of the other subsections), with several sources of heterogeneity, such as the included lesion types, malignancy prevalence, lesion size spectra, PET acquisition and reconstruction settings, quantitation methods, and methodological quality. The overall study quality as assessed by QUADAS and QUIPS was good (Fig.2), but more specific research questions regarding PVC and more rigorous designs are needed. Apart from being a limitation, the small number of retrieved studies applying PVC in oncology is also an important finding, designating the lack of application of PVC in recent decades.

Recommendations

When applying PVC in studies investigating diagnostic accuracy, SUV thresholds should be redefined for corrected data. Also, results on test characteristics should be stratified for lesion size (using a 3cm diameter or 14mL cutoff). In

prognostication studies, partial-volume-corrected SUV may complement rather than substitute uncorrected SUV, and could be included separately in prognostic models. The impact of PVC on PERCIST classifications of response merits further investigation in prospective studies. For now we recommend that lesions ≤10mm in diameter should not be included in quantitative analyses until novel PVC methods proven to be efficacious for these lesions are available. To demonstrate dependency of results on the applied PVC methodology, studies comparing multiple methods in the same sample of patients are highly recommended. Both functional and volumetric semi-quantitative PET metrics should be investigated simultaneously, including SUVs, MATV, and their product TLG (see for example refs. (*27,38,40,42,43*)). Also, when PET is used for therapeutic dosimetry applications, e.g. for nuclide radiotherapy, PVC will likely improve estimates of tracer or radionuclide uptake, and thereby improve estimates of tumour radiation dose.

Conclusion

The accumulated evidence to date does not support routine application of PVC in standard clinical PET studies. In meta-analysis of quantitative diagnostic PET studies, PVC did not increase diagnostic accuracy. Limitations of published studies pertain to lack of analysis stratified for size, limited exploration of the impact of alternative (SUV) thresholds of test positivity on diagnostic accuracy measures and heterogeneity in applied PVC methodologies. For accurate and reproducible results on tumour uptake quantification, consensus on the preferred tumour delineation and PVC methodologies needs to be reached. Partial-volume corrected metrics should be used as adjuncts to, but not yet replace, uncorrected data.

References

1. 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.
2. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of ^{18}F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. ***J Nucl Med.*** 2006;47:1059-1066.
3. 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.
4. Hoekstra CJ, Paglianiti I, Hoekstra OS, 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;27:731-743.
5. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. ***J Nucl Med.*** 2007;48:932-945.
6. Hoetjes NJ, van Velden FH, Hoekstra OS, et al. Partial volume correction strategies for quantitative FDG PET in oncology. ***Eur J Nucl Med Mol Imaging.*** 2010;37:1679-1687.
7. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. ***Phys Med Biol.*** 2012;57:R119-159.
8. Cysouw MC, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial-Volume Correction in Oncological PET/CT Studies. ***J Nucl Med.*** 2016;57:1642-1649.
9. Rahmim A, Qi J, Sossi V. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. ***Med Phys.*** 2013;40:064301.
10. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. ***J Nucl Med.*** 2007;48:802-810.
11. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. ***Ann Intern Med.*** 2011;155:529-536.
12. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. ***Ann Intern Med.*** 2013;158:280-286.
13. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. ***J Clin Epidemiol.*** 2005;58:982-990.
14. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. ***Stat Med.*** 1993;12:1293-1316.
15. Hickeson M, Yun M, Matthies A, et al. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. ***Eur J Nucl Med Mol Imaging.*** 2002;29:1639-1647.
16. Mamede M, Higashi T, Kitaichi M, et al. [^{18}F]FDG uptake and PCNA, Glut-1, and Hexokinase-II expressions in cancers and inflammatory lesions of the lung. ***Neoplasia.*** 2005;7:369-379.
17. Alkhawaldeh K, Bural G, Kumar R, Alavi A. Impact of dual-time-point (18)F-FDG PET imaging and partial volume correction in the assessment of solitary pulmonary nodules. ***Eur J Nucl Med Mol Imaging.*** 2008;35:246-252.
18. Degirmenci B, Wilson D, Laymon CM, et al. Standardized uptake value-based evaluations of solitary pulmonary nodules using F-18 fluorodeoxyglucose-PET/computed tomography. ***Nucl Med Commun.*** 2008;29:614-622.
19. Menda Y, Bushnell DL, Madsen MT, McLaughlin K, Kahn D, Kernstine KH. Evaluation of various corrections to the standardized uptake value for diagnosis of pulmonary malignancy. ***Nucl Med Commun.*** 2001;22:1077-1081.
20. Kim SC, Machac J, Krynyckyi BR, et al. Fluoro-deoxy-glucose positron emission tomography for evaluation of indeterminate lung nodules: assigning a probability of malignancy may be preferable to binary readings. ***Ann Nucl Med.*** 2008;22:165-170.

21. Avril N, Bense S, Ziegler SI, et al. Breast imaging with fluorine-18-FDG PET: quantitative image analysis. ***J Nucl Med.*** 1997;38:1186-1191.

22. Tsujikawa T, Otsuka H, Morita N, et al. Does partial volume corrected maximum SUV based on count recovery coefficient in 3D-PET/CT correlate with clinical aggressiveness of non-Hodgkin's lymphoma? ***Ann Nucl Med.*** 2008;22:23-30.

23. Cremerius U, Fabry U, Neuerburg J, Zimny M, Osieka R, Buell U. Positron emission tomography with ^{18}F-FDG to detect residual disease after therapy for malignant lymphoma. ***Nucl Med Commun.*** 1998;19:1055-1063.

24. Tsujikawa T, Tsuchida T, Imamura Y, et al. Kikuchi-Fujimoto disease: PET/CT assessment of a rare cause of cervical lymphadenopathy. ***Clin Nucl Med.*** 2011;36:661-664.

25. Vesselle H, Turcotte E, Wiens L, et al. Relationship between non-small cell lung cancer fluorodeoxyglucose uptake at positron emission tomography and surgical stage with relevance to patient prognosis. ***Clin Cancer Res.*** 2004;10:4709-4716.

26. Rousseau C, Devillers A, Campone M, et al. FDG PET evaluation of early axillary lymph node response to neoadjuvant chemotherapy in stage II and III breast cancer patients. ***Eur J Nucl Med Mol Imaging.*** 2011;38:1029-1036.

27. Fayad H, Le Pogam A, Lamare F, et al. Influence of partial volume correction in staging of head and neck squamous cell carcinoma using PET/CT. ***Q J Nucl Med Mol Imaging.*** 2014;58:319-328.

28. Sakaguchi Y, Mizoguchi N, Mitsumoto T, et al. A simple table lookup method for PET/CT partial volume correction using a point-spread function in diagnosing lymph node metastasis. ***Ann Nucl Med.*** 2010;24:585-591.

29. Wang YW, Wu CS, Chang CH, et al. Partial volume correction for equivocal retropharyngeal nodal metastases of nasopharyngeal carcinoma with fluorodeoxyglucose positron emission tomography-computed tomography. ***Journal of Medical and Biological Engineering.*** 2015;35:218-225.

30. Mathieu C, Ferrer L, Carlier T, et al. Assessment of Lymph Nodes and Prostate Status Using Early Dynamic Curves with (18)F-Choline PET/CT in Prostate Cancer. ***Front Med (Lausanne).*** 2015;2:67.

31. Lasnon C, Hicks RJ, Beauregard JM, et al. Impact of point spread function reconstruction on thoracic lymph node staging with ^{18}F-FDG PET/CT in non-small cell lung cancer. ***Clin Nucl Med.*** 2012;37:971-976.

32. Bellevre D, Blanc Fournier C, Switsers O, et al. Staging the axilla in breast cancer patients with (1) (8)F-FDG PET: how small are the metastases that we can detect with new generation clinical PET systems? ***Eur J Nucl Med Mol Imaging.*** 2014;41:1103-1112.

33. Kawashima K, Kato K, Tomabechi M, et al. Clinical evaluation of F-fluorodeoxyglucose-positron emission tomography/CT using point spread function reconstruction for nodal staging of colorectal cancer. ***Br J Radiol.*** 2016:20150938.

34. Ozawa Y, Hara M, Shibamoto Y, Tamaki T, Nishio M, Omi K. Utility of high-definition FDG-PET image reconstruction for lung cancer staging. ***Acta Radiol.*** 2013;54:916-920.

35. Um SW, Kim H, Koh WJ, et al. Prognostic Value of F-18-FDG Uptake on Positron Emission Tomography in Patients with Pathologic Stage I Non-small Cell Lung Cancer. ***Journal of Thoracic Oncology.*** 2009;4:1331-1336.

36. Ohtaka K, Hida Y, Kaga K, et al. Outcome analysis of (18)F-fluorodeoxyglucose positron-emission tomography in patients with lung cancer after partial volume correction. ***Anticancer Res.*** 2013;33:5193-5198.

37. van Heijl M, Omloo JM, van Berge Henegouwen MI, van Lanschot JJ, Sloof GW, Boellaard R. Influence of ROI definition, partial volume correction and SUV normalization on SUV-survival correlation in oesophageal cancer. ***Nucl Med Commun.*** 2010;31:652-658.

38. Hatt M, Le Pogam A, Visvikis D, Pradier O, Cheze Le Rest C. Impact of partial-volume effect correction on the predictive and prognostic value of baseline ^{18}F-FDG PET images in esophageal cancer. ***J Nucl Med.*** 2012;53:12-20.

39. Gallivanone F, Canevari C, Gianolli L, et al. A partial volume effect correction tailored for ^{18}F-FDG-PET oncological studies. ***Biomed Res Int.*** 2013;2013:780458.

40. Picchio M, Kirienko M, Mapelli P, et al. Predictive value of pre-therapy (18)F-FDG PET/CT for the outcome of (18)F-FDG PET-guided radiotherapy in patients with head and neck cancer. ***Eur J Nucl Med Mol Imaging.*** 2014;41:21-31.

41. 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.

42. Hatt M, Groheux D, Martineau A, et al. Comparison between ^{18}F-FDG PET image-derived indices for early prediction of response to neoadjuvant chemotherapy in breast cancer. ***J Nucl Med.*** 2013;54:341-349.

43. Hatt M, van Stiphout R, le Pogam A, Lammering G, Visvikis D, Lambin P. Early prediction of pathological response in locally advanced rectal cancer based on sequential 18F-FDG PET. ***Acta Oncol.*** 2013;52:619-626.

44. Maisonobe JA, Garcia CA, Necib H, et al. Comparison of PET metabolic indices for the early assessment of tumour response in metastatic colorectal cancer patients treated by polychemotherapy. ***Eur J Nucl Med Mol Imaging.*** 2013;40:166-174.

45. Anouan KJ, Lelandais B, Edet-Sanson A, et al. ^{18}F-FDG-PET Partial volume effect correction using a modified recovery coefficient approach based on functional volume and local contrast: Physical validation and clinical feasibility in oncology. ***Q J Nucl Med Mol Imaging.*** 2015.

46. Lasnon C, Desmonts C, Quak E, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. ***Eur J Nucl Med Mol Imaging.*** 2013;40:985-996.

47. Schouten CS, Hakim S, Boellaard R, et al. Interaction of quantitative F-FDG-PET-CT imaging parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. ***Head Neck.*** 2014.

48. Vesselle H, Freeman JD, Wiens L, et al. Fluorodeoxyglucose uptake of primary non-small cell lung cancer at positron emission tomography: new contrary data on prognostic role. ***Clin Cancer Res.*** 2007;13:3255-3263.

49. Stefano A, Gallivanone F, Messa C, Gilardi MC, Gastiglioni I. Metabolic impact of partial volume correction of [^{18}F]FDG PET-CT oncological studies on the assessment of tumor response to treatment. ***Q J Nucl Med Mol Imaging.*** 2014;58:413-423.

50. Rousset O, Rahmim A, Alavi A, Zaidi H. Partial Volume Correction Strategies in PET. ***PET Clinics.*** 2007;2:235-249.

51. Bettinardi V, Castiglioni I, De Bernardi E, Gilardi MC. PET quantification: Strategies for partial volume correction. ***Clinical and Translational Imaging.*** 2014;2:199-218.

52. Cheebsumon P, Boellaard R, de Ruysscher D, et al. Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. ***EJNMMI Res.*** 2012;2:56.

53. Guvenis A, Koc A. Optimising delineation accuracy of tumours in PET for radiotherapy planning using blind deconvolution. ***Radiat Prot Dosimetry.*** 2015;165:495-498.

54. Wallsten E, Axelsson J, Sundstrom T, Riklund K, Larsson A. Subcentimeter tumor lesion delineation for high-resolution 18F-FDG PET images: optimizing correction for partial-volume effects. ***J Nucl Med Technol.*** 2013;41:85-91.

Supplemental files

Supplemental Table 7.1: Search strategy in PubMed May 9, 2016 (read from bottom-up).

Set	Search terms	Result
#4	#1 AND #2 AND #3	322
#3	partial-volume effect*[tiab] OR partial volume effect*[tiab] OR partial-volume correction*[tiab] OR partial volume correction*[tiab] OR deconvolution*[tiab] OR recovery coefficient*[tiab] OR point spread function reconstruction*[tiab] OR PSF reconstruction*[tiab] OR PSF-reconstruction*[tiab] OR point-spread-function reconstruction*[tiab] OR resolution model*[tiab] OR resolution recover*[tiab] OR resolution model*[tiab] OR high definition*[tiab] OR high definition*[tiab] OR PSF model*[tiab] OR point-spread-function model*[tiab] OR point spread function model*[tiab] OR HD reconstruction*[tiab]	10752
#2	"Neoplasms"[Mesh] OR oncolog*[tiab] OR cancer*[tiab] OR neoplasm*[tiab] OR tumour*[tiab] OR tumor*[tiab] OR carcinoma*[tiab] OR malignan*[tiab] OR metasta*[tiab] OR lesion*[tiab] OR lymphoma*[tiab]	3931706
#1	"Positron-Emission Tomography"[Mesh] OR positron emission tomograph*[tiab] OR PET[tiab] OR PET/CT[tiab] OR PET-CT[tiab] OR FDG-PET*[tiab] OR 18F-FDG-PET*[tiab]	90587

Supplemental Table 7.2: Search strategy in Embase.com May 9 , 2016 (read from bottom-up).

Set	Search terms	Result
#5	#4 NOT 'conference abstract'/it	335
#4	#1 AND #2 AND #3	614
#3	'neoplasm'/exp OR 'oncology'/exp OR oncolog*:ab,ti OR cancer*:ab,ti OR neoplasm*:ab,ti OR tumour*:ab,ti OR tumor*:ab,ti OR carcinoma*:ab,ti OR malignan*:ab,ti OR metasta*:ab,ti OR lesion*:ab,ti OR lymphoma*:ab,ti	5026012
#2	(partial NEXT/1 volume NEXT/1 effect*):ab,ti OR (partial NEXT/1 volume NEXT/1 correction*):ab,ti OR deconvolution*:ab,ti OR (recovery NEXT/1 coefficient*):ab,ti OR (point NEXT/1 spread NEXT/1 function NEXT/1 reconstruction*):ab,ti OR (psf NEXT/1 reconstruction*):ab,ti OR (resolution NEXT/1 model*):ab,ti OR (resolution NEXT/1 recover*):ab,ti OR (high NEXT/1 definition*):ab,ti OR (psf NEXT/1 model*):ab,ti OR (point NEXT/1 spread NEXT/1 function NEXT/1 model*):ab,ti OR (hd NEXT/1 reconstruction*):ab,ti	13014
#1	'positron emission tomography'/exp OR (positron NEXT/1 emission NEXT/1 tomograph*):ab,ti OR pet:ab,ti OR (pet NEXT/1 ct):ab,ti OR (fdg NEXT/1 pet*):ab,ti	157749

Supplemental Table 7.3: Technical PET details from included studies. Reconstruction settings and resolution data pertain to the uncorrected images.

Ref.	SUV type	Delineation method	PVC method	Scanner	Reconstruction	system FWHM	filter FWHM	final FWHM	voxel-size
[21]	mean	manual	RC	PET	FBP	-	-	-	-
[23]	mean	A50%max	RC	PET	MLA	-	-	-	7*7*7
[19]	max	manual	RC	PET	FBP	6.5	10 (H)	-	-
[15]	max mean (pvc)	manual	CT-volume correction	PET	RAMLA	-	-	-	4*?*?
[16]	max	90%max	RC	PET	OSEM (16s3i)	-	-	-	-
[17]	max	manual	RC	PET	OSEM (8s4i)	-	-	-	-
[18]	max mean	manual	RC	PET/CT	OSEM (8s2i)	-	-	-	-
[22]	max	manual	RC	PET/CT	OSEM (12s4i)	4.5	8 (G)	9.2	2*?*?
[20]	max	manual	RC	PET	OSEM (28s2i)	4.5	8 (G)	-	-
[24]	max	manual	RC	PET/CT	OSEM (14s2i)	-	-	-	-
[25]	max	manual	RC	PET	FBP	-	12 (H)	-	-
[28]	max	manual	RC	PET	OSEM (28s2i)	-	6	-	5.5*5.5*3.3
[26]	max	manual	GTM	PET	OSEM	-	-	-	-
[31]	n.a.	n.a.	PSF reconstruction*	PET/CT	OSEM (8s4i)	-	5 (G)	-	4.1*4.1*4.1
[34]	n.a.	n.a.	PSF reconstruction*	PET/CT	OSEM FORE	-	-	4.2	2*?-?
[32]	n.a.	n.a.	PSF reconstruction*	PET/CT	OSEM (8s4i)	-	5 (G)	6.2 to 6.5	4.1*4.1*5
[27]	max mean	manual	IDC with denoising	PET/CT	RAMLA OSEM (21s2i)	-	-	-	4*4*4 4.7*4.7*3.3
[29]	max	manual	RC	PET/CT	OSEM (15s2i)	-	7	-	5.5*5.5*3.3
[30]	mean	system-specific contrast-oriented algorithm	RC	PET/CT	OSEM (28s2i) OSEM-PSF (21s3i)	-	5.5 2	-	-
[33]	n.a.	n.a.	PSF reconstruction*	PET/CT	OSEM (16s2i)	-	6 (G)	-	4.7*4.7*4.7
[35]	max	scanner-implemented	RC	PET/CT	-	-	-	-	-
[37]	mean	A41%max	Mask-based PVC	PET/CT	OSEM (16s2i)	-	5 (G)	-	-

[38]	max mean peak	fuzzy locally adaptive Bayesian algorithm	IDC with denoising	PET/CT	RAMLA	-	5 (G)	-	4*4*4
[36]	max	-	RC	PET	OSEM FORE (32s1i)	-	-	-	-
[39]	-	60%max	RC	PET/CT	OSEM	-	-	-	4.7*4.7*3.3
[40]	mean	40%max	RC	PET/CT	OSEM (28s2i)	-	-	-	-
[41]	n.a.	manual	RC	PET	-	-	-	10 to 12	-
[42]	max mean peak	fuzzy locally adaptive Bayesian algorithm	IDC with denoising	PET/CT	RAMLA	-	-	-	4*4*4
[43]	max mean	fuzzy locally adaptive Bayesian algorithm	IDC with denoising	PET/CT	OSEM FORE (8s4i)	6	-	-	4.1*4.1*3
[44]	max mean (pvc) peak	system-specific contrast-oriented algorithm	RC + IDC	PET/CT	OSEM (28s2i)	-	5.5 (G)	-	3.9*3.9*4.3
[45]	mean	adaptive iterative thresholding algorithm	RC	PET/CT	OSEM (8s4i)	-	5 (G)	-	4*4*2

RC = recovery coefficient; PSF = point spread function; CT = computed tomography; IDC = iterative deconvolution; GTM = geometric transfer matrix; FBP = filtered backprojection; MLA = maximum-likelihood algorithm; RAMLA = row-action maximum-likelihood algorithm; OSEM = ordered subset expectation maximization; FORE = Fourier rebinning; H = Hanning filter; G = Gauss filter; FWHM = full width at half maximum; n.a. = not applicable.


QI-1-PET

Chapter 8

Prognostic Value of [^{18}F]-Fluoromethylcholine PET-CT before Stereotactic Body Radiation Therapy for Oligometastatic Prostate Cancer

M.C.F. Cysouw, E.W. Bouman-Wammes, O.S. Hoekstra,
A.J.M. van den Eertwegh, M. Piet, R.J.A. van Moorselaar, R. Boellaard,
M. Dahele, D.E. Oprea-Lager

Int J Radiat Oncol Biol Phys. 2018 Jun 1;101(2):406-410.

Abstract

Purpose: Treating oligometastases detected by [^{18}F]-fluoromethylcholine positron-emission tomography/computed tomography (PET/CT) with stereotactic body radiotherapy (SBRT) may postpone initiation of androgen-deprivation therapy, and possibly prolong progression-free survival (PFS). However, better prognostic factors are needed to improve patient selection. We investigated the predictive value of [^{18}F]-fluoromethylcholine-PET/CT-derived parameters on PFS in oligometastatic prostate cancer patients treated with SBRT.

Methods and Materials: In [^{18}F]-fluoromethylcholine PET/CT scans of forty consecutive patients with ≤4 metachronous metastases treated with SBRT we retrospectively measured the number of metastases, standardized uptake values (SUVmean, SUVmax, SUVpeak), metabolically active tumour volume (MATV), and total lesion choline uptake (TLCU). Partial-volume correction was applied using the iterative deconvolution Lucy-Richardson algorithm.

Results: 37 lymph node and 13 bone metastases were treated with SBRT. 33 patients (82.5%) had 1 lesion, 4 (10%) had 2 lesions, and 3 (7.5%) had 3 lesions. After a median follow-up of 32.6 months (IQR 35.5), the median PFS was 11.5 (95%CI 8.4-14.6). Having more than a single lesion was a significant prognostic factor (HR=2.74; p=0.03), and there was a trend in risk of progression for large MATV (HR=1.86; p=0.10). No SUV or TLCU was significantly predictive for PFS, regardless of partial-volume correction. All PET semi-quantitative parameters were significantly correlated with each other (p≤0.013).

Conclusion: The number of choline-avid metastases was a significant prognostic factor for progression after [^{18}F]-fluormethylcholine PET/CT-guided SBRT for recurrent oligometastatic prostate cancer, and there seemed to be a trend in risk of progression for patients with large MATVs. The lesional level of [^{18}F]-fluoromethylcholine uptake was not prognostic for progression.

Introduction

For hormone-sensitive metastatic prostate cancer first-line treatment commonly consists of androgen deprivation therapy (ADT) with or without chemotherapy (*1,2*). While effective in deferring disease progression, side-effects can compromise quality of life (*3*). There is growing interest in local therapy for oligometastatic disease (e.g. stereotactic body radiotherapy, SBRT) with the aim of achieving prolonged progression-free survival (PFS) and postponing or avoiding initiation of ADT (*4-9*). However, reported PFS and ADT-free survival rates are variable and often limited (*5*). Therefore, prognostic or predictive biomarkers are urgently needed.

[^{18}F]-fluoromethylcholine PET/CT may provide prognostic information in prostate cancer patients (*10-12*). We therefore comprehensively explored the potential prognostic value of semi-quantitative [^{18}F]-fluoromethylcholine PET-derived metrics in SBRT-treated oligometastatic patients.

Materials and Methods

Patients

We included 40 consecutive patients with PSA relapse after primary local treatment for prostate adenocarcinoma, with oligometastases (≤4 lesions) detected at [^{18}F]-fluoromethylcholine PET/CT, treated with SBRT between January 2009 and December 2015, without ADT during SBRT. Clinical results have been reported [9]. Standard dose-fractionation schedules were 5x7Gy and 3x10Gy. Biochemical progression after SBRT was defined as PSA rising ≥25% or ≥2.0 ng/ml above baseline or post-SBRT nadir, documented on two consecutive measurements. Our institutional medical ethics committee waived the need for informed consent.

PET/CT

Forty (±9) minutes after injection of 336±68MBq [^{18}F]-fluoromethylcholine (4MBq/kg) patients underwent whole body [^{18}F]-fluoromethylcholine PET/CT (n=37 Gemini/Ingenuity [Philips Healthcare], n=3 Biograph [Siemens]; EARL-accredited), with image reconstruction compliant with EANM standards (*13*). We analysed lesions on volume-of-interest basis, yielding mean standardized-uptake-value (SUV, equation 1) with and without correction for partial-volume effects by

Lucy-Richardson iterative deconvolution (SUV_{mean}, SUV_{pvc}, resp.), SUV_{max}, SUV_{peak}, metabolically active tumour volume (MATV), and total lesion choline uptake ($TLCU_{pvc}$). Within patients, we calculated both highest and total MATV and TLCU in case of multiple lesions (indicated as suffix). Tabular data of uncorrected SUV_{mean}, SUV_{max}, SUV_{peak} and uncorrected TLCU are presented as supplemental data, available at https://www.redjournal.org/.

$$\text{SUV} = \frac{Activity\ Concentration\ (^{Bq}/_{mL})}{Injected\ Dose\ (Bq)/\ bodyweight\ (g)} \qquad \text{Eq.1}$$

Statistical analysis

Survival analysis was performed using the Kaplan-Meier method, univariate Cox-regression, and log-rank test (variables were dichotomized). Besides quantitative PET-parameters, we analysed several other potential prognostic factors (number of choline-avid lesions, lesion type, prior ADT treatment, Gleason score, PSA at metastasis, PSA nadir). We assessed correlations with Spearman rank, and differences between groups with Mann-Whitney U test, setting significance levels at p=0.05. Analyses were performed using SPSS (22.0;IBM).

Results

Table 8.1 presents baseline characteristics. Thirty-three patients had a solitary metastasis on PET/CT (n=25 lymph node, n=8 bone); 4 had two (nodal) metastases, and 3 presented with three metastases, (n=1 all lymph nodes, n=1 all bone, n=1 lymph node and bone). Median follow-up and PFS after SBRT were 32.6 (IQR 14.7-50.3) and 11.5 (95%CI 8.4-14.6) months, respectively. Compared to patients without a post-SBRT PSA nadir, patients with a PSA nadir had a HR for progression of 0.23 (p<0.001).

Median (IQR) $MATV_{highest,}$ $MATV_{total,}$ $SUV_{mean\text{-}pvc}$, $TLCU_{highest\text{-}pvc}$, and $TLCU_{total\text{-}pvc}$ were 2.3 (1.5-3.7), 2.6 (1.6-4.3), 4.3 (3.2-5.8), 9.5 (5.5-17.5), and 10.5 (5.9-22.7), respectively. The lesion-based MATV and $TLCU_{pvc}$ were significantly higher for bone compared to lymph nodes (Table 8.2). The presence of >1 metastasis on [^{18}F]-fluoromethylcholine PET/CT had a HR of 2.74 (p=0.03) for progression. All PET semi-quantitative parameters were significantly correlated (p≤0.013). There was no PET parameter threshold on ROC analysis (supplemental data). No semi-quantitative PET/CT parameter predicted PFS, but a trend in risk of progression for large MATV was noted (HR=1.86; p=0.10) (Table 8.3).

We assessed potential associations of PET quantitative parameters with clinical data (supplemental data). Several SUV and TLCU variants were significantly higher for patients with PSA levels above median (3.75 ng/ml) at PET/CT (p=0.01-0.04). For Gleason score and pre-treatment with ADT we found no significant association with quantitative PET parameters.

Table 8.1: Patient characteristics.

Characteristic	
Age at metastases	
Mean±SD	67±6.7 years
Time from diagnosis to metastases	
Median (IQR)	46.2 (12.0-81.8) months
Gleason score	
5	2 (5%)
6	5 (13%)
7	16 (41%)
8	10 (26%)
9	6 (16%)
TNM-stage at diagnosis	
Stage 2	20 (51%)
Stage 3	16 (41%)
Stage 4	3 (8%)
Primary treatment	
Surgery	25 (63%)
Surgery + radiotherapy	4 (10%)
Radiotherapy	5 (13%)
Radiotherapy + hormone therapy	3 (8%)
Brachytherapy	3 (8%)
Lymph node dissection	
Yes	13 (33%)
No	27 (68%)
PSA at metastases	
Median (IQR)	3.75 (2.43-6.80) ng/ml
Type of metastases	
Bone	9 (23%)
Lymph node	30 (75%)
Bone + Lymph node	1 (3%)

Table 8.2: Lesion-based values of PET parameters. Data are presented as median with IQR.

	All lesions (n=50)	Lymph node (n=37)	Bone (n=13)	p-value[a]
MATV	2.21 (1.58-3.28)	1.92 (1.40-2.74)	3.46 (2.68-8.96)	<0.001
$SUV_{mean\text{-}pvc}$	3.93 (3.01-5.34)	3.72 (3.12-5.25)	4.45 (2.80-5.55)	0.816
$TLCU_{pvc}$	8.39 (5.24-15.43)	7.43 (4.77-13.06)	12.68 (7.52-50.33)	0.009

[a]Lymph node vs. bone

Table 8.3: Results from survival analyses. Significant differences in bold.

		PFS (95%CI)	HR (95%CI)	p-value[b]
PET parameters:				
$MATV_{highest}$	< 2.3 mL ≥ 2.3 mL	14.3 (2.8-25.8) 9.5 (8.5-10.5)	1.86 (0.87-3.97)	0.103
$MATV_{total}$	< 2.6 mL ≥ 2.6 mL	12.1 (4.4-19.9) 9.5 (6.8-12.3)	1.66 (0.78-3.51)	0.181
$SUV_{mean\text{-}pvc}$	< 4.3 ≥ 4.3	11.5 (6.8-16.2) 9.5 (5.8-13.3)	0.99 (0.47-2.00)	0.969
$TLCU_{highest\text{-}pvc}$	< 9.5 ≥ 9.5	9.4 (1.3-17.4) 11.5 (8.6-14.4)	0.98 (0.48-1.99)	0.944
$TLCU_{total\text{-}pvc}$	< 10.5 ≥ 10.5	9.4 (1.2-17.6) 11.5 (8.6-14.4)	0.98 (0.48-2.00)	0.955
number of metastases	=1 >1	11.5 (8.1-15.0) 6.5 (2.6-10.4)	2.74 (1.06-7.11)	**0.031**
lesion type[a]	lymph node bone	11.5 (7.2-15.8) 9.5 (6.3-12.7)	1.37 (0.61-3.09)	0.436
Gleason score	≤ 7 > 7	13.4 (9.6-17.2) 8.5 (3.9-13.2)	1.81 (0.86-3.81)	0.112
PSA at PET/CT	< 3.75 ng/ml ≥ 3.75 ng/ml	9.4 (4.7-14.1) 11.5 (8.7-14.4)	0.86 (0.42-1.74)	0.670
PSA nadir after SBRT	no yes	3.5 (2.6-4.5) 12.1 (9.8-14.5)	0.23 (0.11-0.50)	**<0.001**
pre-SBRT ADT	no yes	11.5 (8.7-14.3) 6.5 (0.0-15.0)	0.85 (0.29-2.43)	0.753

[a]lesion with highest uptake in case of >1 lesion; [b]from log rank-test.

Discussion

The number of metastases at [^{18}F]-fluoromethylcholine PET/CT predicted progression after SBRT for recurrent hormone-sensitive oligometastatic prostate cancer, but other (quantitative) PET measures did not.

In patients receiving ADT for treatment of hormone-sensitive prostate cancer recurrence a similar association between number of metastases and survival has been reported, and at biochemical recurrence after primary therapy the disease burden (number of [^{18}F]-fluoromethylcholine PET/CT positive lesions) was an independent prognostic factor for developing castration-resistant disease (*14-16*).

[^{18}F]/[^{11}C]-labelled choline PET/CT is one of the preferred methods for restaging at biochemical progression (*1,17*), with a prevalence of oligometastatic disease of 40-91% (*14,18-21*). However, other PET tracers (e.g. PSMA-ligands) may detect more metastases, especially at low PSA values (*22,23*). Schwenck et al. observed that in 27% of patients with oligometastases on choline PET/CT, PSMA

detected more lesions, resulting in their re-classification as non-oligometastatic (*23*). In our study, 11/40 patients (27.5%) had no initial PSA response after SBRT (presence of PSA nadir: HR 0.23; p<0.001), which may reflect suboptimal metastasis detection rates for [^{18}F]- fluoromethylcholine PET/CT in this setting. However, whether PSMA-guided management will improve patient outcomes remains to be shown.

Limitations of this study are its retrospective nature and relatively small sample size. A strength is its clinical relevance - and PET/CT imaging, SBRT, and PSA follow-up (tested at least every 3 months) were all performed according to institutional protocols used in clinical practice.

The role of post-SBRT ADT or the relative benefits of SBRT to only PET positive lesion(s) versus larger volume irradiation (e.g. the involved lymph node chain) remains to be investigated. Further research is also warranted to characterize the time-course of activity in treated lesions on post-SBRT [^{18}F]-fluoromethylcholine PET/CT scans, and to avoid over-diagnosing local failure (Figure 8.1).

In conclusion, the number of detected oligometastases seems prognostic. Additional data, including from prospective randomised controlled trials (e.g. NTC01558427 and NTC02680587), will hopefully determine whether the advantage of SBRT for oligometastatic prostate cancer is limited to deferring systemic therapy and identify additional patient or tumour characteristics predictive for disease progression.

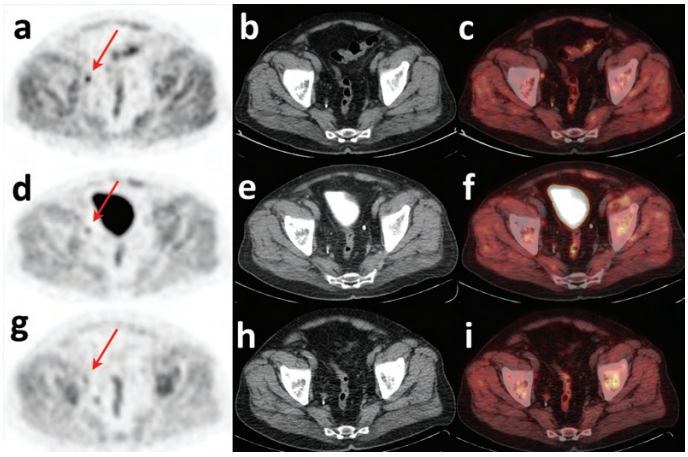


Figure 8.1: [^{18}F]-fluoromethylcholine PET/CT images of a 74-year-old patient with persisting, but decreasing, activity in a right para-iliac lymph node after SBRT. PET, low-dose CT, and fused PET/CT images before (A-C) SBRT, 20 months (D-F) and 32 months (G-I) after SBRT, respectively.

References

1. Heidenreich A, Bastian PJ, Bellmunt J, et al. EAU guidelines on prostate cancer. Part II: Treatment of advanced, relapsing, and castration-resistant prostate cancer. ***Eur Urol.*** 2014;65:467-479.
2. James ND, Sydes MR, Clarke NW, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. ***Lancet.*** 2016;387:1163-1177.
3. Dacal K, Sereika SM, Greenspan SL. Quality of life in prostate cancer patients taking androgen deprivation therapy. ***J Am Geriatr Soc.*** 2006;54:85-90.
4. Tosoian JJ, Gorin MA, Ross AE, Pienta KJ, Tran PT, Schaeffer EM. Oligometastatic prostate cancer: definitions, clinical outcomes, and treatment considerations. ***Nat Rev Urol.*** 2017;14:15-25.
5. Ost P, Bossi A, Decaestecker K, et al. Metastasis-directed therapy of regional and distant recurrences after curative treatment of prostate cancer: a systematic review of the literature. ***Eur Urol.*** 2015;67:852-863.
6. Ost P, Jereczek-Fossa BA, As NV, et al. Progression-free Survival Following Stereotactic Body Radiotherapy for Oligometastatic Prostate Cancer Treatment-naive Recurrence: A Multi-institutional Analysis. ***Eur Urol.*** 2016;69:9-12.
7. Pasqualetti F, Panichi M, Sainato A, et al. [(18)F]Choline PET/CT and stereotactic body radiotherapy on treatment decision making of oligometastatic prostate cancer patients: preliminary results. ***Radiat Oncol.*** 2016;11:9.
8. Ploussard G, Almeras C, Briganti A, et al. Management of Node Only Recurrence after Primary Local Treatment for Prostate Cancer: A Systematic Review of the Literature. ***J Urol.*** 2015;194:983-988.
9. Bouman-Wammes EW, van Dodewaard-De Jong JM, Dahele M, et al. Benefits of Using Stereotactic Body Radiotherapy in Patients With Metachronous Oligometastases of Hormone-Sensitive Prostate Cancer Detected by [18F]fluoromethylcholine PET/CT. ***Clin Genitourin Cancer.*** 2017;15:e773-e782.
10. Schaefferkoetter JD, Wang Z, Stephenson MC, et al. Quantitative (18)F-fluorocholine positron emission tomography for prostate cancer: correlation between kinetic parameters and Gleason scoring. ***EJNMMI Res.*** 2017;7:25.
11. Colombie M, Campion L, Bailly C, et al. Prognostic value of metabolic parameters and clinical impact of (1)(8)F-fluorocholine PET/CT in biochemical recurrent prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2015;42:1784-1793.
12. Kwee SA, Lim J, Watanabe A, Kromer-Baker K, Coel MN. Prognosis Related to Metastatic Burden Measured by (1)(8)F-Fluorocholine PET/CT in Castration-Resistant Prostate Cancer. ***J Nucl Med.*** 2014;55:905-910.
13. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
14. De Bruycker A, Lambert B, Claeys T, et al. Prevalence and prognosis of low-volume, oligorecurrent, hormone-sensitive prostate cancer amenable to lesion ablative therapy. ***BJU Int.*** 2017.
15. Ost P, Decaestecker K, Lambert B, et al. Prognostic factors influencing prostate cancer-specific survival in non-castrate patients with metastatic prostate cancer. ***Prostate.*** 2014;74:297-305.
16. Schweizer MT, Zhou XC, Wang H, et al. Metastasis-free survival is associated with overall survival in men with PSA-recurrent prostate cancer treated with deferred androgen deprivation therapy. ***Ann Oncol.*** 2013;24:2881-2886.
17. Umbehr MH, Muntener M, Hany T, Sulser T, Bachmann LM. The role of 11C-choline and ^{18}F-fluorocholine positron emission tomography (PET) and PET/CT in prostate cancer: a systematic review and meta-analysis. ***Eur Urol.*** 2013;64:106-117.

18. Ceci F, Castellucci P, Graziani T, et al. 11C-choline PET/CT detects the site of relapse in the majority of prostate cancer patients showing biochemical recurrence after EBRT. ***Eur J Nucl Med Mol Imaging.*** 2014;41:878-886.

19. Graziani T, Ceci F, Castellucci P, et al. (11)C-Choline PET/CT for restaging prostate cancer. Results from 4,426 scans in a single-centre patient series. ***Eur J Nucl Med Mol Imaging.*** 2016;43:1971-1979.

20. Parker WP, Davis BJ, Park SS, et al. Identification of Site-specific Recurrence Following Primary Radiation Therapy for Prostate Cancer Using C-11 Choline Positron Emission Tomography/Computed Tomography: A Nomogram for Predicting Extrapelvic Disease. ***Eur Urol.*** 2017;71:340-348.

21. Sobol I, Zaid HB, Haloi R, et al. Contemporary Mapping of Post-Prostatectomy Prostate Cancer Relapse with 11C-Choline Positron Emission Tomography and Multiparametric Magnetic Resonance Imaging. ***J Urol.*** 2017;197:129-134.

22. Afshar-Oromieh A, Zechmann CM, Malcher A, et al. Comparison of PET imaging with a (68)Ga-labelled PSMA ligand and (18)F-choline-based PET/CT for the diagnosis of recurrent prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2014;41:11-20.

23. Schwenck J, Rempp H, Reischl G, et al. Comparison of 68Ga-labelled PSMA-11 and 11C-choline in the detection of prostate cancer metastases by PET/CT. ***Eur J Nucl Med Mol Imaging.*** 2017;44:92-101.

Supplemental files

Supplemental Table 8.1: Patient-based values of PET parameters. Data is presented as median with IQR.

SUV_{mean}	2.9 (2.2-3.7)
SUV_{max}	4.9 (3.4-6.7)
SUV_{peak}	3.1 (2.2-4.7)
$TLCU_{highest}$	6.6 (3.6-12.1)
$TLCU_{total}$	7.1 (4.0-14.8)

Supplemental Table 8.2: Lesion-based values of PET parameters. Data is presented as median with IQR.

	All lesions (n=50)	**Lymph node (n=37)**	**Bone (n=13)**	**p-value***
SUV_{mean}	2.75 (2.10-3.50)	2.55 (2.10-3.31)	3.39 (2.11-4.28)	0.237
SUV_{max}	4.59 (3.07-5.88)	4.50 (3.02-5.61)	5.16 (3.30-7.04)	0.347
SUV_{peak}	2.79 (2.12-4.05)	2.63 (1.99-3.51)	3.89 (2.25-5.43)	0.099
TLCU	5.64 (3.51-10.90)	5.07 (3.29-9.33)	10.09 (5.54-41.79)	**0.004**

*Lymph node vs. bone

Supplemental Table 8.3: Results from ROC analysis at median PFS (11.5 months).

	11.5 month PFS		
	AUC	p	cut-off
$SUV_{mean-pvc}$	0.41	0.34	4.2
SUV_{mean}	0.41	0.34	7.4
SUV_{max}	0.42	0.40	4.6
SUV_{peak}	0.44	0.50	1.7
$MATV_{highest}$	0.51	0.92	1.8
$MATV_{total}$	0.56	0.50	1.8
$TLCU_{highest-pvc}$	0.44	0.51	4.6
$TLCU_{total-pvc}$	0.47	0.77	4.6
$TLCU_{highest}$	0.45	0.57	3.0
$TLCU_{total}$	0.49	0.87	3.0

Supplemental Table 8.4: Results from survival analyses.

		PFS (95%CI)	**HR (95%CI)**	**p-value#**
SUV_{mean}	< 2.9	9.4 (2.1-16.6)	0.87 (0.43-1.76)	0.694
	≥ 2.9	11.5 (7.9-15.2)		
SUV_{max}	< 4.9	9.4 (2.1-16.6)	0.88 (0.43-1.79)	0.722
	≥ 4.9	11.5(7.9-15.2)		
SUV_{peak}	< 3.1	9.4 (2.1-16.6)	0.89 (0.44-1.81)	0.751
	≥ 3.1	11.5 (7.9-15.2)		
$TLCU_{highest}$	< 6.6	9.4 (1.2-17.6)	1.00 (0.49-2.03)	0.990
	≥ 6.6	11.5 (8.6-14.4)		
$TLCU_{total}$	< 7.1	9.4 (1.2-17.6)	1.00 (0.49-2.03)	0.990
	≥ 7.1	11.5 (8.6-14.4)		

#from log-rank

Supplemental Table 8.5: Patient-based subgroup analysis of PET parameters. Data is presented as median with IQR. Significance level was set at p<0.05. Significant differences in bold.

	Gleason score		No. of metastases		ADT pre-SBRT		PSA at FCH-PET/CT	
	≤ 7	> 7	1	> 1	No	Yes	< 3.75	≥ 3.75
$MATV_{highest}$	2.71 (2.70)	2.14 (1.65)	2.16 (2.02)	3.46 (4.99)	2.17 (1.97)	4.94 (8.95)	2.02 (1.41)	3.12 (5.16)
p-value	0.420		0.218		0.118		0.072	
$MATV_{total}$	2.71 (3.92)	2.46 (2.31)	2.16 (2.02)	7.42 (6.46)	2.20 (2.46)	8.06 (12.59)	2.27 (1.50)	4.22 (7.67)
p-value	0.899		**0.002**		0.071		0.174	
SUV_{mean}	3.13 (2.06)	2.58 (1.21)	2.78 (1.79)	3.10 (1.43)	2.85 (1.43)	3.32 (3.46)	2.45 (1.10)	3.44 (2.41)
p-value	0.177		0.507		0.592		**0.023**	
$SUV_{mean-pvc}$	4.45 (2.31)	3.61 (2.24)	4.30 (2.84)	4.24 (2.14)	4.25 (2.55)	4.35 (3.75)	3.70 (1.72)	4.70 (2.59)
p-value	0.159		0.651		0.956		0.056	
SUV_{max}	5.16 (3.41)	3.87 (2.64)	4.97 (3.39)	4.86 (2.84)	4.79 (3.33)	5.53 (4.54)	4.32 (2.15)	5.98 (4.65)
p-value	0.114		0.754		0.868		**0.040**	
SUV_{peak}	3.10 (2.86)	2.83 (1.78)	3.05 (2.44)	3.37 (2.55)	3.06 (2.30)	4.45 (4.38)	2.52 (1.31)	4.45 (3.89)
p-value	0.263		0.363		0.517		**0.030**	
$TLCU_{highest}$	9.00 (18.51)	4.87 (6.49)	5.72 (9.07)	9.85 (37.56)	5.80 (7.44)	25.10 (48.93)	4.63 (4.92)	10.25 (36.41)
p-value	0.168		0.277		0.184		**0.015**	
$TLCU_{highest-pvc}$	12.39 (23.58)	7.99 (7.84)	9.20 (12.37)	12.68 (44.95)	9.31 (10.76)	29.85 (59.93)	7.19 (6.64)	15.09 (44.07)
p-value	0.128		0.277		0.255		**0.014**	
$TLCU_{total}$	9.00 (18.51)	6.72 (10.09)	5.72 (9.07)	18.19 (44.74)	6.72 (9.81)	26.99 (57.69)	5.41 (5.42)	11.16 (36.95)
p-value	0.420		**0.022**		0.159		**0.035**	
$TLCU_{total-pvc}$	12.39 (25.31)	9.38 (13.63)	9.20 (12.37)	24.09 (54.44)	9.53 (12.22)	32.41 (71.45)	8.87 (7.21)	16.57 (45.04)
p-value	0.329		**0.016**		0.184		**0.033**	


QI-1-PET

Chapter 9

Methodological Considerations for Response Assessment using [^{18}F]FDHT and [^{18}F]DCFPyL PET-CT in Castration-Resistant Prostate Cancer: A Clinical Illustration.

M.C.F. Cysouw, B.H.E. Jansen, M. Yaqub, J. Voortman, A.N. Vis, R.J.A. van Moorselaar, O.S. Hoekstra, R. Boellaard, D.E. Oprea-Lager

Adapted from *Mol Imaging Biol (2019).*
https://doi.org/10.1007/s11307-019-01438-y

Background

We read with interest the recently published paper by Werner et al. on the impact of prostate cancer tumour burden on uptake of [18F]DCFPyL, a 2nd generation fluorine-labeled prostate-specific membrane antigen (PSMA) ligand, in normal tissues (*1*). Such studies are essential for future use of theranostic PSMA radioligand therapies (e.g. 177Lu-PMSA), as the presence of a so-called 'sink effect' might require adaptation of therapeutic dosages to intra-patient tumour volumes. The authors performed a secondary analysis on a cohort of 50 prostate cancer patients that underwent [18F]DCFPyL PET-CT for various clinical indications, correlating PSMA-positive tumour volume with uptake values in normal tissues. They concluded that, in their cohort, PSMA-positive tumour volume did not correlate with [18F]DCFPyL uptake in normal organs, such lacrimal glands, parotid glands, submandibular glands, spleen and liver. Of all tissues examined, only left kidney uptake correlated significantly with tumour volume. We would like to compliment the authors for their thorough analysis and appropriate acknowledgement of its limitations. In this reply, we would like to address some additional methodologic aspects and supplement the analysis of Werner et al. with our own experience of [18F]DCFPyL imaging, especially those patients with larger tumour volumes.

The spectrum of the included patients by Werner et al., appears to be consisting largely of patients with low tumour burden. The low prostate cancer burden is evident from several parameters: i) a low median PSA level of 3.2 ng/mL with a maximum of 48 ng/mL, ii) a median of 3 tumour volumes-of-interest (VOI) delineated per patient with a positive PET scan, and iii) a median total PSMA-positive tumour volume of 4.8 ml with a maximum of 98.4 mL. The authors do acknowledge the low tumour burden as a limitation, but counterpoint this with the fact that patients with 'superscans' (i.e. with extensive skeletal involvement) were also included. However, only one of the included patients presented with a tumour volume above 40 mL (see Figure 1 in Werner et al. (*1*)). In order to find a statistically significant correlation between tumour volume and normal tissue PSMA-uptake, including only a few patients with relatively large tumour burden will likely not suffice. Moreover, it is not explicitly mentioned in the manuscript how many castration-resistant patients were included. Seemingly, at least 46% of the included patients underwent [18F]DCFPyL PET-CT for clinical indications in the hormone-sensitive setting, whereas 177Lu-PMSA therapy is primarily indicated for metastatic castration-resistant (mCRPC) patients (*2*). Gaertner and colleagues

did find a significant sink effect of large tumour burden in [^{68}Ga]PSMA PET, including a population that is more representative of patients currently receiving ^{177}Lu-PMSA radioligand therapy (*3*). To compare, their included patients had a mean PSA of 188 ng/mL (ranging up to 2860 ng/mL), and for part of the patients evaluation for ^{177}Lu-PSMA therapy was the indication of [^{68}Ga]PSMA PET-CT. Unfortunately, tumour PSMA burden was assessed visually rather than quantitatively (*3*).

The sink effect

Recently, we reported on a full pharmacokinetic analysis of [^{18}F]DCFPyL on PET-CT, from which we concluded that SUV is not a valid parameter for quantifying [^{18}F]DCFPyL uptake in prostate cancer lesions (*4*). In using SUV, one assumes that input functions (e.g. the time activity curves of [^{18}F]DCFPyL in blood) are equal in shape and size across patients and scale proportionally by injected activity over weight. We concluded that this assumption was invalid for [^{18}F]DCFPyL since two patients with large tumour volumes had lower activity concentrations of [^{18}F] DCFPyL in their blood (see Supplemental Figure 3 in Jansen et al. (*4*)). Moreover, liver SUVmean of the patient with a superscan was below previously reported reference values (1.68; reference range 3.31–8.53) (*5*). The consequence was a poor correlation between SUV and reference pharmacokinetic parameter Ki (R^2 0.47-0.60; see Figure 4 in (*4*)). However, when normalizing tumour uptake (Bq/mL) to activity concentrations in blood (Bq/mL), we derived a Tumour-to-Blood ratio that correlated near perfectly with Ki (R^2 0.96) (*4*). This provided fundamental insight into the effect of intrapatient tumour burden on kinetics of [^{18}F]DCFPyL, indicating that a 'sink effect' can indeed be present for [^{18}F]DCFPyL – as it is for [^{68}Ga]PSMA. In patients with low tumour burden, however, no measurable effect on [^{18}F]DCFPyL input functions was present, conforming to the conclusions of the study by Werner and colleagues

For [^{68}Ga]DOTATATE PET-CT in patients with neuroendocrine tumours (NET) a similar relationship between tumour burden and normal tissue uptake has been observed (*6*). Intriguingly, in a recent pharmacokinetic analysis of this tracer, tumour Ki correlated strongly with Tumour-to-Blood ratio (R^2 0.93), whilst the correlation between tumour Ki and SUV was non- linear and much lower (R^2 0.78) (*7*). Hence, as also noted by Ilan et al., a general effect may be present

for PET quantification of specifically targeted radiotracers (such as [68Ga]/[18F] PSMA, [68Ga]DOTATATE, and [18F]fluorodihydrotestosterone [FDHT](*8*)), as the distribution volumes of such tracers are mostly limited to tissues that express their specific target (*7*). Thus, (at fixed tracer dosages) large tumour volumes are more likely to influence tracer availability in plasma for such tracers and using SUV for quantification can become invalid.

Clinical illustration

To illustrate the sink effect, we hereby present a case of a 72-year old patient with mCRPC who underwent [18F]DCFPyL and [18F]FDHT PET-CT before starting abiraterone treatment, and again after 1 month of treatment (as part of an ongoing IRB-approved prospective study, IRB number 2014.218; the patient provided written informed consent for participation). This patient had been treated sequentially with upfront docetaxel, enzalutamide, and cabazitaxel. Baseline and follow-up PSA levels were 1436 ng/mL and 1936 ng/mL, respectively, indicating progressive disease under abiraterone.

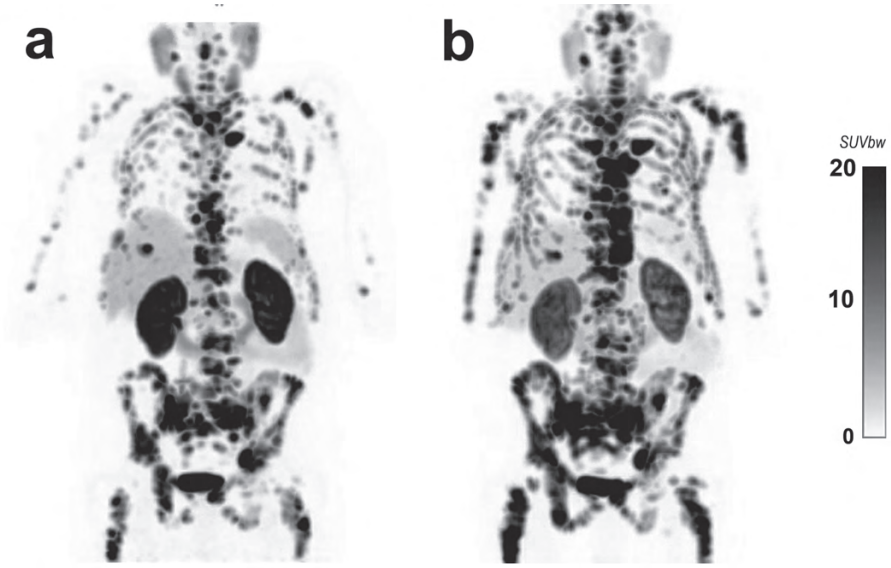


Figure 9.1: [18F]DCFPyL PET-CT images of a 72-year old patient with metastatic castration-resistant prostate cancer scanned (A) before and (B) during abiraterone treatment. A PSA rise from 1436 ng/mL to 1936 ng/mL indicated disease progression (one month after treatment start), and an accompanying increase in PSMA-positive disease burden is evident. PET images are shown as maximum-intensity projections.

Methods

The patient underwent whole body PET-CT imaging on a time-of-flight scanner for both [^{18}F]DCFPyL and [^{18}F]FDHT. Injected dosages of [^{18}F]DCFPyL at baseline and follow-up were 316 MBq and 303 Mbq, with uptake intervals of 124 min and 133 min, respectively. Injected dosages of [^{18}F]FDHT were 232 MBq and 238 Mbq, with uptake intervals of 45 min and 45 min, respectively. For [^{18}F]DCFPyL, we delineated the total tumour burden using a PERCIST SUV threshold based on liver uptake (*9*), which is a feasible approach due to the high tumour-to-background contrast and stable liver uptake between patients. Due to the unfavorable biodistribution of [^{18}F]FDHT, an alternative approach was needed. Therefore, the total tumour burden was delineated using a bone-mask (CT-based) where all uptake with SUV >1.9 was included (*10*). Next, nodal metastases and bone marrow disease was manually delineated. On each scan, we measured the PSMA- or FDHT-positive tumour volume, SUVmean, the Tumour-to-Blood ratio (TBR), and the total lesion uptake for SUV and TBR (SUV/TBR multiplied by the total tumour volume). Also, we measured the SUVmean in normal tissues: blood, kidneys, parotid glands, spleen, and liver for [^{18}F]DCFPyL; blood, kidneys, spleen, muscle, and liver for [^{18}F]FDHT. Then, we calculated the relative differences in those measures between the baseline and follow-up PET-CT scans. As prostate cancer metastases are often very small, we also quantified the total tumour burden using images that were corrected for partial-volume effects using post-reconstruction iterative deconvolution that was calibrated to comply with updated EARL2 guidelines.

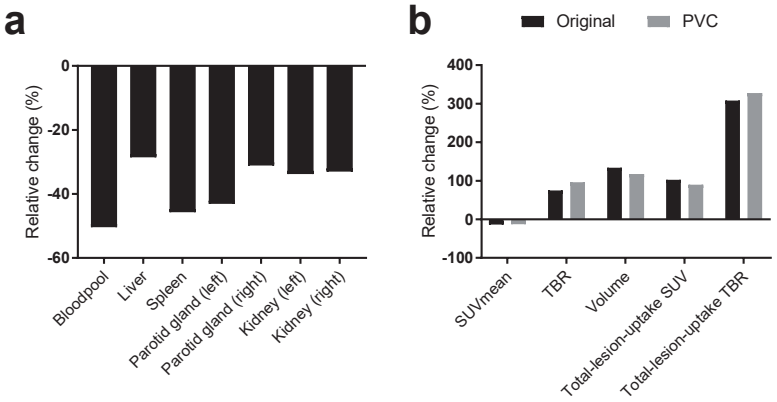


Figure 9.2: Changes in [^{18}F]DCFPyL uptake of (A) normal tissues and (B) the total tumor burden, from the scans shown in Figure 9.1. Changes in normal tissue uptake pertain to SUVmean. The whole body tumor metrics were derived from the total tumor burden as a single VOI. PVC = partial-volume correction.

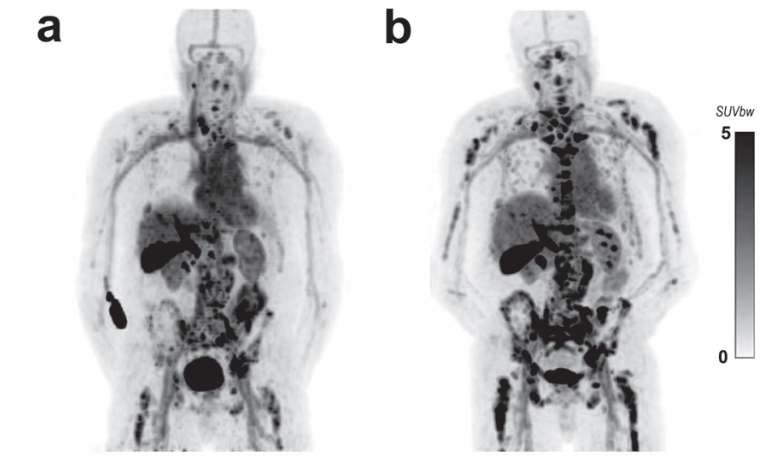


Figure 9.3: [18F]FDHT PET-CT images of a 72-year old patient with metastatic castration-resistant prostate cancer scanned (A) before and (B) during abiraterone treatment. A PSA rise from 1436 ng/mL to 1936 ng/mL indicated disease progression (one month after treatment start), and an accompanying increase in androgen receptor-positive disease burden is evident. PET images are shown as maximum-intensity projections.

Results

Visually, a substantial increase in PSMA-avid disease is apparent, accompanied by clear decreases in uptake in parotid glands, liver, spleen, and kidneys (Figure 9.1). This was in agreement with the substantial increase in measured PSMA-positive tumour volume, which increased by +134%, from 1001.3 mL to 2340.9 mL (Figure 9.2). This was accompanied by an inverse change in blood SUVmean, which decreased by 50.4%, from 0.85 to 0.42. Moreover, the SUVmean in all measured normal tissues consistently decreased by 28.6-45.7% (Figure 9.2). A less pronounced but very similar effect was noted for [18F]FDHT, where there was a clear visual progression of disease, and a decrease in uptake in normal tissue and organs (Figure 9.3). This visual progression conformed to a large increase in FDHT-positive tumour volume, which increased by 61%, from 1093 mL to 1763 mL (Figure 4). Again, this was accompanied by a decrease in blood SUVmean, which decreased by 22.7%. Also, all other measured tissues demonstrated consistent decreases in uptake ranging -14.3% to -21.6% (Figure 9.4).

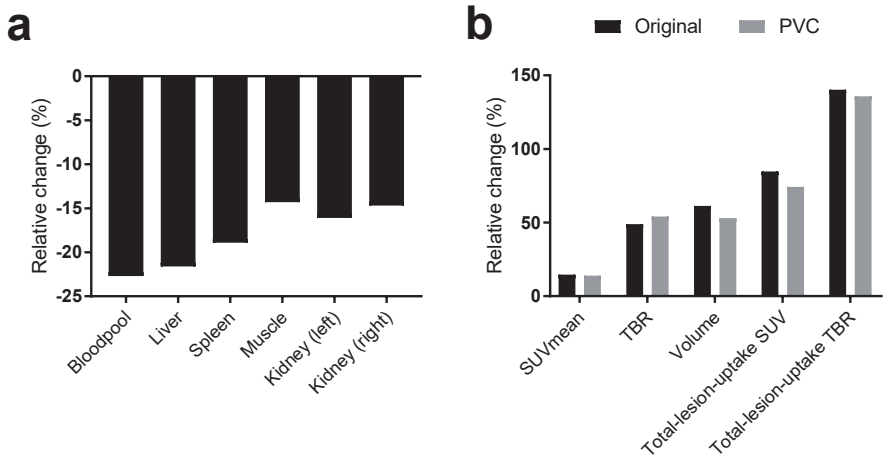


Figure 9.4: Changes in [^{18}F]FDHT uptake of (A) normal tissues and (B) the total tumor burden, from the scans shown in Figure 2. Changes in normal tissue uptake pertain to SUVmean. The whole body tumor metrics were derived from the total tumor burden as a single VOI. PVC = partial-volume correction.

Interpretation

Due to the 'sink effect' that occurred as the result of the increased tumour volume, a clear discrepancy between changes in SUVmean and Tumour-to-Blood ratio between scans is evident for both tracers. At disease progression, tumour SUVmean had decreased by 13% for [^{18}F]DCFPyL and increased by 14.6% for [^{18}F]FDHT, both within repeatability limits. In contrast, the Tumour-to-Blood ratio increased substantially by 75% for [^{18}F]DCFPyL and 49% for [^{18}F]FDHT. Hence, using the TBR seems to correct for the sink effect and adequately reflects the clinically observed disease progression. Though we present a rather extreme case, this does provide further illustration that use of SUV should be avoided for both individual reads and response assessment purposes of [^{18}F]DCFPyL and [^{18}F]FDHT PET in large tumour volume patients. While we cannot exclude an effect of abiraterone on tumoural uptake, it is highly unlikely that it directly impacted normal tissue uptake. While having a clear effect on absolute reads at baseline and follow-up, partial-volume correction did not substantially change relative quantitative reads. This is in agreement with what we observed in Chapter 5, where the total tumour burden metrics seemed robust to reconstruction protocol.

Conclusion

In conclusion and in addition to the results of Werner et al., in patients harboring a small tumour load no relevant sink effect of [18F]DCFPyL is to be expected. However, for patients with larger tumour volumes a sink effect may in fact occur, meaning personalized dosimetry of 177Lu-PSMA radioligand therapy might be needed for these patients. For [18F]FDHT, a similar sink effect of tumour load can be expected. Hence, there indeed seems to be a general effect for radiotracers that have a specific targeting, rendering use of SUV for such tracers invalid. We urge colleagues currently using both [18F]DCFPyL and 177Lu-PSMA radioligand therapy to repeat the study by Werner et al. in patients with a clinical indication for 177Lu-PSMA radioligand therapy. Moreover, since a clear definition of small versus large tumour load is not established, future studies should refrain from using SUV for quantification of tumoural [18F]DCFPyL uptake and resort to using TBR instead (*4*). For [18F]FDHT, use of SUV should be avoided regardless of tumour load, but more specifically in high-volume patients in response assessment settings.

References

1. Werner RA, Bundschuh RA, Bundschuh L, et al. Semiquantitative Parameters in PSMA-Targeted PET Imaging with [(18)F]DCFPyL: Impact of Tumor Burden on Normal Organ Uptake. ***Mol Imaging Biol.*** 2019.
2. Yadav MP, Ballal S, Sahoo RK, Dwivedi SN, Bal C. Radioligand Therapy With (177)Lu-PSMA for Metastatic Castration-Resistant Prostate Cancer: A Systematic Review and Meta-Analysis. ***AJR Am J Roentgenol.*** 2019:1-11.
3. Gaertner FC, Halabi K, Ahmadzadehfar H, et al. Uptake of PSMA-ligands in normal tissues is dependent on tumor load in patients with prostate cancer. ***Oncotarget.*** 2017;8:55094-55103.
4. Jansen BHE, Yaqub M, Voortman J, et al. Simplified Methods for Quantification of (18)F-DCFPyL Uptake in Patients with Prostate Cancer. ***J Nucl Med.*** 2019.
5. Jansen BHE, Kramer GM, Cysouw MCF, et al. Healthy Tissue Uptake of (68)Ga-Prostate Specific Membrane Antigen (PSMA), (18)F-DCFPyL, (18)F-Fluoromethylcholine (FCH) and (18) F-Dihydrotestosterone (FDHT). ***J Nucl Med.*** 2019.
6. Beauregard JM, Hofman MS, Kong G, Hicks RJ. The tumour sink effect on the biodistribution of 68Ga-DOTA-octreotate: implications for peptide receptor radionuclide therapy. ***Eur J Nucl Med Mol Imaging.*** 2012;39:50-56.
7. Ilan E, Velikyan I, Sandstrom M, Sundin A, Lubberink M. Tumor-to-blood ratio for assessment of somatostatin receptor density in neuroendocrine tumors using (68)Ga-DOTATOC and (68)Ga-DOTATATE. ***J Nucl Med.*** 2019.
8. Kramer GM, Yaqub M, Vargas HA, et al. Assessment of Simplified Methods for Quantification of (18)F-FDHT Uptake in Patients with Metastatic Castration-Resistant Prostate Cancer. ***J Nucl Med.*** 2019;60:1221-1227.
9. 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.
10. Venema CM, Mammatas LH, Schroder CP, et al. Androgen and Estrogen Receptor Imaging in Metastatic Breast Cancer Patients as a Surrogate for Tissue Biopsies. ***J Nucl Med.*** 2017;58:1906-1912.


QI-1-PET

Chapter 10

Prostate-Specific Membrane Antigen Expression Measured on Positron Emission Tomography as a Novel Biomarker in Primary Prostate Cancer.

M.C.F. Cysouw, B.H.E. Jansen, D.E. Oprea-Lager, O.S. Hoekstra,
R.J.A. van Moorselaar, T. van de Brug, R. Boellaard, A.N. Vis

To be submitted

Research letter

Risk stratification of patients with primary prostate cancer (PCa) before initial treatment is notoriously challenging. Prostate-Specific Membrane Antigen (PSMA) PET-CT has proven value for detection of metastases in recurrent PCa, but its sensitivity for detection of lymph node involvement (LNI) in primary PCa is limited (*1*). PSMA is a type-II transmembrane protein that is highly overexpressed on PCa cells. Clinically, this characteristic has been exploited exclusively for the visual detection of disease on PET-imaging. However, PSMA is not merely an identifier of PCa presence, but may have a biological role in disease progression through stimulation of oncogenic signalling (*2*).

PET offers a unique opportunity to quantify primary PCa PSMA-expression and with this measure disease aggressiveness. Next, these data can be used for pre-operative risk stratification. Quantification on PET can be done using artificial intelligence-based radiomics analysis, which aims to capture intensity, morphological, and spatial patterns of PSMA-expression within tumours. Basic PET quantification can also be employed, confining analyses to intensity and volumetric data of PSMA-expression. The advantage of PET is that it allows for measurement of the entire tumour, even before radical prostatectomy. In Chapter 6, we performed a PSMA-radiomics analysis in patients with PCa patients who prospectively underwent [^{18}F]PSMA PET-CT imaging prior to robot-assisted radical prostatectomy and extended pelvic lymph node dissection (*3*). In the present chapter, we benchmark these findings to clinical prediction approaches (for example using nomograms) in the same cohort, and compare our results to those of similar studies using the receiver-operator characteristics area under the curve (AUC).

The main finding of our study was that the machine learning models trained using PSMA-radiomics predicted LNI with an AUC of 0.86, which is comparable to the results from Zamboglou et al. and higher than observed by Ferraro et al. (Table 10.1) (*3-5*). These AUCs were similar or even higher than those of several nomograms were (Table 10.1). Adding visual analysis to the radiomics predictions in cross-validation did not improve the LNI prediction AUC. See Figure 10.1 for a clinical case where PSMA-radiomics outperformed the MSKCC nomogram for LNI prediction. Additionally, we found a high AUC in PSMA-radiomics for predicting Gleason score ≥8 in the radical prostatectomy specimen (AUC 0.81), as was similarly observed by Zamboglou et al. (*3*). Interestingly, the PSMA-radiomics outperformed biopsy samples for prediction of Gleason score ≥8 in the

prostatectomy specimens (AUC 0.74 for biopsy Gleason score versus AUC 0.81 for PSMA-radiomics). Lastly, the PSMA-radiomics predicted extracapsular tumour extension at an AUC of 0.76, again outperforming the MSKCC nomogram (AUC 0.69).

Table 10.1: Receiver-operator characteristics AUCs for the prediction of lymph node involvement (LNI), radical prostatectomy Gleason score, and extracapsular extension (ECE) in the radical prostatectomy specimen using PSMA-expression quantified on PET imaging versus classical prediction models using clinical features such as within nomograms.

Study Tracer	n	PSMA-expression on PET			Clinical approaches		
		LNI	Gleason score ≥8	ECE	LNI	Gleason score ≥8	ECE
Cysouw et al.(3) [^{18}F]DCFPyL	72	0.86	0.81	0.76	0.70$^{\#}$	0.74‡	0.69$^{\#}$
Zamboglou et al.(4) [^{68}Ga]-PSMA-11	60	0.85 - 0.87	0.84 - 0.93				
Ferraro et al.(5)† [^{68}Ga]-PSMA-11	60	0.70 - 0.76			0.62 - 0.83*		

*MSKCC nomogram, Yale formula, Roach Formula, Winter nomogram, and Partin tables; $^{\#}$MSKCC nomogram; ‡Biopsy Gleason score. †Basic PET features.

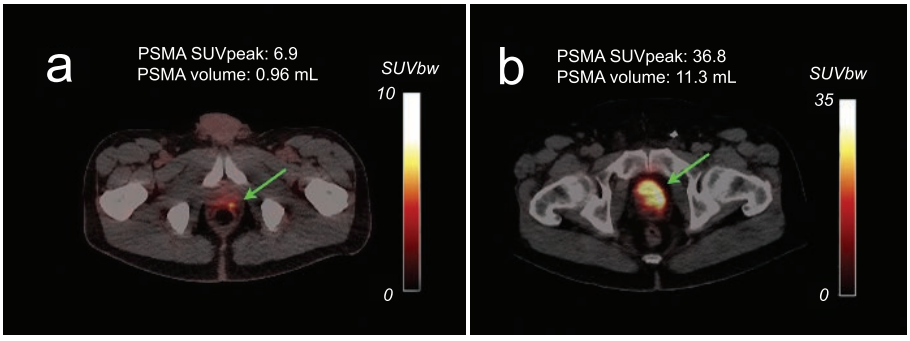


Figure 10.1: Pre-operative [^{18}F]PSMA PET-CT images of (A) a patient with a high MSKCC risk for LNI of 31% (Gleason 4+4, PSA 7.4 ng/ml, cT2b, 50% positive cores) who was pN0 after ePLDN, and (B) a patient with a low MSKCC risk for LNI of 9% (Gleason 3+4, PSA 10.4, cT2a, 67% positive cores) that had pN1 disease after ePLND. A machine learning model trained on [^{18}F]PSMA PET radiomics predicted risks of LNI of 3% for patient (A) and 98% for patient (B), respectively. In this example, the MSKCC nomogram did not accurately predict LNI, in contrast with the [^{18}F]PSMA PET radiomics-based model.

Conclusion

These emerging data may indicate that PSMA-expression might be an additional biomarker to PSA, biopsy pathology characteristics, and tumour stage in the pre-treatment stratification of primary PCa patients. The predictive value of PSMA-expression on PET seems to be valid using both basic PET analysis (*5*) and radiomics analysis (*3,4*), and using both [^{18}F]-labelled and [^{68}Ga]-labelled PSMA-ligands, and may even outperform existing prediction models. In clinical practice, PSMA PET can initially be used in intermediate- to high-risk patients to visually exclude distant metastases, and subsequently to predict the risk of LNI through quantification of PSMA-expression in the primary tumour. We acknowledge that current datasets are relatively small and that external model validation is lacking. Larger multicenter cohort studies are needed to validate these findings and develop robust prediction models.

References

1. van Kalmthout LWM, van Melick HHE, Lavalaye J, et al. Prospective Validation of Gallium-68 PSMA-PET/CT in Primary Staging of Prostate Cancer Patients. ***J Urol.*** 2019:101097JU0000000000000531.
2. Kaittanis C, Andreou C, Hieronymus H, et al. Prostate-specific membrane antigen cleavage of vitamin B9 stimulates oncogenic signaling through metabotropic glutamate receptors. ***J Exp Med.*** 2018;215:159-175.
3. Cysouw MCF, Jansen BHE, Oprea-Lager DE, et al. Machine Learning-based Analysis of [^{18}F]PSMA PET Radiomics for Risk Stratification in Primary Prostate Cancer. Submitted.
4. Zamboglou C, Carles M, Fechter T, et al. Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate- and high-risk prostate cancer - a comparison study with histology reference. ***Theranostics.*** 2019;9:2595-2605.
5. Ferraro DA, Muehlematter UJ, Garcia Schuler HI, et al. (68)Ga-PSMA-11 PET has the potential to improve patient selection for extended pelvic lymph node dissection in intermediate to high-risk prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019.


QI-1-PET

Chapter 11

Summarizing discussion and future perspectives

M.C.F. Cysouw

Summarizing discussion

The present thesis focused on technical validation of quantitative PET-CT as biomarker in primary, recurrent, and castration-resistant prostate cancer (mCRPC) and in non-small cell lung cancer (NSCLC). More specifically, in **Part 1** we investigated technical and methodological aspects of small tumour quantification on PET-CT, including radiomics analysis, followed by evaluation of their clinical application in **Part 2.**

Methodological validation

The validity of a biomarker as a measure of tumour biology or pathology, relies on the accuracy and precision of its measurement. As already introduced in **Chapter 1**, accuracy and precision are two terms that are often used in describing the performance characteristics of a measurement device, such as PET-CT (*1*). To assess the validity of quantitative tumour measurements on PET, it is important that both accuracy and precision are taken into account (*2*). Ideally, a biomarker is both accurate (it measures the truth) and precise (it is reliable). Yet, an accurate biomarker can have low precision and, vice versa, a precise biomarker can be inaccurate.

Using PVC requires careful validation with respect to tumour delineation

In **Chapter 2** we empirically validated the use of PVC to improve the accuracy of small tumour quantification, while taking into account the effect of PVC on its precision (*1*). To assess accuracy, we used a phantom scan and clinically realistic [^{18}F]FDG PET-CT simulations to assess true tumour uptake. Results of both methods are obtained under rather ideal circumstances, as phantom and simulated lesions are spherical and have homogeneous tracer uptake. Nonetheless, it provided valuable insight into the performance of several PVC methods and their susceptibility to tumour delineation accuracy that needs to be taken into account.

We observed that PVC methods have great potential to improve accuracy of small tumour quantification, but that they do not reach full accuracy for lesions ≤1cm in diameter. This means that 1 cm tumour diameter seems to be the limit of current PET systems. Still, as PVE generally started to occur below 3cm diameter, this still indicates a large benefit of PVC to acquire accurate reads. An important finding of this study was that the PVC methods that make use of the pre-defined

tumour boundaries (i.e. the volumes of interest defined on PET) are highly susceptible to delineation accuracy. Therefore, such methods may worsen rather than improve accuracy. We recommend the use of parametric (voxel-wise) PVC methods such as iterative deconvolution, which performed rather well regardless of tumour delineation accuracy. An advantage of an iterative deconvolution algorithm is that it is easily implemented and computationally efficient since it can be applied to standard reconstructed clinical images (*3,4*). It does, however, require that the PSF setting is calibrated for individual PET-CT systems. This might be easily performed using the phantom images that are routinely acquired for the EARL calibration (*5*). A disadvantage of such an algorithm is that it tends to amplify image noise, mandating some form of image denoising when SUVmax is used (*6*).

PVC in dynamic PET: worth the trouble?

In **Chapter 3** we evaluated the impact of PVC in a dynamic pharmacokinetic [^{18}F]FLT PET-CT study in NSCLC patient treated with tyrosine kinase inhibitors (*7*). Full pharmacokinetic studies are needed to technically validate simplified quantitative metrics of tracer uptake on PET (e.g. SUV or TBR) (*8-11*). In this regard, accuracy has a somewhat different meaning than it had it the previous paragraph. Here, the ground truth is the reference pharmacokinetic parameter (distribution volume, V_T), and accuracy pertains to the correlation between V_T and the simplified parameters. PVC may affect both differently, and thus needs to be considered in such pharmacokinetic PET studies. We observed that during a dynamic PET acquisition, the PVE changes over time, depending on tracer kinetics in blood and tumour. Also, the denoising algorithm required specific optimization to prevent that the temporal course of PVE was omitted (*12*). As expected, PVC increased both tumour kinetic parameters and SUV/TBR reads. However, as the effects of PVC on both these parameters were quite similar, it did not substantially change their correlation coefficients (despite having a significant impact on absolute values). Therefore, PVC improved nor worsened the accuracy of SUV and TBR taking V_T as reference for response assessment purposes. This study provided valuable, and perhaps reassuring, insight into the impact of PVE on technical validation of simplified PET metrics for clinical use, indicating that it may not be of value for future full pharmacokinetic response monitoring validation studies in oncology. It should be noted that these findings only apply to relative treatment-induced changes in measured tumour parameters. If absolute values,

for example pre-treatment (baseline) reads, are validated in pharmacokinetic analysis for prognostic or predictive purposes, PVC may have a substantial impact on validation of the preferred simplified metric. Also, in situations where a variable treatment effect on tumour size is expected (i.e. pseudo-progression during immunotherapy), the impact of PVC could be larger. In dynamic PET acquisitions in brain studies, PVC will be of benefit as the image-derived input function derived from carotid arteries is subject to PVE, but this is not an issue in oncological PET (*13*).

Precision must be balanced against accuracy

One of the main focuses of quantitative PET research has been its potential use in response assessment to systemic treatment (*14*). Use of PET for this purpose is attractive since it may provide for an assessment of the cancer burden as a whole, and for assessment of intrapatient heterogeneity of response. Also, the general hypothesis is that functional changes (e.g. reduction in tumour metabolism or proliferation) precede the anatomical tumour changes that are commonly measured on CT (e.g. using RECIST1.1), allowing for earlier positioning of PET in the clinical follow-up (*15,16*). As previously stated, knowledge of the intrinsic variability of PET quantification is necessary to be able to discern actual treatment-induced changes versus day-to-day variations (*17-21*). When aiming to improve the accuracy of small tumour quantification with PVC, or when using novel PSF image reconstructions, a potential negative effect on precision must be acknowledged and investigated as this may negatively affect the response assessment abilities of quantitative PET.

In **Chapters 2, 4, and 5**, we thoroughly investigated how PVC affects the test-retest variability of quantitative PET metrics in patients with NSCLC ([^{18}F]FDG) and metastatic prostate cancer ([^{18}F]FCH, [^{18}F]FDHT, and [^{18}F]DCFPyL). First, we demonstrated that the PVC methods that are prone to yield inaccurate results (i.e. those depending on tumour delineation accuracy) also had a substantial negative impact on the test-retest variability (*1*). In contrast, methods that performed well in terms of accuracy (e.g. iterative deconvolution) had only a marginal effect on the repeatability as expressed in the intra-class correlation coefficient (ICCs >0.9), and was thus chosen for clinical evaluation in **Chapters 4 and 5.**

[^{18}F]FDHT and [^{18}F]DCFPyL (a PSMA-ligand) both are prostate cancer-targeted radiotracers that may provide for PET-derived biomarkers in metastatic

castration-resistant mCRPC and hormone-sensitive prostate cancer (*22-24*). In **Chapter 4** ([^{18}F]FDHT) and **Chapter 5** ([^{18}F]DCFPyL) we applied PVC such that signal recovery complied with novel EARL ('EARL2') guidelines for scanner calibration. The EARL guidelines have recently been updated to accommodate the novel high-resolution reconstructions (or post-reconstruction processing) that are now routinely provided by system vendors (*5*). These 'high-resolution' PET images are often preferred by clinicians due to a higher diagnostic confidence and assumed higher detection rates. However, as we show in both **Chapter 4 and Chapter 5,** the test-retest variability of tumour quantitative measurements tends to increase to a variable extent, depending on the metric used (*25*). For example SUVmax, being most sensitive to image noise levels, became rather unreliable after PVC, rendering it unsuitable for response assessment. SUVpeak, however, seems to be a good alternative, as it is was robust to PVC-induced noise and is similarly observer-independent.

Simplified quantification of both [^{18}F]FDHT and [^{18}F]DCFPyL has been validated in two full pharmacokinetic PET studies (*10,11*). In both studies, SUV measurements proved to be invalid surrogates for the reference pharmacokinetic parameter *Ki*. Consequently, some form of tumour activity normalization to blood pool radioactivity was needed (input function integral for [^{18}F]FDHT, yielding SUVauc; whole blood image-based activity for [^{18}F]DCFPyL, yielding TBR). These metrics, however, have an inherently lower repeatability than SUV measurements due to the added variability in blood activity normalization parameters. Since these parameters are in whole (TBR) or partly (SUVauc) derived from the PET image itself, applying PVC adds additional noise to the measurements, potentially rendering them less suitable for use in response assessment studies with small effect sizes.

A way to at least partially mitigate the PVC-induced noise propagation, is to improve the count density of the images. In theory, this can be achieved in two ways: 1. by increasing the injected dosages, or 2. by extending the image acquisition time. Increasing the injected dosage to improve count statistics assumes that the dosage does not affect the tracer distribution. Extending the image acquisition time to improve count statistics assumes that tracer kinetics are stable (within and between bed positions), and neglects the added noise induced by radioactive decay. In **Chapter 4**, we investigated the repeatability of [^{18}F]FDHT PET in mCRPC as a function of PVC and acquisition time. We hereto split the original list-mode data (i.e. the counts) from an original 3 minutes per bed position to a 1.5 minutes per

bed position. We observed that PVC does have a rather small negative impact on the test-retest repeatability, but that this worsened when reducing the counts in the PET image. Therefore, shortening the acquisition time propagates the negative impact of PVC on PET precision. Extending acquisition times currently used for [^{18}F]DCFPyL and [^{18}F]FDHT will not likely be clinically feasible. Still, it should be maximized as the repeatability of uptake quantification can be substantially affected when count density is reduced, as we show in **Chapter 4** for [^{18}F]FDHT (*25*).

Taken together, for [^{18}F]DCFPyL and [^{18}F]FDHT PET quantification, the original (non-PVC, EARL1-calibrated) images may be preferred for response assessment studies. Clinical protocols should preferably include double image reconstruction: EARL1 for quantification in response assessment, EARL2 for routine visual analysis (*26*). Of note, we did observe that the repeatability of whole body tumour burden assessments, which are usable in ^{177}Lu-PSMA response monitoring, are robust to PVC and require no additional image reconstructions. The relative changes measured during treatment might not necessarily lead to different conclusions between non-PVC and PVC images, as we observed in **Chapter 3** for lung cancer and **Chapter 9** for prostate cancer (*7,27*). The negative impact of PVC on response assessment will of course depend on the treatment-induced effect sizes.

Extracting PET radiomics from small tumours is feasible and valuable

In recent years, artificial intelligence (AI) has gained immense popularity and attention in the field of radiology and nuclear medicine (*28*). While AI models and algorithms have existed for some decades, computational power and large amounts of available data now allow for its potential to be fully revealed (*28*). Specifically in the field of medical imaging, there are high hopes for AI to of significant clinical benefit, as large amounts of data are acquired on a routine basis. Here, AI may be used to perform physician tasks (e.g. detection or classification of disease), or to make objective predictions on patient outcomes (clinical or pathological). The latter may be of particular interest for PET imaging, as it is inherently quantitative and targeted at tumour biology.

In **Chapter 6** we investigated the use of artificial intelligence (AI) in analysis of [^{18}F]DCFPyL PET images to predict prostate cancer risk, combining radiomics with machine learning. From [^{18}F]FDG PET studies, it is well known that PET radiomics features, especially those based on texture analysis, are strongly

influenced by image reconstructions, use of PVC, and tumour segmentation approaches (*21,29,30*). However, few studies have evaluated whether this sensitivity propagates into clinical outcome predictions. Therefore, we optimized the PET methodology as function of the accuracy of the AI-based predictions. We concluded that the machine learning algorithms that were trained using [^{18}F] PSMA-radiomics features could accurately predict presence of metastatic disease and tumour histopathological grade (ROC AUC 0.76-0.86). However, especially for lymph node metastasis prediction, the tumour delineation approach and use of PVC seemed to affect the prediction AUC. This may raise some concerns regarding the robustness of the PSMA radiomics analysis in multicenter setting, especially with regard to heterogeneity in PET image characteristics (*31*). To circumvent the issues in radiomics analysis due to tumour segmentation, it might be worthwhile to investigate the use of 'deep learning' instead of radiomics and machine learning (*30*). In contrast with radiomics, deep learning models require no pre-defined features and segmentations. An in-depth discussion on deep learning for image analysis is beyond the scope of this work (*32*).

In PET radiomics studies, the added value of using radiomics versus using standard PET metrics (SUV, metabolic volume, etc.) should always be determined. If equal in performance, such standard metrics are preferred over radiomics due to their ease of use, widespread availability, and potentially better generalizability. In **Chapter 6** we found that there was predictive value in the basic PET features, but predictions seemed to be less stable compared to when based on radiomics data.

It is often assumed that PET radiomics, especially those based on texture analysis, cannot be reliably extracted from small lesions (*33*). Still, even in small lesions, there can be complementary value of radiomics above basic PET features (*30,34*). Indeed, in our study we mainly included small lesions (inherent to primary prostate cancer) and radiomics still outperformed basic PET features (though non-significantly). Due to the inherently small tumour lesions, we applied PVC, which seemed of particular benefit for the radiomics-based predictions of lymph node metastases. This benefit may partly be due to the improved uptake contrast, which may have improved delineation of low-contrast lesions.

Overall, PET radiomics analysis in [^{18}F]PSMA PET imaging of primary prostate cancer is promising, but the predictions are susceptible to variations in PET methodological factors. Future work should validate the superiority of radiomics above standard PET features, preferably in a multicenter (multivendor,

multisystem) setting. To this end, the benefit of using post-acquisition feature harmonization algorithms should be evaluated (*35*).

Clinical application

From a scientific standpoint, validation of quantitative biomarkers in terms of accuracy and precision is of the utmost importance. However, the ultimate goal of such validation is to define and/or improve the clinical benefit of quantitative PET analysis. The latter is an important aspect that is not often taken into consideration in validation studies. Simply stated: will patients benefit from our validation efforts? To arrive at an answer to this question is complex and not straightforward, as most technical validation studies do not (or cannot) make use of clinical endpoints.

The clinical benefit of partial-volume correction thus far

At writing of this thesis, many quantitative oncological PET studies with clinical endpoints have been performed over several decades. Therefore, in **Chapter 7** we aimed to review the body of literature for PET studies with clinical endpoints that evaluated the benefit of PVC (*36*). To avoid narrative and potentially biased summarization of the literature, we performed a systematic review with a meta-analysis that was not restricted to cancer type.

The first striking finding was that only 31 studies were eligible for inclusion, from which several studies were performed by the same groups. In almost all of the investigated studies, PVC did not change the final conclusions on the clinical value of quantitative PET. A critical note was that the PVC methodology was often confined to a recovery coefficient method, which is the most basic method for PVC (*37*). Indeed, it might be expected that such methods do not benefit clinical predictions or diagnoses. We elaborately discussed the reviewed studies and provided our recommendations for future use of PVC towards a clinical benefit. To summarize, we recommended that i) more sophisticated and better validated PVC methods should be used, and that ii) there should be a consensus amongst investigators regarding the preferred methodologies, and iii) in future studies partial-volume corrected data should be used in parallel to uncorrected data.

Quantification does not add to visual reads in oligometastatic prostate cancer

Oligometastatic prostate cancer has become a newly recognized clinical entity, both at diagnosis (synchronous) and at biochemical recurrence (metachronous)

(*38*). The hypothesis is that these metastases run a more indolent course as they originate from a more hospitable primary tumour environment (*39*), therefore being more amenable to local metastasis-direct treatments, such as stereotactic body radiotherapy (SBRT) (*40*). Still, not all patients have a (lasting) PSA response after SBRT treatment. In **Chapter 8**, we aimed to predict which patients were more likely to have durable treatment responses by quantifying uptake of [^{18}F]FCH in metastatic lesions, but concluded that none of the investigated PET metrics was predictive for biochemical progression-free survival (PFS) (*41*). We hypothesized that PVC might improve these predictions, since the analyzed metastases were inherently small, but this was not the case. The only 'quantitative' measure prognostic for PFS was the number of visually detected [^{18}F]FCH-avid lesions (*41*). This study is an example of a clinical setting where, despite methodological rigor, PET quantification does not add to current practice.

Technical PET validation into clinical practice

In **Chapter 9** we commented on a study that investigated whether [^{18}F]DCFPyL was subject to a 'sink effect' in metastatic prostate cancer (*27,42*). A sink effect pertains to the observation that the distribution of a radiotracer in background tissues (e.g. liver, blood pool, spleen etc.) is inversely related to the tumour burden(*43*). The authors of the study that we commented on did not observe a sink effect, but they only investigated patients with low tumour burden. This provided us with an opportunity to discuss previous findings on the pharmacokinetics of [^{18}F]DCFPyL, where we observed that a high tumour burden affected the plasma and whole blood input functions, resulting in lower activity concentrations over time in blood (*11*). The consequence of this was that SUV as simplified parameter for tumour PSMA uptake was not valid. However, the tumour-to-blood ratio (TBR) strongly correlated with the validated pharmacokinetic parameter *Ki (11)*.

The sink-effect may be a general effect that occurs for PET radiotracers that are targeted to a specific ligand or receptor (*44*). For such tracers, the distribution volume is limited (almost) exclusively to tumours that express their targets. Hence, in patients with a relatively large tumour burden, this might results in a sink-effect. Moreover, large longitudinal changes in tumour volumes within patients will hamper use of SUV for response assessment as this could affect the blood input functions (*27*).

To illustrate the clinical relevance of this effect, and demonstrate the importance of technical PET validation, we presented a clinical case of a patient

with high-volume mCPRC who underwent both [^{18}F]DCFPyl and [^{18}F]FDHT PET before and during systemic treatment. For both tracers, we demonstrated that there was a large tracer sink effect over time, due to substantial increases in tumour volumes. Moreover, changes in TBR evidently correlated with the clinical progression, whereas changes in SUV erroneously indicated stable disease. We also evaluated whether application of PVC in compliance with EARL2 affected these conclusions, but found no substantial change. As we previously discussed, use of TBR for response assessment is often unwanted due to the added noise of blood activity measurements, especially when PVC is applied. However, despite its worse precision, TBR did accurate reflect the true progressive disease during treatment, in contrast with SUV (Figure 9.2 and 9.4). Hence, this clinical example illustrates that, in some cases, high accuracy (use of TBR) is more important than high precision (use of SUV), even in response assessment.

Finally, we must acknowledge that the progression quantified on both [^{18}F]DCFPyL and [^{18}F]FDHT PET was confirmed by basic prostate specific antigen (PSA) measurements. Initial studies on response assessment with PSMA PET have shown a similar correlation between PSA response and PET response, warranting further determination of the true added value of PET-based response assessment in metastatic prostate cancer (*45,46*). Here, the advantage PET-based assessments will be capturing interlesional heterogeneity of response.

PSMA-expression quantified on PET as novel biomarker?

In **Chapter 6** we validated and optimized the use of PET radiomics from [^{18}F]DCFPyL for risk stratification in primary prostate cancer. In this analysis, we found that these radiomics features were highly predictive when optimized according to tumour delineation and PVC, and that these features may outperform basic PET parameters. In **Chapter 10** these results were briefly discussed with respect to clinical relevance for the urologist community. We summarized results from our and two other studies on PSMA-expression quantified on PET for prediction of lymph node involvement (LNI), high Gleason score, and/or extracapsular tumour extension (*47,48*). Interestingly, all studies observed high AUCs for prediction of LNI, high Gleason score, and (in our study) extracapsular tumour extension. We also benchmarked the radiomics AUCs to clinical prediction models, and found that the [^{18}F]PSMA radiomics outperformed these models. All taken together, PSMA-expression quantified on PET is a very promising new player in the group of biomarkers in primary prostate cancer (e.g. PSA, Gleason score). However,

there still are significant challenges that remain to be solved, regarding collection of large harmonized multicenter datasets for model training, external validation of radiomics versus simplified metrics, and incorporation of the available biomarkers (e.g. Gleason score, PSA, tumour stage, and PSMA expression) into a novel prediction model that will fit into clinical workflows.

Future perspectives and recommendations

In this thesis, we performed technical validation of small tumour quantification on PET-CT to derive biomarkers for prognostication and response assessment in prostate and lung cancer. Part of this validation included an investigation into use of artificial intelligence for quantitative PET analysis in prostate cancer. We also investigated and commented on the benefits of clinical application of these biomarkers with respect to clinical endpoints.

As always in medical research, no study fills all knowledge gaps or answers all questions. Even more so, well-performed studies tend to generate more new questions than answers. Hence, using the knowledge gained in this thesis we can now comment on the future perspectives of the investigated radiotracers and use of quantitative PET biomarkers, and provide some recommendations for future research.

Prostate cancer

In **Chapter 2** and **Chapter 8**, we investigated the repeatability and prognostic value of [^{18}F]FCH PET-CT in metastatic prostate cancer (*1,41*). In the process of writing this thesis, choline-based ligands for prostate cancer have been (almost) entirely replaced by the novel PSMA-ligands, due to the higher detection rates for metastatic disease of the latter (*49-52*). Therefore, we think that [^{18}F]FCH PET-CT will have no significant indication in future prostate cancer imaging. It may, however, be of interest for early response assessment of therapeutic PSMA-inhibitors, where PSMA PET-CT will not be feasible due to tracer competition. Whether PSMA-inhibitors will have any future as clinical drugs remains to be investigated, though (*53*).

PSMA-ligands are the most revolutionary and successful (group of) radiotracers in nuclear medicine since the advent of [^{18}F]FDG (*54,55*). Clinically, success of this tracer is greatest in the biochemical recurrence stage of prostate

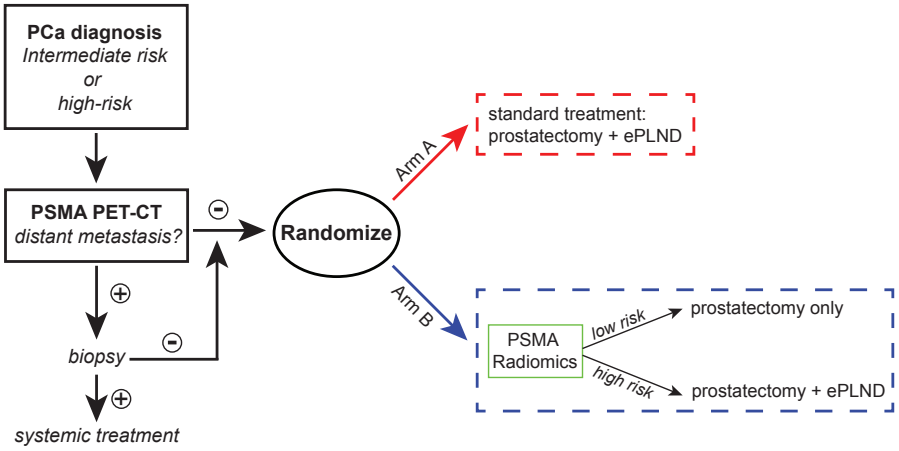


Figure 11.1: Example of a randomized-controlled trial to demonstrate the clinical benefit of using PSMA-radiomics with machine learning in primary prostate cancer patients before surgery. This should be a non-inferiority trail with biochemical recurrence-free survival as the primary endpoint. Secondary endpoints should include time to castration-resistance and prevalence of surgical complications. ePLND = extended pelvic lymph node dissection.

cancer, where detection rates (even at low PSA levels) are higher than for any other imaging modality (*56,57*). For [^{68}Ga]PSMA, a recent study demonstrated its superiority over the currently registered ligand [^{18}F]Fluciclovine, rendering FDA approval of this ligand quite likely (*57*). It remains to be shown whether the ligand investigated in this thesis, [^{18}F]DCFPyL, has the same benefit over [^{18}F]Fluciclovine, but initial results are promising (*58*). Investigations into the true clinical benefit of PSMA PET at biochemical recurrence are ongoing. Quantitative PET-CT analysis will not likely have be useful in this clinical setting to make clinical predictions. Deep learning approaches might, however, support radiologists in image reading.

In primary prostate cancer, the story of PSMA PET-CT has not been glorious, with a sensitivity for lymph node staging that tends to be around 40-50%, with large variations between studies (*59-62*). However, as we show in **Chapter 6**, in this setting there still is a very promising role for PSMA PET-derived quantitative imaging biomarkers. Here, tumour biology and PET technology are perfectly integrated, yielding clinically usable data to predict a patient's risk for metastatic disease. To get quantitative PSMA PET biomarkers, such as radiomics, into the clinical daily practice will require further validation using large pooled multicenter datasets, and thorough cross-center prospective validation studies (*63*). Following such (methodological) validation, the clinical benefit of PSMA radiomics will need to be proven in a randomized controlled

trial, where patients with primary prostate cancer (non-metastatic on PET) are treated based on predictions from clinical nomograms versus predictions from machine learning-based radiomics analysis (Figure 11.1). Such a trial (aimed at proving non-inferiority) would need to utilize a hard clinical endpoint, such as time to biochemical recurrence, and/or time to castration-resistant disease.

In mCRPC, the main current role seems to be selection of patients amenable to PSMA-radioligand therapy (^{177}Lu or ^{225}Ac) and follow-up imaging during and after these treatments (*45,64*). Response assessment is especially promising, since the repeatability of whole body quantitative PET parameters is very high as we show in **Chapter 5**. For other treatments, such as taxane chemotherapy and androgen-axis inhibitors, the benefit of quantitative PSMA PET-CT for response assessment over basic measures such as PSA remains to be shown (*46,65*). This needs to include a more elaborate investigation of the flare phenomenon that maybe present during AR-targeted therapy (*66,67*). Also, the superiority in assessment of treatment response using PSMA PET-CT versus standard imaging modalities (bone scintigraphy and diagnostic CT) has not yet been established.

Finally, we may briefly speculate about the use of PSMA PET-CT in development of drugs that target PSMA, such as the PMPA-2 PSMA inhibitor (*53*). Pre-clinical data have shown that inhibiting PSMA has a therapeutic effect, but these drugs have not yet been clinically tested (*53*). If phase I trials are to be performed, they should use PSMA PET-CT to evaluate treatment targeting, preferably using quantitative analysis to assess heterogeneity of drug targeting between (or even within) lesions.

The treatment paradigm for mCRPC has shifted toward androgen receptor (AR) axis-targeted treatments. As [^{18}F]FDHT specifically targets the AR, it has been shown to be very prognostic and may also be predictive for these treatments(*24*). For [^{18}F]FDHT PET-CT, response assessment in mCPRC is mostly limited to abiraterone as this drug does not directly bind the AR (*68*). Prognostication, on the other hand, can be performed for all treatment types. Especially for AR-targeted treatments, [^{18}F]FDHT PET biomarkers will be useful as it can be used to assess intrapatient heterogeneity of AR-expression, potentially identifying AR-mutant lesions, or lesions harboring AR-splice variants (*69,70*). A limitation of [^{18}F]FDHT PET is its poor biodistribution, presence of metabolites, and the need for dynamic imaging; all rendering routine clinical use unlikely (*10*). We recommend that [^{18}F]FDHT PET quantification should be used as a tool in clinical trials aiming to identify AR-mutant lesions for biopsy targeting to identify novel molecular prostate cancer subtypes.

Lung cancer

In lung cancer, we assessed whether correcting for PET's limited spatial resolution using PVC *i*) hampered the test-retest repeatability of [^{18}F]FDG PET biomarkers (*1*), and *ii)* improved pharmacokinetic parameter estimation and validation of simplified parameters for response assessment of [^{18}F]FLT PET. For both, using PVC did not substantially affect the methodological validity of quantitative PET biomarkers for response assessment needs. Taken together, we think that methodological validation of these tracers for NSCLC response assessment studies have been adequately studied in recent years. In NSCLC, the future of quantitative PET biomarkers will be in prediction of response to immunotherapy using PD1/PD-L1-targeted radioligands, or even with [^{18}F]FDG PET (*71,72*). Here, quantifying intratumoural heterogeneity of target expression (e.g. using radiomics analysis) might be crucial for accurate response prediction, which will again require thorough methodological validation with respect to tumour delineation and PVC.

PET biomarkers in general

To date, full pharmacokinetic PET studies have mainly been used for tracer validation studies. Now, with the advent of total body PET scanners, performing whole body dynamic studies has become a possibility (*73*). Still, even with these new scanners we do not think that dynamic imaging will become the new clinical standard, since it still requires extended imaging acquisitions, and (depending on the tracer) arterial or venous blood sampling. Moreover, the costs of total body PET scanners will limit their availability.

Simplified analysis of PET data, on the other hand, has become routinely available in virtually all clinical PET centers, as system vendors have implemented such quantification analysis into their software. Unfortunately, there are many drawbacks to using simplified analysis (SUV) pertaining to e.g. scanner calibration and harmonization, sensitivity of such parameters to variations in imaging protocols, tumour delineation methodology, and PET image processing (*2,74*). However, as these issues have been elaborately investigated, the routine clinical use of simplified PET metrics is nearby.

In recent years, AI-based analysis of PET images has made an entrance. Especially in molecular imaging, such analysis is attractive since it fully integrates tumour biology and PET technology, and may allow for thorough image-based tumour phenotyping. Still, radiomics or deep learning analyses are challenging in PET, due to the poor resolution, relatively high image noise levels, and difficulty in tumour delineation (*31*). Moreover, robustness of analysis, especially between centers, is a large issue. A main issue here is the heterogeneity between studies/centers regarding image reconstruction parameters (voxel size) and use of post-processing. The Image Biomarker Standardization Initiative (IBSI) has aimed to harmonize radiomics analysis, but evidence of a benefit for external validation remains to be shown. An idea would be to perform an additional 'IBSI-radiomics' reconstruction, besides the standard EARL reconstructions. Such a reconstruction should include isotropic voxels, preferably small (e.g. 2x2x2mm (*75*)) and a use of post-filtering calibrated using on phantom data that harmonizes both recovery and noise. To enhance clinically applicability, system vendors will need to implement standardized radiomics analysis within their software packages. Finally, a step-by-step guideline for initiating, developing, and validating PET radiomic analysis for clinical use needs to be defined (Figure 11.2). Such a guideline could perhaps also mitigate the publication bias present in radiomics research, which is a known issue (*76*). The role of deep learning analysis in PET is still promising, as it avoids known issues in radiomics analysis pertaining to tumour delineation and feature definitions. Still, such models are more complex to develop and train and require large computational power.

Finally, if AI-based prediction algorithms, whether radiomics or deep-learning-based, have been validated to such extent that they can be clinically deployed to have an effect on patient management, several ethical and legal issues arise. For example, the person who is responsible for mistakes made by these AI-models needs to be defined: the physician using the model, or the model developer? Also, legislation will need to be developed to ensure a minimum level of safety and efficacy of such models for routine clinical use. The FDA has already approved several AI-models as Software as a Medical Device (SaMD), and is moving towards a certification framework for modifying approved AI-models (*77*).

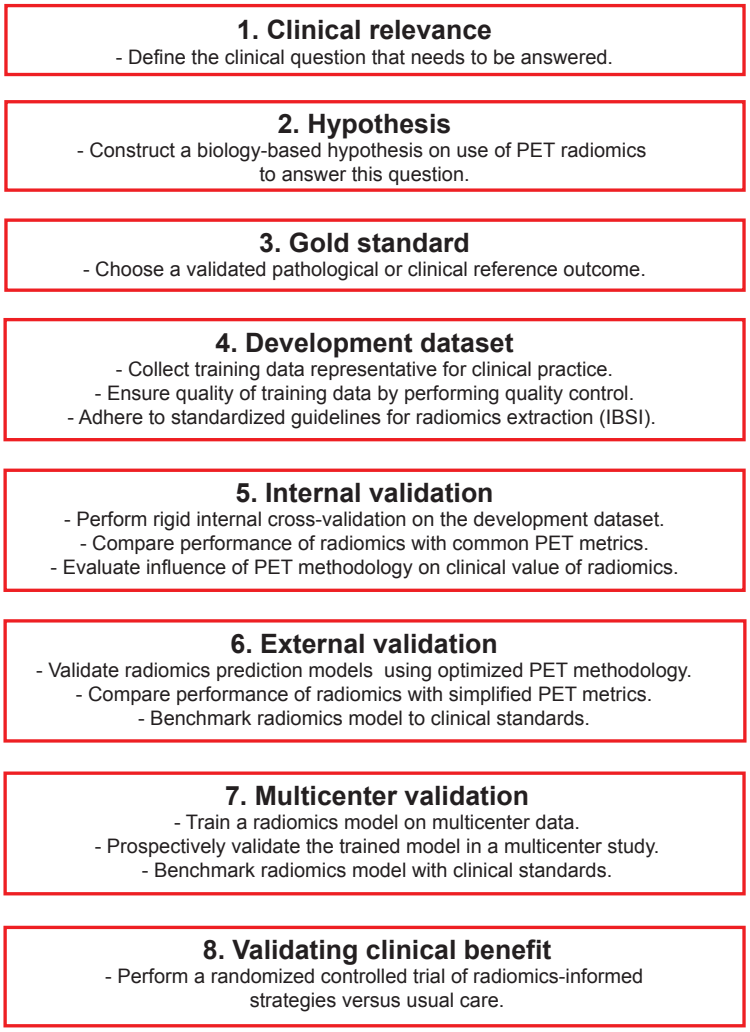


Figure 11.2: A potential guideline for validation of PET radiomics prediction models towards clinical implementation.

Conclusion

In this thesis we optimized and validated quantitative PET biomarkers derived from small tumours in primary, recurrent, and castration-resistant prostate cancer and in non-small cell lung cancer. We observed that despite PET's low resolution, small tumours can be accurately quantified when PVC methods are carefully validated. This comes at a small expense of worsened precision for [^{18}F]FCH, [^{18}F]FDG, [^{18}F]FDHT, and [^{18}F]DCFPyL, which could hamper response assessment using these tracers. Therefore, a balance needs to be struck between accuracy and precision, depending on the specific clinical indication and the expected effect sizes. Pharmacokinetic response assessment validation studies will not likely require any correction for the poor resolution of PET, as we showed for NSCLC using [^{18}F]FLT PET-CT. [^{18}F]FCH quantification did not allow for prediction of progression-free survival in oligometastatic prostate cancer. We validated artificial intelligence-based analysis of [^{18}F]DCFPyL radiomics analysis for risk stratification in prostate cancer, where tumour delineation and PVC seemed to be important determinants for prediction accuracy. Future technical validation studies for quantitative PET biomarkers should preferably include clinical endpoints. Lastly, we made several recommendations for future radiomics development studies towards clinical use.

References

1. Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial-Volume Correction in Oncological PET/CT Studies. ***J Nucl Med.*** 2016;57:1642-1649.
2. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. ***Eur J Nucl Med Mol Imaging.*** 2015;42:328-354.
3. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. ***J Nucl Med.*** 2007;48:802-810.
4. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. ***Neuroimage.*** 2008;39:1570-1584.
5. Kaalep A, Sera T, Rijnsdorp S, et al. Feasibility of state of the art PET/CT systems performance harmonisation. ***Eur J Nucl Med Mol Imaging.*** 2018;45:1344-1361.
6. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. ***Eur J Nucl Med Mol Imaging.*** 2009;36:1064-1075.
7. Cysouw MCF, Golla SVS, Frings V, et al. Partial-volume correction in dynamic PET-CT: effect on tumor kinetic parameter estimation and validation of simplified metrics. ***EJNMMI Res.*** 2019;9:12.
8. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. ***J Nucl Med.*** 2017;58:1019-1024.
9. Verwer EE, Oprea-Lager DE, van den Eertwegh AJ, et al. Quantification of 18F-fluorocholine kinetics in patients with prostate cancer. ***J Nucl Med.*** 2015;56:365-371.
10. Kramer GM, Yaqub M, Vargas HA, et al. Assessment of Simplified Methods for Quantification of (18)F-FDHT Uptake in Patients with Metastatic Castration-Resistant Prostate Cancer. ***J Nucl Med.*** 2019;60:1221-1227.
11. Jansen BHE, Yaqub M, Voortman J, et al. Simplified Methods for Quantification of (18)F-DCFPyL Uptake in Patients with Prostate Cancer. ***J Nucl Med.*** 2019.
12. Christian BT, Vandehey NT, Floberg JM, Mistretta CA. Dynamic PET denoising with HYPR processing. ***J Nucl Med.*** 2010;51:1147-1154.
13. Golla SSV, Lubberink M, van Berckel BNM, Lammertsma AA, Boellaard R. Partial volume correction of brain PET studies using iterative deconvolution in combination with HYPR denoising. ***EJNMMI Res.*** 2017;7:36.
14. 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.
15. Mankoff DA, Katz SI. PET imaging for assessing tumor response to therapy. ***J Surg Oncol.*** 2018;118:362-373.
16. Weber WA. Assessing tumor response to therapy. ***J Nucl Med.*** 2009;50 Suppl 1:1S-10S.
17. Vargas HA, Kramer GM, Scott AM, et al. Reproducibility and Repeatability of Semiquantitative (18) F-Fluorodihydrotestosterone Uptake Metrics in Castration-Resistant Prostate Cancer Metastases: A Prospective Multicenter Study. ***J Nucl Med.*** 2018;59:1516-1523.
18. Kolinger GD, Vallez Garcia D, Kramer GM, et al. Repeatability of [(18)F]FDG PET/CT total metabolic active tumour volume and total tumour burden in NSCLC patients. ***EJNMMI Res.*** 2019;9:14.
19. Kramer GM, Frings V, Hoetjes N, et al. Repeatability of Quantitative Whole-Body 18F-FDG PET/CT Uptake Measures as Function of Uptake Interval and Lesion Selection in Non-Small Cell Lung Cancer Patients. ***J Nucl Med.*** 2016;57:1343-1349.
20. Oprea-Lager DE, Kramer G, van de Ven PM, et al. Repeatability of Quantitative 18F-Fluoromethylcholine PET/CT Studies in Prostate Cancer. ***J Nucl Med.*** 2016;57:721-727.
21. van Velden FH, Kramer GM, Frings V, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. ***Mol Imaging Biol.*** 2016;18:788-795.
22. Beattie BJ, Smith-Jones PM, Jhanwar YS, et al. Pharmacokinetic assessment of the uptake of 16beta-18F-fluoro-5alpha-dihydrotestosterone (FDHT) in prostate tumors as measured by PET. ***J Nucl Med.*** 2010;51:183-192.

23. Szabo Z, Mena E, Rowe SP, et al. Initial Evaluation of [(18)F]DCFPyL for Prostate-Specific Membrane Antigen (PSMA)-Targeted PET Imaging of Prostate Cancer. ***Mol Imaging Biol.*** 2015;17:565-574.

24. Fox JJ, Gavane SC, Blanc-Autran E, et al. Positron Emission Tomography/Computed Tomography-Based Assessments of Androgen Receptor Expression and Glycolytic Activity as a Prognostic Biomarker for Metastatic Castration-Resistant Prostate Cancer. ***JAMA Oncol.*** 2018;4:217-224.

25. Cysouw MCF, Kramer GM, Heijtel D, et al. Sensitivity of (18)F-fluorodihydrotestosterone PET-CT to count statistics and reconstruction protocol in metastatic castration-resistant prostate cancer. ***EJNMMI Res.*** 2019;9:70.

26. Devriese J, Beels L, Maes A, Van de Wiele C, Pottel H. Impact of PET reconstruction protocols on quantification of lesions that fulfil the PERCIST lesion inclusion criteria. ***EJNMMI Phys.*** 2018;5:35.

27. Cysouw MCF, Jansen BHE, Yaqub M, et al. Letter to the Editor re: Semiquantitative Parameters in PSMA-Targeted PET Imaging with [(18)F]DCFPyL: Impact of Tumor Burden on Normal Organ Uptake. ***Mol Imaging Biol.*** 2019.

28. Nensa F, Demircioglu A, Rischpler C. Artificial Intelligence in Nuclear Medicine. ***J Nucl Med.*** 2019;60:29S-37S.

29. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour (1)(8)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. ***Eur J Nucl Med Mol Imaging.*** 2013;40:1662-1671.

30. Hatt M, Le Rest CC, Tixier F, Badic B, Schick U, Visvikis D. Radiomics: Data Are Also Images. ***J Nucl Med.*** 2019;60:38S-44S.

31. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. ***Eur J Nucl Med Mol Imaging.*** 2019.

32. Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. ***Eur J Nucl Med Mol Imaging.*** 2019.

33. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. ***J Nucl Med.*** 2014;55:37-42.

34. Hatt M, Majdoub M, Vallieres M, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. ***J Nucl Med.*** 2015;56:38-44.

35. Orlhac F, Boughdad S, Philippe C, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. ***J Nucl Med.*** 2018;59:1321-1328.

36. Cysouw MCF, Kramer GM, Schoonmade LJ, Boellaard R, de Vet HCW, Hoekstra OS. Impact of partial-volume correction in oncological PET studies: a systematic review and meta-analysis. ***Eur J Nucl Med Mol Imaging.*** 2017;44:2105-2116.

37. Geworski L, Knoop BO, de Cabrejas ML, Knapp WH, Munz DL. Recovery correction for quantitation in emission tomography: a feasibility study. ***Eur J Nucl Med.*** 2000;27:161-169.

38. Tosoian JJ, Gorin MA, Ross AE, Pienta KJ, Tran PT, Schaeffer EM. Oligometastatic prostate cancer: definitions, clinical outcomes, and treatment considerations. ***Nat Rev Urol.*** 2017;14:15-25.

39. Hellman S, Weichselbaum RR. Oligometastases. ***J Clin Oncol.*** 1995;13:8-10.

40. Bouman-Wammes EW, van Dodewaard-De Jong JM, Dahele M, et al. Benefits of Using Stereotactic Body Radiotherapy in Patients With Metachronous Oligometastases of Hormone-Sensitive Prostate Cancer Detected by [18F]fluoromethylcholine PET/CT. ***Clin Genitourin Cancer.*** 2017;15:e773-e782.

41. Cysouw M, Bouman-Wammes E, Hoekstra O, et al. Prognostic Value of [(18)F]-Fluoromethylcholine Positron Emission Tomography/Computed Tomography Before Stereotactic Body Radiation Therapy for Oligometastatic Prostate Cancer. ***Int J Radiat Oncol Biol Phys.*** 2018;101:406-410.

42. Werner RA, Bundschuh RA, Bundschuh L, et al. Semiquantitative Parameters in PSMA-Targeted PET Imaging with [(18)F]DCFPyL: Impact of Tumor Burden on Normal Organ Uptake. ***Mol Imaging Biol.*** 2019.

43. Beauregard JM, Hofman MS, Kong G, Hicks RJ. The tumour sink effect on the biodistribution of 68Ga-DOTA-octreotate: implications for peptide receptor radionuclide therapy. ***Eur J Nucl Med Mol Imaging.*** 2012;39:50-56.

44. Ilan E, Velikyan I, Sandstrom M, Sundin A, Lubberink M. Tumor-to-blood ratio for assessment of somatostatin receptor density in neuroendocrine tumors using (68)Ga-DOTATOC and (68)Ga-DOTATATE. ***J Nucl Med.*** 2019.

45. Grubmuller B, Senn D, Kramer G, et al. Response assessment using (68)Ga-PSMA ligand PET in patients undergoing (177)Lu-PSMA radioligand therapy for metastatic castration-resistant prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019;46:1063-1072.

46. Kuten J, Sarid D, Yossepowitch O, Mabjeesh NJ, Even-Sapir E. [(68)Ga]Ga-PSMA-11 PET/CT for monitoring response to treatment in metastatic prostate cancer: is there any added value over standard follow-up? ***EJNMMI Res.*** 2019;9:84.

47. Ferraro DA, Muehlematter UJ, Garcia Schuler HI, et al. (68)Ga-PSMA-11 PET has the potential to improve patient selection for extended pelvic lymph node dissection in intermediate to high-risk prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019.

48. Zamboglou C, Carles M, Fechter T, et al. Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate- and high-risk prostate cancer - a comparison study with histology reference. ***Theranostics.*** 2019;9:2595-2605.

49. Boustani AM, Pucar D, Saperstein L. Molecular imaging of prostate cancer. ***Br J Radiol.*** 2018;91:20170736.

50. Alonso O, Dos Santos G, Garcia Fontes M, Balter H, Engler H. (68)Ga-PSMA and (11)C-Choline comparison using a tri-modality PET/CT-MRI (3.0 T) system with a dedicated shuttle. ***Eur J Hybrid Imaging.*** 2018;2:9.

51. Afshar-Oromieh A, Zechmann CM, Malcher A, et al. Comparison of PET imaging with a (68)Ga-labelled PSMA ligand and (18)F-choline-based PET/CT for the diagnosis of recurrent prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2014;41:11-20.

52. Morigi JJ, Stricker PD, van Leeuwen PJ, et al. Prospective Comparison of 18F-Fluoromethylcholine Versus 68Ga-PSMA PET/CT in Prostate Cancer Patients Who Have Rising PSA After Curative Treatment and Are Being Considered for Targeted Therapy. ***J Nucl Med.*** 2015;56:1185-1190.

53. Kaittanis C, Andreou C, Hieronymus H, et al. Prostate-specific membrane antigen cleavage of vitamin B9 stimulates oncogenic signaling through metabotropic glutamate receptors. ***J Exp Med.*** 2018;215:159-175.

54. Donin NM, Reiter RE. Why Targeting PSMA Is a Game Changer in the Management of Prostate Cancer. ***J Nucl Med.*** 2018;59:177-182.

55. Schwarzenboeck SM, Rauscher I, Bluemel C, et al. PSMA Ligands for PET Imaging of Prostate Cancer. ***J Nucl Med.*** 2017;58:1545-1552.

56. Wondergem M, Jansen BHE, van der Zant FM, et al. Early lesion detection with (18)F-DCFPyL PET/CT in 248 patients with biochemically recurrent prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019;46:1911-1918.

57. Calais J, Ceci F, Eiber M, et al. (18)F-fluciclovine PET-CT and (68)Ga-PSMA-11 PET-CT in patients with early biochemical recurrence after prostatectomy: a prospective, single-centre, single-arm, comparative imaging trial. ***Lancet Oncol.*** 2019;20:1286-1294.

58. Song H, Harrison C, Duan H, et al. Prospective Evaluation in an Academic Center of (18)F-DCFPyL PET/CT in Biochemically Recurrent Prostate Cancer: A Focus on Localizing Disease and Changes in Management. ***J Nucl Med.*** 2019.

59. Cytawa W, Seitz AK, Kircher S, et al. (68)Ga-PSMA I&T PET/CT for primary staging of prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019.

60. van Kalmthout LWM, van Melick HHE, Lavalaye J, et al. Prospective Validation of Gallium-68 PSMA-PET/CT in Primary Staging of Prostate Cancer Patients. ***J Urol.*** 2019:101097JU0000000000000531.

61. Herlemann A, Wenter V, Kretschmer A, et al. (68)Ga-PSMA Positron Emission Tomography/Computed Tomography Provides Accurate Staging of Lymph Node Regions Prior to Lymph Node Dissection in Patients with Prostate Cancer. ***Eur Urol.*** 2016;70:553-557.

62. Maurer T, Gschwend JE, Rauscher I, et al. Diagnostic Efficacy of (68)Gallium-PSMA Positron Emission Tomography Compared to Conventional Imaging for Lymph Node Staging of 130 Consecutive Patients with Intermediate to High Risk Prostate Cancer. ***J Urol.*** 2016;195:1436-1443.

63. Hatt M, Lucia F, Schick U, Visvikis D. Multicentric validation of radiomics findings: challenges and opportunities. ***EBioMedicine.*** 2019;47:20-21.

64. Heinzel A, Boghos D, Mottaghy FM, et al. (68)Ga-PSMA PET/CT for monitoring response to (177)Lu-PSMA-617 radioligand therapy in patients with metastatic castration-resistant prostate cancer. ***Eur J Nucl Med Mol Imaging.*** 2019;46:1054-1062.

65. Seitz AK, Rauscher I, Haller B, et al. Preliminary results on response assessment using (68)Ga-HBED-CC-PSMA PET/CT in patients with metastatic prostate cancer undergoing docetaxel chemotherapy. ***Eur J Nucl Med Mol Imaging.*** 2018;45:602-612.

66. Aggarwal R, Wei X, Kim W, et al. Heterogeneous Flare in Prostate-specific Membrane Antigen Positron Emission Tomography Tracer Uptake with Initiation of Androgen Pathway Blockade in Metastatic Prostate Cancer. ***Eur Urol Oncol.*** 2018;1:78-82.

67. Vaz S, Hadaschik B, Gabriel M, Herrmann K, Eiber M, Costa D. Influence of androgen deprivation therapy on PSMA expression and PSMA-ligand PET imaging of prostate cancer patients. ***Eur J Nucl Med Mol Imaging.*** 2019.

68. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. ***N Engl J Med.*** 2011;364:1995-2005.

69. Armstrong AJ, Halabi S, Luo J, et al. Prospective Multicenter Validation of Androgen Receptor Splice Variant 7 and Hormone Therapy Resistance in High-Risk Castration-Resistant Prostate Cancer: The PROPHECY Study. ***J Clin Oncol.*** 2019;37:1120-1129.

70. Heinlein CA, Chang C. Androgen receptor in prostate cancer. ***Endocr Rev.*** 2004;25:276-308.

71. Wang Y, Zhao N, Wu Z, et al. New insight on the correlation of metabolic status on (18)F-FDG PET/CT with immune marker expression in patients with non-small cell lung cancer. ***Eur J Nucl Med Mol Imaging.*** 2019.

72. Niemeijer AN, Leung D, Huisman MC, et al. Whole body PD-1 and PD-L1 positron emission tomography in patients with non-small-cell lung cancer. ***Nat Commun.*** 2018;9:4664.

73. Badawi RD, Shi H, Hu P, et al. First Human Imaging Studies with the EXPLORER Total-Body PET Scanner. ***J Nucl Med.*** 2019;60:299-303.

74. 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.

75. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? ***Eur J Nucl Med Mol Imaging.*** 2017;44:151-165.

76. Buvat I, Orlhac F. The dark side of radiomics: on the paramount importance of publishing negative results. ***J Nucl Med.*** 2019.

77. Administration USFD. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback. ***Published online at https://wwwfdagov/media/122535/download.*** 2019.


QI-1-PET

Addendum

Nederlandse samenvatting

Introductie

Kanker is een zeer veel voorkomende en ernstige ziekte, en is wereldwijd een grote oorzaak van morbiditeit en mortaliteit. Ondanks dat er grote vooruitgangen zijn geboekt in het screenen, diagnosticeren, en behandelen van kanker is het nog steeds de grootste doodsoorzaak in Nederland, waar ongeveer een derde van alle overlijdensgevallen kanker-gerelateerd zijn. In dit proefschrift worden twee veelvoorkomende soorten kanker onderzocht: prostaatkanker en longkanker.

Prostaatkanker is de meest voorkomende kankersoort bij mannen, met meer dan 12.000 nieuwe diagnoses per jaar in Nederland. Het is zowel klinisch als biologisch een heterogene ziekte die geclassificeerd kan worden als primair hormoongevoelige ziekte, biochemisch recidiverende hormoongevoelige ziekte, en gemetastaseerde of niet-gemetastaseerde castratie-resistente ziekte. Longkanker komt veel voor bij zowel mannen als (in minder mate) vrouwen, met ongeveer 13.000 nieuwe diagnoses per jaar in Nederland. Bij diagnose is de ziekte helaas bij veel patiënten al gemetastaseerd en is de prognose zeer matig. In dit proefschrift onderzoeken we het niet-kleincellig longcarcinoom (NSCLC).

De behandelingen van kanker zijn (grof gesteld) onder te verdelen in chirurgie, radiotherapie, en systemische therapie (o.a. chemo-, immuno- en hormoontherapie). Bij lokale ziekte wordt meestal gekozen voor chirurgie of radiotherapie (al dan niet voorgegaan door systemische therapie), terwijl bij gemetastaseerde ziekte wordt er veelal gekozen voor systemische therapie. Een uitzondering hierop zijn oligometastasen, waar het beperkte aantal metastasen lokale behandeling mogelijk maakt.

Ondanks dat er veel werkzame behandelingen zijn ontwikkeld, zijn veel middelen niet of matig werkzaam bij een deel van de patiënten. Dit betekent dat deze patiënten ineffectief behandeld worden, onnodige toxiciteit ondergaan, en dat deze zorg niet kosteneffectief is. Het is erg lastig om per patiënt te voorspellen of een behandeling tegen kanker gaat werken, of het effect van behandeling betrouwbaar vast te kunnen stellen. Van oudsher werden kankerpatiënten behandeld volgens vaststaande en rigide richtlijnen, waarin geen of weinig patiënt-specifieke informatie werd meegenomen. In de huidige medische zorg wordt geprobeerd om hier zo veel mogelijk van af te stappen, en de behandeling waar mogelijk af te stemmen op de individuele patiënt.

Om dit te kunnen doen zijn zogeheten *biomarkers* nodig. Een biomarker kan (vrij vertaald) gedefinieerd worden als een "karakteristiek dat objectief gemeten en geëvalueerd wordt als maat van normale biologische processen, pathogene processen, of farmacologische respons op een therapeutische interventie". Biomarkers kunnen dus gebruikt worden om per patiënt een rationele keuze te maken voor een bepaalde behandeling, of vroegtijdig ineffectieve behandeling te stoppen. Ze kunnen onder andere bepaald worden in bloed, weefsel (verkregen via biopten), of uit beeldvorming.

Het voordeel van beeldvorming is dat biomarkers zowel op patiënt- als op laesie-niveau geëxtraheerd en longitudinaal geëvalueerd kunnen worden. Een beeldvormingsmethode die zich in oncologie uitermate goed leent voor het verkrijgen van in-vivo kwantitatieve biomarkers is positron emissie tomografie (PET). In PET-beeldvorming worden radioactief-gelabelde stoffen (tracers) geïnjecteerd, waarvan de distributie vervolgens 3-dimensionaal gevisualiseerd en gekwantificeerd kan worden. Deze tracers zijn gericht op specifieke karakteristieken van maligne tumoren, en de opname ervan kan op PET gekwantificeerd worden om te voorspellen of te meten of een behandeling effectief is of zal zijn. We onderzoeken in dit proefschrift de tracers [^{18}F]fluoromethylcholine, [^{18}F]DCFPyL, en [^{18}F]fluorodihydro-testosteron ([^{18}F]FDHT) bij hormoongevoelig en castratie-resistent prostaatkanker, en [^{18}F]fluorodeoxyglucose ([^{18}F]FDG) en [^{18}F]fluorothymidine ([^{18}F]FLT) bij NSCLC. Door met PET te meten hoe veel van deze radioactieve tracers wordt opgenomen in tumoren, of hoe deze zich verdelen bínnen tumoren, kunnen we proberen te voorspellen hoe de ziekte zal verlopen of te meten of de behandeling effectief is. Echter, voordat deze kwantitatieve PET biomarkers in de klinische praktijk geïmplementeerd kunnen worden, is biologische, methodologisch, en klinische validatie noodzakelijk.

Het kwantificeren van traceropname in kleine tumoren is op PET een lastig probleem door de lage spatiele resolutie. In **deel 1** van dit proefschrift richten we ons op methodologische validatie van kwantitatieve PET biomarkers, waarbij we met name onderzoeken óf en hóe we kleine tumoren het beste kunnen kwantificeren en wat voor invloed dat heeft op de test-hertest betrouwbaarheid. Ook onderzoeken we hoe kunstmatige intelligentie ons kan helpen in het voorspellen van gemetastaseerde ziekte vanuit kleine prostaattumoren. In **deel 2** van dit proefschrift kijken we vervolgens naar de klinische toepassing van de PET-kwantificatie die we in deel 1 gevalideerd hebben.

Deel I – Methodologische validatie

In **hoofdstuk 2** onderzochten we de effectiviteit van *partial-volume correction* (PVC) methoden, die als doel hebben de PET-kwantificatie van traceropname in kleine laesies (<2 tot 3cm diameter) te corrigeren voor de beperkte resolutie van PET. Dit onderzochten we met gebruik van [^{18}F]FDG PET simulaties van longkankerpatiënten en fantoomscans (aldus in tamelijk ideale omstandigheden). We vonden dat deze PVC-methoden veel potentie hebben om de metingen nauwkeuriger te maken voor laesies tot 1cm diameter, soms wel tot 100% nauwkeurigheid, maar dat er een aantal methoden zijn die erg gevoelig zijn voor de tumorsegmentatie methode. Ook observeerden we dat deze gevoelige PVC methoden een sterk negatief effect konden hebben op de test-hertest betrouwbaarheid van [^{18}F]FCH opname in gemetastaseerd prostaatkanker en [^{18}F]FDG opname in NSCLC tumoren.

In **hoofdstuk 3** onderzochten we of PVC op voxel-niveau met en zonder gebruik van *denoising* in dynamische [^{18}F]FLT PET-CT scans bij NSCLC patiënten een effect had op de validatie van versimpelde maten van tracer opname voor het meten van respons op behandeling. De versimpelde maten waren de S*tandardized Uptake Value (SUV)* en de *Tumour-To-Blood ratio (TBR)*. We zagen dat de beperkte resolutie van PET over de tijd (tijdens de scan) variabel effect had op de tumor kwantificatie, en dat dit effect niet volledig gelijk was aan het effect op de versimpelde PET-maten. Echter, PVC had geen evident effect op de validatie van deze versimpelde PET maten voor het vaststellen van respons op behandeling. Het lijkt bij dit soort respons monitoring studies dus niet nodig om te corrigeren voor de beperkte resolutie van PET, zelfs bij kleine longtumoren en metastasen, als het volume van de tumor over de tijd niet stabiel blijft.

In **hoofdstuk 4** onderzochten we de invloed van ruis en PVC (conform de vernieuwde scanner-kalibratie richtlijn; EARL2) op de test-hertest betrouwbaarheid van [^{18}F]FDHT PET in gemetastaseerd castratie-resistent prostaatcarcinoom (mCRPC). Het doel was om te onderzoeken of we korter konden scannen met [^{18}F]FDHT PET, omdat patiënten met deze ziekte vaak pijnlijke uitzaaiingen hebben in het skelet en dus niet lang comfortabel in de scanner kunnen liggen. Ook wilden we de invloed van de nieuwe scanner-kalibratie richtlijn EARL2 op de betrouwbaarheid onderzoeken, omdat deze de komende jaren in de praktijk ingevoerd gaat worden. Deze vernieuwde richtlijn houdt in dat er bij PET-beeldreconstructie gebruikt wordt gemaakt van PSF-reconstructies, welke in principe een vorm van PVC zijn. We vonden dat zowel het verhogen

van de beeldruis van de PET scans (korter scannen) als het gebruik van PVC de test-hertest betrouwbaarheid vermindert, en [^{18}F]FDHT PET daarmee minder geschikt maakt voor het meten van respons op behandelingen. Dit effect was met name aanwezig voor SUVmax, wat de meest gebruikte maat voor tumor opname is wereldwijd. Echter, SUVpeak was veel minder aangedaan door de ruizigheid en toepassing van PVC, en zou dus bij voorkeur gebruikt moeten worden voor dit soort [^{18}F]FDHT PET scans.

Een andere nieuwe tracer voor PET beeldvorming van prostaatkanker is [^{18}F]DCFPyL. [^{18}F]DCFPyL is een prostaat-specifiek membraan antigeen (PSMA) ligand, dat zeer specifiek gericht is op prostaatkanker cellen. Door te kwantificeren hoe veel [^{18}F]DCFPyL-opname er is in tumoren kan (bijvoorbeeld) gemeten worden of een behandeling effectief is. Net zoals bij [^{18}F]FDHT, moet hiervoor de test-hertest betrouwbaarheid worden onderzocht, wat we in **hoofdstuk 5** gedaan hebben. We vonden bij patiënten met hormoongevoelig en castratie-resistent prostaatcarcinoom dat de verschillende maten voor PET-kwantificatie zeer betrouwbaar waren, en deze tracer dus geschikt maken voor gebruik in responsmetingen. Ook vonden we dat de totale tumor opname-maten (alle tumoren binnen patiënten gecombineerd) het meest betrouwbaar waren, wat deze maten erg geschikt maken voor het meten van effect van systemische therapie (bijvoorbeeld ^{177}Lu-PSMA-radioligandtherapie). Ook waren deze totale tumor opname-maten én SUVpeak ongevoelig voor gebruik van PVC (conform EARL2), wat de bruikbaarheid van deze maten in de klinische praktijk vergroot.

In **hoofdstuk 6** onderzochten we of we met *machine learning*, een vorm van kunstmatige intelligentie, en *radiomics* verkregen uit primaire prostaatkanker-lesies op [^{18}F]DCFPyL PET het risico op lymfekliermetastasen en hoog-risico factoren konden voorspellen. Radiomics betreft het uitvoerig kwantificeren van traceropname uit tumoren (in dit geval op PET), waarbij honderden maten van intensiteit, morfologie, en heterogeniteit verkregen worden. In dit onderzoek keken we daarbij specifiek naar de invloed van tumorsegmentatie en gebruik van PVC op de nauwkeurigheid van de voorspellingen van het machine learning model. We vonden dat de PSMA-radiomics zeer voorspellend waren voor de aanwezigheid van lymfekliermetastasen en hoge tumorgraad (Gleason score ≥8). Ook was het toepassen van PVC en een wat hogere afkapwaarde voor tumorsegmentatie (65-70% van SUVpeak) van positieve invloed op het kunnen voorspellen van lymfekliermetastasen. De resultaten uit deze studie zijn zeer veelbelovend, en toekomstige prospectieve studies met grotere cohorten dienen

zich richten op externe validatie om uit te wijzen of deze methode betrouwbaar genoeg is voor klinische implementatie.

Deel II – Klinische toepassing

Methodologische validatie van kwantitatieve PET biomarkers met betrekking tot hun nauwkeurigheid en betrouwbaarheid is van groot belang voor valide klinisch gebruik. Echter, evaluatie van de toegevoegde waarde van hun klinische toepassing is minstens net zo belangrijk. In de meeste technische PET-validatiestudies worden er geen klinische eindpunten meegenomen, omdat deze niet bij de doelstellingen van deze studies horen, of simpelweg niet mogelijk zijn door het design van de studie (bijvoorbeeld in een test-hertest studie). In dit deel van het proefschrift evalueerden we het voordeel van toepassing van PET-kwantificatie in verscheidene klinische settings.

In **hoofdstuk 7** hebben we de gepubliceerde literatuur over het gebruik van PVC in PET biomarker studies in oncologie systematisch samengevat en kritisch beoordeeld. Hierbij evalueerden we studies die keken naar het effect van PVC op het stellen van diagnose, de stadiëring van ziekte, het voorspellen van prognose, en het monitoren van respons op behandeling. Ook deden we een meta-analyse van studies naar het gebruik van [^{18}F]FDG PET voor het maken van onderscheid tussen benigne en maligne nodules in de long. We concludeerden dat, tot heden, het gebruik van PVC geen klinische toegevoegde waarde heeft gehad, behalve voor het voorspellen van prognose in patiënten met hoofdhalskanker. Een andere bevinding was dat de meeste studies gebruik maakten van een *recovery coefficient* methode voor PVC, welke de meest simpele methode is die vele aannames doet en enkel op laesie niveau kan corrigeren (en niet op voxelniveau). Naar aanleiding van deze systematische review deden we tevens aanbevelingen voor toekomstige studies, onder andere aangaande *i)* het gebruik van betere en meer uitvoerig gevalideerde PVC-methoden (conform hoofdstuk 2), *ii)* het bereiken van een consensus in de wetenschappelijke gemeenschap over welke methode geprefereerd is zodat resultaten van studies vergeleken kunnen worden, en *iii)* het gebruik van PVC-gecorrigeerde data niet als vervanging van ongecorrigeerde data, maar eerder als toevoeging hierop.

In **hoofdstuk 8** onderzochten we of we bij patiënten met oligometastasen van prostaatkanker bij biochemisch recidief konden voorspellen hoe lang zij progressie-vrij zouden blijven na stereotactische bestraling van deze metastasen. Oligometastasen worden gedefinieerd als een beperkt aantal metastasen waarbij

lokale behandeling mogelijk is. Op [^{18}F]FCH PET scans die vóór bestraling gemaakt werden kwantificeerden we de traceropname in de metastasen. Omdat de metastasen (met name lymfekliermetastasen) erg klein waren, hebben we PVC toegepast om de nauwkeurigheid van de metingen in deze laesies te vergroten. We concludeerden dat, zowel met als zonder gebruik van PVC, er geen significante voorspellende waarde zat in de intensiteit en metabole volumina van [^{18}F]FCH op PET. Echter, interessant genoeg was het aantal visueel detecteerbare metastasen wel voorspellend voor progressie-vrije overleving (in univariaat analyse). Deze studie is een voorbeeld van een setting waar, ondanks degelijke methodologische validatie, er geen klinische waarde van PET-kwantificatie is boven simpele klinisch gebruikte maten.

Uit een recente studie weten we dat de farmacokinetiek van [^{18}F]DCFPyL op PET gevoelig is voor grote tumorvolumina binnen patiënten. Dit betekent dat bij uitgebreide metastasering de inputfunctie in het blood anders is dan bij patiënten met enkele metastasen. De consequentie is dat SUV geen valide maat voor kwantificatie is, in tegenstelling tot TBR omdat deze corrigeert voor eventuele variaties in de inputfunctie. Ook betekent dit dat in studies naar responsmeting met [^{18}F]DCFPyL PET het gebruik van relatieve verschillen in SUV niet valide is als er grote behandelingseffecten zijn op de tumorvolumina binnen patiënten. Het nadeel van TBR is echter dat de betrouwbaarheid lager is dan SUV, omdat er toegevoegde variabiliteit in de bloedmetingen zit, wat met name aanwezig is bij gebruik van PVC beelden (EARL2). Dit zou verhinderen dat we met [^{18}F]DCFPyL PET subtiele verschillen tijdens behandeling kunnen meten. Eenzelfde observatie is eerder ook voor [^{18}F]FDHT gemaakt, waar SUV ook geen valide maat voor traceropname is. Gezien beide tracers zeer specifiek gericht zijn op een bepaalde eigenschap van de tumoren, lijkt dit dus een algemeen effect te zijn voor dit soort tracers. In **hoofdstuk 9** bespraken we dit effect, waarbij we ons commentaar leverden op een recent gepubliceerde studie die geen relatie vond tussen [^{18}F]DCFPyL opname in achtergrondweefsel en het totale tumor volume. Echter, in deze studie werden enkel patiënten met lage tumor volumina geïncludeerd. We bespraken deze resultaten in het licht van de eerder onderzochte farmacokinetiek van [^{18}F]DCFPyL. Ook lieten we met een klinisch voorbeeld zien dat dit duidelijke consequenties kan hebben voor het monitoren van respons op behandeling bij een patiënt met mCRPC die behandeld werd met abirateron, en zowel [^{18}F]DCFPyL als [^{18}F]FDHT PET onderging voor en tijdens behandeling. Bij beide tracers liet SUV globaal stabiele ziekte zien, terwijl TBR duidelijk progressieve ziekte aangaf welke klinisch bevestigd werd.

In **hoofdstuk 6** hebben we een PSMA-radiomics met machine learning analyse voor risicostratificatie van patiënten met primair prostaatkanker geoptimaliseerd en gevalideerd. Op basis van deze resultaten vonden we dat het meten van PSMA in de primaire tumor een veelbelovende nieuwe biomarker is, naast de gebruikelijke tumorgradering (Gleason score), PSA, en klinisch tumorstadium. Het zou dan ook mogelijk de risicostratificatie van deze patiënten kunnen verbeteren, zodat minder patiënten onnodig een pelviene lymfeklierdissectie hoeven te ondergaan. In **hoofdstuk 10** bespraken we deze bevindingen kort, gericht op de urologische gemeenschap. Hierbij bespraken we enkele vergelijkbare studies, die dezelfde positieve resultaten lieten zien voor het voorspellen van lymfekliermetastasen op basis van PSMA-expressie in de primaire tumor op PET. Ook vergeleken we de voorspellende waarde van deze PSMA-radiomics ten opzichte van de standaard klinisch gebruikte modellen (bijvoorbeeld nomogrammen), en vonden dat (in ons cohort) de PSMA-radiomics beter waren in het voorspellen van kliermetastasen, hoge Gleason score, en kapseldoorbraak van de tumor. In een van de andere onderzoeken was gebleken dat de PSMA-expressie op PET ongeveer even goed kon voorspellen of er kliermetastasen waren dan de klinische modellen, maar in dit onderzoek werden simpele PET maten gebruikt en geen radiomics. Al met al is het meten van PSMA-expressie op PET een veelbelovende nieuwe speler in de biomarker-groep van prostaatkanker, maar toekomstige studies moeten uitwijzen wat de exacte toegevoegde waarde is ten opzichte van bestaande modellen, en hoe betrouwbaar deze nieuwe biomarker is in een multicenter setting.

Conclusie

In dit proefschrift hebben we onderzocht of en hoe we kleine tumoren nauwkeurig en betrouwbaar kunnen kwantificeren op PET in primair, recidiverend, en castratie-resistent prostaatkanker en niet-kleincellig longkanker. Ook evalueerden we wat de toepassing hiervan oplevert in klinische settings. We vonden dat ondanks de lage resolutie van PET we traceropname in tumoren ≥1cm nauwkeurig konden kwantificeren, als PVC methoden zorgvuldig gevalideerd waren. In literatuur vonden we dat PVC zeer beperkt en veelal matig gevalideerd werd toegepast. De nauwkeurigheid van gevalideerde PVC had een beperkt negatief effect op de betrouwbaarheid van de PET kwantificatie voor [^{18}F]FCH, [^{18}F]FDG, [^{18}F]FDHT, en [^{18}F]DCFPyL, wat subtiele responsevaluatie met deze tracers mogelijk kan verhinderen. In farmacokinetische validatie van

responsmetingen in longkanker met versimpelde PET-maten met dynamische PET bleek het toepassen van PVC geen toegevoegde waarde te hebben. [^{18}F]FCH PET kwantificatie, met of zonder PVC, bleek niet van toegevoegde waarde in het voorspellen van progressie-vrije overleving in patiënten met oligometastasen van prostaatkanker. In het voorspellen van lymfekliermetastasen op basis van [^{18}F] DCFPyL-radiomics met kunstmatige intelligentie, bleken zowel toepassing van PVC als het optimaliseren van tumorsegmentatie de voorspellingen te verbeteren. Toekomstige methodologische PET validatiestudies zouden bij voorkeur ook klinische eindpunten moeten incorporeren.

Dankwoord

Op de voorzijde van dit proefschrift pronkt een prachtige Volvo Amazon uit 1967. Net als de PET-scanner die in dit proefschrift onderzocht werd, staat deze auto voor betrouwbare en vooruitstrevende techniek. Zoals een klassieke Volvo niet zonder brandstof kan en doelloos is zonder bestuurder, kan ik dit proefschrift niet zonder de nodige hulp en samenwerking hebben gerealiseerd. Daarom wil ik graag een aantal mensen in het bijzonder bedanken voor hun bijdrage aan de totstandkoming van dit proefschrift.

Otto Hoekstra, beste Otto, dank dat je mij de wereld van de (radiologie en) nucleaire geneeskunde hebt binnengehaald. Je bevlogenheid en passie voor dit vak en de wetenschap zijn aanstekelijk, en ik bewonder je kunde om dit over te brengen op anderen!

Ronald Boellaard, beste Ronald, bedankt voor alle begeleiding en de vele uren waarop je beschikbaar was om resultaten te bespreken. Je hebt de unieke gave om complexe zaken voor ieder inzichtelijk te maken, en daar waar onderzoeken vastlopen toch net die éne uitkomst te vinden.

Daniela Oprea-Lager, beste Daniela, sinds het begin van mijn onderzoek ben jij nauw betrokken geweest bij mijn werk en heb ik zeer plezierig met je samengewerkt. Met jouw enthousiasme en optimisme is geen probleem te groot!

Jens Voortman, beste Jens, bedankt voor de prettige samenwerking (al was het met name op juridisch, ethisch, en administratief gebied). Hopelijk kunnen we in de toekomst nog veel samenwerken!

Geachte prof.dr. Jeroen van Moorselaar, prof.dr. Fons van den Eertwegh, prof.dr. Hugo de Jong, prof.dr. Sigrid Stroobants, prof.dr. Kees van Kuijk, en dr. Bart de Keizer, dank voor het lezen en kritisch beoordelen van mijn proefschrift. Prof.dr. Irene Buvat, merci d'avoir pris le temps de lire et d'évaluer ma thèse!

Gem Kramer, beste Gem, toen ik als jonge student op de afdeling binnen kwam heb jij me onder jouw vleugels genomen en alle ins en outs van het 'onderzoeker-zijn' bijgebracht, vaak onder het genot van een biertje! Dank daarvoor en voor de fijne samenwerking!

Bernard Jansen, beste Bernard, bedankt dat we zulke mooie dingen hebben kunnen doen, zowel op wetenschappelijk gebied als daarbuiten. Hopelijk kunnen we deze trend in de toekomst doorzetten!

Thomas Koopman, beste Thomas, de afgelopen jaren hebben we samen vele kopjes koffie en biertjes mogen nuttigen, en mooie relativerende discussies gevoerd over zowel de zinnige als onzinnige kanten van ons werk. Bedankt daarvoor!

Fiona Heeman, beste Fiona, ik heb met veel plezier de kamer met je gedeeld. Bedankt voor deze leuke tijd! Daarnaast ben je de beste kattenoppas die ik ooit heb gehad!

Bart de Vries, beste Bart, samen brachten we de sfeer op onze kamer tot een hoger niveau (soms tot ongenoegen van bovenstaande)! Je wordt vast een goed technisch geneeskundige, en ik hoop niet dat je mijn vak gaat automatiseren.

Cemile Karga, beste Cemile, bedankt (en sorry) voor het oneindige (uit)lenen van je nietmachine, perforator, ontnieter, post-its, schaar, enzovoorts!

Maqsood Yaqub, beste Maqsood, bedankt voor al je hulp en geduld bij de ondersteuning van de PET-analyses en het gebruik van al die verschillende (error-gevoelige) tools.

Roland Martens, beste Roland, dank voor de prettige samenwerking.

Corinne, Yvonne, Coreline, Dennis, Sandeep, Sanne, Hayel, Eline, Julia, Marijke, Marc, en alle andere collega's op afdeling 1F en de J-vleugel: bedankt voor alle gezellige lunches, koffiepauzes, en borrels!

Beste Amanda Kroonenberg-Kalwij, Greet Stolker-Bouknecht, en alle andere medewerkers van het secretariaat Radiologie en Nucleaire Geneeskunde, bedankt voor de vele hulp met het regelen en inplannen van afspraken, aanstellingen, financiële zaken, enzovoorts!

Judith van Es, bedankt voor al je hulp in het regelen van scanprotocollen, reconstructies, en planningen. Ook wil ik alle laboranten bedanken voor hun uitstekende werk.

Beste André Vis, Jeroen van Moorselaar, en Yves Bodar, bedankt voor de prettige en goede samenwerking op het gebied van prostaatkanker imaging.

Prof.dr. Morris, dear Michael, thank you for the collaboration for the PCF project.

Ben Zwezerijnen, bedankt voor je loopbaanadviezen!

Beste collega's van de afdeling Medische Oncologie, bedankt voor het aanhoren van mijn PET-praatjes. Ramsha Iqbal, bedankt voor je adviezen!

Riekie de Vet, bedankt voor al je inzet en advies bij het schrijven van de systematische review!

Tim van de Brug, bedankt voor de statistische adviezen bij het uitvoeren van de machine learning analyses!

Voor zover nog niet benoemd wil ik al mijn co-auteurs bedanken voor het meedenken en meeschrijven aan de artikelen in dit proefschrift.

~

Aan mijn vrienden en familie: bedankt voor het verzorgen van de benodigde geestelijke ontspanning in de afgelopen jaren.

Lieve Arianne, Arjan, en kleine Thijs, bedankt voor alle gezellige momenten samen.

Lieve papa en mama, zonder jullie was ik niet gekomen waar ik nu ben. Bedankt voor jullie steun en alles wat jullie mij gegeven hebben.

Liefste Elise, jij bent degene die mij in goede banen leidt. Dank voor al je bemoedigende woorden en ondersteuning in de afgelopen jaren.

List of publications

Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial-Volume Correction in Oncological PET/CT Studies. *J Nucl Med.* 2016;57(10):1642-1649.

Bouman-Wammes EW, van Dodewaard-de Jong, Dahele M, **Cysouw MCF** et al. Benefits of using stereotactic body radiotherapy in patients with metachronous oligometastases of hormone-sensitive prostate cancer detected by [^{18}F]fluoromethylcholine PET/CT. *Clin Genitourin Cancer.* 2017;15(5):e773-e782.

Cysouw MCF, Kramer GM, Frings V, et al. Baseline and longitudinal variability of normal tissue uptake values of [^{18}F]-fluorothymidine-PET images. *Nucl Med Biol*. 2017;51:18-24.

Cysouw MCF, Kramer GM, Schoonmade LJ, et al. Impact of partial-volume correction in oncological PET studies: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. 2017;44(12):2105-2116.

Cysouw MCF, Bouman-Wammes EW, Oprea-Lager DE, et al. Prognostic value of [^{18}F]-fluoromethylcholine positron-emission tomography before stereotactic body radiotherapy for oligometastatic prostate cancer. *Int J Radiat Oncol Biol Phys*. 2018 Jun 1;101(2):406-410.

Jansen BHE, Kramer GM, **Cysouw MCF**, et al. Healthy Tissue Uptake of ^{68}Ga-Prostate Specific Membrane Antigen (PSMA), ^{18}F-DCFPyL, ^{18}F-Fluoromethylcholine (FCH) and ^{18}F-Dihydrotestosterone (FDHT). *J Nucl Med*. 2019 Jan 10.

Cysouw MCF, Golla SVS, Frings V, et al. Partial-volume correction in dynamic PET-CT: effect on tumor kinetic parameter estimation and validation of simplified metrics. *EJNMMI Res*. 2019 Feb 4;9(1):12.

Jansen BHE, Yaqub M, Voortman **J, Cysouw MCF**, et al. Simplified Methods for Quantification of ^{18}F-DCFPyL Uptake in Patients with Prostate Cancer. *J Nucl Med*. 2019 Apr 18.

Cysouw MCF, Kramer GM, Heijtel D, et al. Sensitivity of ^{18}F-fluorodihydrotestosterone PET-CT to count statistics and reconstruction protocol in metastatic castration-resistant prostate cancer. *EJNMMI Res.* 2019 Jul; 9:70.

Jansen BHE, Yaqub MM, **Cysouw MCF**, et al. Reply to: Quantification of ^{18}F-DCFPyL Uptake: TBR versus Patlak's Analysis. *J Nucl Med.* 2019 Aug 26.

Cysouw MCF, BHE Jansen, Yaqub MM, et al. Letter to the Editor Re: Semiquantitative Parameters in PSMA-Targeted PET Imaging with [^{18}F]DCFPyL: Impact of Tumor Burden on Normal Organ Uptake. *Mol Imaging Biol.* 2019 Oct.

Jansen BHE, **Cysouw MCF**, Vis AN, et al. Repeatability of Quantitative ^{18}F-DCFPyL PET/CT Measurements in Metastatic Prostate Cancer. *J Nucl Med.* 2020 Jan 10.

Cysouw MCF, Jansen BHE, van de Brug T, et al. Machine learning-based analysis of [^{18}F]DCFPyL PET radiomics for risk stratification in primary prostate cancer. *Eur J Nucl Med Mol Imaging.* 2020 Jul 31.

Abstracts and presentations

Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial Volume Correction in Oncological PET/CT Studies. Poster presentation. Society of Nuclear Medicine & Molecular Imaging Annual Meeting, San Diego, USA, 2016.

Cysouw MCF, Kramer GM, Hoekstra OS, et al. Accuracy and Precision of Partial Volume Correction in Oncological PET/CT Studies. Oral presentation. Nederlandse Vereniging voor Klinische Fysica Annual Meeting, Zeist, the Netherlands, 2016.

Cysouw MCF, Golla SVS, Frings V, et al. Partial-volume correction in dynamic PET-CT: effect on tumor kinetic parameter estimation and validation of simplified metrics. Oral presentation. Cancer Center Amsterdam Annual Retreat, Noordwijk aan Zee, the Netherlands, 2019.

Cysouw MCF, Kramer GM, Heijtel D, et al. Impact of scan statistics on accuracy, repeatability and lesion detectability of whole body [^{18}F]-FDHT PET-CT in metastatic castration resistant prostate cancer. Oral presentation. Society of Nuclear Medicine & Molecular Imaging Annual Meeting, Anaheim, USA, 2019.

Cysouw MCF, Jansen BHE, van der Zande KC, et al. Radiomics from [^{18}F]PSMA PET-CT with machine learning as a novel biomarker in primary prostate cancer. Oral presentation. European Association of Nuclear Medicine Annual Meeting, Barcelona, Spain, 2019.

Cysouw MCF, Jansen BHE, Oprea-Lager DE, et al. Radiomics from ^{18}F-PSMA PET-CT analysed with machine learning for risk stratification in primary prostate cancer patients. Nederlandse Vereniging voor Nucleaire Geneeskunde Annual Meeting, Amersfoort, the Netherlands, 2019.

List of abbreviations

[^{18}F]DCFPyL	^{18}F-bound PSMA-ligand
[^{18}F]FCH	^{18}F-fluoromethylcholine
[^{18}F]FDG	2-deoxy-2-(^{18}F)fluoro-D-glucose
[^{18}F]FDHT	^{18}F-fluorodihydrotestosterone
[^{18}F]FLT	^{18}F-fluorothymidine
[^{18}F]PSMA	^{18}F-prostate-specific membrane antigen
AC	activity concentration
ADT	androgen deprivation therapy
AI	artificial intelligence
AR	Androgen receptor
ARSi	androgen receptor signalling inhibitor
AUC	area-under-the-curve
BLOB-OS-TF	iterative time-of-flight reconstruction algorithm
BPND	non-displaceable binding potential
CNR	contrast-to-noise ratio
COV	coefficient of variation
CT	computed tomography
DNA	deoxyribonucleic acid
EANM	European Association of Nuclear Medicine
EARL	EANM Research Ltd.
ECE	extracapsular tumor extension
EGFR	endothelial growth factor receptor
ePLND	extended pelvic lymph node dissection
FWHM	full-width-at-half-maximum
GLCM	grey-level co-occurrence matrices
GLDZM	grey-level distance zone matrices
GLRLM	grey-level run length matrices
GLSZM	grey-level size zone matrices
GS	Gleason score
HR	hazard ratio
HYPR	highly-constrained backprojection
IBSI	image biomarker standardization initiative
ICC	intraclass correlation coefficient
IDIF	image-derived input function

IQR	interquartile range
KBq	Kilobecquerel
Kg	kilogram
LNI	lymph node involvement
LOR	line-of-response
LR	Lucy-Richardson
MATV	metabolically active tumor volume
MBq	Megabecquerel
mCRPC	metastatic castration-resistant prostate cancer
ML	machine learning
MRI	magnetic resonance imaging
NGLDM	neighboring grey-level dependence matrices
NGTDM	neighborhood grey-tone difference matrices
NSCLC	non-small cell lung cancer
PCa	Prostate cancer
PCA	principal component analysis
PERCIST	PET response assessment criteria
PET	positron emission tomography
PFS	progression-free survival
PP	parent plasma
PSA	prostate-specific antigen
PSF	point-spread function
PSMA	prostate-specific membrane antigen
PVC	partial-volume correction
PVE	partial-volume effect
RC	repeatability coefficient
ROC	receiver-operator characteristics
SBRT	stereotactic body radiotherapy
SD	standard deviation
SMOTE	Synthetic Minority Oversampling Technique
SUV	Standardized Uptake Value
TBR	tumor-to-blood ratio
TKI	tyrosine kinase inhibitor
TLCU	total lesion choline uptake
TLG	total lesion glycolysis
TLU	total lesion uptake

TOF	time-of-flight
TTB	total tumor burden
TTV	total tumor volume
VOI	volume of interest
V_T	Volume of distribution

Curriculum Vitae

Matthijs C.F. Cysouw was born in Terneuzen, the Netherlands, on the 19th of February 1993. After graduating bilingual secondary school at the Zeldenrust Steelantcollege in 2011, he moved to Amsterdam to study at the Vrije Universiteit Amsterdam. After one year of studying Biomedical Sciences, he started his Medicine study at the VUmc in 2012.

Due to his interest in technology and medical imaging, he joined the Honours Programme in his Bachelor's and started doing scientific research at the department of Radiology and Nuclear Medicine under supervision of prof.dr. Otto Hoekstra and prof.dr. Ronald Boellaard. His research focussed on quantitative imaging of oncological PET studies, primarily investigating the use of molecular imaging in prostate cancer.

After successfully finishing his Honours Programme at VUmc, Matthijs ensued with an MD-PhD Programme, continuing his research at the department of Radiology and Nuclear Medicine. During this period, he combined his clinical Master's internships with a PhD project. He did his pre-doctor internship at the department of Medical Oncology and Hematology at the Spaarne Gasthuis. In September 2018, Matthijs obtained his medical degree and started working on his research full-time to finish the PhD project that resulted in the present thesis.

In January 2020, Matthijs started his residency in Radiology and Nuclear Medicine at the Amsterdam UMC (location VUmc).