Quantitative Characterisation of Solid Tumours by $^{18}$F-FDG PET: What’s in a Number?
Quantitative Characterisation of Solid Tumours by $^{18}$F-FDG PET

What’s in a Number?

Dennis Vriens
Quantitative Characterisation of Solid Tumours by $^{18}$F-FDG PET: *What’s in a Number?*; thesis, Radboud University, Nijmegen, the Netherlands.

The research presented in this thesis was performed at the dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands and was additionally financially supported by the Netherlands Organisation for Health Research and Development (ZonMW), grant 9200.3552. Printing of this thesis was financially supported by the Radboud University, Nijmegen, the Netherlands. All funding is gratefully acknowledged.

**Cover illustration:**
Artist impression of the 3D static whole-body $^{18}$F-FDG PET/CT of the 75-year-old female patient with a 45 mm moderately differentiated squamous cell carcinoma without lymphnode metastases ($pT_2aN_0cM_0$) centrally located in the right lower lobe, also described in *figure 7 (page xxvi)*. Mean and maximum intensity projections of $^{18}$F-FDG PET and $^{18}$F-FDG PET fused with CT are displayed. Numerical matrices of the same images are shown.

Cover design: Dennis Vriens
Lay-out and editing: Dennis Vriens (typeset using \LaTeX)
Printing: Gildeprint - the Netherlands (first edition, 500 prints)

© 2015, D. Vriens, Nijmegen, the Netherlands

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronically, mechanically, by photo-copying, recording or otherwise, without the prior written permission of the holder of the copyright. The copyright of the articles that have been published has been transferred to the respective journals. Copyright of published work: Springer (*chapter 1*), Minerva Medica (*chapter 2, 4*), Lippincott Williams & Wilkins (*chapter 3*), Society of Nuclear Medicine & Molecular Imaging (*chapter 5, 7*), Elsevier (*chapter 6, 8*), John Wiley & Sons (*chapter 9*) and the Endocrine Society (*chapter 10*).
Quantitative Characterisation of Solid Tumours by $^{18}$F-FDG PET

*What’s in a Number?*

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 30 april 2015
om 12.30 uur precies

door

**Dennis Vriens**
geboren op 15 april 1981
te Tilburg
Promotoren:
Prof. dr. W.J.G. Oyen
Prof. dr. L.F. de Geus-Oei (Leids Universitair Medisch Centrum / Universiteit Twente)

Copromotor:
Dr. E.P. Visser

Manuscriptcommissie:
Prof. dr. W.R. Gerritsen (voorzitter)
Prof. dr. R. Boellaard (VU Medisch Centrum)
Dr. J. Bussink
Quantitative Characterisation of Solid Tumours by $^{18}$F-FDG PET

What’s in a Number?

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. Th.L.M. Engelen,
according to the decision of the Council of Deans
to be defended in public on Thursday, April 30, 2015
at 12.30 hours

by

Dennis Vriens
Born on April 15, 1981
in Tilburg (the Netherlands)
Supervisors:
Prof. W.J.G. Oyen, MD, PhD
Prof. L.F. de Geus-Oei, MD, PhD (Leiden University Medical Centre / University of Twente)

Co-supervisor:
E.P. Visser, PhD

Doctoral Thesis Committee:
Prof. W.R. Gerritsen, MD, PhD (chair)
Prof. R. Boellaard, PhD (VU University Medical Centre)
J. Bussink, MD, PhD
to my parents
PRELUDE

“Nothing exists until it is measured.”
- Niels Bohr, 1930
Table of Contents
# Table of Contents

<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prologue</td>
<td>xv</td>
</tr>
<tr>
<td>$^{18}$F-FDG PET/CT in Solid Tumours</td>
<td>xvii</td>
</tr>
<tr>
<td>Aims, Scope &amp; Outline of the Doctoral Thesis</td>
<td>xxix</td>
</tr>
<tr>
<td>I  Methodological Considerations &amp; Clinical Relevance</td>
<td>1</td>
</tr>
<tr>
<td>1 Methodological Considerations in Quantification of $^{18}$F-FDG PET studies</td>
<td>3</td>
</tr>
<tr>
<td>2 Tailoring Therapy in Colorectal Carcinoma by $^{18}$F-FDG PET/CT</td>
<td>29</td>
</tr>
<tr>
<td>II Optimisation of Quantification Methodology</td>
<td>55</td>
</tr>
<tr>
<td>3 Evaluation of Different Normalisation Procedures for the Calculation of the Standardised Uptake Value in Therapy Response Evaluation Studies</td>
<td>57</td>
</tr>
<tr>
<td>4 Comparison of Two Volume-of-Interest Definition Methods for Metabolic Response Evaluation with $^{18}$F-FDG PET</td>
<td>71</td>
</tr>
<tr>
<td>5 A Curve-Fitting Approach to Estimate the Arterial Plasma Input Function for the Assessment of Glucose Metabolic Rate and Response to Treatment</td>
<td>87</td>
</tr>
</tbody>
</table>
# Table of Contents

## III Tissue Heterogeneity & Therapy Response Evaluation 101

6 Quantitative Assessment of Heterogeneity in Tumour Metabolism using \(^{18}\)F-FDG PET 103

7 Chemotherapy Response Evaluation of Colorectal Liver Metastases by DCE-MRI Perfusion Parameters and 4D Dynamic \(^{18}\)F-FDG PET Metabolic Rate 115

8 Vascular and Metabolic Response to Bevacizumab-Containing Regimens in Two Patients with Colorectal Liver Metastases Measured by DCE-MRI and 4D Dynamic \(^{18}\)F-FDG PET 131

## IV Clinical Impact of Molecular Imaging 141

9 The Role of \(^{18}\)F-FDG PET in Thyroid Nodules with Indeterminate Fine-Needle Aspiration Cytology: Systematic Review and Meta-Analysis of the Literature 143

10 Cost-Effectiveness of \(^{18}\)F-FDG PET/CT for Cytologically Indeterminate Thyroid Nodules: a Decision Analytic Approach 161

## Epilogue 183

General Discussion & Future Prospects 185

Summary & Conclusions 205

Samenvatting & Conclusies in het Nederlands voor niet-ingewijden (Layman’s Summary & Conclusions in Dutch) 215

## Appendices 227

Glossary 229

References 239

List of Tables 203

List of Figures 205

Curriculum Vitae 267

List of Publications 271

Acknowledgements 277
18F-FDG PET/CT in Solid Tumours

Unpublished

D. Vriens1

1dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands
Impact of cancer from a societal perspective

Since 2008, cancer has surpassed cardiovascular diseases as primary cause of death in the Western World. In the Netherlands in 2012, 27% of all deaths were attributable to cardiovascular disease and almost 32% were due to malignancies. This is mainly due to a steep decline in mortality related to cardiovascular diseases as compared to a relatively slow decline in cancer-related mortality. It is estimated that over 40% of people will be diagnosed with any type of malignancy during their lives, half of them before the age of 66 years [1].

The economic burden of cancer is enormous: it was estimated that this group of diseases costs the European Union €126 billion in 2009. This number is based on healthcare costs (€51 billion), loss of productivity due to early death (almost €43 billion), loss of working days (almost €10 billion) and informal care costs (over €23 billion) [2].

The most common new cases of cancer are solid tumours, carcinomas of the prostate, breast, lung, colon or rectum and melanoma being the top five. Together, lung carcinoma and colorectal carcinoma have the highest absolute mortality whereas the prognosis of prostate and breast carcinoma is much more favourable [3].

Imaging in staging & therapy response evaluation of cancer

Once a diagnosis of malignancy is established, treatment and prognosis of a particular cancer are mainly determined by both patient- and tumour-associated factors. Main tumour-associated factors are tumour grade (differentiation) and its extent (stage) either being in situ, localised, regional or distant. The extent of most solid tumours is categorised using the TNM (Tumour, Node, Metastasis) staging system. Based on history taking, physical examination and laboratory investigation of tumour markers, together with imaging, the stage of the disease is determined. Frequently used imaging modalities include endoscopy, ultrasonography (US), conventional radiography (CR), computed tomography (CT), Magnetic Resonance Imaging (MRI) and molecular imaging techniques, such as Single Photon Emission Tomography (SPECT) and Positron Emission Tomography (PET).

Morphological imaging such as US, CR, CT and MRI are regularly performed for specific regions of the body only and are thereby especially helpful in determination of the local tumour and regional lymphnode stages. For determination of distant metastases, morphological imaging techniques are targeted to most common metastasis locations, e.g., a liver US and chest CR for colorectal carcinoma.

Compared to molecular imaging, morphological imaging techniques have their drawbacks: often, functional alterations precede morphological changes, both during development from benign to cancerous tissue and in tumour therapy response. Furthermore, morphological changes can be non-specific, especially after treatment has been initiated. Finally, most morphological imaging techniques are not whole-body modalities.
Molecular imaging can be performed by using tracer amounts of substrates, providing images for qualitative assessment, but also quantitative information on molecular kinetics, receptor expression or tissue metabolite composition. The latter can be measured by nuclear magnetic resonance spectroscopy (MRS).

Tracers can be visualised and quantified either by labelling them with a fluorophore for optical imaging or to a radioactive isotope for scintigraphy. Optical imaging has the inherent drawback of a limited penetration depth; for most tissues this is only a few millimetres. Therefore tracers are labelled with isotopes that either emit photons (\(\gamma\)- or X-rays; \(e.g., \) \(^{67}\)Ga, \(^{81m}\)Kr, \(^{99m}\)Tc, \(^{111}\)In, \(^{123}\)I, \(^{133}\)Xe or \(^{201}\)TI) or positrons (anti-electrons or \(\beta^+\)-particles; \(e.g., \) \(^{11}\)C, \(^{13}\)N, \(^{15}\)O, \(^{18}\)F, \(^{64}\)Cu, \(^{68}\)Ga, \(^{82}\)Rb, \(^{89}\)Zr or \(^{124}\)I). Photon emitting tracers are tomographically imaged using SPECT. PET is used to localise the photon-pair arising from the annihilation of an electron with the positron derived from positron-emitting radiopharmaceuticals. SPECT and PET have much higher detection sensitivities (picomolar-nanomolar range) than MRS. Their major drawbacks are a relatively poor spatial resolution in the multi-millimetre range, radiation exposure to patients and personnel and costs. Opposed to these drawbacks, their advantages are a high detection sensitivity and their ability of whole-body imaging.

Overall, PET has a much higher detection sensitivity and spatial resolution than SPECT. Also PET is fully quantitative, whereas SPECT still has to overcome technical difficulties. However, PET-isotopes may require resolution of chemical and logistical challenges. Currently SPECT and PET are generally performed in combination with morphological imaging (CT or MRI). This hybrid imaging has the additional benefits of improving localisation of lesions, diagnostic sensitivity and specificity. Finally, the addition of morphological imaging techniques aids in correcting for tissue photon attenuation.

Both morphological and molecular imaging are currently being employed for therapy response evaluation. Morphological therapy response evaluation involves the measurement of the diameter of target lesions. In January 2009 an updated version of the widely used ‘response evaluation criteria in solid tumours’ (RECIST) was published. This document provides the criteria to define a disease as either partially or completely responding to treatment or being stable or progressive despite treatment. This latest version allows incorporation of \(^{18}\)F-FDG PET to complement CT scanning in assessment of disease progression only. At almost the same time the PET Response Criteria in Solid Tumours (PERCIST) were published, succeeding the 1999 European Organisation for Research and Treatment of Cancer (EORTC) standard. These new criteria propose a draft framework for standardised therapy response evaluation using quantitative PET/CT. Based on the changes in \(^{18}\)F-FDG uptake, the outcome of cancer treatment is classified a complete metabolic response, a partial metabolic response, stable disease or progressive metabolic disease.

Molecular imaging is not limited to nuclear medicine imaging techniques, as significant advances have recently been made in developing imaging techniques that utilise US, CT, MRI and optical imaging.
Positron emission tomography using a glucose analogue

In molecular imaging of cancer, the most commonly used tracer is the non-metabolisable glucose analogue 2-(18)F)fluoro-2-deoxy-D-glucose (18F-FDG) that is injected intravenously. 18F-FDG consists of a tracer (2-deoxy-D-glucose, DG) radio-labelled with a positron emitting radionuclide (18F), which can be visualised with PET (figure 1, page xxi). 18F-FDG PET reflects in vivo whole-body glucose metabolism and therefore is a highly sensitive investigation for many types of malignancies.

The metabolic fate of DG is reviewed elsewhere [9] and briefly described here (figure 2, page xxii). DG, like D-glucose, enters the cell by the facilitative membrane-bound sodium-independent glucose transporters (GLUT). Once inside the cell, both D-glucose and DG are catabolised by the glycolytic pathway. D-glucose is first catabolised in multiple steps to pyruvate in the cytosol. Thereafter, dependent on the oxygen supply of the cell, pyruvate is either catabolised in the mitochondrion using oxidative phosphorylation by the tricarboxylic acid (Krebs) cycle to carbon dioxide and water (aerobic glycolysis) or in the cytosol using fermentation to lactate (anaerobic glycolysis). The energy yield of the former is much higher than the latter; theoretically 36-38 molecules of adenosine triphosphate (ATP) are formed per molecule glucose during aerobic glycolysis vs only two molecules of ATP during anaerobic glycolysis. Anaerobic glycolysis is faster, but less efficient.

DG enters the same metabolic pathway: it is taken up by cells using the GLUTs, followed by phosphorylation to DG-6-phosphate by hexokinase. DG-6-phosphate is itself an allosteric inhibitor of hexokinase. As the selective phosphoglucone isomerase enzyme has no affinity for DG-6-phosphate, opposed to D-glucose-6-phosphate, this metabolite cannot be catabolised further. Once phosphorylated, DG-6-phosphate cannot escape the cell and conversion back to DG by glucose-6-phosphatase (gluconeogenesis) is in most cases impossible due to a very low concentration of this enzyme in most tissues and most tumours. Therefore, DG-6-phosphate accumulates in the cytosol and with it, the bound 18F. In short, 18F-FDG accumulation reflects cellular energy needs.

Cancer cells tend to favour the inefficient anaerobic glycolysis over aerobic glycolysis, even under normoxic conditions. This is a result of the mutations that occur during malignant transformation of a cell and is known as the Warburg effect, postulated by Otto Warburg in 1956 [10]. The high energy expenditure of tumour cells

![Figure 1: Structure of D-glucose (left), 2-deoxy-D-glucose (middle) and 2-(18)F)fluoro-2-deoxy-D-glucose (right).](image-url)
FDG PET/CT in Solid Tumours

Figure 2: Differences in metabolic fate of D-glucose and 2-((18)F)fluoro-2-deoxy-D-glucose. ATP: adenosine triphosphate; CO₂: carbon dioxide; EES: extravascular extracellular space; 18F-FDG: 2-(18F)fluoro-2-deoxy-D-glucose; GLUT: facilitative membrane-bound sodium-independent glucose transporters; H₂O: water; K₁: transmembranous influx rate of GLUT; k₂: transmembranous efflux rate of GLUT; k₃: hexokinase rate; k₄: glucose-6-phosphatase rate; k₅/k₆: phosphoglucose isomerase rate; O₂: oxygen; -6P: -6-phosphate; TCA: tricarboxylic acid cycle.

combined with inefficient fermentative glycolysis causes malignant, rapidly growing tumour cells to typically have glycolytic rates up to 200 times higher than those of their normal tissues of origin. This phenomenon is the cause of different 18F-FDG accumulation in healthy and malignant tissues.

The process of 18F-FDG accumulation is however not specific. Also inflammatory conditions cause accumulation of 18F-FDG. 18F-FDG PET/CT therefore is also used for infectious of inflammatory conditions [11], but inflammation is also a notorious pitfall when interpreting 18F-FDG PET/CT scans.

The role of 18F-FDG PET/CT in oncology

Due to the pharmacokinetics of 18F-FDG, tissues with high metabolic need tend to accumulate 18F-FDG faster than other tissues. This feature is exploited when performing 18F-FDG PET in oncology: measuring the 18F-FDG distribution and uptake after a fixed amount of time after injection, the higher 18F-FDG accumulation rate of cancerous tissues is reflected by a higher uptake level.

xxii
The role of $^{18}$F-FDG PET/CT in oncology

Figure 3: Example of an $^{18}$F-FDG PET/CT used for tumour staging. CT (left), fused $^{18}$F-FDG PET/CT (middle) and $^{18}$F-FDG PET (right) at the lower cervical level (upper panels) and the pelvic level (lower panels). A summary MIP is provided on the right side. This case represents a 66-year-old woman in whom during surgery for a sigmoid adenocarcinoma peritoneal metastases were found. In the work-up for hyperthermic intraperitoneal chemotherapy (HIPEC) an $^{18}$F-FDG PET/CT was performed showing the sigmoid adenocarcinoma with mesenteric lymphnode metastases and peritoneal metastases (lower panels). Also a lesion in the left thyroid lobe was found, which proved to be a benign thyroid adenoma at hemithyroidectomy (upper panels). No liver, lung or bone metastases were established, thus HIPEC could be performed instead of palliative chemotherapy. The $^{18}$F-FDG uptake in the distal oesophagus was considered inflammatory due to recurrent gastric reflux. Colour scale is in $SUV_{BW}$ [g·cm$^{-3}$]. CT: low-dose computed tomography without contrast-enhancement; $^{18}$F-FDG PET: $2-(^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; MIP: maximum intensity projection; $SUV_{BW}$: standardised uptake value normalised for bodyweight.

For many cancer types, $^{18}$F-FDG PET/CT has a role in distinguishing benign from malignant tissues [12], finding an unknown primary in case (locoregional) spread has been established [13], guiding biopsies in suspected lesions [14] or staging disease extent by detecting lymphnodes or distant metastases [15]. $^{18}$F-FDG uptake correlates with disease grade in many cancer types. Therefore, $^{18}$F-FDG uptake may reflect the prognosis of the patient, independent of tumour stage [16,17]. $^{18}$F-FDG PET/CT can be used to select individual patient treatment options (figure 3, page xxiii), not only by distinguishing locoregional and distantly spread disease, but also by excluding disease in other organs, e.g., in patients with liver or lung oligometastases planned for resection [18]. Currently $^{18}$F-FDG PET/CT is also used for determination of the target volume for radiotherapy and its potential in dose escalation are being explored [19].

$^{18}$F-FDG PET/CT is also used during treatment (figure 4, page xxiv). It can e.g., be used to determine the early response to neoadjuvant treatment, to identify the patients who benefit from this treatment and those who preferably should undergo surgery at an earlier time point [20]. Also the response to palliative systemic therapy can be evaluated using $^{18}$F-FDG PET/CT [21,22]. A special case is the treatment of gastro-intestinal stromal tumours with imatinib mesylate, a tyrosine kinase inhibitor targeting c-KIT. As glucose metabolism and thus $^{18}$F-FDG uptake, is directly regulated by this pathway, it inhibition will lead to immediate reduction of $^{18}$F-FDG uptake. Imatinib mesylate resistance, often due to c-KIT mutation, can therefore
Figure 4: Example of an $^{18}$F-FDG PET/CT for therapy response evaluation. Maximum intensity projections (MIPs) of a 19-year-old male diagnosed with a desmoplastic small round cell sarcoma in the left soleus muscle with an ipsilateral inguinal lymphnode metastasis and bilateral pulmonary metastases. The symmetric uptake cervical, along the shoulders and thoracic spine was considered activated brown adipose tissue (left panel). Over 4 months later, after 6 cycles of vincristine, doxorubicin, ifosfamide & dactinomycin chemotherapy, there is a 74% reduction in the $\text{SUV}_{\text{peak}}$ of the primary lesion (right panel). The lung metastases have disappeared. This is compatible with a partial metabolic response. Symmetric intense bone marrow uptake is noted, compatible with repopulation after chemotherapy. Colour scale is in $\text{SUV}_{BW}$ [g·cm$^{-3}$]. $^{18}$F-FDG PET/CT: 2-($^{18}$F)fluoro-2-deoxy-d-glucose positron emission tomography/computed tomography; $\text{SUV}_{BW}$: standardised uptake value normalised for bodyweight; $(\Delta)\text{SUV}_{\text{peak}}$: (relative change in) mean $\text{SUV}_{BW}$ in the hottest cluster.

be detected by persistence of $^{18}$F-FDG uptake after starting with this drug [23]. In many cancers, the degree of metabolic therapy response is associated with the patient’s prognosis [21,22,24]. After completion of treatment, follow-up $^{18}$F-FDG PET/CT can help in distinguishing scar tissue from viable tumour tissue (e.g., postradiotherapy in rectal carcinoma) [25], can aid in detecting the location of the disease in biochemical recurrence [26], in restaging [27] and can be of use when dedifferentiation is suspected [28]. Given the increasing number of indications for $^{18}$F-FDG PET/CT, it is not unexpected that the number of investigations in Western Europe exceeded 900,000 per year in 2011 and was rising by 21% year$^{-1}$ between 2005 and 2010 [29]. Even though costs for the radiopharmaceuticals have declined substantially, the costs for the pur-
chase and maintenance of a PET/CT scanner with a limited technological life span make a PET/CT scan more expensive than CT alone. Moreover the radiation burden (less than 8 mSv for a whole-body $^{18}$F-FDG PET/CT examination) and the potential false-positive findings as $^{18}$F-FDG is a nonspecific tracer, call for proper research to its additional value in tailoring patient treatment. As costs for current highly targeted treatments are increasing exponentially, appropriate use of $^{18}$F-FDG PET/CT can still be very cost-effective.

**Optimisation of quantitative $^{18}$F-FDG PET/CT**

One of the most important advantages of PET using $^{18}$F-FDG is that it is fully quantitative, i.e., $^{18}$F-FDG uptake can be expressed as a number. For this purpose, the standardised uptake value, or $SUV$, is most commonly used. Other quantitative parameters do exist, such as tumour-background ratio, tumour-liver ratio, tumour-blood ratio, total lesion glycolysis, glucose metabolic rate and pharmacokinetic parameters.

Many biological and technical factors have a major influence on $^{18}$F-FDG uptake. These factors include patient preparation, management of diabetics, waiting time after injection, correction for residual radiopharmaceutical in the syringe, scanner calibration and scan speed. Also processing factors such as correction for photon attenuation and scatter, reconstruction algorithms and settings, partial volume corrections influence the resulting value for $^{18}$F-FDG uptake. All these sources of heterogeneity and potential error complicate comparison between different measurements and result in difficulties in the interpretation of results obtained in multi centre studies. Even in case all data are acquired according to a standardised methodology, different postprocessing factors may cause significant heterogeneity. As stated above, there are many different quantitative parameters for $^{18}$F-FDG uptake. Different normalisation factors are being used to correct for $^{18}$F-FDG volume of distribution, such as bodyweight, lean body mass, body surface area. Also the uptake is determined in various volumes-of-interest (VOI), such as the maximum of a lesion, the most active cluster and volumes derived by manual or threshold- or CT-based contouring. Finally different summary measures for these volumes are used: either the maximum, peak, mean or median value.

As shown in figure 5 (page xxvi), the number of PubMed indexed articles on $^{18}$F-FDG PET and oncology between 1982 and 2015 follows approximately a quadratic growth, expected to exceed 2,000 articles this year, many describing their individual methodology. To exploit the full power of quantitative $^{18}$F-FDG PET, standardisation must occur, as heterogeneity in study methodology blurs its true potential. For this purpose, choices have to be made, which methods are most strongly correlated to biological outcome measures such as survival or response rates (validity), least operator dependent (reliability or reproducibility) and least prone to error thus being as simple as possible as every additional factor propagates in overall uncertainty. To aid in harmonisation, European procedure guidelines for quantitative $^{18}$F-FDG PET have been formulated in 2010 [30] and updated in 2015 [31]. However, even when data is acquired in a highly standardised way, the day-to-day variability of $^{18}$F-FDG uptake is over 30% [32].
FDG PET/CT in Solid Tumours

Dynamic $^{18}$F-FDG PET/CT

In clinical and most research settings, a 3D static $^{18}$F-FDG PET/CT of the whole-body is acquired, approximately 60 minutes after intravenous injection of $^{18}$F-FDG. For specific research questions however, dynamic PET/CT is acquired directly after a standardised injection of the tracer. Practically, it is only feasible to acquire only one bed position typically up till one hour after injection of the radiopharmacon, covering at most 258 mm transaxially (figure 6 page xxvii). By adding the time-dimension, the full time-activity concentration curve can be determined of blood and tissue. In the case of $^{18}$F-FDG PET/CT, $^{18}$F-FDG influx ($K_i$, [ml·g$^{-1}$·min$^{-1}$]) and the glucose metabolic rate ($MR_{glc}$, [nmol·min$^{-1}$·cm$^{-3}$]) tissue can be calculated [33]. This requires the assumption of an underlying compartment model, comparable to figure 2 (page xxviii). Moreover, the blood volume fraction ($V_B$) and the pharmacokinetic rate constants of this compartment model, representing the enzymatic reaction rate constants of GLUT ($K_1$, $k_2$), hexokinase ($k_3$) and, when non-negligible, glucose-6-phosphatase activity ($k_4$) can be determined [34,35].

Currently, dynamic $^{18}$F-FDG PET/CT is not considered standard clinical care as its advantage of being able to quantify uptake dynamics does not outweigh its disadvantages of not being a whole-body investigation, requiring a sophisticated acquisition protocol and requiring the patient to remain in the scanner bore for a long time span.

---

Figure 5: PubMed indexed new references per year concerning $^{18}$F-FDG PET in oncology. 2015 figures are extrapolated. $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-d-glucose positron emission tomography.

* * *
Figure 6: Example of a 4D dynamic $^{18}$F-FDG PET/CT of a 75-year-old female patient with non-small cell lung carcinoma centrally located in the right lower lobe. Pulmonary lobectomy revealed a 45 mm moderately differentiated squamous cell carcinoma without lymphnode metastases ($pT_2a\,N_0\,cM_0$). Prior to the investigation, the patient fasted for 6 hours (serum glucose: 5.0 mmol·l$^{-1}$). A 60-min 4D dynamic $^{18}$F-FDG PET/CT of the thorax was performed after a standardised injection of 177 MBq $^{18}$F-FDG in the left median cubital vein. During the investigation $^{18}$F-FDG continues to accumulate in the lesion but decreases in the blood pool (maximum intensity projections: left panel). Time-activity concentration curves of the aortic blood and the tumour can be derived (right upper panel) from which tumour pharmacokinetic rate constants can be computed, assuming an underlying compartment model. Using a Gjedde-Patlak plot (right lower panel) tumour glucose metabolic rate can be computed directly. Inset: maximum intensity projection of a parametric MR$\text{glc}$ image of this patient. For this patient $SUV_{\text{peak}}$ was 17.9 g·cm$^{-3}$, mean tumour $SUV$ was 7.4 g·cm$^{-3}$ and mean tumour $MR_{\text{glc}}$ was 215 nmol·min$^{-1}$·cm$^{-3}$. Mean tumour values are computed from a volume-of-interest with a threshold of 50% of the maximum tumour uptake. The whole-body static $^{18}$F-FDG PET/CT of this patient is displayed on the cover of this thesis showing no suspicion of distant metastases.

$C_{\text{plasma}}(t)$: the time activity concentration of $^{18}$F-FDG-6-phosphate in arterial plasma; $C_{\text{tissue}}(t)$: the time activity concentration of $^{18}$F-FDG-6-phosphate in tissue; $^{18}$F-FDG PET: 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography; GLUT: facilitative membrane-bound sodium-independent glucose transporters; $K_1$: GLUT influx rate constant; $k_2$: GLUT efflux rate constant; $k_3$: hexokinase phosphorylation rate constant; $k_4$: glucose-6-phosphatase dephosphorylation rate constant; $MR_{\text{glc}}$: glucose metabolic rate; $R^2$: Pearson product-moment correlation coefficient squared; $SUV$: standardised uptake value normalised for bodyweight; $V_B$: blood volume fraction.
Aims, Scope & Outline of the Doctoral Thesis

Unpublished

D. Vriens\textsuperscript{1}

\textsuperscript{1}dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands
Aims & scope of this thesis

Three diagnostic challenges will be addressed in this thesis, (1) the influence of different methods for data-analysis on measured metabolic therapy response and their effect on its relation with clinical outcome, (2) metabolic and vascular characterisation of lesions and (3) clinical and economical impact of the use of $^{18}$F-FDG PET/CT for tissue characterisation.

These issues were studied in colorectal carcinoma (CRC), non-small cell lung carcinoma (NSCLC) and thyroid nodules suspected for differentiated thyroid carcinoma.

Outline of this thesis

This thesis consists of four parts. Part I provides a general introduction to quantitative $^{18}$F-FDG PET/CT. Chapter 1 reviews the methodological issues that should be addressed when performing quantitative $^{18}$F-FDG PET/CT. This chapter provides an overview of the many factors that potentially introduce bias or random error in quantitative $^{18}$F-FDG PET/CT from patient inclusion and preparation, to data acquisition, processing and analysis. Chapter 2 discusses the current role of (quantitative) $^{18}$F-FDG PET/CT, using CRC as a clinical example. This chapter serves to support the relevance of quantitative $^{18}$F-FDG PET/CT in the daily practice for patients and physicians.

In the remainder of this thesis, the three diagnostic challenges mentioned above are investigated. In part II, three of the methodological issues raised in chapter 1 are being explored. Chapter 3 investigates four different normalisation procedures for the volume of distribution of $^{18}$F-FDG for their predictive ability of overall (OS) and progression-free survival (PFS). These normalisations are based on bodyweight, lean body mass, body surface area and a combination of bodyweight and plasma glucose level. This study evaluates chemotherapy response in patients with CRC and NSCLC. Chapter 4 describes two different methods of defining a volume-of-interest on pre- and postchemotherapy $^{18}$F-FDG PET scan for metabolic therapy response evaluation. One method quantifies tumour metabolism at the two time points, the other also takes into account a change in tumour volume. Again, OS and PFS serve as primary outcome measures. The last chapter of this section, chapter 5, investigates the requirement to measure the arterial plasma time activity curve in dynamic $^{18}$F-FDG PET. Based on a learning population, this study models a generalised curve, which can be individualised either based on administered dose and patient initial volume of distribution or a single arterial sample. These two non-invasive curves together with an image-derived curve were compared with the gold standard of full arterial sampling challenging the need for invasive artery cannulation.

In part III, the technique of quantitative $^{18}$F-FDG PET is used to explore heterogeneity in metabolic tissue and the combined use of metabolic and vascular response evaluation. Chapter 6 looks into the intratumoural regional differences in metabolic characteristics of the tumour most active ‘core’ and its surrounding layers. This chapter describes cancer metabolism on a two orders of magnitude deeper level: (i) going from whole patient/lesion to separate intratumoural clusters and (ii) going from a gross glucose metabolic rate to pharmacokinetic microparameters of the enzymatic
processes described above. In chapter 7 and chapter 8, the predictive effect of therapy response evaluation in patients with CRC liver metastases is described with OS and PFS as primary outcome measures. In both of these studies metabolic, vascular and morphological response to therapy is measured using, dynamic $^{18}$F-FDG PET, dynamic gadopentetate dimeglumine contrast-enhanced MRI (DCE-MRI) and morphological imaging, respectively. In chapter 7 these patients are treated with cytotoxic chemotherapy. Chapter 8 describes patients additionally treated with anti-angiogenic targeted drugs.

In part IV of this thesis, the clinical impact of a new indication for $^{18}$F-FDG PET/CT is explored: characterisation of cytological indeterminate thyroid nodules. Chapter 9 critically reviews all published patient series on this topic and provides an individual patient meta-analysis. In the last chapter, chapter 10, a decision analysis is performed, supporting the potential cost-effectiveness of $^{18}$F-FDG PET/CT for this indication compared to current standards in Europe and the USA.

The thesis concludes with a general discussion on the studies presented in this work, including future prospects.

***
Part I

Methodological Considerations & Clinical Relevance
Methodological Considerations in Quantification of $^{18}$F-FDG PET studies


D. Vriens$^1$, 
E.P. Visser$^1$, 
L.F. de Geus-Oei$^1$ and 
W.J.G. Oyen$^1$

$^1$dept. of Nuclear Medicine, 
Radboud University Medical Centre, Nijmegen, the Netherlands
1. Methodological Considerations in FDG PET

Abstract

Purpose This review aims to provide insight into the factors that influence quantification of glucose metabolism by $^{18}$F-FDG PET images in oncology as well as their influence on repeated measures studies (i.e., therapy response evaluation), offering improved understanding both for clinical practice and research.

Methods Structural PubMed searches have been performed for the many factors affecting quantification of glucose metabolism by $^{18}$F-FDG PET. Review articles and references lists have been used to supplement the search findings.

Results Biological factors such as fasting blood glucose level, $^{18}$F-FDG uptake period, $^{18}$F-FDG distribution and clearance, patient motion (breathing) and patient discomfort (stress) all influence quantification. Acquisition parameters should be adjusted to maximise the signal-to-noise ratio without exposing the patient to a higher than strictly necessary radiation dose. This is especially challenging in pharmacokinetic analysis, where the temporal resolution is of significant importance. The literature is reviewed on the influence of attenuation correction on parameters for glucose metabolism, the effect of motion, metal artefacts and contrast agents on quantification of CT attenuation-corrected images. Reconstruction settings (analytical vs iterative reconstruction, postreconstruction filtering and image matrix size) all potentially influence quantification due to artefacts, noise levels and lesion size dependency. Many volume-of-interest definitions are available, but increased complexity does not necessarily result in improved performance. Different methods for the quantification of the tissue of interest can introduce systematic and random inaccuracy.

Conclusions This review provides an up-to-date overview of the many factors that influence quantification of glucose metabolism by $^{18}$F-FDG PET.

Keywords Computer-Assisted Image Processing · 2-$^{(18)}$Ffluoro-2-deoxy-D-glucose · Pharmacokinetics · Positron Emission Tomography · Reference Standards
1.1 Introduction

Positron emission tomography (PET) using 2-(18F)fluoro-2-deoxy-d-glucose (18F-FDG) is an established imaging modality in oncology [36]. Although in daily practice visual inspection of 18F-FDG PET images is used for diagnosis and assessment of response to therapy, it has been shown that (semi-)quantitative analysis allows an objective complement to visual interpretation of lesions [37–39]. Results of this analysis might be used for individual tailoring of therapy, since increased 18F-FDG uptake usually corresponds to a dismal course of the disease. Repeated measurements can be used in early response assessment, valuable for further individualisation of therapy [40]. Generally, lesions are quantified using the standardised uptake value of 18F-FDG (SUV, i.e., the 18F-FDG activity concentration at a single time point normalised to the administered activity (AA) and a measure for distribution volume) and therapy response is evaluated by the relative change between a baseline and a follow-up scan during the course of therapy (ΔSUV). Although the results of the first trials have been published in which the performance of individually tailored therapy based on early response to neoadjuvant treatment is investigated [20] and new trials are currently being undertaken, the multiple factors influencing the quantification of glucose metabolism by 18F-FDG PET are still under discussion [6,41,42].

Besides the influence that these factors have on the quantification of parameters of glucose metabolism, a variety of factors also influence the reproducibility of these parameters [43,44]. Quantification of glucose metabolism by 18F-FDG PET is not only dependent on biological properties of the disease under investigation, but also on methodological aspects of patient preparation, image acquisition, reconstruction, volume-of-interest (VOI) definition and methods of parameter computation. To be able to perform multicentre studies or meta-analysis, but also to apply results of studies in clinical practice, the influence of these factors should be minimised by standardisation. This has led to the development of consensus recommendations by the European Organisation for Research and Treatment of Cancer (EORTC) [5], the National Cancer Institute (NCI) [45] and the Netherlands Society of Nuclear Medicine (NEDPAS) [46]. The Society of Nuclear Medicine has agreed on procedure guidelines for tumour imaging but conclude that optimal methods for semiquantitative measurements need further elucidation [47].

This review aims to give a theoretical background illustrated by up-to-date publications on the influence of methodological factors influencing quantification of 18F-FDG PET. It will not merely focus on the semiquantitative parameter SUV, but also include fully quantitative parameters such as the glucose metabolic rate (MR_{glc}) and the pharmacokinetic rate constants of two-compartment model analysis. Hardware issues influencing scanner sensitivity, such as detector crystal material, photon energy window, coincidence timing window, detector ring diameter and axial length of the field-of-view (FoV), are not addressed in this review. Several other factors are considered outside the scope of this study; these are: methodological errors, such as invalid cross-calibration, asynchronous clocks, omission of decay correction for the time period between calibration and start of the PET scan, low precision of plasma glucose measurement, failure to measure residual activity of the infusion system or paravenous infiltration of 18F-FDG and factors inextricably linked to the non-specific targeting of 18F-FDG (e.g., infection, postradiotherapy inflammation).
1.2 Patient preparation & image acquisition

Biological factors affecting quantification

Several biological factors affecting quantification, such as fasting plasma glucose level, uptake period, $^{18}$F-FDG distribution and clearance, patient motion (breathing) and patient discomfort (stress), all deserve attention at the time of patient preparation, $^{18}$F-FDG administration and distribution and image acquisition.

Blood glucose level

High blood glucose levels, due to a non-fasting state or diabetes mellitus, interfere with $^{18}$F-FDG uptake in malignant lesions. The transmembranous glucose transport facilitators (GLUT), albeit overexpressed in many cancers, can be saturated by an excess of unlabelled glucose. This diminishes $^{18}$F-FDG uptake as glucose and $^{18}$F-FDG both compete for the binding sites of transporters and enzymes, leading to zero-order kinetics. In patients without any known form of glucose intolerance it is shown in two consecutive scans that the SUV, using body weight as a measure of distribution volume, is significantly lower in the loaded state (serum glucose $> 8.0$ mmol·l$^{-1}$) in both head and neck carcinoma ($SUV_{BW} = 6.9$ vs $4.0$ g·cm$^{-3}$, $p < 0.02$) [48] and bronchial carcinoma ($SUV_{BW} = 5.07$ vs $2.84$ g·cm$^{-3}$, $p < 0.001$) [49] compared to the fasting state ($< 6.0$ mmol·l$^{-1}$). In contrast to the reduction of tumour uptake, skeletal muscle accumulates more $^{18}$F-FDG, resulting in blurring of tumour margins and less clear localisation of lesions [48]. Also Patlak-based net influx rate constants ($K_i$) of the lesions decrease markedly in the glucose loaded state (mean -25%, $p < 0.05$), while $MR_{glc}$ is on average 36% higher. This paradoxical increase in $MR_{glc}$ might be due to the fact that the authors assumed the same value for the relative affinity of the biological system to $^{18}$F-FDG and glucose (‘lumped constant’, $LC$) in both states to compute $MR_{glc}$ from $K_i$, which is likely untrue [50].

The effect of hyperglycaemia has clear impact on visual interpretation of $^{18}$F-FDG PET images as it results in a reduced detection rate of malignancies [51], but it also has a major effect on quantification. Therefore, hyperglycaemia should be a reason to reschedule the procedure. Correction of hyperglycaemia by insulin directly prior to the $^{18}$F-FDG injection is dissuaded, because hyperinsulinaemia increases the translocation of GLUT4 thereby rapidly and efficiently shunting $^{18}$F-FDG to organs with a high density of insulin receptors (e.g., skeletal and cardiac muscles) [51]. When $^{18}$F-FDG injection is postponed to 1 h after insulin administration (up to a serum glucose below 8.0 mmol·l$^{-1}$) in hyperglycaemic diabetic patients, no differences between normoglycaemic non-diabetics and hyperglycaemic (insulin-corrected) diabetics are found in SUV, using lean body mass as measure for distribution volume ($SUV_{LBM}$), in lungs, liver, muscles, myocardium or suspected pulmonary lesions [52]. However, recently a standardised protocol of intravenous insulin at least 1 h prior to $^{18}$F-FDG injection led to an unacceptable biodistribution in 25% of diabetic cancer patients (increased muscle uptake and decreased liver uptake). In these patients the interval between insulin and $^{18}$F-FDG injection was significantly shorter than in the patients with a normal biodistribution (65.7 vs 80.2 min, $p < 0.01$) [53]. Metformin strongly increases the SUV of the small and large intestines, potentially decreasing
the detection sensitivity for malignant lesions, but the influence on lesion quantification is not described [54]. To prevent hyperglycaemia in non-diabetic patients, guidelines [5,45,46] recommend that patients should have fasted at least 4 h, but preferably 6 h before administration of $^{18}\text{F}$-FDG. When blood glucose levels are well outside physiological ranges (e.g., > 11 mmol·l$^{-1}$ [46]) the scan should be postponed. We use a stricter cut-off for research purposes (8.0 mmol·l$^{-1}$).

1.2. Patient preparation & image acquisition

Uptake period

Typically, PET acquisition is performed 45-60 min after $^{18}\text{F}$-FDG injection, based on the fact that $^{18}\text{F}$-FDG activity concentrations become constant within the first hour after tracer injection in normal tissues. In malignancies, however, it has been shown that for some tumour types $^{18}\text{F}$-FDG concentration continues to increase up to 4-5 h after injection and that constant $^{18}\text{F}$-FDG activity concentrations are rarely reached within the first hour after injection. Between 60 and 90 min postinjection, it is not uncommon for a lesion to further increase $SUV$ as much as one tenth of the value reached 10 h postinjection [55]. Since a further increase in $^{18}\text{F}$-FDG activity concentration 1 h after injection is rare in normal tissues [56], it has been hypothesised that dual-time-point imaging (45 and 90 min postinjection) might improve detection rates of lesions with high glucose metabolism by increasing tumour-background ratios. Therefore, for therapy response evaluation studies using $SUV$s it is highly important to keep the uptake period of $^{18}\text{F}$-FDG within narrow limits (55-65 min) [45,46].

$^{18}\text{F}$-FDG distribution & clearance

To optimise the distribution of $^{18}\text{F}$-FDG throughout the body, international guidelines recommend prehydration (0.5 l) of the patient to ensure excretion of $^{18}\text{F}$-FDG from healthy background tissue, which may be further enhanced by furosemide-induced forced diuresis. This might be of special value when the pelvis or kidney regions are of interest [5,45,46]. Median $SUV$s of tumours are lower if furosemide is used. However, the total fraction of excreted activity is not different if diuretics are used, but is significantly higher early after injection for the furosemide group leading to improved image quality and reducing patient radiation exposure [57]. Therefore, it seems important when comparing parameters for glucose metabolism that the use of (loop) diuretics should be taken into account. A practical disadvantage of using forced diuresis during dynamic acquisition is that measures should be taken to avoid premature termination of the acquisition due to urinary urgency.

Patient (periodic) motion

Apart from exercise-induced increased muscle uptake during the uptake period, the effect of motion during acquisition has consequences for lesion localisation (e.g., spatial mismatch around the diaphragm due to breathing) and causes smearing of the lesion activity concentration within the volume of movement. Consequently, the lesion metabolic volume is overestimated and the $SUV$ is underestimated. Moreover, tissue inhomogeneity is similarly smeared, leading to loss of spatial heterogeneity. The magnitude of the decrease of recovered activity concentrations depends most
markedly on lesion size and amplitude of motion and to a lesser extent on the motion frequency. Recovered activity concentrations can be increased by better lesion volume estimation by a motion correction algorithm. Verified in nine lung carcinoma patients, this algorithm reduces the estimated lesion volume by 15% leading to an increase of the mean $SUV$ in the VOI ($SUV_{\text{mean}}$) by 5% \[58\]. Different other techniques may be applied to improve recovery of activity concentration in periodically moving lesions such as gated PET/CT (in which data are only acquired during a certain respiratory phase), respiratory correlated dynamic PET/CT (by summing the sinograms of images of a particular breathing phase selected using a point source) or deep inspiration breath-hold PET/CT \[59\].

During dynamic acquisition, non-periodic patient movement can have major influence on measured parameters, since the lesion can “move into” the VOI usually defined in the last time frame(s). Consequently, activity concentrations of lesions are underestimated for the period before patient movement due to a mispositioned VOI. In the time frames in which the lesion of interest is visible the VOI can be realigned, but during the early time frames this is often impossible due to noisy images or insufficient tissue to blood contrast. Patient movement therefore should be prevented and monitored during acquisition, the original position should be restored as quickly as possible and any motion should be noted including time at which it occurred.

If CT-based attenuation correction (AC) is used, any spatial mismatch will lead to incorrect amplification of measured activity concentrations, leading to unreliable values for lesion quantification (differences of around 10% in $SUV$ \[60,61\] due to breathing, smallest for free breathing CT \[60\]).

**Patient discomfort**

Patient stress or cold during the distribution period increases uptake of $^{18}$F-FDG in muscle and brown adipose tissue (BAT) \[62-64\]. This can lower detection sensitivity (by lowering contrast) and specificity (by increasing the number of false-positive lesions) of the images and might affect quantification \[46,65\]. Factors known to increase $^{18}$F-FDG uptake in BAT such as exposure to cold \[63,64,66,68\] should be prevented by exposing patients to thermoneutral conditions. Medication such as the beta-adrenergic blocking agent propranolol \[63,68,71\], reserpine \[68\] and the opiate fentanyl \[72\] all prevent $^{18}$F-FDG uptake in BAT, but the effect of benzodiazepines seems doubtful \[67,68,72,73\]. The NEDPAS guidelines \[46\] suggest considering the use of benzodiazepines to reduce muscle uptake when tumours in the head and neck region are expected and state that there is no place for it to reduce uptake in BAT.

**Technical factors affecting quantification**

At the time of data acquisition, several factors influence quantification. These include scan acquisition parameters (such as acquisition mode, scan duration, bed overlap and administered $^{18}$F-FDG activity), time frame duration in dynamic acquisitions, factors related to attenuation correction (such as motion and the use of CT contrast agents) and other forms of data correction. Most of these settings are based on the performance of the scanner and should be measured according to the NEMA (National Electrical Manufacturers Association) NU2-2007 standard \[74\] (table \[1.7\] page \[5\].
1.2. Patient preparation & image acquisition

<table>
<thead>
<tr>
<th>Manufacturer: model</th>
<th>Crystal material</th>
<th>Axial Transaxial</th>
<th>Sensitivity in centre of FoV [cps·kBq⁻¹]</th>
<th>Peak NEC(1R) [kcps]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE: Discovery PET/CT 690</td>
<td>BGO</td>
<td>157</td>
<td>4.8</td>
<td>6.7</td>
</tr>
<tr>
<td>Philips: PET/CT Gemini TF</td>
<td>LYSO</td>
<td>180</td>
<td>4.8</td>
<td>6.6</td>
</tr>
<tr>
<td>Siemens: Biograph mCT</td>
<td>LSO</td>
<td>162 (218)</td>
<td>(4.1)</td>
<td>(8.1)</td>
</tr>
</tbody>
</table>

*Average measured FWHM 1 cm from centre of the FoV.†The Siemens Biograph mCT has an optional extension from three to four detector rings. Properties for this extended FoV are given in parenthesis.

**Table 1.1:** NEMA properties of currently commercially available modern PET/CT scanners. BSO: bismuth germanium oxide; FoV: field-of-view; FWHM: full-width at half-maximum; GE: General Electric; L(Y)SO: lutetium (yttrium) oxyorthosilicate; NEC(1R): noise equivalent count rate computed for a noise-free estimation of randoms; NEMA according to the national electrical manufacturers association standard NU2-2007 [74]; PET/CT: positron emission tomography / computed tomography; *Average measured FWHM 1 cm from centre of the FoV.†The Siemens Biograph mCT has an optional extension from three to four detector rings. Properties for this extended FoV are given in parenthesis.

**Acquisition parameters**

The choices of acquisition parameters are closely interrelated and are based on trade-offs between signal-to-noise ratio (SNR) and radiation safety, financial issues (e.g., patient throughput, costs of ¹⁸F-FDG), patient comfort and logistics. Increasing the sensitivity of the PET scanner by using 3D acquisition (instead of 2D acquisition using lead or tungsten septa between crystal rings), increasing scan duration, enlarging bed overlap and raising the administered ¹⁸F-FDG activity (below the amount that results in count rates exceeding the maximum count rate capabilities of the scanner) all improve SNR. Removing the interplane septa from the PET scanner (i.e., changing from 2D to 3D mode) typically leads to a four- to eightfold increase in sensitivity but also increases the number of detected scattered and random photons. Generally, the peak noise equivalent count rate (NEC) is obtained at a lower activity concentration in 3D than in 2D as expected from the behaviour of the (increased) true coincidence rate. As a result the SNR (for a uniform cylinder \( SNR \propto \sqrt{NEC} \) [78]) will be higher when operating in 3D vs 2D mode for the low activity concentration range. Recommendations for administered ¹⁸F-FDG activity therefore should be based on acquisition mode (2D or 3D), scan duration and bed overlap, rather than on pre-determined diagnostic reference levels [5]. The NCI guidelines recommend a total administered ¹⁸F-FDG activity of 5.18-7.77 MBq·kg⁻¹ [45] and the NEDPAS recommendations advise 27.5 MBq·min·kg⁻¹ (for 2D and ≤ 25% bed overlap), 13.8 MBq·min·kg⁻¹ (for 3D and ≤ 25% bed overlap) or 6.9 MBq·min·kg⁻¹ (for 3D and 50% bed overlap). For a 3D acquisition of a 70-kg patient with 4 min per bed position (with 25% bed overlap), it recommends approximately 242 MBq (±10%) to be administered, which corresponds to an effective absorbed dose of 4.6 mSv of the ¹⁸F-FDG (0.019 mSv·MBq⁻¹ [79]). Reducing scan duration therefore necessitates increased administered ¹⁸F-FDG activity to maintain SNR. Increasing acquisition time effectively maintains the quality (SNR) of ¹⁸F-FDG PET scans of heavier patients, agreeing with the hypothesis that increasing body weight increases the fractions of
1. Methodological Considerations in FDG PET

In the acquisition parameters for dynamic 18F-FDG PET in oncology of different study groups:

- **Strauss (2008)**
  - 10-30 s, 5-60 s, 5-120 s, 8-300 s
  - **AA**: ~310 MBq bolus +
  - **Scanner**: Siemens ECAT EXACT HR+ (2D)
  - **Reconstruction**: OSEM 6i2s no filter
  - **Pixel size**: 2.277 mm
  - **Input function**: IDIF.

- **Krak (2008)**
  - 6-5 s, 6-10 s, 3-20 s, 5-30 s, 5-60 s, 8-150 s, 6-300 s
  - **AA**: ~370 MBq
  - **Scanner**: Siemens ECAT EXACT HR+ (2D)
  - **Reconstruction**: FBP 0.5 Hanning filter
  - **Resolution**: ~7 mm FWHM
  - **Input function**: IDIF verified by venous blood samples.

- **deGeus-Oei (2006)**
  - 10-30 s, 3-300 s, 3-600 s
  - **AA**: ~200 MBq in 60 s
  - **Scanner**: Siemens ECAT EXACT 47 (2D)
  - **Reconstruction**: FBP 4 mm Gaussian filter
  - **Resolution**: ~6 mm FWHM
  - **Input function**: Arterial sampling & IDIF.

**Figure 1.1: Dynamic framing duration in three groups [83–85] performing dynamic oncological 18F-FDG PET for pharmacokinetic two-compartment modelling.**

**AA**: administered activity; FBP: filtered backprojection; 18F-FDG PET: 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography; FWHM: full-width at half-maximum; IDIF: image-derived input function; OSEM: ordered subsets expectation maximisation. *Duration of a bolus is a few seconds.

Random and scatter coincidences [80,81]. Increasing administered 18F-FDG activity does not improve SNR in a study, explained by the fact that it was likely to saturate count rates of the equipment [82].

**Time frame duration**

For dynamic acquisition, activity concentration changes are largest during the first minutes of acquisition (bolus transit) but depend on the rate of 18F-FDG infusion, distribution and clearance. Therefore, when partitioning list-mode data or using predefined time frames, one should consider a high temporal resolution for this period, especially important for pharmacokinetic analysis of two-compartment models. Improvement of temporal resolution inherently decreases the count accuracy per time frame, since the (approximate) Poisson distribution of count statistics (for prompt coincidences) dictates a relative standard deviation dependent on the inverse square root of the observed counts. As the choice of framing and the duration of 18F-FDG infusion are related, it is impossible to give general recommendations. Generally, for scanners with low sensitivity, time frame durations should be longer and consequently the rate of 18F-FDG infusion slower. The choices of different groups using dynamic acquisition for oncological purposes can be found elsewhere [83,85] (figure 1.1, page 10).
1.2. Patient preparation & image acquisition

Attenuation correction

Correction for photon attenuation is necessary since positron annihilation photons interact with surrounding tissues (mainly Compton scattering, to a lesser extent photoelectric effect). Before the introduction of combined PET/CT, attenuation correction was applied by transmission imaging of the subject with an external rotating isotope source, either positron emitters (e.g., $^{68}$Ga using its mother isotope $^{68}$Ge) or single photon emitters (e.g., $^{57}$Co or $^{137}$mBa using its mother isotope $^{137}$Cs). Since photon attenuation is dependent on the photon energy, the latter two require scaling to obtain transmission sinograms adapted to 511 keV photons. The attenuation sinogram, the ratio of a transmission scan and a scan without an object in the FoV (blank scan), is used to correct the emission sinogram, which is subsequently reconstructed to an image. Since the introduction of PET/CT, the spatial distribution of the linear attenuation coefficients can be obtained using the CT scanner. Advantages of CT AC are: reduction in total scan time, lower noise in the transmission sinograms and no replacement of isotope sources necessary due to decay. Finally, they can be acquired after tracer injection, choosing the energy window around the CT photon peak to exclude the 511 keV photons (radionuclide-based transmission scans suffer from contamination by emission photons unless the transmission data are acquired before the PET tracer is administered to the patient). CT AC however leads to extra challenges: the use of polyenergetic Bremsstrahlung photons with energy much lower than 511 keV (X-ray tube potential difference usually $\sim$130 kV leading to photon energy ranging from 40 to 130 keV, effective energy $\sim$70-80 keV), artefacts due to spatial mismatches with PET or metal objects, effects of CT-enhancing contrast and artefacts due to truncation and beam hardening [86].

Conversion of the X-ray to 511 keV linear attenuation coefficients can be performed by either segmentation, scaling or dual X-ray imaging. By segmentation, regions of different material types are identified for which linear attenuation coefficients for 511 keV photons are known. Linear scaling of X-ray linear attenuation coefficients can be performed as long as Compton scattering is the major determinant of attenuation. However, at CT photon energies the photoelectric effect plays a significant role for high atomic number elements (e.g., calcium in bone, iodine or barium in contrast agents with atomic numbers of 20, 53 and 56, respectively) in which case hybrid or bilinear scaling is appropriate [86]. Finally, linear attenuation coefficients can be measured at two tube potential differences, enabling the separation of attenuation due to the Compton scattering and the photoelectric effect. In practice, no significant difference in quantification of malignant lesions is found comparing CT-based attenuation correction with segmented AC (SAC) using transmission imaging [60] or between measured (MAC) and segmented (SAC) attenuation correction using transmission imaging [61].

CT artefacts due to metal objects and attenuation due to iodine-containing CT contrast agents might influence lesion quantification. Multiple studies have found a negligible change of the SUV in a variety of malignancies varying from $+2.8\%$ to $+4\%$ [87,90]. This small effect surprisingly disappears after chemotherapy [87]. The $SUV_{mean}$ significantly increases in the (normal) aorta (+15%), kidneys (+13%), liver (+11%), spleen (+10%) and inferior caval vein (+12%) [87]. This might be of relevance when the plasma time-activity concentration curve for pharmacokinetic analysis...
1. Methodological Considerations in FDG PET

is image-derived (IDIF): a positive bias in the input function leads to underestimation of $MR_{glc}$. Around metal prosthesis, CT AC leads to an underestimation of $^{18}$F-FDG activity concentration by $< 6\%$ [60].

Finally, truncation artefacts, due to a smaller transaxial FoV of CT than that of PET, lead to incorrect attenuation correction if there are tissues outside the CT FoV. Therefore, it is recommended to scan a patient with arms up except for head and neck imaging. Usually the degree of truncation is small and therefore no major effects on quantification are expected. Lowering the arms also causes beam hardening (structures of higher density cause attenuation of low-energy photons of the polyenergetic X-ray bundle) and scatter-induced artefacts. These can influence quantification up to 11-15% [86].

Other data corrections

Apart from attenuation correction, other corrections must be applied before an image can be quantified. These are: normalisation, correction for random coincidences, correction for scattered radiation, correction for dead time and cross-calibration with a dose calibrator or well counter.

Remaining differences in detection sensitivity that have not been corrected in the setup procedure are corrected by normalisation. During normalisation all detectors are exposed to the same amount of radiation by a rotating source or a cylinder with a uniform activity concentration. The measured counts of each detector pair (line of response, LOR), which should all be equal, are corrected by normalisation factors, which are used to scale the number of counts of each detector pair to correct the emission sinogram. Using 3D acquisition poses a new challenge for this, since the number of LORs are high (order of $10^8$); therefore, to obtain enough counts per LOR long normalisation procedures seem necessary. For this reason modified methods have been developed. Instead of using the acquired number of counts in each LOR separately, a factorisation into different components is being made (component-based normalisation). These components are the individual crystal detection efficiencies, geometrical factors which account for differences in the distance between detectors and their exposed face and the position of crystals in a detector block [91,92].

Random coincidences arise when photons of two unrelated positron annihilations are detected within the coincidence timing window and are recorded as a single pair. This leads to incorrectly positioned LORs and thus adds a relatively uniform background to the reconstructed images. An estimate of the random coincidence rate can be obtained by delaying the coincidence timing window by an interval relatively large to its width. Delayed coincidences cannot be true or scattered events, since photons of the same positron annihilation will always be detected within a few nanoseconds of each other, but the rate of random coincidences will be the same in the delayed and the original window. This number can be subtracted in real time from the total number of coincidences for the detector pair to correct for randoms. This, however, will increase the statistical noise level of the corrected images, since the variance of the number of random counts is added to the variance of the uncorrected counts. Other methods use ‘smoothed delays’, in which the delayed coincidences are acquired in a separate and smoothed sinogram or estimations based on singles rates [93].
One or both photons arising from one positron annihilation might be Compton scattered by the tissue within the gantry, leading to mispositioned LORs. This leads to a hazy background activity concentration, generally highest in the centre of the image. The percentage of scattered events detected in 3D PET acquisition might approach 60-70% of all coincidences. The attenuation data can be used to estimate the number of scattered photons as for 511 keV photons virtually all attenuation is due to the Compton effect [86]. Another method is by extrapolation of the projection profiles immediately outside the object (determined from the attenuation data), which only represent scattered photons as almost no positron annihilations occur in air. This scatter distribution profile can be subtracted from the projections prior to image reconstruction [94]. A study of cerebral $^{18}$F-FDG PET describes that all the pharmacokinetic rate constants of glucose metabolism are overestimated due to photon scatter up to 10-30% (in decreasing magnitude: $K_1$, $K_i$, $k_2$, $k_3$ and $k_4$). $MR_{glc}$ is 12-30% higher when no scatter correction is applied [95].

At high count rates effects of system recovery after detection of a photon are piled up leading to dead time. Systems can either be non-paralysable (i.e., any event within the dead time will not be counted) or paralysable (i.e., this event will furthermore restart the dead time). Empirical dead time models can be used in which the observed count rate as a function of radioactivity concentration is measured for a range of object sizes at different energy window widths. Other methods are under investigation [96].

If all previously mentioned corrections are applied, the number of counts per voxel in the reconstructed images is directly proportional to the activity concentration of that voxel. Calibration to absolute concentrations can be performed by scanning e.g., a cylinder containing a uniform solution of known activity concentration. The obtained calibration factor can be used for absolute quantification of the images.

It can be concluded that standardisation of PET acquisition is highly important. Images should be acquired in a normoglycaemic (fasting) state. When using insulin to reverse hyperglycaemia, this should be injected at least but preferably longer than 1 h before the $^{18}$F-FDG. The distribution period should be held within narrow limits (55-65 min). The effect of prehydration and diuretics led to recommendations advocating standardisation. For dynamic acquisition, however, the practicability of forced diuresis should be considered. Patient motion leads to underestimation in the measured $SUV$ or $MR_{glc}$ and should therefore be prevented. To prevent muscle uptake and uptake in BAT, the waiting room should be kept warm and the patient should be instructed to minimise exercise. The effect of benzodiazepines is doubtful and recent literature suggests the use of beta-blocking agents when uptake in BAT is interfering interpretation of the images.

The administered $^{18}$F-FDG activity should be standardised, dependent on acquisition method and patient body weight and kept within narrow limits (< ±10%). When using CT attenuation correction, quantification in areas around (metal) artefacts should not be performed. Likewise, quantification of lesions in cases of spatial mismatch between PET and CT should not be performed. Even though the effect of contrast agents on $SUV$ may be small, the bias introduced in the blood pool used for pharmacokinetic modelling is large and therefore its use is discouraged. Whenever possible, acquisition should be performed with the arms outside the $FoV$ to prevent...
effects of truncation, beam hardening and increased photon scattering. For reliable quantification, the scanner should be normalised, (cross-)calibrated and corrections for photon attenuation, randoms, scatter and dead time should be performed.

1.3 Tomographic reconstruction

Analytical algorithms (e.g., filtered backprojection, FBP) are almost completely replaced by iterative statistical algorithms (e.g., ordered subset expectation maximisation, OSEM) for tomographic reconstruction of acquired coincidence events to quantifiable images. Iterative reconstruction has a number of potential advantages that make it attractive in comparison with analytical methods. Analytical algorithms have the intrinsic, limiting assumption that measured data are perfectly consistent with the object, which is never true in practice due to noise and other physical factors (e.g., attenuation). In contrast, iterative algorithms can incorporate a priori information such as the statistical distribution of the coincidences and position-dependent spatial resolution. Important adjustable parameters for tomographic reconstruction are the matrix size (the number of voxels in the transaxial plane of the reconstructed images) and the reconstruction smoothing filter. For iterative algorithms also the number of iterations (the number of times the estimate of the real object is updated) and the number of subsets (the number of projections updated simultaneously in each iteration, in OSEM iterative reconstruction only) need to be defined.

Analytical vs iterative reconstruction

Studies comparing the difference of FBP to OSEM on quantification of glucose metabolism all report higher SUVs for OSEM as compared to FBP. This might not only be due to the reconstruction algorithm but also caused by different attenuation correction techniques (SAC with OSEM vs MAC with FBP) and by different reconstruction filters used. Similar or slightly lower (2.3%) SUV_{mean} for different tissues are found in OSEM compared to FBP, using the same method of attenuation correction. This negligible effect might be due to the higher noise levels in FBP reconstructed images, leading to different VOIs defined on FBP compared to OSEM reconstructed images: when the same VOIs are used on both images, no significant differences are found. Apart from noise, also a difference in resolution may cause the dissimilarity in quantification: higher uptake values and MR_{gltc} are found in tumours (+14%), brain (+2 to +4%) and heart (+15 to +21%) for OSEM than for FBP, which is almost completely reversible by equalizing the image resolution by smoothing with a 5-mm FWHM (full-width at half-maximum) Gaussian kernel.

When specifically looking at the IDIF of the left ventricle and ascending aorta, good agreement between OSEM and FBP is observed for the first 5 min of the scan, but for the last 30 min of a 1-h scan, OSEM-derived IDIFs result in 30% higher activity concentrations for the ascending aorta compared to those derived by FBP. It was concluded that OSEM causes bias in regions located within a hotter background, especially relevant for determination of an IDIF. Activity concentrations of IDIFs in the ascending or descending aorta of FBP reconstructed images are within 5% of the arterially sampled, leading to similar results for Patlak MR_{gltc} using either this IDIF or arterial sampling.
1.4. VOI definition

Parameters of iterative reconstruction

The number of subsets has only a small effect on the SUV in phantom experiments when the product of iterations and subsets is kept constant. A low number of iterations (one or two) results in poor recovery of tumour activity concentrations. Further increase of the number of iterations does not improve the accuracy of quantification of glucose metabolism of lesions with an SUV higher than 5, but mainly results in an increase of image noise. When the SUV is lower than 5, large variability in SUV is seen as a function of the number of iterations \[1]. In a study of 50 oncological patients, the SUV\textsubscript{mean} in images reconstructed with 28 subsets and a varying number of iterations was systematically increased. This effect was very small after five iterations (\(< 1\%\) change between five and 40 iterations) \[41\].

Image matrix size influences both noise (smaller number of counts per voxel in larger matrices) and spatial resolution in phantom experiments \[42\]. The recovery of the activity concentration in the spheres of a phantom is better when the matrix size is increased from 128·128 (voxel size: \(~5 \cdot 5\) mm) to 256·256 (voxel size: \(~2.5 \cdot 2.5\) mm). This dependency is smaller when the smoothing kernel is larger (8 vs 5 mm FWHM). This can be explained since the image resolution (5 or 8 mm FWHM) is not at least twice the size of the voxel size (violation of the Nyquist principle), which can be solved by increasing the matrix to 256·256.

It can be concluded that iterative reconstruction can be used for quantification in oncology. For pharmacokinetic analysis of two-compartment models, the IDIF of the left ventricle seems to be sensitive to spill-in and the use of the aorta is preferred. Care should be taken not to overiterate, which only adds noise, which is especially cumbersome in pharmacokinetic rate constant estimation. Too few iterations however will lead to loss of high spatial frequency features, such as resolution and heterogeneity. The matrix size should be chosen not to violate the Nyquist criterion.

1.4 VOI definition

\(^{18}\text{F-FDG}\) uptake or metabolism is determined in a VOI (3D) or region-of-interest (ROI, 2D), which can be the hottest voxel within the lesion (VOI\textsubscript{max}), but can also be based on an absolute threshold (\(e.g., \)an SUV\textsubscript{BW} of 4.0 g·cm\(^{-3}\): VOI\textsubscript{4.0}), on a relative threshold (\(e.g., > 50\%\) of maximum voxel value within the VOI: VOI\textsubscript{50\%}), on a manually placed fixed volume (\(e.g., 1\) ml sphere: VOI\textsubscript{sphere: 1 ml}) or adaptive (\(e.g., \)relative threshold level: VOI\textsubscript{RTL}). Further refinements can be introduced such as applying a relative threshold of the background subtracted maximum voxel value (\(e.g., > 50\% \cdot (\text{maximum background}) \text{ above background}: \text{VOI}_{50\% \cdot (B+\text{max})}\)). The best VOI is dependent on its goal: when used to quantify tumour \(^{18}\text{F-FDG}\) uptake or metabolism, the VOI that yields the best correlation with clinical outcome is preferred. In other situations (\(e.g., \)radiotherapy target planning), the exact volume and position of the VOI are more important. Since all but the fixed volume VOI are dependent on the maximum voxel value, the shape and size of these VOIs are dependent on all earlier mentioned factors influencing the SNR.
The influence of the VOI definition on the accuracy of the SUV was determined in a phantom study comparing VOI \(_{50\%}\), VOI \(_{70\%}\), VOI \(_{50\%} \cdot (B+\max)\), VOI \(_{\max}\) and VOI \(_{\text{square}: 15 \text{ mm}}\). As expected, the recovery coefficient increases for all VOI definitions as a function of the lesion size. There is a strong dependency of SUV on SNR, with a high SNR leading towards a positive bias for the maximum voxel value and thus overestimation of SUVs calculated with a VOI dependent on this maximum voxel value (all but VOI \(_{\text{square}: 15 \text{ mm}}\)). The effects of the VOI method on the accuracy of SUV determination are trivial: VOI \(_{70\%}\) and VOI \(_{50\%}\) are about 15% and 30% lower than VOI \(_{\max}\), respectively and VOI \(_{50\%} \cdot (B+\max)\) is in between: close to VOI \(_{50\%}\) for high tumour-background ratios and close to VOI \(_{70\%}\) for a low tumour-background ratio. Overall, VOI \(_{50\%}\) seems to be most accurate for high-resolution (noisy) data and VOI \(_{\max}\) seems to be most accurate for smoothed (low-noise) data. The fixed volume VOI (VOI \(_{\text{square}: 15 \text{ mm}}\)) performs worst since it includes a significant number of nontumour voxels, especially in smaller lesions.

In a VOI definition study of response assessment of lung carcinoma comparing VOI \(_{\text{manual}}\), VOI \(_{\max}\), VOI \(_{\text{circle}: 15 \text{ mm}}\), VOI \(_{75\%}\) and VOI \(_{50\%}\) it is mentioned that VOI \(_{50\%}\) in lesions with low uptake are often discarded on postchemotherapy scans because of inclusion of nontumour tissue. Nevertheless, an excellent test-retest reproducibility of the volume of VOI \(_{50\%}\) on two consecutive days is reported (intraclass correlation coefficient \(ICC = 0.99\)). The fixed volume-based VOI (VOI \(_{\text{circle}: 15 \text{ mm}}\)) shows best reproducibility with respect to SUV \(_{\text{mean}}\) (\(ICC = 0.95\)), but the reproducibility of VOI \(_{50\%}\) is also very high (\(ICC = 0.91\)).

Due to the partial volume effect (PVE), the isocontour level for proper whole-lesion delineation is lesion size dependent. Previously, in phantom measurements the exponential relation between lesion volume and threshold (percentage of maximum value) was determined for five different contrast levels and used on metastatic lung lesions yielding highly correlated VOI and CT volumes (correlation coefficient: \(R = 0.999\)). A more sophisticated approach was promoted later in which a 3D sphere was convolved with a symmetric trivariate Gaussian point spread function. This equation for background-subtracted relative threshold level (RTL) as a function of tumour radius and image resolution is independent of the tumour-background ratio and is used iteratively on PET data until the measured PET volume and threshold match. In patients the technique seems feasible and similar dimensions are achieved as pathological examination of liver metastases.

In conclusion, a number of methods for definition of the volume-of-interest are described in the literature. The maximum voxel value is preferable in data with limited noise, does not require specialised algorithms and does not suffer from inter-observer variability. A threshold-based VOI can provide reproducible quantification of glucose metabolism with better accuracy in noisy images. The NEDPAS guidelines recommend the use of a VOI \(_{41\%} \cdot (B+\max)\), but to increase the (background-corrected) threshold when no meaningful tumour volume definitions are provided in lesions with a low signal to background ratio. They stress that the maximal voxel value should always be noted.
1.5 Quantification

Quantification of $^{18}$F-FDG uptake or metabolism of the tissue within the VOI can be performed on several levels of complexity: semiquantitatively or quantitatively (from 4D dynamic $^{18}$F-FDG PET acquisitions) using pharmacokinetic analysis of two-compartment models.

**Semiquantitative methods**

The simplest method to quantify tracer uptake is by calculating tumour-nontumour ratios ($T/N$) using a reference tissue, which may be difficult to define or may have limited uptake (resulting in a relatively high noise level). Furthermore, uptake in this reference tissue may be influenced by factors such as therapy and therefore $T/N$ ratios can change without change in tumour biology.

Since tracer uptake is directly related to body volume of the patient and the AA present in the subject at start time of the scan, measured uptake [Bq cm$^{-3}$] should be normalised for these factors. This results in an SUV and is in older papers being referred to as differential absorption (or uptake) ratio ($DAR$, $DUR$). The least complicated normalisation is by AA and patient body weight ($SUV_{BW}$). For uniform tracer distribution the SUV is 1.0 and an SUV $> 1$ implies tracer accumulation. As sceptically reviewed by Keyes [107], body composition and habitus are a source of variability since fat has a much lower uptake of $^{18}$F-FDG than other tissues. Therefore, other definitions for volume of distribution of $^{18}$F-FDG are proposed. It is observed that $SUV_{BW}$ is still positively correlated to body weight for normal tissues, leading to $SUV_{s_{BW}}$ in heavy weighted patients up to twice that of normal patients. However, the $SUV_{LBM}$ is weight independent [108]. Others [109–111] provide evidence that the overestimation of $SUV_{BW}$ of liver tissue in heavy oncological patients can be prevented using BSA (body surface area) as normalisation factor. See table 1.2 (page 17) for definition of various types of SUV.

<table>
<thead>
<tr>
<th>Parameter representing $V_D$</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodyweight (BW)</td>
<td>$SUV_{BW} = \frac{AC_{FDG}}{AA} \cdot BW \cdot 1,000$</td>
</tr>
<tr>
<td>Body surface area (BSA)</td>
<td>$SUV_{BSA}[10^{-6} \text{ m}^{-1}] = \frac{AC_{FDG}}{AA} \cdot (0.007184 \cdot BW^{0.425} \cdot h^{0.725})^*$</td>
</tr>
<tr>
<td>Lean body mass (LBM)</td>
<td>$\varphi$: $SUV_{LBM} = \frac{AC_{FDG}}{AA} \cdot (48.0 + 1.06 \cdot (h - 152)) \cdot 1,000$</td>
</tr>
<tr>
<td></td>
<td>$\varphi$: $SUV_{LBM} = \frac{AC_{FDG}}{AA} \cdot (45.5 + 0.91 \cdot (h - 152)) \cdot 1,000$</td>
</tr>
<tr>
<td>Bodyweight &amp; plasma glucose</td>
<td>$SUV_{BW + glc} = \frac{AC_{FDG}}{AA} \cdot C_{glc} \cdot BW \cdot 1,000$</td>
</tr>
</tbody>
</table>

Table 1.2: Various methods for normalisation of $^{18}$F-FDG uptake for different parameters representing distribution volume ($V_D$). AA: administered activity, corrected for decay occurring between time of measurement and start time of scan [Bq]; $AC_{FDG}$: average measured $^{18}$F-FDG activity concentration in VOI [Bq cm$^{-3}$]; BW: bodyweight [kg]; $C_{glc}$: glucose concentration in venous plasma [mmol·l$^{-1}$]; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-d-glucose; $h$: height [cm]; SUV: standardised uptake value [g·cm$^{-3}$] (often SUV is presented dimensionless since it is based on the density of water ($\rho = 1.0$ g·cm$^{-3}$); VOI: volume-of-interest. $^*$ $SUV_{BSA}$ has a different unit (i.e., $10^{-6}$ m$^{-1}$), therefore, it is to be rescaled according to the ratio $SUV_{BW} \cdot SUV_{BSA}^{-1}$, which is 3.42·0.083$^{-1}$ = 41.2 kg·m$^{-2}$ in one study [109].
Another factor related to SUV is the plasma glucose level. It is unlikely that the effects of variations in glucose level only hold true for hyperglycaemic conditions. The decreased uptake values during hyperglycaemia can be adjusted for by normalising the SUV for plasma glucose divided by the population average (100 mg·l⁻¹ (≈5.6 mmol·l⁻¹)) [49]. The NEDPAS guidelines [46] support a normalisation for glucose concentration based on plasma glucose concentration [mmol·l⁻¹] divided by the population average (5.0 mmol·l⁻¹).

Quantitative measures

Analysis of two-compartment models by non-linear regression

4D Dynamic PET studies provide the opportunity to perform pharmacokinetic analysis of two-compartment models of glucose metabolism. In addition to the PET signal \( C_{PET}(t) \), the tracer concentration in the arterial blood plasma should also be measured \( C_{plasma}(t) \), which is a drawback compared to static methods in which only the SUV is of interest. Another drawback of dynamic acquisition is that only one FoV (typically ~15-20 cm) can be taken into account; therefore, in metastasised disease not all lesions can be quantified simultaneously.

Since 2-deoxy-d-glucose-6-phosphate, in contrast to glucose-6-phosphate, cannot be catabolised further and does not diffuse across cell membranes, metabolism can be simplified to a two-compartment model (figure 1.2, page 19) with four rate constants. Tracer kinetic modelling is based on several key assumptions including the tracer principle (i.e., negligible concentration of the tracer), steady-state assumption (i.e., metabolic processes are at steady state during measurement), tissue homogeneity and instantaneous mixing assumption (i.e., homogeneous tracer distribution within each compartment), linearity assumption (i.e., rate constants are independent of tracer concentration: first-order kinetics) and the tracer dynamic assumption (i.e., the tracer behaves similarly to the substance under investigation) [112]. Moreover, it is assumed that the extraction of \(^{18}\)F-FDG from the plasma normally is low enough for the delivery of \(^{18}\)F-FDG to be independent of blood flow.

According to the Michaelis-Menten hypothesis, an intermediate complex is formed between the substrate (S) and the transporter or enzyme, which is then converted to the chemical product (P) with release of the transporter or enzyme. The reaction rates of these processes are described by \( k_{S \rightarrow P} = V_{max} \cdot [S] \cdot ([S] + K_m)^{-1} \) in which \( V_{max} \) is the maximum rate of the reaction and \( K_m \) (the Michaelis constant) is that concentration of the substrate ([S]) which leads to 0.5 \( \cdot V_{max} \). This is clearly a non-linear relation, but it is still possible to use linear tracer compartment models if an alternative substrate (\( S^* \)) is competing for the transporter/enzyme, the concentration of which is of a much lower value than of S [113][114]. In this case, the reaction rate of the original substrate is (approximately) unaltered and the reaction rate of the tracer can be described as: \( k_{S \rightarrow P^*} \cong V_{max} \cdot [S^*] \cdot K_m \cdot (K_m \cdot ([S] + K_m))^{-1} \) which is a linear function of \([S^*]\) as long as \([S^*] \ll [S]\). Therefore, the two-compartment model with four pharmacokinetic rate constants (Phelps 4K model) can be expressed by the following two differential equations [35]:

1. Methodological Considerations in FDG PET
1.5. Quantification

Figure 1.2: The two-compartment model for $^{18}$F-FDG catabolism. The vertical dotted line symbolises the cell membrane. $C_{\text{bound}}(t)$: the intracellular activity concentration of $^{18}$F-FDG-6-phosphate; $C_{\text{free}}(t)$: the intracellular activity concentration of free $^{18}$F-FDG; $C_{\text{PET}}(t)$: the measured PET signal which is a combination of $C_{\text{free}}(t)$, $C_{\text{bound}}(t)$ and a fraction (shaded area, $V_B$) of $C_{\text{plasma}}(t)$; $C_{\text{plasma}}(t)$: the activity concentration of $^{18}$F-FDG in the arterial blood plasma; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-D-glucose; GLUT: facilitative membrane-bound sodium-independent glucose transporters; $K_1$: GLUT influx rate constant; $k_2$: GLUT efflux rate constant; $k_3$: hexokinase phosphorylation rate constant; $k_4$: glucose-6-phosphatase dephosphorylation rate constant; PET: positron emission tomography; $V_B$: blood volume fraction.

\[
\begin{align*}
\frac{dC_{\text{free}}(t)}{dt} &= K_1 \cdot C_{\text{plasma}}(t) - (k_2 + k_3) \cdot C_{\text{free}}(t) + k_4 \cdot C_{\text{bound}}(t) \\
\Rightarrow C_{\text{free}}(t) &= \frac{K_1}{\alpha_2 - \alpha_1} \cdot \left[ (k_4 - \alpha_1) \cdot e^{-\alpha_1 \cdot t} + (\alpha_2 - k_4) \cdot e^{-\alpha_2 \cdot t} \right] \otimes C_{\text{plasma}}(t) \\
\frac{dC_{\text{bound}}(t)}{dt} &= k_3 \cdot C_{\text{free}}(t) - k_4 \cdot C_{\text{bound}}(t) \\
\Rightarrow C_{\text{bound}}(t) &= \frac{K_1 \cdot k_3}{\alpha_2 - \alpha_1} \cdot \left( e^{-\alpha_1 \cdot t} - e^{-\alpha_2 \cdot t} \right) \otimes C_{\text{plasma}}(t)
\end{align*}
\]

Where ‘$\otimes$’ stands for the operation of convolution and:

\[
\alpha_{1,2} = \frac{1}{2} \cdot \left[ k_2 + k_3 + k_4 \mp \sqrt{(k_2 + k_3 + k_4)^2 - 4 \cdot k_2 \cdot k_4} \right]
\]

The sum of the activity concentrations in both compartments plus a fraction of the plasma activity concentration ($V_B \cdot C_{\text{plasma}}(t)$, with $V_B$ the blood volume fraction) are measured within the VOI by the dynamic PET acquisition:

\[
C_{\text{PET}}(t) = (1 - V_B) \cdot (C_{\text{free}}(t) + C_{\text{bound}}(t)) + V_B \cdot C_{\text{plasma}}(t)
\]

Therefore, when the plasma input function is known, all five free parameters of $^{18}$F-FDG metabolism ($K_1$, $k_2$, $k_3$, $k_4$ and $V_B$) can be estimated using non-linear least squares fitting of $C_{\text{PET}}(t)$. 

\[19\]
1. Methodological Considerations in FDG PET

The original irreversible model (Sokoloff 3K model) did not incorporate $k_4$\cite{34} since the rate of hydrolysis of $^{18}$F-FDG-6-phosphate by glucose-6-phosphatase activity is negligible in mammalian tissues, except for liver tissue\cite{115}. This is verified in further studies\cite{116}, which compare the residual sum of squares of fits with and without $k_4$ by the Akaike Information Criterion\cite{117} and Schwarz Criterion\cite{118}. These statistics reward goodness of fit but depreciate the number of free parameters of the model (the lowest value denotes the model that best explains the data with a minimum of free parameters). Moreover, simulation studies caution that a $k_4$ might result from tissue heterogeneity rather than real dephosphorylation\cite{119}. Therefore, currently most studies use the simplified three-rate constant (Sokoloff 3K) model. With $k_4 = 0$ min$^{-1}$, equations 1.1-1.2 (page 19) simplify to:

$$\frac{dC_{\text{free}}(t)}{dt} = K_1 \cdot C_{\text{plasma}}(t) - (k_2 + k_3) \cdot C_{\text{free}}(t)$$

$$\Rightarrow C_{\text{free}}(t) = K_1 \cdot e^{-(k_2+k_3)t} \otimes C_{\text{plasma}}(t) \quad (1.5)$$

$$\frac{dC_{\text{bound}}(t)}{dt} = k_3 \cdot C_{\text{free}}(t)$$

$$\Rightarrow C_{\text{bound}}(t) = \frac{K_1 \cdot k_3}{k_2 + k_3} \cdot \left(1 - e^{-(k_2+k_3)t}\right) \otimes C_{\text{plasma}}(t) \quad (1.6)$$

The ratio of the phosphorylation rates of $^{18}$F-FDG ($MR_{FDG}$) and glucose ($MR_{glc}$) equals both the ratio of fluxes between the compartments of free and bound substrates (equation 1.6, page 20) and the ratio of the $V_{\text{max}}$, $K_m$ and concentration of both the intracellular free tracer and the natural substrate:

$$\frac{MR_{FDG}}{MR_{glc}} = \frac{k_{3,FDG} \cdot C_{\text{free,FDG}}(t)}{k_{3,glc} \cdot C_{\text{free,glc}}(t)} = \frac{V_{\text{max,FDG}} \cdot K_{m,glc} \cdot C_{\text{free,FDG}}(t)}{V_{\text{max,glc}} \cdot K_{m,FDG} \cdot C_{\text{free,glc}}(t)} \quad (1.7)$$

In contrast to analogue tracers (such as $^{18}$F-FDG) when direct isotopic substitution labelling of glucose is used (e.g., 1-(11)C)-d-glucose), the $V_{\text{max}}$ and $K_m$ for both substrates are essentially the same and equation 1.7 (page 20) would reduce to a simple ratio of concentrations of $^{18}$F-FDG and glucose. In the condition where $C_{\text{plasma}}(t)$ of glucose and $^{18}$F-FDG is constant, equation 1.7 (page 20) can be written as:

$$\frac{MR_{FDG} \cdot (C_{\text{plasma,FDG}} \cdot F)^{-1}}{MR_{glc} \cdot (C_{\text{plasma,glc}} \cdot F)^{-1}} = \frac{V_{\text{max,FDG}} \cdot K_{m,glc} \cdot \lambda_{FDG}}{V_{\text{max,glc}} \cdot K_{m,FDG} \cdot \lambda_{glc}} = LC_{FDG} \quad (1.8)$$

In which $F$ is the blood flow and $\lambda_{FDG}$ and $\lambda_{glc}$ are the partition coefficients ($C_{\text{free,FDG}} \cdot C_{\text{plasma,FDG}}^{-1}$) of both the substrates. This ratio is called the lumped constant of $^{18}$F-FDG ($LC_{FDG}$) or the steady-state ratio of the net extraction of $^{18}$F-FDG to that of glucose at constant plasma levels of both substrates\cite{34}. The full operational equation is often written as: $LC_{FDG} = V_{\text{max,FDG}} \cdot K_{m,glc} \cdot \lambda \cdot (V_{\text{max,glc}} \cdot K_{m,FDG} \cdot \varphi)^{-1}$. 

20
In this equation λ denotes the ratio of partition coefficients and ϕ is the fraction of d-glucose-6-phosphate that continues down the Embden-Meyerhof pathway (i.e., regular glycolysis), which is normally quite close to 1.0. The value of the $LC_{FDG}$ can be determined by simultaneous measurement of $MR_{glc}$ (e.g., by 1-$(1^{1}C)$-d-glucose) and $MR_{FDG}$ [120], independent determination of all six parameters of the LC or by the ratio of fractional arteriovenous differences for the two substrates [34][121].

In the steady-state condition of the compartment of free intracellular $^{18}$F-FDG, the flux into this compartment is balanced by the flux out ($dC_{\text{free,FDG}}(t) \cdot (dt)^{-1} = 0$), the $MR_{glc}$ can be estimated from equations 1.6 (page 20) and 1.8 (page 20):

$$MR_{glc} = \frac{dC_{\text{bound,glc}}(t)}{dt} = \frac{MR_{FDG}}{C_{\text{plasma,FDG}}} \cdot \frac{C_{\text{plasma,glc}}}{LC_{FDG}} = \left(\frac{K_{1} \cdot k_{3}}{k_{2} + k_{3}}\right)_{FDG} \cdot \frac{C_{\text{plasma,glc}}}{LC_{FDG}}$$

(1.9)

Primary, glucose metabolism (and thus the $LC_{FDG}$) are determined in the normal (rat) brain [50][113][120]. In this field it has been shown that the use of a single $LC_{FDG}$ suits only considering the conditions during which it was determined. The $LC_{FDG}$ is dependent on the time between tracer injection and measurement, surely changes in any disease with an enzymatic component and probably varies with regional glucose concentrations and in conditions of ischaemia, hypoglycaemia, the method of anaesthesia, age (adult or developing) and species of the subject under investigation. Due to the heterogeneity of neoplasia, these limitations are far greater and for many tumours the $LC_{FDG}$ is not known. As perfectly summarised elsewhere, calculations of $MR_{FDG}$ can be used if one is interested in comparative measurements (e.g., metabolic changes over time, during treatment, between a diseased or normal tissue or between different physiological states), assuming the $LC_{FDG}$ of the tissue under investigation remains unchanged. In cases in which the true $MR_{glc}$ needs to be determined from $^{18}$F-FDG results, then one must measure the $LC_{FDG}$ in the particular experimental setup [50]. Studies of the $LC_{FDG}$ in oncology are very limited and reports of determination of the $LC_{FDG}$ of non-glioma malignant tissues are lacking to this date.

**Patlak graphical method**

Patlak et al. [33] derived a graphical method (often called Patlak analysis, Patlak-Rutland or Gjedde-Patlak plot) that uses linear regression to analyse pharmacokinetics described by any compartment model with at least one irreversible transport step or reaction (‘trapping’ of $^{18}$F-FDG due to $k_{4} = 0$ min$^{-1}$). It further assumes that all the reversible compartments must be in equilibrium with the plasma, which in practice only occurs when $dC_{\text{plasma}}(t) \cdot (dt)^{-1}$ is small enough for these tissue compartments to follow. Combining equations 1.4-1.6 (page 19-20), with $K_{i} = K_{1} \cdot k_{3} \cdot (k_{2} + k_{3})^{-1}$ leads to:
1. Methodological Considerations in FDG PET

\[
\frac{C_{PET}(t)}{C_{plasma}(t)} = (K_i \cdot (1 - V_B)) \cdot \left( \frac{\int_0^t C_{plasma}(\tau) \, d\tau}{C_{plasma}(t)} \right) + \left( \frac{(1 - V_B) \cdot K_1 \cdot k_2}{(k_2 + k_3)^2} + V_B \right)
\] (1.10)

Linear regression of the plot \( C_{PET}(t) \cdot C_{plasma}(t)^{-1} \) vs \( \int_0^t C_{plasma}(\tau) \, d\tau \cdot C_{plasma}(t)^{-1} \) (“Patlak space” or “funny time”) results in slope: \( K_i \cdot (1 - V_B) \), from which, with an estimated \( V_B \), \( LC_{FDG} \) and a known \( C_{plasma, glc} \), \( MR_{glc} \) can be computed using equation 1.9 (page 27). Simplification of the problem of solving differential equations by non-linear optimisation to an approach amenable to linear regression avoids many problems inherent in the former approach: sensitivity to noise in the time-activity concentration curves, parameter covariance, local minima in the approximate solution to the differential equations and dependence of parameter estimates on starting guesses. As a trade-off only the \( K_i \), but not the individual kinetic parameters, is estimated (figure 1.3, page 23).

Linearity in Patlak space is reached in conditions where the plasma \(^{18}\)F-FDG concentration is constant, but since \( C_{plasma}(t) \) continues to decrease due to irreversible cellular uptake and renal clearance, this situation is never fully met. In good approximation, onset of linearity is usually attained 10-15 min after the bolus injection. It is shown that \( MR_{glc} \) determined by Patlak analysis over different intervals is highly similar compared to the Sokoloff (3K) model (all \( R^2 \geq 0.951 \)). VOIs were defined on the last three time frames (45-60 min), which is a limitation of this study; in shortened protocols, tumour VOIs have to be drawn at earlier time points, in which the contrast is lower, leading to different VOIs which likely will produce less accurate \( MR_{glc} \) [116]. In a study with 20 patients with non-small cell lung carcinoma using the Sokoloff (3K) model, it was shown that parameter estimation by non-linear least squares fitting of 30-min 4D dynamic data yielded essentially the same results for \( K_1, k_2, k_3 \) and \( K_i \), \( (R^2 = 0.918, 0.937, 0.785 \) and 0.924, respectively, with mean relative differences varying 9-25%) compared to a 60 min protocol [122]. Another way of shortening acquisition duration is by combination of an initial 10-min 4D dynamic scan with a single 3D static time frame 56-60 min after tracer injection [123]. Highly similar values for \( K_i \) between full dynamic and a shortened dynamic acquisition are described (nonlinear regression: \( R^2 = 0.815 \)). In practice the VOIs need to be repositioned since the patient leaves the gantry between both acquisitions.

Adaptations to pharmacokinetic models

Variations to the previously mentioned pharmacokinetic models have been reported frequently [124][130]. A variation to the Sokoloff 3K model with six parameters uses a reference region outside the tumour to account for the normal tissue within the tumour VOI (2 VOIs 6P model). In simulation studies it was found to adequately describe
1.5. Quantification

Figure 1.3: Quantitative analysis of a 4D dynamic $^{18}$F-FDG PET/CT of a pT$_2$ adenocarcinoma of the right superior lung lobe (left top). Right top: time-activity concentration curves. Left bottom: analysis of two-compartment models by both Sokoloff 3K and Phelps 4K model. These non-linear fits resulted in $K_1 = 0.153$ ml·g$^{-1}$·min$^{-1}$, $k_2 = 0.481$ min$^{-1}$, $k_3 = 0.0644$ min$^{-1}$, $V_B = 0.122$ and $K'_1 = 0.0181$ ml·g$^{-1}$·min$^{-1}$ ($AIC = 390$, Sokoloff 3K) and $K_1 = 0.178$ ml·g$^{-1}$·min$^{-1}$, $k_2 = 0.640$ min$^{-1}$, $k_3 = 0.0898$ min$^{-1}$, $k_4 = 0.00853$ min$^{-1}$ and $V_B = 0.116$ ($AIC = 372$, Phelps 4K). Right bottom: Gjedde-Patlak graphical analysis. The linear fit line “$y = 0.0148 \cdot x + 0.401$” with $R^2 = 0.997$ results in the parameter $K_i \cdot (1 - V_B) = 0.0148$ ml·g$^{-1}$·min$^{-1}$. $AIC$: Akaike’s an information criterion; $C_{PET}(t)$: the measured PET signal which is a combination of intracellular free $^{18}$F-FDG, intracellular bound $^{18}$F-FDG-6-phosphate and a fraction ($V_B$) of $C_{plasma}(t)$; $C_{plasma}(t)$: the activity concentration of $^{18}$F-FDG in the arterial blood plasma; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-d-glucose; GLUT: facilitative membrane-bound sodium-independent glucose transporters; $K_1$: GLUT influx rate constant; $k_2$: GLUT efflux rate constant; $k_3$: hexokinase phosphorylation rate constant; $k_4$: glucose-6-phosphatase dephosphorylation rate constant; $K_i$: $^{18}$F-FDG net influx constant; PET/CT: positron emission tomography / computed tomography; $V_B$: blood volume fraction.

tumour pharmacokinetic rate constants of two-compartment models regardless of the amount of normal tissue within the VOI. A drawback of the technique is that reference tissue must be available [124].

Many variations to the Patlak graphical method are reported [125–130]. Wong et al. [125] describe a technique in which they create three sequential images (each 15 min duration) starting 10 min after $^{18}$F-FDG injection combined with three arterialised venous samples taken at the mid-time of each scan. This method pre-empts the need for continued sampling and showed $K_i$ to be within 3-4% of the values obtained by regular Patlak analysis. Another variation computes lesion $K_i$ from the mean of the per voxel $K_i$ for which the correlation coefficient of the Patlak plot is above a threshold [126]. This technique was further adapted by defining a 2D VOI on a summed image of all correlation coefficient-constrained planes containing the lesion. This VOI was propagated over the planes containing the lesion in which the paramet-
ers of interest were determined. The authors conclude that this technique (total lesion evaluation method, TLE) is less VOI dependent and also incorporates therapy-related volume changes and is especially suitable for therapy response evaluation [127]. A simplified kinetic method (SKM) is advocated [128], assuming a uniform input function, based on fitting the input function of control patients with a triexponential decay function. The SKM allows $MR_{glc}$ to be calculated from a static image and one late venous blood sample. It was shown in lung carcinoma patients that the SKM was an improvement over the $SUV$ and approaches the $MR_{glc}$, calculated by non-linear least squares. A hybrid method of Patlak and SKM shows less bias and variability than either technique alone. For this method six time frames (each 5 min, 25-55 min postinjection) are acquired. By using every other time frame (three data points) or every third time frame (two data points) this method can be used to up to three fields-of-view [129]. The last method mentioned is based on the relation between $K_i$ and $SUV$, average plasma clearance rate and the initial distribution volume of $^{18}$F-FDG (Sadato method). Estimated $K_i$ and $SUV$ are compared with Patlak $K_i$, leading to the conclusion that the estimated, non-invasively determined $K_i$ is a better indicator of tumour uptake than the $SUV$ [130].

Hoekstra et al. reviewed [131] and compared [116] 34 variations of previously mentioned methods to obtain glucose consumption (two $T/N$, 12 $SUV$, two Sadato, two SKM, five Patlak, ten TLE and one “2 VOI, 6P” variations) on 30 randomly selected 4D dynamic $^{18}$F-FDG PET scans in 19 lung carcinoma patients. Since incorporation of $k_4$ did not improve fits, the Sokoloff 3K model was used as gold standard. The reliability of the gold standard was considered high since the test-retest variability ($ICC = 0.95$) and intra- and interobserver variability ($ICC = 0.98$ for both) were small. Of the 34 models tested, ten met the required minimal correlation with the gold standard ($R^2 > 0.95$). They concluded that the best simplified options are $SUV_{BSA+glucose}$ (40-60 min or 50-60 min postinjection), Sadato method based on $BSA$ and Patlak graphical $MR_{glc}$ analysis (10-60 min postinjection). Similarly, Lammertsma et al. [132] pooled the results of three studies comparing methodological aspects of therapy response evaluation in lung [116], breast [133] and gastroesophageal [134] carcinoma (in total 170 $^{18}$F-FDG PET studies) and show excellent correlation with the Sokoloff 3K model of Patlak $MR_{glc}$ ($ICC = 0.98$), SKM ($ICC = 0.94$), $SUV_{BSA+glucose}$ ($ICC = 0.91$) and $SUV_{BSA}$ ($ICC = 0.91$). They conclude that although Patlak $MR_{glc}$ has the best correlation with the gold standard, it remains to be proved that these findings are of clinical relevance. Changes found by $SUV$ estimation may still represent a relevant response [21][22].

**Input functions**

For all previously described fully quantitative measures of glucose metabolism, the arterial plasma input function (IF or $C_{plasma}(t)$) of the tumour should be known. Different approaches to obtain this IF have been described in the literature. Since it is impossible to sample the artery directly vascularising the tumour, the gold standard is serial arterial sampling of a superficial artery (e.g., the radial artery) after which the activity concentration of the plasma derived by centrifugation can be determined [135]. Alternatives to serial arterial sampling with less complications [136] or less
1.5. Quantification

Radiation burden to the personnel are: (arterialised) venous blood sampling \cite{137}, use of an IDIF \cite{83,138}, modelling of a population-based IF \cite{139,141} or extraction of IF using mathematical segmentation methods \cite{142}.

To overcome the time-dependent ratio of arterial and venous blood activity concentration, the heated hand procedure, which shunts the arterial blood to the venous system, can be performed \cite{137}. In Patlak analysis, the use of arterialised venous blood shows a net effect of \(\sim 10\%\) overestimation of \(K_i\) and \(\sim 5\%\) overestimation of \(MR_{glc}\) compared to the gold standard of arterial sampling. IDIFs of the ascending \((ICC = 0.98)\) or abdominal \((ICC = 0.96)\) aorta show better Patlak \(MR_{glc}\) correlation than IDIFs of the left ventricle \((ICC = 0.94)\) \cite{83,138}, but due to the PVE the recovery of activity concentrations in the aorta is less than one, causing an underestimation of the \(^{18}\)F-FDG activity concentration. A left ventricle IDIF shows positive bias, since the PVE caused a hot myocardium to spill-in activity in later time frames \cite{83}, leading to an underestimation of \(MR_{glc}\) of \(16.2-17.5\%\) \cite{143}.

Another drawback of the IDIF is that it is a measure of whole-blood activity concentration, which is known to be lower than in plasma and not constant over time \cite{137}. Population-based curves reduce the need for blood sampling and moreover can be used in body areas where IDIFs are not feasible. They can be based on averaging normalised blood curves or on fitting to a proposed equation \cite{139,140}. Patlak-based \(MR_{glc}\) obtained by population-based curves overall show high correlation with the gold standard \((R^2 > 0.984)\) \cite{139}. Usage of mathematical tissue segmentation results in whole-blood activity concentrations as well \cite{142,144}, but it leads to an image-derived whole-blood IF with similar drawbacks as an IDIF.

For estimation of pharmacokinetic rate constants of two-compartment models, the exact timing (time delay) and shape (dispersion) of the IF are needed. When it is measured by sampling \cite{115}, this does not necessarily reflect the supply of the tracer to the lesion. The time delay can be corrected for, but the dispersion of the IF results from the impulse response characteristics of the distributing system of the patient and is therefore difficult to predict \cite{114}. These factors complicate the use of pharmacokinetic analysis of two-compartment models and therefore frequently IDIFs of a large blood pool close to the tumour are used \cite{84,85}. The effect of time delay and dispersion is of negligible relevance for Patlak \(MR_{glc}\) estimation, since the \(C_{plasma}(t)\) in the Patlak equation is relatively constant in the period of linear regression and the integral of \(C_{plasma}(t)\) is almost not affected.

Concluding, for clinical practice and most research into the value of quantification of \(^{18}\)F-FDG PET, the \(SUV_{max}\) and \(SUVs\) of isocontour-based VOIs are most relevant. They can be performed whole-body, pre-empt the need of an IF and can be obtained from static images. In special situations \(^{18}\)F-FDG metabolism can be quantified using 4D dynamic acquisition, preferably using IDIFs of the ascending aorta. Even though these result in absolute quantification of \(^{18}\)F-FDG influx, they are limited in calculation of ‘real’ \(MR_{glc}\) due to the indefinite value of the \(LC_{FDG}\).
1.6 Repeated measures studies

Quantitative $^{18}$F-FDG PET is not only useful as a single measurement (e.g., for treatment stratification), but repeated measurements before, during and after treatment may be used for early response assessment of therapy. For this purpose a new quantification parameter is introduced, defined as the product of the $SUV$ and (metabolic) tumour volume (total lesion glycolysis, $TLG$) [145]. Even though the $\delta TLG$ is used in a small number of recent studies [146,147], this response parameter needs further evaluation before it can be recommended for routine use, since it does not perform better than $\Delta SUV$ alone in all studies [146].

Measured tumour response is nearly independent of the VOI definition, since most factors contributing to bias and noise cancel out in calculation of the relative change [42,44]. However, when the tumour volume changes significantly, the PVE can play a role which cannot be cancelled by measuring relative effects [42]. For matter of reproducibility, threshold-based VOIs are recommended for repeated measures studies, with a test-retest reproducibility of the $SUV$ of $1 − 13 ± 6 − 12\%$ [44,148,150], of $K_i$ of $10 ± 8\%$ and of $K_1$, $k_2$ and $k_3$ of $24 − 42 ± 13 − 31\%$ [149]. Due to the magnitude of the day-to-day variation, changes in quantification of $SUV ± 15 − 20\%$ (i.e., 2 standard deviations) are within the reproducibility limits of this method of quantification and therefore should be considered as stable values.

The robustness of variations of the $\Delta SUV [\%]$ was shown in gastric carcinoma patients neoadjuvantly treated with cisplatin-based chemotherapy [43]. The authors conclude that in gastric carcinomas the prediction of response to chemotherapy on the basis of relative tumour $SUV$ changes is not essentially influenced by any of the methodological variations investigated (incubation period 40 or 90 min, reconstruction method FBP or OSEM and various $SUV$ normalisations). However, phantom experiments reveal a difference between institutions for $SUV$ quantification up to 30%, which can be improved by calibration [61]. This is especially important when baseline and follow-up scans are performed in different settings or institutions, underlining the need for standardisation [46].

1.7 Conclusions

Apart from hardware issues and sources of error, many methodological and biological factors influence quantification in $^{18}$F-FDG PET. For multicentre investigation these parameters must be standardised and intercentre calibration has to be performed. In general, the relative simplicity of semiquantitative quantification by $SUV$ seems to outweigh its drawbacks, providing that the process from acquisition till quantification is standardised. Repeated measures studies seem less dependent on most factors influencing quantification as they cancel out in calculation of relative treatment effects providing that the methodology at the baseline scan is repeated for the follow-up scan. Pharmacokinetic quantification is a sophisticated and rather complicated method and is therefore mainly applied in a research setting.
Tailoring Therapy in Colorectal Carcinoma by $^{18}$F-FDG PET/CT

*Q J Nucl Med Mol Imaging, 2009 Apr;53(2):224-44*
(adapted and reprinted with permission)

D. Vriens$^1$,  
L.F. de Geus-Oei$^1$,  
W.T.A. van der Graaf$^2$ and  
W.J.G. Oyen$^1$

dep. of $^1$Nuclear Medicine and  
$^2$Medical Oncology, Radboud University Medical Centre,  
Nijmegen, the Netherlands
Abstract

Positron emission tomography (PET) using 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose (\(^{18}\)F-FDG) has an added value in the clinical management of patients with colorectal carcinoma (CRC). This includes restaging patients before surgical resection or local recurrence of liver metastases, assessment whether residual lesions are scar or recurrence and in pinpointing recurrence in case of unexplained increase in serum levels of carcinoembryonic antigen.

At present, there is an increasing interest in new roles for \(^{18}\)F-FDG PET, especially for characterisation of lesions, for prognosis and response prediction and for early evaluation of therapy response to commenced therapy. \(^{18}\)F-FDG PET may lead to better selection of patients for different therapeutic options or to early individual adjustment of current treatment.

This systematic review aims to provide an up-to-date overview of literature on the current and potential value of \(^{18}\)F-FDG PET in CRC patients by addressing staging and recurrence detection, prognosis and response prediction and evaluation of preoperative (chemo)radiotherapy for primary rectal carcinoma, ablative treatment for unresectable liver metastases and chemotherapy for advanced CRC.

Keywords Colorectal Neoplasms · 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose · Management · Positron Emission Tomography · Prognostic Stratification · Therapy · Therapy Response Evaluation
2.1 Introduction

Colorectal carcinoma (CRC) is the third most common malignancy and third leading cause of cancer-related deaths in the United States. Even though the annual age-adjusted incidence rates and death rates are slowly declining in the last two decades, it remains a large health problem worldwide [151]. According to the National Cancer Institute [151,152] the age-adjusted incidence rate in the United States (2001-2005) is 50.6 per 100,000 per year, with a cancer-related death rate of 18.8 per 100,000 per year and an overall five-year survival rate of 64.4%. Approximately 5.3% of people will develop CRC during their lives and it is estimated that in 2008, 148,810 people were diagnosed with and 49,960 people died from CRC in the United States.

At time of diagnosis, approximately 40% of CRC is confined to the primary site, 36% has spread locoregionally and 19% of patients are suffering from metastasised disease (for 5% in this registration the stage was unknown) [151,152]. Progress has been made in improvement of patient prognosis with the introduction of hepatic resection for treatment of isolated liver metastasis and with the development of effective chemotherapeutic and targeted agents [153,154].

Positron emission combined with computed tomography (PET/CT) with 2-(\(^{18}\)F)fluoro-2-deoxy-D-glucose (\(^{18}\)F-FDG) has proven a useful diagnostic modality in different phases of CRC management. This comprehensive review discusses the current and potential future applications of \(^{18}\)F-FDG PET in management decisions of patients with CRC. The literature is systematically reviewed on the (potential) role of \(^{18}\)F-FDG PET in changing individual CRC patient management by addressing 1) the impact of \(^{18}\)F-FDG PET on staging disease and detection of recurrence on individual management, 2) the prediction of individual patient prognosis and therapy response and 3) the evaluation of therapy response.

2.2 Search strategy & selection criteria of literature

References for this review were identified by systematic searches in PubMed, EMBASE (OvidSP), MEDLINE (OvidSP) and the Cochrane Library up to December 31, 2008. The strategy of Mijnhout et al. [155] was adapted for our research question. The construct of the query was: “(PET OR PET/CT) AND colorectal AND cancer”, using medical subject headings, synonyms and truncations for all three building blocks of the search question (table 2.1 page 32).

Only articles in English were included. A total of 1,595 articles were retrieved and screened. Case-reports, small series (< 15 patients), research by questionnaires, reviews, reports from meetings, abstracts of poster presentations, editorial comments or letters-to-the editor were excluded. Papers on disease (re)staging or recurrence detection which failed to describe the implications for clinical management or without verification of the results by histology or follow-up were excluded. Papers on therapy response without fixed outcome-parameters (e.g., histological or morphological response, patient survival) were excluded from further analysis. Studies using PET-tracers other than \(^{18}\)F-FDG were excluded. Results of the search strategy were supplemented by the references from included articles. In total 86 articles were considered suitable for further discussion in this review.
2. Tailoring CRC Therapy by PET/CT

<table>
<thead>
<tr>
<th>PET OR PET/CT</th>
<th>Colorectal</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positron emission tomography (MeSH)</td>
<td>Colorectal neoplasms (MeSH)†</td>
<td>Carcinom*</td>
</tr>
<tr>
<td>PET</td>
<td>Colorect*</td>
<td>Adenocarc*</td>
</tr>
<tr>
<td>PET/[*]</td>
<td>Colon</td>
<td>Cancer</td>
</tr>
<tr>
<td>PETscan*</td>
<td>Rectal*</td>
<td>Neoplas*</td>
</tr>
<tr>
<td>PET/CT*</td>
<td>Rectum*</td>
<td>Tumor</td>
</tr>
<tr>
<td>PET-CT*</td>
<td></td>
<td>Tumour</td>
</tr>
<tr>
<td>CT/PET*</td>
<td></td>
<td>Tumours</td>
</tr>
<tr>
<td>CT-PET*</td>
<td></td>
<td>Tumours</td>
</tr>
<tr>
<td>(Positron* AND emission* AND tomograph*)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Search strategy. Rows are combined with ‘OR’ and columns with ‘AND’. † in EMBASE the corresponding subject heading is “colorectal cancer”. CT: computed tomography; MeSH: medical subject heading; PET: positron emission tomography.

When considered appropriate, results have been pooled using fixed-effects modelling, by weighting effect magnitudes (i.e., fraction management change) by their inverse variance. For calculation of the variance and confidence intervals (CIs) of proportions the β-distribution has been used since the commonly used asymptotic normal approximation only holds true for observed frequencies of ≥ 5 [156]. It should be noted, however, that the variation of results of individual papers is largely attributable to heterogeneity of the study populations.

2.3 Results

Impact of $^{18}$F-FDG PET on individual management in staging disease and detection of recurrence

Many articles address the impact of $^{18}$F-FDG PET during initial staging of primary CRC or in detection of recurrence. Recurrence of CRC can be suspected due to several findings during routine clinical follow-up: abnormalities on morphological imaging and rise of the serum tumour-marker carcinoembryonic antigen (CEA). $^{18}$F-FDG PET may have a pivotal role in patient management specifically in case of equivocal radiological studies, unexplained CEA rise or determination of resectability of local recurrence or colorectal liver metastases.

Staging primary CRC by $^{18}$F-FDG PET

In staging of primary rectal carcinoma, $^{18}$F-FDG PET may influence on management in 12% [157] to 27% [158]. Heriot et al. [159] showed in 46 patients with histology proven stage II-IV rectal carcinoma that the use of $^{18}$F-FDG PET after routine staging by abdominal CT and pelvic MRI and/or transrectal endo-ultrasonography (TREUS) before neoadjuvant therapy changed previously proposed management in 17% of patients. Of these eight cases, surgery was cancelled in six cases due to identification of metastatic disease and in two the neoadjuvant radiotherapy field was altered to include common iliac lymphadenopathy as identified by PET. Gearhart et al. [158] prospectively compared abdominal spiral CT and $^{18}$F-FDG PET after TREUS or pelvic MRI in 37 patients with previously untreated biopsy proven adenocarcinoma of the rectum and found discordant findings in 38%, leading to changes
in the previously proposed treatment plan in 27%. Of these ten cases, in five pa-
tients additional lymphnode metastases were found not detected by CT alone leading
to neoadjuvant treatment or extension of the radiotherapy field. In two cases, CT-
positive lymphnodes proved negative on PET leading to cancellation of neoadjuvant
treatment or radiotherapy. In three additional cases more extensive surgical resection
was performed. Bassi et al. [160] showed that additional staging by $^{18}$F-FDG PET
in 25 T$_{3-4}$ rectal carcinoma patients who were candidates for neoadjuvant chemora-
diotherapy (CRT) prior to surgery led to treatment changes in 16%. $^{18}$F-FDG PET
identified unknown nodal involvement and undiagnosed liver metastases. Another
study in 83 patients performed by Davey et al. [157] showed that staging $^{18}$F-FDG
PET/CT could lead to management changes in 12% of primary rectal carcinoma pa-
tients. Of these ten cases, surgery was cancelled in six patients due to unexpected
metastases, in three neoadjuvant CRT was considered necessary due to identification
of pelvic nodal spread and in one patient neoadjuvant CRT was cancelled because iliac lymphnode metastases on CT appeared to be false-positive.

In staging of primary CRC, $^{18}$F-FDG PET may influence on management in 2%
[161] to 27% [162]. Kantorova et al. [163] found a change of treatment in 16% of
38 patients with histologically proven primary CRC which were prospectively staged
by conventional imaging and $^{18}$F-FDG PET (8% treatment modality change, 13%
change in range of surgery). Park et al. [162] studied 100 patients with primary
CRC (45 colon and 55 rectum carcinoma: three stage I, 23 stage II, 25 stage III,
49 stage IV) planned for surgery with $^{18}$F-FDG PET/CT who had increased CEA
or showed equivocal signs of metastases on CT. In 27% of the patients, proposed
treatment plan was modified: nine had treatment modality changes, ten received
more extensive surgery and in eight unnecessary procedures could be avoided. The
reason for a large proportion of patients in whom management changed might be
that they only included patients with a relatively high likelihood of metastasised
disease due to equivocal radiological findings or increased CEA levels. Veit-Haibach
et al. [164] performed $^{18}$F-FDG PET/CT in 47 patients with suspicious lesions at
colonoscopy (50 sites: 13 rectum and 37 colon carcinoma). They found that $^{18}$F-
FDG PET/CT compared to CT alone led to management changes in 9% of the
patients. Another study, by Llamas-Elvira et al. [165] showed that using $^{18}$F-FDG
PET/CT next to CT changed management in 12% of the 104 patients with histology
proven CRC (56 rectum and 48 colon carcinoma) referred for surgery. In seven
cases surgery was cancelled for extensive disease not detected by CT and in five
the therapeutic approach was altered. In contrast, Furukawa et al. [161] showed no
impact of $^{18}$F-FDG PET/CT over whole-body CT alone, as PET/CT only changed
management in one of 44 patients (2%) with histologically proven primary CRC
(38 rectum carcinoma). They attributed discordance with other authors to the less
advanced disease stage in their patient series.

Pooling the data of the four studies on rectal carcinoma leads to a weighted mean
change in management of 15.7% (95% − $CI$: 10.8-21.7%), as shown in figure [2.1](page
34). For the studies with both colon and rectal carcinoma the weighted mean change
in management is 10.7% (95% − $CI$: 7.6-14.5%). Apparently due to the limited
influence of $^{18}$F-FDG on management in these patients, the latest editions of the
European Society for Medical Oncology (ESMO) guidelines for rectal carcinoma [166]
2. **Tailoring CRC Therapy by PET/CT**

Figure 2.1: Forest plot of management changes due to $^{18}$F-FDG PET with corresponding confidence intervals for staging of primary rectal (upper) and primary colorectal carcinoma (lower) described in nine references [157–165]. The size of the squares denotes the weight for calculation of the pooled average. CRC: colorectal carcinoma; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography.

or CRC [167] do not recommend to use this technique routinely for staging primary disease. They do, however, note that for staging in advanced CRC [168] $^{18}$F-FDG PET can have a role.

### Influence of $^{18}$F-FDG PET in suspected recurrence

Many studies have been performed to investigate whether $^{18}$F-FDG PET changes management in patients in whom recurrent CRC was expected based on equivocal lesions on conventional diagnostic follow-up, rising CEA levels with normal radiologic findings or during restaging before surgical treatment of local recurrence or metastases.
2.3. Results

Figure 2.2: Example of an $^{18}$F-FDG PET/CT in a male patient with a pT$_3$N$_0$M$_1$ rectosigmoid carcinoma treated by rectosigmoid resection in combination with chemotherapy for synchronous liver metastases. During follow-up with $^{18}$F-FDG PET/CT a local recurrence in the pelvis near the rectal stump (arrowhead) was detected that was equivocal on CT. CT: contrast-enhanced computed tomography; $^{18}$F-FDG PET/CT: 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography / computed tomography.

Equivocal radiologic findings suggestive for recurrence. After radiotherapy or surgery of the primary tumour, most patients develop a region of scar tissue in the surgical bed. These changes complicate the detection of local recurrence by ultrasonography, CT or MRI. $^{18}$F-FDG PET can distinguish metabolic active disease (tumour) from less active disease (scar tissue). An example is shown in figure 2.2 (page 35).

Of the 35 patients with a history of resected primary CRC described by Beets et al. [169] eight cases were included for pre-sacral masses with uncertain CT findings. $^{18}$F-FDG PET correctly classified them as recurrence in five cases (62.5%) causing change in management in these patients. In the series of Simo et al. [170] patient management was altered in 14 of 31 patients (45%) with inconclusive imaging during follow-up after surgical resection of primary CRC. In all patients treatment changed from local therapy to systemic treatment for disseminated disease. Scott et al. [171] described 93 patients with residual structural after surgery for primary CRC suggestive of recurrence. Treatment changed in 66% of these patients.

One study [172] examined the influence of $^{18}$F-FDG PET during routine follow-up after a history of CRC when there was no sign of recurrence as physical examination, CT, MRI and CEA were all normal. $^{18}$F-FDG PET changed individual management in only two of 31 cases (6%). The first was an omental metastasis and the second was a false-positive PET causing unnecessary laparotomy. Therefore, follow-up of CRC by $^{18}$F-FDG PET without any signs pointing to recurrence seems to be of limited value.
Unexplained rise in CEA. When serum CEA levels rise during postoperative surveillance in asymptomatic patients and history taking, physical examination and imaging do not lead to a distinct cause, treatment decisions are difficult to be made. With the aid of $^{18}$F-FDG PET, localisation of the source of increased serum CEA levels may lead to a change in management in the majority of patients.

$^{18}$F-FDG PET in patients with a history of CRC and CEA rise with normal ($n = 31$) or equivocal ($n = 19$) findings on conventional work-up (including abdominal CT and chest X-ray or CT) detected recurrent disease in 68% of patients in the study of Flamen et al. [173], thereby changing management from observation to start of a new treatment (curative surgery for resectable disease or finding of non-resectable disease). Of the 56 lesions, 20% were local recurrences, 27% liver metastases, 9% lung metastases, 36% other abdominal lesions and 9% were non-pulmonary extra-abdominal lesions. In the subgroup of eight patients included in an earlier study by Flamen et al. [174] with a rising CEA level, but negative findings on morphologic imaging, PET led to change in management in three (37.5%), due to detection of one local recurrence, one liver metastasis and one lymph node metastasis. Valk et al. [175] described 18 patients with a rise of CEA without abnormal findings on abdominal CT-scanning. Of these, 12 patients had detectable disease by $^{18}$F-FDG PET (67%), which was subsequently histologically confirmed to be recurrent CRC. Simo et al. [170] described a subset of 58 patients with rise of CEA with normal findings on conventional imaging. With $^{18}$F-FDG PET, they found the cause in 34 patients (59%), resulting in change of management. Of these, 18 could be treated with curative surgery, whereas the remaining 16 were treated with systemic therapy. Even-Sapir et al. [176] mentioned 16 cases with occult rising of CEA in which 13 (81%) tumour recurrences were detected. Of these patients, nine were treated with chemotherapy and four with surgery. Only one patient had a negative PET despite an intraluminal recurrence at repeat colonoscopy. Shen et al. [177] reported that PET had influence on individual management in 41 of 50 patients (82%) with suspicion of recurrent CRC based on asymptptomatically elevated serum levels of CEA.

Restaging for resectable local recurrence. When local recurrence is confirmed, resectability of disease is assessed by restaging the patient. $^{18}$F-FDG PET can be useful by detecting distant disease, which would make surgery futile.

In case of presumed resectable pelvic recurrence of rectal carcinoma, Faneyte et al. [178] found management changes in 14% of 32 cases due to discrepant PET findings after conventional imaging. These five cases caused cancellation of surgery for extensive disease in one, less extensive surgery in three and surgery instead of palliative therapy in one.

Flamen et al. [174] described a subset of 23 patients with recurrent locoregional CRC which was presumed resectable based on clinical and radiological findings. In eight of these patients (35%) management was altered due to unexpected findings of additional tumour sites in five and the exclusion of disease in three patients. Valk et al. [175] described a subgroup of 78 patients with a history of CRC with local recurrence considered resectable based on conventional diagnostic work-up. $^{18}$F-FDG PET showed additional lesions in 23 patients (29%) rendering these recurrences unresectable. In contrast, PET did not show any signs of recurrence in six patients,
of whom two showed malignant local lesions during laparotomy. Kalff et al. [179] asked the attending clinicians to assign a treatment plan to 102 consecutive patients with recurrent local CRC presumed being resectable based on conventional imaging. This treatment plan was then compared with that based on incremental information supplied by $^{18}$F-FDG PET. In 54 cases treatment plan was altered due to unexpected PET-findings and in six more cases referring oncologists would not commit to a management plan without access to PET-information (59%). Of all these cases, one false-positive result was due to a pelvic abscess and in four the extent of metastatic disease was underestimated by PET.

Restaging for resectable liver metastases. $^{18}$F-FDG PET can be used to restage disease in case of presumed resectable liver metastases to confirm resectability in these patients and to avoid futile liver surgery.

Wiering et al. [202] performed a systematic review and a meta-analysis of 32 studies published up to 2003 concerning patients selected for surgical treatment for liver metastases. They found pooled sensitivity and specificity for $^{18}$F-FDG PET to be 88% and 96.1% for hepatic lesions and 91.5% and 95.4% for extra-hepatic lesions. Pooling results of CT-scanning resulted in 82.7%, 84.1% (hepatic lesions), 60.9% and 91.1% (extra-hepatic lesions), respectively, underlining the higher sensitivity of $^{18}$F-FDG PET for extra-hepatic lesions as compared to CT. Detection of extra-hepatic lesions may lead to management changes such as a different surgical approach or cancellation of surgery for extensive disease and starting of palliative chemo(radio)therapy. They noted that only 18 of 32 studies mentioned the change in patient management due to $^{18}$F-FDG PET findings, the pooled value being 32% (range: 20-58%).

Our search query resulted in 25 papers [169,170,174,180–201] in which $^{18}$F-FDG PET was used in restaging patients prior to surgery for liver metastases (figure 2.3, page 38). Management changes were reported in 11% [186] to 70% [170]. The authors of the paper with lowest management change (11%) [186] noted that in their population in only 5.5% of patients intra-abdominal unexpected extrahepatic metastases were present, a number which is exceptionally low. They attributed this to improvement of accuracy of conventional diagnostic imaging. The high percentage of management changes noted by Denecke et al. [198] (52%) was possibly due to the fact that they included patients with recurrence after LASER induced thermotherapy of unresectable liver metastases. A high number of unexpected extrahepatic lesions was found. In addition, when calculating management changes, two of eleven cases were included, in which false-positive $^{18}$F-FDG PET results led to inadequate conclusions and unnecessary interventions. For Simo et al. [170] the high proportion of management changes (seven of ten patients who were restaged before liver surgery) might be due to inclusion of three cases in which changes were limited to surgical planning (use of radiofrequency ablation (RFA) or resection of additional lesions).

The pooled mean management change in these 1,060 patients is 22.3% (95% − CI: 19.8-24.9%). An example of how $^{18}$F-FDG PET can influence management in liver metastases is displayed in figure 2.4 (page 40). The results of Scott et al. [171] (management changes in 49% of 98 patients) were not used for calculation of pooled
2. Tailoring CRC Therapy by PET/CT

Figure 2.3: Forest plot of management changes with corresponding confidence intervals for preoperative restaging in colorectal liver metastases by $^{18}$F-FDG PET described in 25 references [169,170,173,180,201]. The size of the squares denotes the weight for calculation of the pooled average. $^{18}$F-FDG PET: $^{18}$F-fluoro-2-deoxy-D-glucose positron emission tomography.
management change, since they included both patients with potentially resectable hepatic and pulmonary CRC metastases. It was not possible to derive how many of these had liver lesions only.

The consequence of restaging these patients by $^{18}$F-FDG PET prior to liver surgery is described in two cohorts of in total 203 patients who were selected for hepatic surgery [203]. Patients staged by $^{18}$F-FDG PET (group A, $n = 100$) were compared to patients staged by CT alone (group B, $n = 103$). Although futile laparotomy ratios were similar for both groups (19.4% vs 28%; $p = 0.186$), significantly less extrahepatic disease was seen during surgery in the cohort of patients staged by $^{18}$F-FDG PET (1.9% vs 10%; $p = 0.017$). Most of these futile laparotomies were due to too extensive hepatic disease, but no difference between both cohorts were seen (17.4% vs 17.0%; $p = 1$). Pawlik et al. [204] on the other hand did show in a retrospective analysis of 461 patients surgically treated for liver metastases in the same period, that the rate of unnecessary laparotomies was significantly lower in patients restaged by $^{18}$F-FDG PET compared to patients who did not have $^{18}$F-FDG PET (5.6% vs 12.4%; $p = 0.009$).

Fernandez et al. [205] described improved overall survival in patient selected for surgery for liver metastases by $^{18}$F-FDG PET. In their study, the outcome of 100 patients selected for resection of hepatic metastases by $^{18}$F-FDG PET had better five-year overall survival rates than in 19 reviewed similar studies (including 6,066 patients) not using functional imaging. Five-year overall survival in this study was 58.6% (95% − CI: 45.6-71.6%), which was higher than in the 19 other studies (30%; range: 12-41%). Strasberg et al. [185] noted that in their series of 35 patients restaged by PET before liver surgery, three-year overall survival was 77% and this proportion was higher than any of the 13 similar articles they reviewed that used conventional restaging (range: 30-58%). However, care must be taken to compare results with historic data, since the improvements in CT scans has led to stage migration and thus survival benefit [206]. Wiering et al. [203] found no differences in both overall survival (three-year: 57.1% vs 60.1%; $p = 0.678$) and disease-free survival (three-year: 29.9% vs 29.2%, $p = 0.656$) between the group restaged by $^{18}$F-FDG PET and the group without $^{18}$F-FDG PET. They explain this discrepancy by stating that they used well-matched control group in contrast to the others who used a historical control group. The effect of $^{18}$F-FDG PET on overall survival seemed lower than reported in these other studies. They concluded that tumour biology, resectability and chemotherapy response seem to be the major determinants of survival and that their results suggest that the intraoperative surgical approach to disease control and postoperative care in both groups were similar.

It should be noted that previous chemotherapy lowers sensitivity of $^{18}$F-FDG PET when restaging patients before liver surgery for liver metastases. Akhurst et al. [207] stated that sensitivity of $^{18}$F-FDG PET in the detection of colorectal metastases during preoperative staging was decreased in patients pre-treated by neoadjuvant chemotherapy due to downregulation of hexokinase activity (lesion detection sensitivity: 63% vs 77%). No lesions larger than 1.2 cm were missed in the untreated group, but lesions up to 3.2 cm were missed after neoadjuvant chemotherapy. Interpretation of $^{18}$F-FDG PET data should be done with caution in the context of concomitant chemotherapy. In this respect both the specific cytostatic agent(s) prescribed as the timing of PET scanning after neoadjuvant treatment are of relevance.
Figure 2.4: Example of an $^{18}$F-FDG PET/CT in a male patient with a pT$_3$N$_0$M$_1$ rectosigmoid carcinoma treated by rectosigmoid resection in combination with chemotherapy. After an initially good response, resection of liver metastasis was performed. During follow-up with ultrasonography, CT and $^{18}$F-FDG PET/CT a recurrence in the liver (arrowhead) was detected by PET/CT which was not detected by ultrasonography or CT alone. CT: contrast enhanced computed tomography; $^{18}$F-FDG PET/CT: 2-$(^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography / computed tomography.

Overall effect on treatment decisions in suspected recurrence. The remaining papers papers [171,176,208–215] dealt with heterogeneous populations of patients in whom during follow-up of CRC any recurrence or metastasis was suspected based on clinical findings, CEA increase or conventional diagnostic imaging (tables 2.2-2.4 pages 41-42). In these studies the detection ratio of local recurrence (sensitivity) by $^{18}$F-FDG PET varied from 90% to 100%, which is higher compared to CT (71-88%). Specificity of PET and CT in these studies was similar (86-92% for PET vs 85-89% for CT). For hepatic metastases sensitivity and specificity for PET vs CT were 89-100% vs 45-100% and 91-100% vs 60-100%, respectively. For extra-hepatic metastases
sensitivity and specificity were 94-100% vs 64-74% and 40-100% vs 50-96% for PET vs CT, respectively. Management changes in these ten papers varied from 6-30%.

Huebner et al. performed a meta-analysis of 11 similar articles up to 1999 and found a pooled management change in 29% (95% CI: 25-34%).

The consequence of PET-tailored management in follow-up of patients after curative resection of colon or rectal carcinoma was investigated by Sobhani et al. They stratified and randomised 130 patients in a group with a standardised follow-up consisting of history taking, physical examination, biomarker assays and conventional imaging (ultrasonography, thorax X-ray, abdominal CT) and a group in which this follow-up included a whole-body 18F-FDG PET after nine and 15 months. They found the time to recurrence-detection was significantly shorter in the 18F-FDG PET arm (12.1 vs 15.4 months; p = 0.01) associated with more curative resections of recurrences (65% vs 9.5%; p < 0.01).

The added value of fusing 18F-FDG PET and CT was assessed by Fukunaga et al., who compared fused PET/CT with separate PET and CT in patients with suspected local recurrence after curative resection of rectal carcinoma. They reported improved accuracy of diagnosis of fused PET/CT over PET or CT alone of 93%, 79% (p = 0.0138) and 88% (p = 0.0156), respectively. Nakamoto et al. investigated 63 patients with suspected recurrent CRC, but failed to show a significant improvement of diagnostic accuracy of 18F-FDG PET/CT (CT alone: 78%, 18F-FDG PET alone: 79%, 18F-FDG PET and CT: 84% and fused 18F-FDG PET/CT: 92%; p = 0.13). Even-Sapir et al. showed in 62 patients with suspected recurrence or metastases after rectal carcinoma and preoperative staging for rectal carcinoma, that the specificity of fused PET/CT is higher than PET alone (89% vs 74%; p < 0.05) with similar sensitivity (96% vs 88%).

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>n</th>
<th>Design</th>
<th>Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delbeke et al.</td>
<td>1997</td>
<td>52</td>
<td>Retrospective</td>
<td>Suspected recurrence in CRC follow-up</td>
</tr>
<tr>
<td>Ruhlmann et al.</td>
<td>1997</td>
<td>59</td>
<td>Retrospective</td>
<td>Suspected primary, screening follow-up, suspected recurrence in CRC</td>
</tr>
<tr>
<td>Imdahl et al.</td>
<td>2000</td>
<td>71</td>
<td>Prospective</td>
<td>Suspected recurrence or metastases in CRC follow-up</td>
</tr>
<tr>
<td>Whiteford et al.</td>
<td>2000</td>
<td>105</td>
<td>Prospective</td>
<td>Suspected recurrence or metastases</td>
</tr>
<tr>
<td>Arulampalam et al.</td>
<td>2001</td>
<td>42</td>
<td>Prospective</td>
<td>Suspected or confirmed recurrence</td>
</tr>
<tr>
<td>Deai et al.</td>
<td>2003</td>
<td>42</td>
<td>Prospective</td>
<td>Follow-up or preoperative staging with resectable disease on CT</td>
</tr>
<tr>
<td>Even-Sapir et al.</td>
<td>2004</td>
<td>62</td>
<td>Retrospective</td>
<td>Suspected recurrence/metastases or preoperative staging in rectal carcinoma</td>
</tr>
<tr>
<td>Nakamoto et al.</td>
<td>2007</td>
<td>63</td>
<td>Prospective</td>
<td>Suspected recurrence or screening follow-up PET/CT</td>
</tr>
<tr>
<td>Akiyoshi et al.</td>
<td>2009</td>
<td>63</td>
<td>Prospective</td>
<td>Suspected metastases or pre-chemoradiotherapy staging</td>
</tr>
</tbody>
</table>

Table 2.2: Management changes and test characteristics in recurrent CRC by 18F-FDG PET. CDW: conventional diagnostic work-up; CRC: colorectal carcinoma; CT: computed tomography; 18F-FDG: 2-(18F)fluoro-2-deoxy-D-glucose; N/A: not applicable; PET: positron emission tomography; Se: sensitivity, Sp: specificity. *Including one patient in whom an incidental second tumour type was found. †Test characteristics for extrahepatic metastases are based on detection of recurrent and metastatic CRC. ‡PET/CT compared to PET alone. §Only locoregional lymphnode metastases were included.
### Table 2.3: Management changes and test characteristics in recurrent CRC by $^{18}$F-FDG PET (continued). For abbreviations see table 2.2 (page 41)

<table>
<thead>
<tr>
<th>Author</th>
<th>Disease free / benign</th>
<th>Observation to start new treatment</th>
<th>Cancel treatment for no disease</th>
<th>Cancel surgery for extensive disease</th>
<th>Cancel other treatment for extensive disease</th>
<th>Modifications to treatment plan</th>
<th>Inconclusive CDW</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delbeke et al. [208]</td>
<td>6 / 52 (histology or follow-up &gt; 12 months)</td>
<td>2 2 9 4 17 (33%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruhlmann et al. [209]</td>
<td>14 / 59 (histology or suggested by CDW)</td>
<td>2 2 9 4 17 (10%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imdahl et al. [210]</td>
<td>20 / 71 (histology or suggested by CDW)</td>
<td>7 9 16 (23%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whiteford et al. [211]</td>
<td>22 / 105 (histology or follow-up &gt; 6 months)</td>
<td>4 8 14 4 30 (29%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arulampalam et al. [212]</td>
<td>12 / 42 (histology or follow-up)</td>
<td>9 3 2* 2 16 (38%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desai et al. [213]</td>
<td>0 / 42 (histology or follow-up)</td>
<td>17 17 (40%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Even-Sapir et al. [176]</td>
<td>19 / 62 (histology or follow-up &gt; 6 months)</td>
<td>13 8 5 3 29 (47%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakamoto et al. [214]</td>
<td>27 / 63 (histology or follow-up &gt; 6 months)</td>
<td>20 3 1 1 25 (40%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akiyoshi et al. [215]</td>
<td>0 / 63 (histology)</td>
<td>4 6 10 (16%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2.4: Management changes and test characteristics in recurrent CRC by $^{18}$F-FDG PET (continued). For abbreviations see table 2.2 (page 41)

It can be concluded that $^{18}$F-FDG PET results in modification of individual patient management in situations where conventional diagnostic work-up shows equivocal findings in CRC patients (45-66%), in patients with unexplained rise in CEA (37.5-82%), in preoperative restaging of resectable local recurrence (29-59%) and assessment
of patients before surgery for liver metastases (11-70%). Interpretation of $^{18}$F-FDG PET images for detection of metastases should be carried out with caution during or soon after administration of chemotherapy, since the sensitivity for detection of metastases is lower than normal. Use of $^{18}$F-FDG PET in patients with a history of CRC without clinical, biochemical or radiological signs of recurrence, appears of limited additional value.

The value of metabolic imaging next to morphologic imaging seems to have high sensitivity for local disease with similar specificity as CT. Especially in detection of extrahepatic metastases, the application of $^{18}$F-FDG PET is superior to CT alone. In a population with suspected recurrence, management will change in about 6-30% due to $^{18}$F-FDG PET findings leading to earlier detection of recurrences and to more curative resections. Combined PET/CT is superior to PET alone in recurrence detection.

**Prognostic stratification and therapy response prediction by PET**

**Determination of prognosis by PET**

The use of (semi-)quantitative parameters for tracer uptake before start of treatment, such as the mean standardised uptake value ($S_UV_{mean}$), maximum standardised uptake value ($S_UV_{max}$) or parameters of $^{18}$F-FDG metabolism in tumour lesions, can be related to overall patient outcome (prognosis). This might help in selecting the appropriate treatment for an individual patient (table 2.5, page 44).

Calvo et al. [219] performed a study in primary rectal carcinoma treated by neoadjuvant CRT (45-50.4 Gy combined with 5-FU/FA or tegafur) followed by surgical resection showed that the three-year overall survival ratio in patients with a (arbitrarily chosen) $S_UV_{max}$ of 6 or lower on baseline $^{18}$F-FDG PET was significantly higher than for higher values for the $S_UV_{max}$ (92% vs 60%; $p = 0.04$) (table 2.5, page 44). The above-mentioned paper of Scott et al. [171] showed the prognostic potential of $^{18}$F-FDG PET in a subgroup of 93 patients with residual structural lesions during follow-up of CRC after primary surgery. Significantly better one-year progression-free survival was found when no additional lesions were detected by PET as compared to patients in whom PET showed additional lesions (60.5% vs 36.2%; $p = 0.04$).

Most studies which used baseline PET to predict patient outcome used individuals with metastasised disease treated either with surgery or chemotherapy. Dimitrakopoulou-Strauss et al. [220] published a paper on patients with metastatic CRC treated with second line folinic acid/5-FU/oxaliplatin (FOLFOX). They used $^{18}$F-FDG $S_UV_{mean}$, fractal dimensions and pharmacokinetic rate constants combined in a discriminant analysis in 25 patients. $S_UV_{mean}$ correctly classified one-year overall survival in 67% and the pharmacokinetic parameters in 76%. Unfortunately, their discriminant functions with coefficients were not provided, which makes implementation of their model in different subsets of patients difficult. Our group used the $^{18}$F-FDG $S_UV_{mean}$ in 152 patients with CRC metastases treated by resection or pyrimidine-based chemotherapy. The 76 patients with a $S_UV_{mean}$ lower than 4.26 had a longer median overall survival than the rest of the patients (32 months vs 19 months; $p = 0.017$) [16]. Riedl et al. [221] performed a similar experiment in surgically treated liver metastases and found overall survival benefit in subgroups of
Table 2.5: Prognostic stratification by baseline $^{18}$F-FDG PET. 5-FU: 5-fluorouracil; CCR: correct classification rate; CDW: conventional diagnostic work-up; CRC: colorectal carcinoma; CRLM: CRC liver metastases; CRT: chemoradiotherapy; FA: folinic acid; FD: fractal dimensions; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-d-glucose; FOLFOX: FA/5-FU/oxaliplatin; $K_1-k_4$: two-compartment rate constants; OS: overall survival; PET: positron emission tomography; PFS: progression-free survival; sign.: statistical significance; SUV: standardised uptake value; $V_B$: blood volume fraction.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>n</th>
<th>Inclusion → Therapy</th>
<th>PET- &amp; outcome parameters</th>
<th>Favourable criteria</th>
<th>Results (sign.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvo et al. (2004)</td>
<td>25</td>
<td>cT2-4Nx primary rectal carcinoma → CRT (45-50.4 Gy, 5-FU/FA or tegafur) + resection</td>
<td>$SUV_{\text{max}}$, three-year OS</td>
<td>$SUV_{\text{max}} \leq 6$</td>
<td>92% vs 60% (p = 0.04)</td>
</tr>
<tr>
<td>Dimitrakopoulou Strauss et al. (2004)</td>
<td>20</td>
<td>CRC metastases → 2nd line FOLFOX</td>
<td>$SUV_{\text{mean}}$, one-year OS</td>
<td>$SUV_{\text{mean}}$</td>
<td>CCR = 67%</td>
</tr>
<tr>
<td>de Geus-Oei et al. (2006)</td>
<td>152</td>
<td>CRC metastases → resection of pyrimidine based chemotherapy</td>
<td>$SUV_{\text{mean}}$, median OS</td>
<td>$SUV_{\text{mean}} \leq 4.26$</td>
<td>32 months vs 19 months (p = 0.017)</td>
</tr>
<tr>
<td>Riedl et al. (2007)</td>
<td>90</td>
<td>CRLM → resection</td>
<td>$SUV_{\text{max}}$, median OS</td>
<td>$SUV_{\text{max}} &lt; 5$</td>
<td>&gt;72 months vs 48 months (p = 0.014)</td>
</tr>
<tr>
<td>Scott et al. (2008)</td>
<td>91</td>
<td>Suspected recurrence CRC based on CDW → various</td>
<td>Additional sites detected, one-year PFS</td>
<td>No additional lesions</td>
<td>60.5% vs 36.2% (p = 0.04)</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>Resectable liver or pulmonary metastases → various</td>
<td>Additional sites detected, one-year PFS</td>
<td>No additional lesions</td>
<td>65.9% vs 39.2% (p = 0.01)</td>
</tr>
</tbody>
</table>

patients with lowest $SUV_{\text{max}}$ for a range of cut-off values (table 2.5, page 44). Scott et al. [171] showed prognostic ability of $^{18}$F-FDG PET in another subgroup of 98 patients restaged before resection of presumable resectable hepatic and pulmonary metastases. In this group, significant better one-year progression-free survival was found when no additional lesions were detected by PET compared to patients in whom PET did show additional lesions (65.9% vs 39.2%; p = 0.01).

**Prediction of therapy response by PET**

The imaging of glucose uptake of CRC lesions before start of treatment might indicate which patients are more likely to respond to therapy. For patients that are less likely to respond to the opted treatment, a different therapeutic approach might be beneficial. For this purpose, baseline (semi-)quantitative parameters of tracer uptake, such as the $SUV_{\text{mean}}$ are related to individual patient outcome (table 2.6, page 45).

For primary rectal carcinoma treated with 50 Gy of neoadjuvant radiotherapy prior to surgery, Oku et al. [222] found a negative correlation between the shrinkage rate on CT and the baseline $SUV_{\text{mean}}$ measured by $^{18}$F-FDG PET of the lesion (i.e., the
2.3. Results

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>n</th>
<th>Inclusion → Therapy</th>
<th>PET- &amp; outcome parameters</th>
<th>Favourable criteria</th>
<th>Results (sign.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oku et al. (2002) [222]</td>
<td>40</td>
<td>cT2,4 and/or N1,3 primary rectal carcinoma → radiotherapy (50 Gy)</td>
<td>$SUV_{mean}$, CT shrinkage rate at 3-5 weeks</td>
<td></td>
<td>correlation: $-0.162$ ($p = 0.326$)</td>
</tr>
<tr>
<td>Dimitrakopoulou-Strauss et al. (2003) [223]</td>
<td>28</td>
<td>CRC metastases → 2nd line FOLFOX</td>
<td>$SUV_{mean}$, clinical response (WHO-guidelines): PDP, SD or PR</td>
<td>$SUV_{mean}$</td>
<td>CCR: 96% (PD), 47% (SD) and 0% (PR)</td>
</tr>
</tbody>
</table>

Table 2.6: Therapy response prediction by baseline $^{18}$F-FDG PET. 5-FU: 5-fluorouracil; CCR: correct classification rate; CRC: colorectal carcinoma; FA: folinic acid; CT: computed tomography; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-D-glucose; FOLFOX: FA/5-FU/oxaliplatin; PD: progressive disease; PET: positron emission tomography; PR: partial remission; SD: stable disease; sign.: statistical significance; SUV: standardised uptake value; WHO: world health organisation.

larger the $SUV_{mean}$ prior to treatment the larger the treatment induced reduction in lesion size). However, this correlation was very weak and not significant (correlation coefficient: $-0.162; p = 0.326$). They found that only follow-up $SUV_{mean}$ correlated with morphological changes and reasoned that a high $SUV_{mean}$ at follow-up indicated both a high $SUV_{mean}$ at baseline and a low reduction of uptake during treatment.

Dimitrakopoulou-Strauss et al. [223] published a paper on patients with metastatic CRC treated with second line FOLFOX. They showed by discriminant analysis that the pretreatment $SUV_{mean}$ correctly identified nonresponders (96% of 28 patients with progressive disease). The same limitations to implementation of their model apply as described in the previous paragraph.

Functional imaging of lesions before start of treatment can identify patients with poor prognosis or who are less likely to respond to treatment. These patients might benefit from treatment-modification, if alternatives exist. Using a baseline $^{18}$F-FDG PET, it appears feasible to stratify patients with different prognosis and possible resistance to treatment in CRC. So far, no prospective randomised controlled trials have been published using baseline $^{18}$F-FDG PET for determination of individual treatment strategy.

Therapy follow-up by PET-response evaluation

After localised or during systemic treatment, tracer uptake in lesions can be monitored. Uptake during follow-up can be compared to baseline uptake, can be used to assess the effect of therapy and might be used as (early) prediction of therapy response or be related to patient survival for prognostic purposes. Thereby it might contribute to early change in management, if alternative treatment options are available. Therapy response evaluation is most often performed by morphologic imaging according to RECIST (response evaluation criteria in solid tumours) [224]. In an era where new cytotoxic treatment is cytostatic rather than cytoreductive, metabolic changes may precede anatomical changes, PET-imaging might aid in early therapy response evaluation.
Response evaluation of radiotherapy in rectal carcinoma

Engenhart et al. [225] were the first to address the effect of irradiation on inoperable pre-sacral recurrent rectal carcinoma in 21 patients. They noticed a small but significant decrease in $^{18}$F-FDG uptake during radiotherapy (2.3 to 1.6 uptake-units after six months; $p = 0.002$). Oku et al. [222] described a significant negative ($R = -0.383; p = 0.014$) correlation between shrinkage rate on CT (fractional decrease in tumour axial diameter) and $^{18}$F-FDG SUV ratio (fraction remaining SUV of follow-up compared to baseline) after 50 Gy of radiotherapy to primary rectal carcinoma in 40 patients.

Two studies compare TNM tumour (T-)stage, relating baseline TREUS T-stage with follow-up histological T-stage. In 25 patients with local invasive primary rectal carcinoma treated with neoadjuvant CRT prior to surgery, Calvo et al. [219] described a significant difference in $SUV_{max}$-reduction ($\Delta SUV_{max}$) between patients with reduction in T-stage compared to nonresponders (-3.3 vs -1.9 $SUV$-units; $p = 0.03$). Denecke et al. [226] showed that a cut-off for $\Delta SUV_{max}$ of -36% was able to separate responders from nonresponders: 76% of $^{18}$F-FDG PET-responders demonstrated T-downstaging and 100% of $^{18}$F-FDG PET-nonresponders did not show T-downstaging ($p = 0.002$) in 23 patients with advanced rectal carcinoma treated with neoadjuvant CRT in combination with hyperthermia. Results of PET were superior to CT or MRI in therapy response prediction.

Many studies compared visual $^{18}$F-FDG PET response [231,233,236], $\Delta SUV_{max}$ [228–231,234,235,238–240], $\Delta SUV_{mean}$ [232,241], SUV-ratio [237] and $\delta TLG$ (change in total lesion glycolysis: the product of metabolic volume and SUV) at different intervals after radiotherapy, varying from 12 days [232] up to seven weeks [231,236] and all found a significant relation with semi-quantitative histological response (table 2.7, page 48). Depending on response criteria, predictive values of $^{18}$F-FDG PET response (NPV) ranged from 83% [241] to 100% [235] and predictive values of $^{18}$F-FDG PET nonresponse (PPV) varied from 77% [234] to 100% [228,232]. The worst results were found by Engenhart et al. [225] and Melton et al. [235] who used rigorous criteria for definition of therapy response (complete $SUV$ normalisation [225] and $\Delta SUV_{max}$ $\geq$ 70% determined by ROC analysis [235]) and found a PPV of 20% and 58% respectively, which means 42-80% of the patients without response on $^{18}$F-FDG PET, clinically did show local control [225] or regression score during histopathological examination [235]. Kristiansen et al. [236] on the other hand used a very strict criterion for pathological response (defined as no histological detectable residual carcinoma) and, therefore, 40% false $^{18}$F-FDG PET therapy responses were found (accuracy 53%). This confirmed the data of Guillem et al. [227], who found an accuracy of 60% for PET to define the extent of pathological response. Rosenberg et al. [241] attributed their low PPV of $^{18}$F-FDG PET (PPV = 64%) to be due to influx of inflammatory cells.

The prognostic value of the metabolic response of rectal carcinoma to neoadjuvant CRT has been described in a few studies [229,233]. Guillem et al. [229] dichotomised $^{18}$F-FDG PET results in responders and nonresponders in 15 patients with locally advanced primary rectal carcinoma treated by preoperative CRT. Responders were defined as those with only focal or diffuse $^{18}$F-FDG uptake with a maximum $SUV_{max}$ of 1.4 on the follow-up PET-scan. They showed higher median overall survival (> 54
weeks vs 39 weeks; \( p = 0.08 \) and higher median disease-free survival (> 54 weeks vs 26 weeks; \( p = 0.02 \)) of responders compared to nonresponders. The corresponding five-year overall survival percentages were 91% vs 70% (\( p = 0.024 \)) and five-year disease-free survival percentages were 81% vs 62% (\( p = 0.003 \)). In another study, in 34 patients treated with neoadjuvant CRT before curative surgery, visual PET response was categorised into complete remission, partial remission and stable or progressive disease before results of pathology were available. This PET-based response stratification showed three-year overall survival percentages of 100%, 79% and 0% respectively (\( p < 0.0001 \)) and three-year disease-free survival percentages of 100%, 47% and 0% respectively (\( p < 0.0001 \)) [233].

A drawback of postradiotherapy 18F-FDG PET is the radiation-induced inflammation [25]. This causes influx and activation of macrophages, neutrophils, fibroblasts and granulation tissue, that can accumulate approximately 25% of 18F-FDG uptake [242]. On the other hand, direct effect of radiation may induce tumour cell dormancy (“stunning”) which mimics glucose metabolic response [228]. Siegel et al. [238] saw metabolic response (\( \Delta \text{SUV}_{\text{max}} = -39.3\% \)) after short course radiotherapy (25 Gy in five days) at day nine, but could not correlate these results to histopathology (negligible morphological response) which might partly be explained by the latter effect or due to the fact that surgery is performed immediately after radiotherapy in contrast to CRT where surgery is postponed six weeks.

18F-FDG PET allows prediction of pathological response in patients treated by neoadjuvant CRT. Moreover it seems able to predict overall and disease-free survival after surgery with curative intent. However, the clinical consequences remain unclear. It does not seem possible to select patients who have no advantage of surgery after neoadjuvant treatment. It seems possible to demonstrate functional response after short-course radiotherapy [238], but there are no data that shortening of neoadjuvant treatment prior to surgery (i.e., when no histopathological data is available) might benefit the subgroup of patients with early PET-response to CRT or that escalation of chemoradiation (e.g., increasing the radiation dose, adding regional hyperthermia, decision for intraoperative radiation) might benefit the subgroup of patients with no PET-response. Discrimination of responders to neoadjuvant chemoradiation from nonresponders could also be used for preoperative selection for individualised surgery. This could include sphincter-saving surgery in deep-seated tumours, less aggressive treatment in limited disease or planning of intraoperative radiation therapy. However, no studies provide definitive evidence. Apart from selection for individualised surgery, 18F-FDG PET-response might theoretically help to decide which patient benefits from adjuvant chemotherapy after surgery, but no studies concerning this have been published either, nor compared this to the use of histopathological data obtained during surgery.

**Effect of local ablative therapy on liver metastases**

Qualitative assessment of 18F-FDG PET imaging after cryosurgical or RFA of unresectable liver metastases has advantages over conventional imaging such as CT, MRI and ultrasonography since latter techniques do not easily identify treatment failures at an early stage. This is due to hyperechogenicity and rim-like contrast enhancement caused by regional hyperperfusion, resembling residual tumour [243]. The NPV
## 2. Tailoring CRC Therapy by PET/CT

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>n</th>
<th>Inclusion → Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engenhart et al. (1992)</td>
<td>21</td>
<td>Unresectable recurrence → 40 Gy photons, 14 Gy neutrons</td>
</tr>
<tr>
<td>Guillem et al. (2000)</td>
<td>15</td>
<td>cT3 and/or N1 → neoadjuvant CRT (50.4 Gy, 5-FU/FA)</td>
</tr>
<tr>
<td>Oku et al. (2002)</td>
<td>40</td>
<td>cT2-4 and/or N1-3 → neoadjuvant RT (50 Gy)</td>
</tr>
<tr>
<td>Amthauer et al. (2004)</td>
<td>20</td>
<td>cT3,NxM0 → neoadjuvant CRT (45 Gy, 5-FU/FA) + RH</td>
</tr>
<tr>
<td>Calvo et al. (2004)</td>
<td>25</td>
<td>cT2-4,Nx → neoadjuvant CRT (45-50.4 Gy, 5-FU/FA or tegafur)</td>
</tr>
<tr>
<td>Guillem et al. (2004)</td>
<td>15</td>
<td>cT3 and/or N1 → neoadjuvant CRT (50.4 Gy, 5-FU/FA)</td>
</tr>
<tr>
<td>Denecke et al. (2005)</td>
<td>23</td>
<td>cT3,NxM0 → neoadjuvant CRT (45 Gy, 5-FU/FA) + RH</td>
</tr>
<tr>
<td>Koniski et al. (2005)</td>
<td>20</td>
<td>uT3-4 and/or N1 → neoadjuvant CRT (45-55 Gy ± fluoropyrimidine)</td>
</tr>
<tr>
<td>Capirci et al. (2006)</td>
<td>88</td>
<td>cT3-4 and/or N1-3M0 → neoadjuvant CRT (50-56 Gy, 5-FU)</td>
</tr>
<tr>
<td>Cascini et al. (2006)</td>
<td>33</td>
<td>cT3-4 and/or N1 → neoadjuvant CRT (45 Gy, FOLFOX + raltitrexed)</td>
</tr>
<tr>
<td>Kalff et al. (2006)</td>
<td>34</td>
<td>cT3,NxM0 → neoadjuvant CRT (50.4 Gy, 5-FU/FA ± oxaliplatin/carboplatin)</td>
</tr>
<tr>
<td>Capirci et al. (2007)</td>
<td>45</td>
<td>cT3-4 → neoadjuvant CRT (50-56 Gy ± fluoropyrimidine)</td>
</tr>
<tr>
<td>Melton et al. (2007)</td>
<td>21</td>
<td>cT3-4 and/or N1 → neoadjuvant CRT (50.4 Gy ± fluoropyrimidine or FOLFOX)</td>
</tr>
<tr>
<td>Kristiansen et al. (2008)</td>
<td>30</td>
<td>cT3-4 → neoadjuvant CRT (60 Gy, uracil, tegafur, FA)</td>
</tr>
<tr>
<td>Nakagawa et al. (2008)</td>
<td>29</td>
<td>uT3-4 → neoadjuvant RT (50 Gy)</td>
</tr>
<tr>
<td>Siegel et al. (2008)</td>
<td>32</td>
<td>uT2Nx or uT3Nx → neoadjuvant short course CRT (25 Gy, 5-FU)</td>
</tr>
<tr>
<td>Vliegen et al. (2008)</td>
<td>20</td>
<td>cT3-4 → neoadjuvant CRT (50.4 Gy, capecitabine)</td>
</tr>
<tr>
<td>Koniski et al. (2009)</td>
<td>53</td>
<td>cT3,4,Nx/1 → neoadjuvant CRT (50.4-54 Gy, fluoropyrimidine, mitomycin-C)</td>
</tr>
<tr>
<td>Rosenberg et al. (2009)</td>
<td>29</td>
<td>uT3,NxM0 → neoadjuvant CRT (45 Gy, 5-FU)</td>
</tr>
</tbody>
</table>

*Table 2.7: Radiotherapy and multimodality (neoadjuvant) therapy response evaluation in locally advanced rectal carcinoma by $^{18}$F-FDG PET. 5-FU: 5-fluorouracil; CRT: chemoradiation; CT: computed tomography; DFS: disease-free survival; FA: folinic acid; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-d-glucose; FOLFOX: FA/5-FU/oxaliplatin; MRI: magnetic resonance imaging; N/A: not applicable; NPV: negative predictive value (fraction of responder on PET that are clinical responders); OS: overall survival;
### 2.3. Results

<table>
<thead>
<tr>
<th>PET- &amp; outcome parameters</th>
<th>Response criteria</th>
<th>Results (sign.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta SUV_{\text{mean}} ) at 8-9 w</td>
<td>Complete response (( SUV_{\text{mean}} ) normalisation to background)</td>
<td>PPV: 3/15 (0.20)</td>
</tr>
<tr>
<td>Local control</td>
<td></td>
<td>NPV: 4/6 (0.67)</td>
</tr>
<tr>
<td>VRS, ( \Delta SUV_{\text{mean}} ), ( \delta TLG ) at 5 w</td>
<td>Any decrease</td>
<td>PPV: 15/15 (1.00)</td>
</tr>
<tr>
<td>Histological response at 4-6 w</td>
<td></td>
<td>NPV: 0/0 (N/A)</td>
</tr>
<tr>
<td>( SUV_{\text{mean}} )-ratio at 3-5 w</td>
<td>Histological response at 6-8 w</td>
<td>PPV: 6/6 (1.00)</td>
</tr>
<tr>
<td>CT shrinkage rate at 3-5 w</td>
<td></td>
<td>NPV: 13/14 (0.93) (( p &lt; 0.001 ))</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 2-4 w</td>
<td>Histological response at 4-5 w</td>
<td>Responser: -3.3</td>
</tr>
<tr>
<td>Histological response at 6-8 w</td>
<td>Histological T-stage decrease</td>
<td>Non-responders: -1.9 (( p = 0.03 ))</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ), ( \delta TLG ) at 5 w</td>
<td>Any T-stage decrease or ( \Delta SUV_{\text{max}} &lt; -36% )</td>
<td>PPV: 13/17 (0.76)</td>
</tr>
<tr>
<td>Median ( OS ) and ( DFS ) in w</td>
<td>Pathological complete response</td>
<td>NPV: 6/6 (1.00) (( p = 0.002 ))</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 3-4 w</td>
<td>( \Delta SUV_{\text{max}} )</td>
<td>( \Delta SUV_{\text{max}} = -74.1% ) vs ( -52.1% ) (( p = 0.24 ))</td>
</tr>
<tr>
<td>Histological response at 6-8 w</td>
<td>Pathological complete response</td>
<td></td>
</tr>
<tr>
<td>VRS, ( (SUV_{\text{max}}) ) at 7 w</td>
<td>( \delta TLG \geq 69.5% )</td>
<td></td>
</tr>
<tr>
<td>Five-year ( OS ) and ( DFS )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( SUV_{\text{mean}} ) at 12 days</td>
<td>( SUV_{\text{mean}} )</td>
<td>( \Delta SUV_{\text{max}} ) vs ( \Delta SUV_{\text{max}} = -63% ) vs (-22% ) (( p &lt; 0.0001 ))</td>
</tr>
<tr>
<td>Histological response at 8 w</td>
<td></td>
<td>PPV: 13/13 (1.00)</td>
</tr>
<tr>
<td>( VRS ) at 3-4 w</td>
<td>( VRS )</td>
<td>OS: 100%, 79%, 0% (( p &lt; 0.0001 ))</td>
</tr>
<tr>
<td>Three-year ( OS ) and ( DFS )</td>
<td>CR, PR or SD/PD</td>
<td>DFS: 100%, 47%, 0% (( p &lt; 0.0001 ))</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 4-5 w</td>
<td>( \Delta SUV_{\text{max}} )</td>
<td>( \Delta SUV_{\text{max}} &lt; 0% ) vs (-46.9% ) (( p &lt; 0.0001 ))</td>
</tr>
<tr>
<td>Histological response at 8-10 w</td>
<td></td>
<td>PPV: 17/22 (0.77)</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 4-6 w</td>
<td>( \Delta SUV_{\text{max}} )</td>
<td>( \Delta SUV_{\text{max}} &lt; 72% ) vs (-44% ) (( p &lt; 0.0001 ))</td>
</tr>
<tr>
<td>Histological response at 8-10 w</td>
<td></td>
<td>PPV: 7/12 (0.58)</td>
</tr>
<tr>
<td>Visual response at 7 w (PET/CT)</td>
<td>Complete histological response</td>
<td>NPV: 9/9 (1.00)</td>
</tr>
<tr>
<td>Histological response at 8 w</td>
<td>Negative PET on follow-up</td>
<td>PPV: 10/12 (0.83)</td>
</tr>
<tr>
<td>( SUV_{\text{mean}} )-ratio at 2-3 w</td>
<td>Grade 0-3 histological response</td>
<td>NPV: 6/8 (0.13)</td>
</tr>
<tr>
<td>Histological response at 2-3 w</td>
<td>( SUV_{\text{mean}} )-ratio &lt; 1</td>
<td>Significant correlation (( p = 0.047 )) Significantly longer (( p = 0.0121 ))</td>
</tr>
<tr>
<td>Survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 9 days</td>
<td>Histological down-staging</td>
<td>No correlation</td>
</tr>
<tr>
<td>Histological response at 2 w</td>
<td>( \Delta SUV_{\text{max}} )</td>
<td>( \Delta SUV_{\text{max}} \geq -40% ) (( p &lt; 0.001 ))</td>
</tr>
<tr>
<td>( TRG \leq 2 )</td>
<td></td>
<td>PPV: 7/12 (0.58)</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 14 days and at 5 w</td>
<td>Histological response at 10 w</td>
<td>NPV: 14/17 (0.82)</td>
</tr>
<tr>
<td>Histological response at 10 w</td>
<td>( \Delta SUV_{\text{mean}} ) at 14 days</td>
<td>PPV: 7/11 (0.64)</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 5 w</td>
<td>Grade 1 histological response</td>
<td>NPV: 15/18 (0.83)</td>
</tr>
<tr>
<td>( \Delta SUV_{\text{max}} ) at 5 w</td>
<td>Grade 1 histological response</td>
<td></td>
</tr>
</tbody>
</table>

PD: progressive disease; PET: positron emission tomography; PPV: positive predictive value (fraction of nonresponder on PET that are clinical nonresponders); PR: partial remission; RH: radiofrequency hyperthermia; RT: radiotherapy; SD: stable disease; sign.: statistical significance; SUV: standardised uptake value; TLG: total lesion glycolysis; TRG: tumour regression grade; T-stage: TNM-classification tumour-stage; VRS: visual response score; w: weeks.
2. Tailoring CRC Therapy by PET/CT

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>n</th>
<th>Therapy &amp; outcome parameters</th>
<th>Response outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Langenhoff et al. (2002)</td>
<td>22</td>
<td>CSA or RFA Visual interpretation within 3 weeks Histopathology, follow-up &gt; 9 months, CT</td>
<td>PET-negative NPV: 17/17 (1.00) PPV: 4/5 (0.80)</td>
</tr>
<tr>
<td>Donckier et al. (2003)</td>
<td>17</td>
<td>RFA Visual interpretation at 1 week Follow-up &gt; 3 months by CT Visual interpretation at 3 months Follow-up &gt; 3 months by CT</td>
<td>PET-negative NPV: 7/13 (0.54) PPV: 4/4 (1.00) PPV: 6/8 (0.75) PPV: 8/8 (1.00)</td>
</tr>
<tr>
<td>Joosten et al. (2005)</td>
<td>43</td>
<td>CSA or RFA Visual interpretation within 3 weeks Follow-up &gt; 3 months by CT</td>
<td>PET-negative NPV: 35/36 (0.97) PPV: 6/7 (0.87)</td>
</tr>
<tr>
<td>Denecke et al. (2007)</td>
<td>21</td>
<td>LITT Visual interpretation at suspected progression (MRI or CEA) Histopathology, follow-up &gt; 12 months by MRI SUVmax at suspected progression (MRI or CEA) Histopathology, follow-up &gt; 12 months by MRI</td>
<td>PET-negative NPV: 24/25 (0.96) PPV: 28/29 (0.97)</td>
</tr>
</tbody>
</table>

Table 2.8: Response evaluation for local ablative treatment for unresectable CRC liver metastases by 18F-FDG PET. CEA: carcinoembryonic antigen; CRC: colorectal carcinoma; CSA: cryosurgical ablation; CT: computed tomography; 18F-FDG: 2-(18F)fluoro-2-deoxy-d-glucose; LASER: light amplification by stimulated emission of radiation; LITT: LASER induced thermotherapy; MRI: magnetic resonance imaging; NPV: negative predictive value (fraction of patients with response on PET having clinical response); PET: positron emission tomography; PPV: positive predictive value (fraction of patients with no response on PET having no clinical response); RFA: radiofrequency ablation; SUV: standardised uptake value.

of 18F-FDG PET three weeks after local ablative therapy varied from 97% to 100% and of a positive PET from 80% to 86% in two studies treating 81 patients with 237 hepatic lesions. The lower PPV is caused by false-positive results due to liver abscesses occurring after local ablative therapy (table 2.8 page 50).

Donckier et al. compared the predictive value of 18F-FDG PET one week and three months after RFA of unresectable liver metastases. They found that an early negative PET (after one week) is less suitable for determining the disease-free status as compared to a late negative PET (NPV: 54% vs 75%) whereas both early and late positive PET show 100% PPV for residual tumour or recurrence. Apparently the optimal time frame to judge the effect of local ablative therapy is somewhere between one week and three months after surgery.

Denecke et al. used 18F-FDG PET after LASER induced thermotherapy ablation (LITT) of unresectable liver metastases when tumour progression was suspected on MRI or obscure rising serum levels of CEA. Standardised visual interpretation of the images based on consensus of two blinded nuclear medicine physicians led to a NPV of PET of 96% and a PPV of 97%. Using the SUVmax they found a NPV of 96% and a PPV of 93%. They concluded that 18F-FDG PET is a promising tool for assessment of local control and whole-body restaging in patients with clinical suspicion of tumour progression after local ablative treatment for liver metastases. Timing of follow-up though was highly variable in this study, since 18F-FDG PET
was performed only when tumour progression was suspected and 11% of follow-up scans were performed immediately (1-3 days post-LITT), 50% at short term (within 6 months post-LITT) and 39% after more than six months post-LITT.

$^{18}$F-FDG PET seems a promising modality to select patients with incomplete tumour ablation or with early relapse after local ablative treatment for unresectable liver metastases. Assessment of local control by $^{18}$F-FDG PET three weeks after local ablative treatment seems advisable.

**Evaluation of response to chemotherapy**

Depending on the drug and regimen, cytotoxic treatment has a clear influence on CRC adenocarcinoma cell lines. Oxaliplatin, 5-FU and irinotecan cause decreased $^{18}$F-FDG uptake after 72 h due to decrease in glucose transport, a decrease in hexokinase activity \[247\]. This effect can be evaluated by $^{18}$F-FDG PET during systemic treatment of liver metastases.

Findlay et al. \[248\] were the first to report the effect of 5-FU with or without interferon-α chemotherapy on liver metastases > 3 cm in diameter with $^{18}$F-FDG PET. They used morphological CT response as outcome measure and found that a reduction in tumour-liver ($T/L$) ratio of tissue activity concentrations of 15% or more has a sensitivity of 100% with a specificity of 75% to predict morphological response. They noted that patients with an increase in $T/L$ 1-2 weeks subsequently had a reduction in $T/L$ at 4-5 weeks, suggesting that this “flare” phenomenon is caused by infiltration of macrophages as response to tumour cell kill (i.e., early inflammatory reaction).

The two above-mentioned papers by Dimitrakopoulou-Strauss et al. \[220,223\] also evaluated therapy response by including follow-up $^{18}$F-FDG PET in 28 and 25 patients treated with second line FOLFOX. The prognostic value of the metabolic response of liver metastases to chemotherapy is described in three papers \[22,220,249\]. Dimitrakopoulou-Strauss et al. \[220\] showed prognostic aspects of dynamic $^{18}$F-FDG PET response in 25 patients treated with second line FOLFOX. Discriminant analysis based on $SUV_{mean}$ correctly classified overall survival in 69% of the patients and based on two-compartment rate constants, vascular fraction and fractal dimensions correctly classified overall survival in 78% of the patients \[220\]. As stated before they showed the accuracy of their model, but by omitting the model coefficients, it is not transferable to other patient populations. In a different study by our group \[22\] it was shown that the percentage decrease in $SUV_{mean}$ ($\Delta SUV_{mean}$) and glucose metabolic rate ($MR_{glc}$, derived from two-compartment kinetic analysis) are both able to distinguish subgroups with different median overall survival in 50 patients with stage IV CRC treated by various schedules of chemotherapy. Using a cut-off for $\Delta SUV_{mean}$ of -20% distinguished subgroups with a median survival of 25 weeks from 15 weeks ($p = 0.009$). Using a $\Delta MR_{glc}$ cut-off of -65% distinguished subgroups with a median survival of 32 weeks from 18 weeks ($p = 0.021$). These cut-offs were selected after analysis of a range of cut-off values. Small et al. \[249\] show similar results using $^{18}$F-FDG PET/CT to qualitatively evaluate therapy response of 54 patients with liver metastases by FOLFOX/FOLFIRI (folinic acid/5-FU/irinotecan) with or without bevacizumab. Univariate analysis showed the hazard ratio ($HR$) of PET/CT response (defined as absence or reduced uptake of $^{18}$F-FDG) during treatment was not signific-
2. Tailoring CRC Therapy by PET/CT

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>n</th>
<th>Inclusion → Therapy</th>
<th>PET-come &amp; outcomes</th>
<th>Response criteria</th>
<th>Results (sign.)</th>
</tr>
</thead>
</table>
| Findlay et al. (1996) | 18 | CRLM ≥ 3 cm → 5-FU±IFα | ΔT/L at 4-5 weeks | CT response: CR or PR | HR (95%) | 3.127, 95% – CI: 0.874-11.187, but was significant associated with disease-free survival (HR: 3.826; 95% – CI: 1.39-10.534). CT-response alone did predict overall survival (HR: 4.584; 95% – CI: 1.133-18.536) and progression-free survival (HR: 2.925; 95% – CI: 1.078-7.937). Results of all above-mentioned studies are displayed in table 2.9 (page 52).

Thus, the degree of metabolic response during treatment appeared to correlated with both pathological response and survival. However, it is of concern that the test-retest reproducibility of the SUVmean using a semi-automatically delineation of the lesion (50% of maximum value) is limited. When comparing the SUVmean of 28 lung carcinoma patients determined on the same system set-up on two consecutive days, Krak et al. [44] found a standard deviation of the relative differences of both SUVsmean of 11%. In 26 cancer patients in whom an 18F-FDG PET was repeated within 1-5 days, the standard deviation of the relative differences of both SUVsmean was 7% [150]. These standard deviations are larger when using different set-ups (multicentre trials). This suggests that measured tumour responses of less than ~15-20% (two standard deviations) might be regarded as within the reproducibility limits of the method used and should thus be interpreted as no actual change in SUV.

Table 2.9: Chemotherapy response evaluation by 18F-FDG PET for CRC metastases.

1. CRLM: colorectal carcinoma; CT: computed tomography; DFS: disease-free survival; FA: folinic acid; FD: fractal dimensions; 18F-FDG: 2-(18F)fluoro-2-deoxy-D-glucose; FOLFIRI: FA/5-FU/irinotecan; FOLFOX: FA/5-FU/oxalipatin; IFα: interferon-α; K1-k4: two-compartment rate constants; HR: hazard ratio; MRglc: glucose metabolic rate; OS: overall survival; PET: positron emission tomography; PD: progressive disease; PR: partial remission; Se: sensitivity; Sp: specificity; SUV: standardised uptake value; T/L: tumour-liver ratio; V_B blood volume fraction; WHO: world health organisation.

2. metastases → various schemes
3. CRLM → neoadjuvant FOLFOX / FOLFIRI ± bevacizumab
4. CRLM ≥ 3 cm → 5-FU ± IFα
Therapy decisions on PET-response seem feasible. However, before use in routine daily practice future randomised controlled trials are necessary to prove its value. Optimal cut-offs are not only dependent on patient, treatment setting and cytostatic agent, but are dependent on follow-up timing and reproducibility limits of this technique [46,61].

2.4 Conclusions

18F-FDG PET has limited added value in staging primary CRC. It has a convincing role in detection of local recurrence when conventional imaging fails to distinguish scar tissue from recurrent or residual tumour and can influence management decisions in about 38-82% of patients with unexplained rise in CEA. During pre-surgical restaging of local recurrence or metastases it may provide relevant information by detecting additional disease that renders laparotomy futile. The use of concomitant chemotherapy should be taken into account since it lowers sensitivity of PET.

Based on baseline 18F-FDG PET, patients with recurrent or metastasised CRC treated by chemotherapy or surgical resection can be stratified in different subgroups based on prognosis or predicted therapy effect. This risk assessment might be used for individualised treatment assignment. The first studies that use this PET-based stratification for survival to select patients for different treatment schedules are currently being undertaken.

The relation between 18F-FDG PET response to preoperative neoadjuvant (chemo)radiotherapy in rectal carcinoma, seems well-related to both histopathology and survival. The exact clinical significance of PET-based response evaluation for this group of patients remains to be investigated. Unfortunately, early response prediction of radiotherapy seems less feasible since the combined effects of stunning, proliferation and inflammation may cause incorrect 18F-FDG PET findings. In assessment of local ablative treatment of liver metastases 18F-FDG PET seems suitable. The high predictive ability of negative 18F-FDG PET provides evidence that it can be used to assess radicality of tumour ablation. On the other hand, a positive 18F-FDG PET might point to intensification of follow-up or even additional treatment. In case of chemotherapy response evaluation in metastasised CRC, there is a clear relation the results of 18F-FDG PET early after start of treatment with pathological response and survival. This could improve patient management by reducing morbidity, efforts and costs of ineffective treatment in nonresponders. Moreover in an era where new and expensive drugs have limited morphologic effect, it may provide an alternative to anatomy-based evaluation of therapy response. However, it appears impossible to give one single definition of metabolic response, since cut-off values depend on type of treatment, timing of evaluation and tumour type.

***
Part II

Optimisation of Quantification Methodology
Evaluation of Different Normalisation Procedures for the Calculation of the Standardised Uptake Value in Therapy Response Evaluation Studies

*Nucl Med Commun, 2009 Jul;30(7):550-7*  
(adapted and reprinted with permission)

D. Vriens¹,  
L.F. de Geus-Oei¹,  
H.W.M. van Laarhoven²,  
J.N.H. Timmer-Bonte²,  
P.F.M. Krabbe³,  
E.P. Visser¹ and  
W.J.G. Oyen¹

dept. of ¹Nuclear Medicine,  
²Medical Oncology and ³Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands
Abstract

Purpose The aim of this prospective study was to assess the influence of different normalisation procedures on relative changes in standardised uptake values (SUV) of 2-({\textsuperscript{18}}F)fluoro-2-deoxy-D-glucose positron emission tomography (\textsuperscript{18}F-FDG PET) for the assessment of chemotherapy response in patients with colorectal carcinoma (CRC) and non-small cell lung carcinoma (NSCLC).

Methods In 97 patients with CRC (n = 48) or NSCLC (n = 49), \textsuperscript{18}F-FDG PET was performed before and during the course of chemotherapy. Relative changes in SUV (\Delta SUV) were determined after correction for injected dose and bodyweight, lean body mass, body surface area or a combination of bodyweight and plasma glucose. The predictive value for overall and progression-free survival with respect to the different normalised \Delta SUVs was assessed.

Results In both CRC and NSCLC, no differences were seen in the degree of change between the four SUV-normalisations during chemotherapy. Cox regression analysis for overall survival showed significant hazard ratios of 1.14-1.16 per 10\% \Delta SUV change in CRC and 1.10-1.13 in NSCLC and for progression-free survival hazard ratios of 1.15 per 10\% \Delta SUV change in CRC and 1.10-1.12 in NSCLC.

Conclusions Relative changes in SUV is a strong predictor for survival in both CRC and NSCLC. None of the four normalisation methods showed statistical advantage over the other. Therefore, simplifying the methods for analysis of \textsuperscript{18}F-FDG PET data can improve the incorporation of \textsuperscript{18}F-FDG PET in clinical therapy response evaluation and may facilitate application in multicentre trials.

Keywords Chemotherapy · Colorectal Carcinoma · 2-({\textsuperscript{18}}F)fluoro-2-deoxy-D-glucose · Non-Small Cell Lung Carcinoma · Positron Emission Tomography · Standardised Uptake Value · Survival · Therapy Response Evaluation
3.1 Introduction

Functional imaging by $\text{18F-}$fluoro-2-deoxy-d-glucose positron emission tomography ($\text{18F-FDG PET}$) has an established role in the standard care of patients with both colorectal carcinoma (CRC) and non-small cell lung carcinoma (NSCLC). Interest in the application of $\text{18F-FDG PET}$ for prediction and evaluation of tumour response to therapy is growing. When using morphological imaging techniques such as computed tomography (CT) or magnetic resonance imaging it may be difficult to reliably distinguish between necrosis, scar tissue and recurrent or residual tumour in CRC [175,250,251] and NSCLC [252]. Furthermore, metabolic alterations in tumour cells, indicative of tumour response to therapy, may arise earlier than changes in size [253]. Moreover, some new antitumour therapies are cytostatic rather than cytoreductive. Therefore success of treatment cannot reliably be measured on morphologic imaging modalities alone. Early detection of tumour progression during chemotherapy can prevent unbefitting and potentially harmful treatment and provides the opportunity to modify the treatment at an early treatment stage.

In CRC [16] as well as NSCLC [17,254,255], a significant prognostic value for overall survival ($OS$) and progression-free survival ($PFS$) by semiquantitative measures of $\text{18F-FDG PET}$ data analysis has been shown. In addition, the response to chemotherapy, evaluated as relative change in $\text{18F-FDG PET}$-assessed tumour metabolism in CRC [22,223] and NSCLC [21,255], proved to be significantly associated with $OS$ and $PFS$. In these studies, it was suggested that less demanding semiquantitative parameters such as the standardised uptake value ($SUV$) might perform as well as complex dynamic imaging protocols necessary for Patlak analysis [22,116], which would facilitate broad introduction in clinical practice by improving patient compliance. Another advantage of the $SUV$ is that it can be calculated from static, whole-body $\text{18F-FDG PET}$ studies, which depict all metastases. In quantitative dynamic scans only one axial field-of-view ($FoV$, 15-20 cm) can be studied which could exclude metastases, which respond differently to therapy [22].

This study aims at further standardisation, validation and simplification of the methods necessary for metabolic response assessment. For this purpose four distinct normalisation methods for relative changes in $SUV$ ($\Delta SUV$) were evaluated in two patient populations (CRC and NSCLC). $SUV$'s were determined from scans performed on one bed position only, as this study was part of a larger project, which included dynamic acquisition. This single axial $FoV$ was chosen to include as many lesions as possible, based on the baseline whole-body staging $\text{18F-FDG PET}$, acquired at an earlier time point.

$OS$ and $PFS$ were used as outcome measures. For absolute $SUV$ measurement the different normalisation procedures are already extensively compared in former studies [131], of which some seem superior to others [45,116] when compared with full parameter pharmacokinetics by non-linear regression. These studies, however, address the prognostic value of $SUV$ by its absolute value. In contrast, in this study, the relative value will be examined, which is of importance in therapy response evaluation.
3.2 Material & methods

Patient eligibility criteria

Patients eligible for the $^{18}$F-FDG PET chemotherapy response evaluation studies on CRC [22] and NSCLC [21] were included in the present study. In all patients, treatment decisions were based on current guidelines and made by a multidisciplinary team including medical oncologists, oncological and cardio-thoracic surgeons, pulmonologists, radiation oncologists, pathologists, radiologists and nuclear medicine physicians. All clinicians were blinded to the results of the $^{18}$F-FDG PET scans. The study was approved by the Institutional Review Board of the Radboud University Medical Centre, Nijmegen, the Netherlands and written informed consent was obtained from each patient.

$^{18}$F-FDG PET

$^{18}$F-FDG PET was performed at baseline and after two months of treatment (CRC) or after the second or third cycle of chemotherapy (NSCLC), depending on the chemotherapy regimen. Patients were fasted for at least 6 h before imaging. Intake of sugar-free liquids was permitted. A dose of approximately 200 MBq $^{18}$F-FDG (Covidiem, Petten, the Netherlands) was injected intravenously. All scans were acquired between 40 and 50 min postinjection on an ECAT-EXACT47 PET scanner (Siemens Healthcare, Erlangen, Germany) in septa-extended (2D) mode. The position of the patient in the scanner’s FoV (162 mm in 47 planes) was based on whole-body $^{18}$F-FDG PET and CT scans performed for routine clinical work-up, including as many measurable tumour lesions as possible. Only one FoV was scanned, as this study was part of a larger project, which included dynamic scanning [21,22]. A 20-min transmission scan was performed, using the internal $^{68}$Ge/$^{68}$Ga sources, to correct for photon attenuation, the duration of which was chosen to provide a higher signal-to-noise ratio. The emission and transmission sinograms were corrected for randoms and decay. Scatter correction based on measured scatter fractions as implemented in the ECAT 7.2.1. software for 2D reconstructions was used. Attenuation corrected images were reconstructed in $128 \times 128 \times 47$ matrices using filtered backprojection with a Gaussian filter of 4 mm full-width at half-maximum (FWHM). This resulted in voxels of $3.432 \times 3.432 \times 3.375$ mm and a spatial resolution of 6 mm FWHM in the reconstructed images.

$^{18}$F-FDG PET scans were evaluated semiquantitatively by SUV analysis. Tumour volumes-of-interest (VOI) were obtained semiautomatically using a threshold of 50% of the maximum pixel value within the lesion. Four different SUVs, based on injected dose and bodyweight ($SUV_{BW}$), lean body mass ($SUV_{LBM}$), body surface area ($SUV_{BSA}$) and a combination of bodyweight and plasma glucose ($SUV_{BW+glc}$) were calculated (table 3.2, page 31). The injected dose ($ID$) was calculated by subtraction of the residual $^{18}$F activity of the infusion system from the $^{18}$F activity delivered by the laboratory, corrected for decay to time of injection: $ID = (DD \times e^{-\lambda(t-\tau)}) - (RD \times e^{-\lambda(t-\theta)})$ where $DD$ is the delivered dose, $RD$ is
the residual dose, $\lambda$ is the decay constant of $^{18}\text{F}$ ($=0.006314 \text{ min}^{-1}$), $(t - \tau)$ is the time interval between delivery and injection and $(t - \theta)$ is the (negative) time interval between activity measurement of the infusion system and injection.

When multiple lesions were quantified in one patient, a patient’s mean SUV was calculated weighting every lesion by its volume by the equation:

$$SUV_{\text{patient}} = \frac{\sum_{i=1}^{n} (SUV_i \cdot \text{volume}_i)}{\sum_{i=1}^{n} \text{volume}_i}$$ (3.1)

$\Delta SUV$ between the baseline and second $^{18}\text{F}$-FDG PET was calculated ($\Delta SUV = \frac{(SUV_{\text{follow-up}} - SUV_{\text{baseline}})}{SUV_{\text{baseline}}} \cdot 100\%$). All lesions which were completely visible in the FoV were included. There was no maximum to the number of lesions per patient.

**Clinical follow-up**

During and after the treatment, patients were followed up with clinical and radiological examination and laboratory tests at regular intervals. Morphologic tumour response was routinely evaluated according to Response Evaluation Criteria in Solid Tumours (RECIST) [224] without the knowledge of the results of the $^{18}\text{F}$-FDG PET studies. These criteria define progression as a 20% increase in the sum of longest diameters of target lesions or the appearance of new lesion [224]. When disease progression was suspected or proven, patients were always restaged by the previously mentioned multidisciplinary team. The date of local or distant progression was defined as the earliest date at which disease progression was confirmed, either clinically or by imaging or biopsy.

In patients who were progression-free at the close-out date (April 2008) or who had died from any nontumour related cause, the time to progression was censored at that date. OS was measured from the date of the baseline $^{18}\text{F}$-FDG PET scan to the date of death. In patients who were alive at the close-out date, survival was censored at that date.

**Statistical analysis**

Normality of the data was assessed by the Shapiro-Wilk statistic. For normal distributions mean ($\pm$SD) is presented and the paired $t$ test was used for comparison. For non-normal distributions, median (interquartile range, IQR) is presented as measures for central tendency and dispersion and the Wilcoxon signed-rank test was used for comparison. Comparing medians between dependant groups, Friedman two-way analysis of variance by ranks was performed. The different normalised $\Delta SUV$s were compared using a threshold for significant difference of 20%. In studies comparing the test-retest characteristics of $^{18}\text{F}$-FDG uptake, a test-to-test variability of 15-20% is found, meaning that a change in SUV approximately between -20 and +20% is within the reproducibility limits of the test [44][150].
3. SUV Normalisation Procedures for Therapy Response Evaluation

OS and PFS served as the standard of reference. Death of any cause was defined as an event in OS analysis. Survival time was defined as the time between first 18F-FDG PET and event (death or progression). Kaplan-Meier analysis was performed to determine median survival. Dichotomisation was performed at the median metabolic response and strata were compared by the log rank test. Cox’s proportional hazards model was used to assess the predictive value of response evaluation with 18F-FDG PET, as expressed in the ∆SUV’s between the 18F-FDG PET at baseline and follow-up. Hazard ratios (HR) are presented with their 95%-confidence interval (95% – CI, Wald’s χ² test) together with the ratio in median survival (median survival ratio). Finally, to investigate how a prognosis-driven threshold (as contrast to metabolic response-driven threshold) might prove of additional value, a cut-off for ∆SUV was determined for which more than 90% of the patients showed one-year OS.

Analysis was performed with the Statistical Package for Social Sciences (SPSS) version 16.0.2 for Mac (SPSS Inc., Chicago, IL, USA) and GraphPad Prism version 5.0a for Mac (GraphPad Software Inc., La Jolla, CA, USA). Statistical tests were based on a two-sided significance level and the level of significance was set at p = 0.05 for all tests.

3.3 Results

Patient characteristics

One hundred and twenty consecutive eligible patients were included in this prospective study (61 advanced CRC, 59 NSCLC). After the baseline 18F-FDG PET, 23 patients (13 CRC, 10 NSCLC) were excluded for several reasons: owing to technical issues (n = 7), refusal to undergo a second 18F-FDG PET (n = 3), death before the second 18F-FDG PET (n = 3) and early discontinuation of chemotherapy as a result of a significant decline in performance status (n = 10). Therefore, complete datasets of two 18F-FDG PET’s were available in 97 patients (48 CRC and 49 NSCLC) for analysis of therapy response. Patient characteristics are summarised in table 3.1 (page 63). No patients were lost during follow-up. Results of survival analysis by Kaplan-Meier are displayed in table 3.2 (page 63).

In CRC, mean BW was significantly higher at the baseline scan compared to the follow-up scan (78.3 vs 77.1 kg; p = 0.012; individual range -15 to +3 kg). Mean serum glucose (5.4 vs 5.5 mmol·l⁻¹; p = 0.115; individual range -1.1 to +3.2 mmol·l⁻¹) and injected dose (199 vs 204 MBq; p = 0.592) did not differ. In NSCLC, mean serum glucose was significantly higher at the baseline scan compared with the follow-up scan (5.5 vs 5.2 mmol·l⁻¹; p = 0.029; individual range -2.0 to +2.6 mmol·l⁻¹). Mean BW (73.9 vs 73.7 kg; p = 0.637; individual range -9.0 to +10 kg) and injected dose (202 vs 212 MBq; p = 0.215) did not differ.

Semiquantitative changes in 18F-FDG uptake

Median interval between baseline and follow-up 18F-FDG PET was 8.9 weeks (IQR 7.0-9.6) in the CRC group and 7.0 weeks (IQR 5.9-8.6) in the NSCLC group. Box and whisker plots of ∆SUV’s are displayed in figure 3.1 (page 64). In both CRC and
3.3. Results

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CRC</th>
<th>NSCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients</td>
<td>48</td>
<td>49</td>
</tr>
<tr>
<td>Mean age [year] (range)</td>
<td>61.0 (44.7-78.9)</td>
<td>59.6 (38.5-76.2)</td>
</tr>
<tr>
<td>Men [%]</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>Location of Metastases [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>86</td>
<td>14</td>
</tr>
<tr>
<td>Lung</td>
<td>29</td>
<td>37</td>
</tr>
<tr>
<td>Lymphnodes</td>
<td>8</td>
<td>67</td>
</tr>
<tr>
<td>Bone</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Brain</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>Adrenals</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>None</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Histology [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>94</td>
<td>45</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>-</td>
<td>43</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Mucinous adenocarcinoma</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Bronchoalveolar cell carcinoma (adenocarcinoma in situ)</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Clear cell carcinoma</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Tumour Differentiation [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Very poor</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Poor</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intermediate</td>
<td>60</td>
<td>12</td>
</tr>
<tr>
<td>Well</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Mucinous</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Unspecified</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>Tumour Stage [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>-</td>
<td>2 (B)</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>2 (A)</td>
</tr>
<tr>
<td>III</td>
<td>-</td>
<td>16 (A), 20 (B)</td>
</tr>
<tr>
<td>IV</td>
<td>100</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 3.1: Characteristics of patients with CRC or NSCLC. CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma.

NSCLC the decline in median of all four $SUV_s$ between first and second $^{18}$F-FDG PET was statistically significant ($p < 0.001$). No significant differences between the four compared $\Delta SUV_s$ could be found in CRC ($p = 0.143$) and NSCLC ($p = 0.059$).

<table>
<thead>
<tr>
<th>Event-free (%)</th>
<th>Median (95% – CI) [weeks]</th>
<th>one-year (95% – CI) [%]</th>
<th>two-year (95% – CI) [%]</th>
<th>three-year (95% – CI) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CRC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OS</em></td>
<td>7 (15)</td>
<td>85 (70-100)</td>
<td>77 (65-89)</td>
<td>34 (21-48)</td>
</tr>
<tr>
<td><em>PFS</em></td>
<td>1 (2)</td>
<td>24 (19-29)</td>
<td>17 (6-27)</td>
<td>4 (0-10)</td>
</tr>
<tr>
<td><strong>NSCLC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OS</em></td>
<td>7 (14)</td>
<td>62 (42-82)</td>
<td>57 (43-71)</td>
<td>39 (25-53)</td>
</tr>
<tr>
<td><em>PFS</em></td>
<td>4 (8)</td>
<td>30 (17-42)</td>
<td>30 (17-42)</td>
<td>19 (8-30)</td>
</tr>
</tbody>
</table>

Table 3.2: Outcome of follow-up of CRC and NSCLC patients using Kaplan-Meier analysis. 95% – CI: 95%-confidence interval; CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma; OS: overall survival; PFS: progression-free survival.
3. SUV Normalisation Procedures for Therapy Response Evaluation

Figure 3.1: Box-and-whisker plots of four normalisations for percentage change in standardised uptake value (ΔSUV) between baseline and follow-up 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography. The boxes are constructed by the mean, first and third quartiles. The whisks extend 1.5 times the interquartile range above and below the 25th and 75th percentiles. Outliers are displayed as open circles (the number labels represent case IDs). CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma.

In CRC differences in metabolic response between ΔSUV<sub>LBM</sub>, ΔSUV<sub>BSA</sub> and ΔSUV<sub>BW+glc</sub> compared with ΔSUV<sub>BW</sub> was more than the reproducibility limit in one, zero and five of 48 cases, respectively. In NSCLC this was zero, zero and five of 49 cases, respectively.

Prediction of survival by 18F-FDG PET

Cox’s proportional hazards model for CRC and NSCLC showed high predictive significance for OS and PFS for all four ΔSUV normalisations between baseline and follow-up 18F-FDG PET (table 3.3, page 65). To assess for predictive ability of the four normalisation methods of metabolic therapy response as to OS and PFS, patients were dichotomised to the median ΔSUV. Results are displayed in table 3.4 (page 65).

In CRC using a cut-off for ΔSUV of -33% for BW, LBM and BSA and -22% for BW + glc separated patients who had 90% one-year OS from those with lower OS rates. In NSCLC these cut-offs were -56% for BW, -57% for LBM and BSA and -53% for BW + glc. The group defined as metabolic responders consisted of 24-31% of the patients. In contrast to CRC, in NSCLC these numbers were very different from the medians (approximately -37%). This prognosis-driven dichotomisation for NSCLC for OS resulted in median survival ratios of 0.28 (HR 2.8; log rank p = 0.001) for ΔSUV<sub>BW</sub> and ΔSUV<sub>BW+glc</sub> and 0.22 (HR 3.1; log rank p < 0.001) for ΔSUV<sub>LBM</sub> and ΔSUV<sub>BSA</sub>. 
3.4 Discussion

In this study, we showed for the first time that the method for SUV normalisation does not influence the predictive value of $^{18}$F-FDG PET in CRC and NSCLC in chemotherapy response evaluation. The predictive value of $^{18}$F-FDG PET for CRC \cite{16} and NSCLC \cite{256,262} has been studied extensively and is an established predictor for survival. Chemotherapy response evaluation in CRC was performed by different groups, all using absolute \cite{220,223,248} or relative \cite{22} SUV$_{BW}$ differences as a measure for change in metabolic activity. In NSCLC, $\Delta$SUV as result of therapy have...

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
& & OS & & PFS \\
& & HR & 95% – CI & HR & 95% – CI \\
\hline
\textit{CRC} & & & & & \\
$\Delta$SUV$_{BW}$ & 1.16$^\dagger$ & (1.04-1.29) & 1.15$^\dagger$ & (1.04-1.28) \\
$\Delta$SUV$_{LBM}$ & 1.14$^\dagger$ & (1.04-1.26) & 1.15$^\dagger$ & (1.04-1.27) \\
$\Delta$SUV$_{BSA}$ & 1.15$^\dagger$ & (1.04-1.27) & 1.15$^\dagger$ & (1.04-1.27) \\
$\Delta$SUV$_{BW+glc}$ & 1.14$^\dagger$ & (1.03-1.24) & 1.15$^\dagger$ & (1.04-1.26) \\
\hline
\textit{NSCLC} & & & & & \\
$\Delta$SUV$_{BW}$ & 1.12$^\dagger$ & (1.03-1.20) & 1.12$^\dagger$ & (1.04-1.20) \\
$\Delta$SUV$_{LBM}$ & 1.10$^\dagger$ & (1.03-1.17) & 1.10$^\dagger$ & (1.04-1.18) \\
$\Delta$SUV$_{BSA}$ & 1.12$^\dagger$ & (1.04-1.27) & 1.12$^\dagger$ & (1.04-1.27) \\
$\Delta$SUV$_{BW+glc}$ & 1.13$^\dagger$ & (1.04-1.21) & 1.12$^\dagger$ & (1.03-1.20) \\
\hline
\end{tabular}
\caption{Results of univariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta$SUVs between the scan at baseline and evaluation. 95% – CI: 95%-confidence interval; BSA: normalisation for body surface area; BW: normalisation for bodyweight; BW + glc: normalisation for bodyweight and plasma glucose; CRC: colorectal carcinoma; HR: hazard ratio; LBM: normalisation for lean body mass; NSCLC: non-small cell lung carcinoma; OS: overall survival; PFS: progression-free survival; $\Delta$SUV: relative change in standardised uptake value between baseline and follow-up scan. $^\ast$Per 10% change. $^\dagger$Significant by Wald’s $\chi^2$ test.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
& & OS & & PFS \\
& & Median survival ratio & HR & 95% – CI & HR & 95% – CI \\
\hline
\textit{CRC} & & & & & \\
$\Delta$SUV$_{BW}$ & 0.60 & 1.95$^\dagger$ & (1.03-3.69) & 0.59 & 1.92$^\dagger$ & (1.07-3.44) \\
$\Delta$SUV$_{LBM}$ & 0.82 & 1.33 & (0.71-2.47) & 0.65 & 1.57 & (0.88-2.80) \\
$\Delta$SUV$_{BSA}$ & 0.60 & 1.95$^\dagger$ & (1.03-3.69) & 0.59 & 1.92$^\dagger$ & (1.07-3.44) \\
$\Delta$SUV$_{BW+glc}$ & 0.60 & 2.48$^\dagger$ & (1.26-4.89) & 0.52 & 2.08$^\dagger$ & (1.16-3.74) \\
\hline
\textit{NSCLC} & & & & & \\
$\Delta$SUV$_{BW}$ & 0.48 & 2.12$^\dagger$ & (1.14-3.92) & 0.37 & 2.97$^\dagger$ & (1.56-5.66) \\
$\Delta$SUV$_{LBM}$ & 0.48 & 2.12$^\dagger$ & (1.14-3.92) & 0.37 & 2.97$^\dagger$ & (1.56-5.66) \\
$\Delta$SUV$_{BSA}$ & 0.48 & 2.12$^\dagger$ & (1.14-3.92) & 0.37 & 2.97$^\dagger$ & (1.56-5.66) \\
$\Delta$SUV$_{BW+glc}$ & 0.70 & 1.48 & (0.80-2.74) & 0.45 & 1.60 & (0.89-2.88) \\
\hline
\end{tabular}
\caption{Results of univariate Kaplan-Meier analysis for overall and progression-free survival using $\Delta$SUVs between the scan at baseline and evaluation dichotomised at their median for response or nonresponse. 95% – CI: 95%-confidence interval; BSA: normalisation for body surface area; BW: normalisation for bodyweight; BW + glc: normalisation for bodyweight and plasma glucose; CRC: colorectal carcinoma; HR: hazard ratio; LBM: normalisation for lean body mass; NSCLC: non-small cell lung carcinoma; OS: overall survival; PFS: progression-free survival; $\Delta$SUV: relative change in standardised uptake value between baseline and follow-up scan. $^\ast$Per 10% change. $^\dagger$Significant by log rank test.}
\end{table}
been observed in numerous studies. Some used \( BW \) normalised \( SUV \) \cite{21,263,265}, others used normalisation by \( BSA \) and serum glucose \cite{255}. However, for metabolic response assessment none so far has investigated which of the different normalisations is optimal as compared with \( OS \) and \( PFS \).

In both CRC and NSCLC, all four normalised \( SUV \)s showed a significant decrease between baseline and follow-up \(^{18}\text{F-FDG}\) PET. Although it was shown that \( BW \) (in CRC) and serum glucose (in NSCLC), the major factors for normalisation of \( SUV \), were significantly different between both baseline and follow-up scan, we did not find mutual differences between the four \( \Delta SUV \) normalisations. This suggests that there is no preference for either any of the four normalisations in \( SUV \) when therapy response is measured by \( \Delta SUV \)s between baseline and follow-up \(^{18}\text{F-FDG}\) PET.

Krak et al. \cite{17} describe that the standard deviation between two consecutive scans (made on two consecutive days) was 11\% using the same definition of VOI as in this study. Therefore \( SUV \)s on two consecutive days may vary approximately \( \pm 20\% \) because of reproducibility limits of the test. Using this as a cut-off for a significant different \( \Delta SUV \) between the two normalisations, we found that especially in \( \Delta SUV_{BW+glc} \) more often a change in metabolism was found more than \( 20\% \) different from \( \Delta SUV_{BW} \). The addition of the extra parameter plasma glucose level caused in five CRC and five NSCLC patients a highly different conclusion about therapy response. To verify the clinical significance of this different magnitude of metabolic response, it was related to survival.

Choosing a cut-off value of metabolic response for 90\% one-year \( OS \) led to a median survival ratio of 0.22-0.28 for metabolic nonresponders vs responders. In CRC, the effect of abovementioned cut-off was similar to the median \( SUV \) change (and thus selects approximately 50\% best responders).

We observed higher \( HR \) for NSCLC than CRC. This suggests that similar reduction in \( SUV \) in CRC has less effect on patient \( OS \) than in NSCLC. This may be caused by biological differences and differences in chemosensitivity of both types of cancer. However, changes in glucose metabolic rate, are not only dependent on biological behaviour of the tumour to the given treatment, but are also dependent on the treatment protocol and the timing of follow-up scanning. Both these biological differences and variation in treatment and follow-up protocol hinder determination of optimal cut-offs. Therefore standardisation of response measurement protocols are necessary and cut-offs should be dependent of tumour type, antitumour treatment and timing of evaluation.

Earlier publications \cite{21,22} showed that \( SUV \) is a sufficiently robust measure for therapy response evaluation and can reliably replace more complex, invasive and time-consuming measures such as Patlak analysis, which determines glucose metabolic rate \( (MR_{glc}) \). Apart from being a less time-consuming method, \( SUV \) can be calculated from a whole-body \(^{18}\text{F-FDG}\) PET study including all metastatic lesions. Moreover, no input function is required for \( SUV \) determination. The advantage of relative \( SUV \) determination compared with absolute \( SUV \)s, however, is that it is easier to combine data from different studies because they are less sensitive to introduction of errors because of noise, image resolution and VOI definition compared with absolute \( SUV \)s \cite{42,44,61,107}. If the patient is scanned in the same hospital on the same scanner, some scanner-related factors and patient related factors could be ignored \cite{61}.
Different normalisation methods for absolute $SUV$ values have been addressed by others \cite{43, 133, 266, 268}, reporting varying results. Some have shown that $SUV$ corrected for $BSA$ proved to be more accurate, in adult \cite{266} and paediatric \cite{268} patients, compared with the gold standard $MR_{glc}$. It was suggested that normalisation is necessary because of variability in $SUV$ owing to body composition and habitus as well as plasma glucose \cite{55, 107}. Menda et al. \cite{267} observed no advantage of any $SUV$ correction for accuracy of diagnosis of pulmonary malignancy. In contrast with our study, these studies did not address therapy response and therefore did not take into consideration the measurement of $\Delta S$.  

Krak et al. \cite{133} compared different normalisations for relative $SUV$ change in 20 women with locally advanced or metastasised breast carcinoma with the gold standard of non-linear regression after one, two or six courses of chemotherapy. They concluded that of the investigated normalisation methods, $\Delta S$ corrected for both $LBM$ and glucose showed highest correlation with relative change in $MR_{glc}$ as calculated by non-linear regression. This study did not correlate $SUV$ corrections with clinical outcome parameters such as patient survival.

Stahl et al. \cite{43} compared histological response in 43 patients with locally advanced gastric carcinomas with the same four normalisations for $\Delta S$ as in our study, between baseline and follow-up $^{18}$F-FDG PET (after two weeks, during first cycle of platinum-based chemotherapy). They, too, concluded that no normalisation method of $\Delta S$s has an advantage for response prediction and that the theoretical benefits of the measures to reduce the dependency of the $SUV$ on $BW$ or plasma glucose do not translate into a detectable clinical benefit.

Our study showed no preference of any $SUV$ normalisation in therapy response evaluation in CRC and NSCLC. This could be caused by the fact that $SUV$s used in therapy response evaluation are compared in individual patients and not between different patients. It can be easily derived that an increase in $BW$ alone by 10%, increases $\Delta S_{BW}$ and $\Delta S_{BW+glc}$ by 10%, $\Delta S_{BSA}$ by 4.1% ($= (1.1)^{0.425} - 1$) and has no effect on $\Delta S_{LBM}$. An increase of plasma glucose alone by 10% increases $\Delta S_{BW+glc}$ by 10%. The effect of chemotherapy and disease on body composition seems to be relatively small. Therefore it is possible to select the $SUV$, which is simplest, without introduction of extra parameters like $BW$, length or serum glucose. All these parameters necessitate calibration and for that reason may potentially introduce extra uncertainties. Moreover, simplifying $^{18}$F-FDG PET methodology in therapy response evaluation could facilitate its introduction in routine clinical practice.

The variety of applied analytical methods is vast, which hampers multicentre research. Clear methods for standardisation of acquisition, reconstruction, VOI definition and $SUV$ normalisation are need to be determined \cite{61}. In fact, methodology of metabolic response evaluation needs to be standardised in evidence-based multidisciplinary international guidelines. This has been attempted previously by the European Organisation for Research and Treatment of Cancer \cite{5} and the National Cancer Institute \cite{45}. However, they provide consensus-based recommendations, rather than advice based on scientific proof. Evidence-based guidelines would be of utmost importance for the interpretation and comparison of multicentre trials.
3. SUV Normalisation Procedures for Therapy Response Evaluation

3.5 Conclusions

In chemotherapy response evaluation in both patients with CRC and NSCLC, $\Delta SUV$ in $^{18}$F-FDG PET have high predictive value for patient survival. Using relative changes in tumour $^{18}$F-FDG uptake for a patient, no normalisation for body habitus in $SUV$ calculation seems superior, as all perform equally well for prediction of survival in both types of cancer. Therefore theoretical advantages of one normalisation method over another, do not translate into clinical relevant changes. This finding in combination with the fact that relative (rather than absolute) changes are less dependent on quantitative PET acquisition protocols [42,44], will facilitate determination of response to therapy. This could be a step towards standardising the results of therapy evaluation which might be used for pooling of data in multicentre trials. Moreover, it may enable integration of metabolic response measurement in the development of clinical guidelines by giving evidence-based definitions of partial response, stable disease and progressive disease for different cancer types.

***
Comparison of Two Volume-of-Interest Definition Methods for Metabolic Response Evaluation with $^{18}$F-FDG PET

*Q J Nucl Med Mol Imaging, 2010 Dec;54(6):677-88*
(adapted and reprinted with permission)

D. Vriens$^1$,  
L.F. de Geus-Oei$^1$,  
H.W.M. van Laarhoven$^2$,  
H.F.M. van der Heijden$^3$,  
P.F.M. Krabbe$^4$,  
E.P. Visser$^1$ and  
W.J.G. Oyen$^1$

dep. of $^1$Nuclear Medicine,  
$^2$Medical Oncology,  
$^3$Pulmonology and $^4$Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands
4. VOI definitions for Metabolic Response Evaluation

Abstract

Purpose In therapy response evaluation by 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose positron emission tomography (\(^{18}\)F-FDG PET), different tumour delineations are used, resulting in different values for change in glucose metabolic rate (\(\Delta MR_{g lc}\)). We propose a technique to compare metabolic rates in a volume-of-interest (VOI) based on fixed volumes rather than on fixed thresholds. This method involves change in lesion size.

Methods In 49 patients with colorectal carcinoma (CRC) and 50 patients with non-small cell lung carcinoma (NSCLC) scheduled for chemotherapy, \(^{18}\)F-FDG PET was performed at baseline and during chemotherapy. A VOI\(_{fixed\, thresholds}\) was determined by using a 50% threshold on both baseline and follow-up \(^{18}\)F-FDG PET. A VOI\(_{fixed\, volumes}\) was determined by using a 50% threshold, determined on the series with the largest tumour volume. This VOI\(_{fixed\, volumes}\) is used on consecutive scans. Predictive effects of both methods were investigated by survival analysis for OS and PFS.

Results In CRC, only VOI\(_{fixed\, volumes}\) based \(\Delta MR_{g lc}\) showed significant predictive ability. In NSCLC, both techniques showed significant predictive ability. During multivariate analysis, VOI\(_{fixed\, volumes}\) determined \(\Delta MR_{g lc}\) was an independent predictor for both overall (OS and progression-free survival (PFS)) in NSCLC whereas VOI\(_{fixed\, thresholds}\) determined \(MR_{g lc}\) was not. After dichotomisation at the median \(\Delta MR_{g lc}\), median survival ratio was higher in VOI\(_{fixed\, volumes}\) than VOI\(_{fixed\, thresholds}\) for CRC (OS: 1.78 vs 1.25, PFS: 1.57 vs 1.21) and NSCLC (OS: 2.01 vs 2.01, PFS: 2.93 vs 2.13).

Conclusions VOI\(_{fixed\, volumes}\) based \(\Delta MR_{g lc}\) shows better correlation with survival than \(\Delta MR_{g lc}\) determined from a VOI\(_{fixed\, thresholds}\).

Keywords Chemotherapy · Colorectal Carcinoma · Drug Monitoring · 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose · Neoplasms, Therapy · Non-Small Cell Lung Carcinoma · Positron Emission Tomography · Survival
4.1 Introduction

Functional imaging with 2-\(^{18}\)F\)fluoro-2-deoxy-D-glucose positron emission tomography (\(^{18}\)F-FDG PET) has an established role in the standard care for patients with colorectal carcinoma (CRC) or non-small cell lung carcinoma (NSCLC) by staging of the disease. Growing interest in the application of \(^{18}\)F-FDG PET for prediction and evaluation of tumour response to therapy has risen, since morphologic imaging techniques such as computed tomography (CT) or magnetic resonance imaging (MRI) may lead to incorrect conclusions about therapy response. Due to the fact that it proves difficult to reliably distinguish between therapy induced fibrosis, tumour necrosis and recurrent or residual tumour in both CRC and NSCLC, favourable and adverse alterations may be indistinguishable. Furthermore, metabolic changes in tumour cells, indicative of response to therapy, may occur earlier than changes in lesion size, especially as some new antitumour therapies are cytostatic rather than cytoreductive. Early detection of tumour progression during chemotherapy by \(^{18}\)F-FDG PET might therefore prevent unbeneficial, perhaps even harmful treatment.

It has already been demonstrated that pre- and posttherapy parameters for tumour glucose metabolism are of prognostic value in CRC and NSCLC. In addition, changes in the rate of glucose metabolism predict overall (OS) and progression-free survival (PFS) in both CRC and NSCLC.

Since the value of change in glucose metabolic activity is highly dependent on the definition of tumour volume-of-interest (VOI), an exact and reproducible definition of VOI methodology is important. Institution-dependent VOI definitions may lead to variations in the classification of metabolic response and may hinder the integrated or comparative interpretation of results of multiple centres. In literature, many different methods have been used to define tumour VOI: a thresholded 3D isocontour (using 50% or 70% of the maximum voxel value within the lesion), the maximum voxel value or a fixed dimensions-method (e.g., 15 \(\times\) 15 mm\(^2\) around maximum value). The first method determines treatment induced changes in tumour metabolic activity in two different volumes, since using fixed thresholds, glucose metabolic rate (\(MR_{glc}\)) is determined in the metabolic active volume only.

The use of the fixed thresholds-based methodology is attractive since tumour delineation on \(^{18}\)F-FDG PET is practical, easier to perform and more reproducible than other methods. A disadvantage of the fixed thresholds technique is that the threshold level is chosen rather arbitrarily and that only \(MR_{glc}\) in residual, metabolic active tumour is taken into account. As a result, a lesion that decreases in size, but preserves the same baseline metabolic activity will not be considered to respond to treatment with this VOI definition. Moreover, a sole decrease in lesion volume might artefactually decrease the measured metabolic activity due to the partial volume effect, especially in case of residual tumour less than \(\sim 15\) mm in diameter. When using the same VOI during therapy response evaluation (VOI\(_{\text{fixed volumes}}\)), both the change in metabolic activity as well as the change in metabolic tumour volume are taken into account. Since a decrease in lesion volume causes peritumoural tissue with normalised \(^{18}\)F-FDG accumulation that is incorporated in the VOI resulting in a reduction of its \(MR_{glc}\).
4. VOI definitions for Metabolic Response Evaluation

The hypothesis is that using the same VOI-volume for MR\textsubscript{glc} determination, both at baseline and during follow-up is more indicative for therapy response. Present study introduces two distinct VOI methods for metabolic response evaluation assessed prospectively in both CRC and NSCLC patients. One technique solely evaluates metabolic response and the other incorporates the change in volume. Both techniques are correlated with patient survival.

4.2 Material & methods

Patient eligibility criteria

Between March 2002 and December 2005 patients in the Radboud University Medical Centre, Nijmegen, the Netherlands, with metastatic CRC (stage IV), who were scheduled to undergo palliative chemotherapy and patients with any stage of NSCLC, who were scheduled to undergo induction chemotherapy or palliative chemotherapy were asked to participate in this study. Exclusion criteria were diabetes mellitus and pregnancy.

In all patients, treatment decision-making was done by a multidisciplinary team including medical oncologists, surgeons (CRC), cardiothoracic surgeons (NSCLC), pulmonologists (NSCLC), radiation oncologists, pathologists, radiologists and nuclear medicine physicians. All clinicians were blinded to the results of the serial \textsuperscript{18}F-FDG PET scans. The study was approved by the Institutional Review Board of the Radboud University Medical Centre and written informed consent was obtained from each patient.

One hundred and twenty-one consecutive eligible patients could be included in this prospective study (61 advanced CRC, 60 NSCLC). After the baseline \textsuperscript{18}F-FDG PET, 22 patients (12 CRC, 10 NSCLC) were excluded for several reasons: due to technical issues (\(n = 4\)), refusal to undergo a second \textsuperscript{18}F-FDG PET (\(n = 5\)), death before the second \textsuperscript{18}F-FDG PET (\(n = 3\)) and early discontinuation of chemotherapy due to a significant decline in performance status (\(n = 10\)). Therefore, complete data-sets of two \textsuperscript{18}F-FDG PET were available in 99 patients (49 CRC and 50 NSCLC) for analysis of therapy response. Patient characteristics for both tumour types are summarised in table 4.1 (page 75).

Patient treatment

Of the CRC patients, 26 patients received first line chemotherapy, 16 in second line, six in third line and one in fourth line. Chemotherapy regimens were based on fluoropyrimidines (capecitabine and 5-fluorouracil) with or without oxaliplatin and irinotecan or monoclonal antibodies (bevacizumab and cetuximab).

Of the NSCLC patients, 14 patients were treated with induction chemotherapy and the remaining 36 received chemotherapy in a palliative setting (32 in first line and four in second line). Chemotherapy regimens were based on platinum-containing alkylating agents, gemcitabine, etoposide, vinorelbine or docetaxel. Patients receiving induction chemotherapy were subsequently treated with radical radiotherapy (\(n = 9\))
4.2. Material & methods

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CRC</th>
<th>NSCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Mean age [year] (range)</td>
<td>60.5 (44.7-78.9)</td>
<td>59.7 (41.2-76.3)</td>
</tr>
<tr>
<td>Men [%]</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td><strong>Location of CRC Primary Tumour [%]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>20</td>
<td>N/A</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>47</td>
<td>N/A</td>
</tr>
<tr>
<td>Rectum</td>
<td>18</td>
<td>N/A</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td>14</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Location of CRC Metastases [%]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>86</td>
<td>N/A</td>
</tr>
<tr>
<td>Lung</td>
<td>31</td>
<td>N/A</td>
</tr>
<tr>
<td>(Retro)peritoneal lymphnodes</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Bone</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Histology [%]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>94</td>
<td>46</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>-</td>
<td>42</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Mucinous adenocarcinoma</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Bronchoalveolar cell carcinoma (adenocarcinoma in situ)</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Clear cell carcinoma</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td><strong>Tumour Differentiation [%]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Very poor</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Poor</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Intermediate</td>
<td>63</td>
<td>14</td>
</tr>
<tr>
<td>Well</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Unspecified</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td><strong>Tumour Stage [%]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>-</td>
<td>2 (B)</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>2 (A)</td>
</tr>
<tr>
<td>III</td>
<td>-</td>
<td>16 (A), 18 (B)</td>
</tr>
<tr>
<td>IV</td>
<td>100</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 4.1: Characteristics of patients with CRC or NSCLC. CRC: colorectal carcinoma; N/A: not applicable; NSCLC: non-small cell lung carcinoma.

or curative surgery ($n = 2$). The remaining three patients were treated with palliative radiotherapy, because of progression during induction chemotherapy based on CT-criteria.

No patients were lost during follow up. Survival of patients is displayed in table 4.2 (page 76).

$^{18}$F-FDG PET

Quantitative 4D dynamic $^{18}$F-FDG PET data acquisition and reconstruction

4D dynamic $^{18}$F-FDG PET was performed at baseline and after two months of treatment (CRC) or after the second or third cycle of chemotherapy (NSCLC), depending on the chemotherapy regimen. Patients were fasted for at least 6 h before imaging. Intake of sugar-free liquids was permitted. Blood glucose levels (hexokinase method, Aeroset, Abbott diagnostics, Abbott Park, IL, USA) were determined. All scans
### 4. VOI definitions for Metabolic Response Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Event-free (%)</th>
<th>Median (95% − CI) [weeks]</th>
<th>one-year (95% − CI) [%]</th>
<th>two-year (95% − CI) [%]</th>
<th>three-year (95% − CI) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CRC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>8 (16)</td>
<td>89 (74-104)</td>
<td>78 (66-89)</td>
<td>38 (24-52)</td>
<td>16 (5-27)</td>
</tr>
<tr>
<td>PFS</td>
<td>1 (2)</td>
<td>25 (21-30)</td>
<td>18 (8-29)</td>
<td>4 (0-10)</td>
<td>2 (0-6)</td>
</tr>
<tr>
<td><strong>NSCLC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>7 (14)</td>
<td>58 (25-92)</td>
<td>56 (42-70)</td>
<td>36 (23-49)</td>
<td>22 (10-34)</td>
</tr>
<tr>
<td>PFS</td>
<td>3 (6)</td>
<td>26 (14-39)</td>
<td>27 (13-39)</td>
<td>17 (6-27)</td>
<td>8 (0-16)</td>
</tr>
</tbody>
</table>

Table 4.2: Outcome of follow-up of CRC and NSCLC patients using Kaplan-Meier analysis.

95% − CI: 95%-confidence interval; CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma; OS: overall survival; PFS: progression-free survival.

were acquired on an ECAT-EXACT47 PET scanner (Siemens Healthcare, Erlangen, Germany). The position of the patient in the scanner’s field-of-view (162 mm in 47 planes) for 4D dynamic acquisition was based on the whole-body $^{18}$F-FDG PET and CT scans performed in every patient for routine clinical work-up. In case not all lesions could be acquired in one field-of-view, the patient was positioned to include as many measurable tumour lesions as possible. This position was used for follow-up scanning as well. A 20-min transmission scan was made, using the internal $^{68}$Ge/$^{68}$Ga sources, to correct for photon attenuation. Approximately 200 MBq $^{18}$F-FDG (Covidien, Petten, the Netherlands) was injected intravenously using a constant infusion remote-controlled pump (Medrad Inc.; Bayer Schering Pharma AG, Leverkusen, Germany). The 4D dynamic data acquisition, performed in septa-extended (2D) mode, was started simultaneously with the injection of $^{18}$F-FDG and consisted of 16 time frames with parameter duration (10·30 s, 3·300 s, 3·600 s) for a total time of 50 min. Corrections for decay, randoms and scatters was performed. Attenuation-corrected images were reconstructed in 128·128·47 matrices, using filtered back projection with a Gaussian filter of 4 mm full-width at half-maximum (FWHM). This resulted in voxels of 3.432·3.432·3.375 mm (39.8 µL) and a spatial resolution of 6 mm FWHM in the reconstructed images.

**Plasma time-activity concentration curves**

Plasma time-activity concentration curves were derived from 17 manually taken arterial blood samples (~2 ml) from a 20 G cannula in the radial artery. Seven samples were drawn at 15 s intervals, followed by samples at 135 s, 165 s, 225 s, 285 s, 7.5 min, 12.5 min, 17.5 min, 25 min, 35 min and 45 min after injection. Radioactivity in the plasma (obtained by centrifugation) was determined in a well-type $\gamma$-counter (Wallac 1480 Wizard, Perkin Elmer Lifescience, Zaventem, Belgium). This procedure is extensively described before [83]. When arterial cannulation was contraindicated or not feasible (in 44 of 98 CRC scans and 41 of 100 NSCLC scans), an image-derived input function (IDIF) was determined by measuring $^{18}$F-FDG counts in a VOI over the ascending aorta or the abdominal aorta, that accurately represents $^{18}$F-FDG blood levels [83].
4.2. Material & methods

Tumour time-activity concentration curves & volume-of-interest methodology (fixed thresholds, fixed volumes)

Tumour time-activity concentration curves were obtained by placing 3D VOIs over the tumour and each metastasis using two different techniques: the fixed thresholds and fixed volumes method. The locations of the lesions were evaluated visually on the transaxial, coronal and sagittal images in summed late time frames (frame 14-16) yielding a static image of 30 min and a scan mid-time of 35 min postinjection.

The VOI\textsubscript{fixed volumes} were semi-automatically determined in the summed late time frames using isocontours with a threshold of 50\% of the maximum voxel value within the lesion on the \textsuperscript{18}F-FDG PET on which the lesion was largest. In case of partial or complete metabolic response (decrease of threshold based volume) the VOI\textsubscript{fixed volumes} was determined on the pretreatment \textsuperscript{18}F-FDG PET and copied to the second \textsuperscript{18}F-FDG PET and in case of progressive disease (increase of threshold based volume or appearance of new lesions) the VOI\textsubscript{fixed volumes} was determined on the second \textsuperscript{18}F-FDG PET and copied to the pretreatment \textsuperscript{18}F-FDG PET. The copied VOI was manually translated in all three-dimensions using anatomical landmarks of surrounding structures. This VOI\textsubscript{fixed volumes} was used on each time frame of the scan (figure 4.1A and 4.1B, page 78).

The VOI\textsubscript{fixed thresholds} were semi-automatically determined in the summed late time frames using an isocontour with a threshold of 50\% of the maximum voxel value within the lesion on the pretreatment scan. This VOI\textsubscript{fixed thresholds} was copied to each time frame of this scan. This process was repeated on the follow-up scan (figure 4.1C and 4.1D, page 78). In case of complete metabolic response, which implies that no tumour contour could be determined by a certain threshold, the VOI of the baseline scan was copied. Thereby, the VOI included ‘background’ tissue with the same \textsuperscript{18}F-FDG uptake as surrounding tissue (liver or lung).

Patlak analysis

Patlak analysis was used to compute the \( MR\text{glc} \) [\( \mu \text{mol} \cdot \text{min}^{-1} \cdot \text{cm}^{-3} \)] in each lesion using both VOIs and the plasma time-activity concentration curve (or IDIF), as described in detail\textsuperscript{33,83}. \( MR\text{glc} \) in the VOI was calculated by multiplication of the slope of the Patlak-plot (\( i.e., \ K_i \)) with the blood glucose level, thereby implicitly assuming a lumped constant of 1.0. When multiple lesions were quantified in one patient, mean \( MR\text{glc} \) was calculated weighting every lesion by its VOI volume by the equation:

\[
MR\text{glc}_{\text{patient}} = \frac{\sum_{i=1}^{n} (MR\text{glc}_{i} \cdot volume_i)}{\sum_{i=1}^{n} volume_i} \quad (4.1)
\]

The relative change in \( MR\text{glc} \) (\( \Delta MR\text{glc} \)) between the baseline and follow-up \textsuperscript{18}F-FDG PET was calculated (\( \Delta MR\text{glc} = \frac{(MR\text{glc}_{\text{follow-up}} - MR\text{glc}_{\text{baseline}})}{MR\text{glc}_{\text{baseline}}} \cdot 100\% \)).
4. VOI definitions for Metabolic Response Evaluation

**Figure 4.1: Example of different VOI definitions with $^{18}$F-FDG PET.** This case represents a 70-year-old male patient with squamous cell NSCLC, $T_2N_2M_1$ treated with two cycles of first line carboplatin/gemcitabine chemotherapy. Planes were anatomically correlated. $\Delta MR_{\text{glc}}$ for VOI\textsubscript{fixed volumes} was -82% and for VOI\textsubscript{fixed thresholds} was -54%. Patient was alive after three years of treatment. VOI\textsubscript{fixed volumes} at baseline (panel A) and during the course of chemotherapy (panel B). VOI\textsubscript{fixed thresholds} at baseline (panel C) and during the course of chemotherapy (panel D). $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; $\Delta MR_{\text{glc}}$: relative change in glucose metabolic rate between baseline and follow-up $^{18}$F-FDG PET; NSCLC: non-small cell lung carcinoma; VOI: volume-of-interest.

**Clinical follow-up**

During and after treatment, patients were followed with clinical examination at regular intervals, chest CTs (NSCLC, every six months in CRC), abdominal CT/MRI (every three months in CRC), chest X-rays (NSCLC), CEA measurement (CRC), routine laboratory tests and other imaging studies as clinically indicated. Morphologic tumour response was evaluated on CT, MRI or conventional chest X-ray according to response evaluation criteria in solid tumours (RECIST)\textsuperscript{224} without knowledge of the results of the $^{18}$F-FDG PET studies. These criteria define progression as a 20% increase in the sum of longest diameters of target lesions or the appearance of new lesion\textsuperscript{224}. When recurrence was suspected or proven, patients were always restaged. The progression and relapse pattern and cause of death were determined in all cases. The date of local or distant progression was defined as the earliest date at which disease progression was confirmed, either clinically or by imaging or biopsy. In patients who were progression-free at the closeout date (April 2008) or who had died from any cause, the time to progression was censored at that date. Survival was measured from the date of the baseline $^{18}$F-FDG PET scan to the date of death. In patients who were alive at the closeout date survival was censored to that date.
4.3. Results

Statistical analysis

All continuous parameters were assessed for normality by the Shapiro-Wilk statistic. For non-normal distributions, median and interquartile range (IQR) are presented as measures for central tendency and dispersion, the non-parametric Wilcoxon signed-rank test was used for comparison and correlation between both VOI methods is displayed as Spearman’s \( \rho \). Metabolic rate changes during therapy were classified as reduced (\( \Delta MR_{\text{glc}} < -20\% \)), increased (\( \Delta MR_{\text{glc}} > +20\% \)) or stable (\( |\Delta MR_{\text{glc}}| \leq +20\% \)). These cut-offs were based on the test-retest reproducibility of \( MR_{\text{glc}} \) which has a 95%-confidence interval of \( \pm 15\%-20\% \) \([44,150]\).

\( OS \) and \( PFS \) served as the standard of reference. The \( OS \) and \( PFS \) with respect to the different \( \Delta MR_{\text{glc}} \) were calculated using Kaplan-Meier analysis.

Cox’s proportional hazards model was used to assess the predictive value of response evaluation with \( ^{18}\text{F-FDG PET} \), as expressed in the \( \Delta MR_{\text{glc}} \) between the \( ^{18}\text{F-FDG PET} \) at baseline and at follow-up. As candidate covariates patient age, sex and tumour staging (NSCLC) are used in a forward model based on likelihood ratios \( (p < 0.05 \text{ for covariate entry, } p > 0.1 \text{ for covariate removal}) \). Hazard ratios \( (HR) \) are presented, representing the ratio of odds that a metabolic responder will survive a certain amount of time compared to a metabolic nonresponder. Statistical significance of each model parameter was assessed using Wald’s \( \chi^2 \) test. As a measure of time-to-effect, both median survival times and their ratio (median survival ratio) are presented, assessed using the log rank test. As cut-off median \( \Delta MR_{\text{glc}} \) was chosen to avoid the effect that expected lower values for \( VOI_{\text{fixed volumes}} \)-determined \( \Delta MR_{\text{glc}} \) account for differences. The median survival ratios present the ratio of time that 50% of the metabolic responder will survive compared to the time 50% of metabolic non-responders survive \([270]\). Finally, another cut-off for \( \Delta MR_{\text{glc}} \) was used for which >90% of the patients show one-year \( OS \) to investigate how a prognosis-driven threshold (as contrast to metabolic response-driven threshold) might prove of additional value.

Analysis was performed with the Statistical Package for Social Sciences (SPSS) version 14.0.2 (SPSS Inc., Chicago, IL, USA) and GraphPad Prism version 5.0a (GraphPad Software Inc., La Jolla, CA, USA). Statistical tests were based on a two-sided significance level and the level of significance was set at \( p = 0.05 \) for all tests.

4.3 Results

Quantitative changes in glucose metabolism

Median interval between baseline and follow-up \( ^{18}\text{F-FDG PET} \) was 63 days \( (IQR: 55-68) \) in CRC and 48.5 days \( (IQR: 41-60) \) in NSCLC. Bland-Altman plots \([271]\) of \( \Delta MR_{\text{glc}} \) of both VOI definitions and both cancer types are displayed in figure \([4.2]\) \( (page \[80]\). In both CRC and NSCLC the decline in median of all \( MR_{\text{glc}} \) between first and second \( ^{18}\text{F-FDG PET} \) was statistically significant \( (p < 0.001) \). Median \( \Delta MR_{\text{glc}} \) for \( VOI_{\text{fixed thresholds}} \) was -29.6% vs -51.9% for \( VOI_{\text{fixed volumes}} \) in CRC patients \( (p = 0.01) \). For NSCLC this was -37% vs -50% \( (p = 0.01) \). The difference in \( \Delta MR_{\text{glc}} \) based on both VOI-methods varied from -232% to +107% (CRC) and from -29% to +172% (NSCLC). Outliers were especially seen when VOI-volumes assessed by \( VOI_{\text{fixed thresholds}} \) were small, rendering their (mean) \( MR_{\text{glc}} \) more sensitive.
VOI definitions for Metabolic Response Evaluation

Figure 4.2: Bland-Altman plots for CRC (panel A) and NSCLC (panel B). The means of ∆MR\textsubscript{glc} (horizontal axis) and differences between ∆MR\textsubscript{glc} computed by both VOI-techniques (vertical axis) for each tumour type and for each patient are displayed. The solid line states the mean difference; the dotted lines the mean±2 standard deviations. CRC: colorectal carcinoma; \textsuperscript{18}F-FDG PET: 2-\textsuperscript{(18)}F)fluoro-2-deoxy-d-glucose positron emission tomography; ∆MR\textsubscript{glc}: relative change in glucose metabolic rate between baseline and follow-up \textsuperscript{18}F-FDG PET; NSCLC: non-small cell lung carcinoma; VOI: volume-of-interest.

to noise. According to previously mentioned definitions of response in MR\textsubscript{glc} from VOI\textsubscript{fixed} volumes, 33 of 49 CRC patients showed reduced, seven showed increased and the remaining nine showed stable metabolism. In NSCLC, these numbers were 35, six and nine, respectively. Concordance between previously mentioned ∆MR\textsubscript{glc} categories by either VOI definition was seen in 36 of 49 CRC cases and in 45 of 50 NSCLC cases (table 4.3-4.4, page 80). Correlation between ∆MR\textsubscript{glc} based on both VOI definitions was 0.879 (CRC) and 0.932 (NSCLC) (both \(p<0.001\)).

<table>
<thead>
<tr>
<th>∆MR\textsubscript{glc} of VOI\textsubscript{fixed} thresholds</th>
<th>Decreased</th>
<th>Stable</th>
<th>Increased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆MR\textsubscript{glc} of VOI\textsubscript{fixed} volumes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased</td>
<td>26</td>
<td>7</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Stable</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Increased</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>16</td>
<td>5</td>
<td>49</td>
</tr>
</tbody>
</table>

Table 4.3: Number of colorectal carcinoma patients with decreased (∆MR\textsubscript{glc} < −20%), stable (∆MR\textsubscript{glc} | +20%) or increased (∆MR\textsubscript{glc} > +20%) metabolism based on both VOI definitions. ∆MR\textsubscript{glc}: change in glucose metabolic rate between baseline and follow-up scan; VOI: volume-of-interest.

<table>
<thead>
<tr>
<th>∆MR\textsubscript{glc} of VOI\textsubscript{fixed} thresholds</th>
<th>Decreased</th>
<th>Stable</th>
<th>Increased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆MR\textsubscript{glc} of VOI\textsubscript{fixed} volumes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased</td>
<td>32</td>
<td>3</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Stable</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Increased</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>12</td>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4.4: Number of non-small cell lung carcinoma patients with decreased (∆MR\textsubscript{glc} < −20%), stable (∆MR\textsubscript{glc} | +20%) or increased (∆MR\textsubscript{glc} > +20%) metabolism based on both VOI definitions. For abbreviations, see table 4.3 (page 80).
4.3. Results

Prediction of survival by glucose metabolic rate

Cox proportional hazards model for CRC showed significant predictive ability of OS and PFS only for $\Delta MR_{glc}$ in VOI$_{fixed}$ volumes between baseline and second $^{18}$F-FDG PET. No contributing confounders were found in the candidate covariates (table 4.5, page 81).

The same type of analysis was used for NSCLC and showed significance predictive ability for OS as well as for PFS for both $\Delta MR_{glc}$ calculated in VOI$_{fixed}$ volumes and VOI$_{fixed}$ thresholds (table 4.5, page 81). Tumour staging showed strong predictive value in univariate analysis ($HR = 1.723$ in OS and 1.942 in PFS).

Only tumour stage was found to be a significant confounding covariate in NSCLC in multivariate Cox proportional hazards analysis. After correction for this confounder, only $\Delta MR_{glc}$ calculated from a VOI$_{fixed}$ volumes showed significant predictive value for both OS and PFS. VOI$_{fixed}$ thresholds based $\Delta MR_{glc}$ showed significant prediction of only OS after correction for tumour stage (table 4.6, page 82).

Using medians as cut-offs shows that the median survival ratio for OS in CRC was 1.78 ($HR = 2.02$; $p = 0.03$) in VOI$_{fixed}$ volumes and 1.25 in VOI$_{fixed}$ thresholds ($HR = 1.75$; $p = 0.082$). For PFS the median survival ratio is 1.57 ($HR = 1.24$; $p = 0.294$) and 1.21 ($HR = 1.23$; $p = 0.384$), respectively. In NSCLC the median survival ratio for OS is 2.01 ($HR = 2.17$; $p = 0.012$) in VOI$_{fixed}$ volumes and 2.01 in VOI$_{fixed}$ thresholds ($HR = 2.19$; $p = 0.01$) and for PFS 2.93 ($HR = 2.02$; $p = 0.015$) and 2.94 ($HR = 2.13$; $p = 0.009$).

In CRC, a $\Delta MR_{glc}$ of -50% in VOI$_{fixed}$ volumes and -34% in VOI$_{fixed}$ thresholds separates patients who have 90% one-year OS from those with lower OS rates. In NSCLC, these values are -67% for VOI$_{fixed}$ volumes and -64% for VOI$_{fixed}$ thresholds. In NSCLC, this is highly different from the medians. Using these values for prognosis-driven cut-offs in patients with NSCLC instead of the median value, the median survival ratio

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OS</th>
<th>PFS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$HR$</td>
<td>95% − CI</td>
</tr>
<tr>
<td><strong>CRC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI$_{fixed}$ volumes)</td>
<td>1.026</td>
<td>(1.001-1.051)</td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI$_{fixed}$ thresholds)</td>
<td>1.014</td>
<td>(0.987-1.042)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.959</td>
<td>(0.479-1.923)</td>
</tr>
<tr>
<td>Age</td>
<td>0.976</td>
<td>(0.936-1.016)</td>
</tr>
<tr>
<td><strong>NSCLC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI$_{fixed}$ volumes)</td>
<td>1.069</td>
<td>(1.024-1.116)</td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI$_{fixed}$ thresholds)</td>
<td>1.133</td>
<td>(1.046-1.227)</td>
</tr>
<tr>
<td>Stage</td>
<td>1.723</td>
<td>(1.204-2.467)</td>
</tr>
<tr>
<td>Sex</td>
<td>1.142</td>
<td>(0.557-2.342)</td>
</tr>
<tr>
<td>Age</td>
<td>1.009</td>
<td>(0.975-1.043)</td>
</tr>
</tbody>
</table>

Table 4.5: Results of univariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta MR_{glc}$ between the $^{18}$F-FDG PET at baseline and follow-up. 95% − CI: 95%-confidence interval; CRC: colorectal carcinoma; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; $HR$: hazard ratio; $\Delta MR_{glc}$: relative change in glucose metabolic rate between baseline and follow-up $^{18}$F-FDG PET; NSCLC: non-small cell lung carcinoma; OS: overall survival; PFS: progression-free survival; VOI: volume-of-interest. *Per 10% change. †Wald’s $\chi^2$ test. ‡Significant at the $p = 0.05$ level.
4. VOI definitions for Metabolic Response Evaluation

**4. VOI definitions for Metabolic Response Evaluation**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OS</th>
<th>PFS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSCLC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI fixed volumes)</td>
<td>1.051 (1.004-1.100)</td>
<td>1.073 (1.014-1.135)</td>
</tr>
<tr>
<td>Stage</td>
<td>1.624 (1.129-2.337)</td>
<td>1.769 (1.208-2.593)</td>
</tr>
<tr>
<td>$\Delta MR_{glc}^*$ (VOI fixed thresholds)</td>
<td>1.091 (1.003-1.187)</td>
<td>1.942 (1.327-2.841)</td>
</tr>
<tr>
<td>Stage</td>
<td>1.581 (1.093-2.287)</td>
<td>– – –</td>
</tr>
</tbody>
</table>

Table 4.6: Results of multivariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta MR_{glc}$ between the $^{18}$F-FDG PET at baseline and follow-up for NSCLC. 95% CI: 95%-confidence interval; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; $HR$: hazard ratio; $\Delta MR_{glc}$: relative change in glucose metabolic rate between baseline and follow-up $^{18}$F-FDG PET; NSCLC: non-small cell lung carcinoma; $OS$: overall survival; $PFS$: progression-free survival; VOI: volume-of-interest. *Per 10% change. †Wald’s $\chi^2$ test. ‡Significant at the $p = 0.05$ level.

for $OS$ was 4.81 ($HR$: 5.5; $p < 0.001$) in VOI fixed volumes and 4.51 in VOI fixed thresholds ($HR$: 3.68; $p = 0.008$). These cut-offs separate the 18-20% of best responding patients from the rest. Results are presented in figure 4.3 (page 83).

4.4 Discussion

$^{18}$F-FDG PET is a promising imaging modality for therapy response assessment. The variety of analytical methods is vast, thus multicentre research is hardly feasible. For that, standardisation of acquisition, reconstruction, VOI-determination and normalisation of the standardised uptake values ($SUV$) needs to be accomplished [46,61]. In the present study, two methods for determination of VOIs, used for Patlak $MR_{glc}$-estimations, were compared using survival as primary outcome measure. Only a small number of studies have addressed the effect of VOI definition on $^{18}$F-FDG uptake [42][44][61], mostly using $SUV$ instead of the gold standard $MR_{glc}$, of which is known that high correlations exist for both CRC [22] and NSCLC [21]. To the best of our knowledge, most of the studies were performed on anthropomorphic phantoms and the studies on patients did not use patient survival as the gold standard nor compared the use of a VOI fixed volumes to that of the more commonly used VOI fixed thresholds method.

Experiments with an anthropomorphic thorax phantom using $SUV$s determined in different VOIs, report that ratios of posttreatment and pretreatment $SUV$s, used for response evaluation, were only slightly dependent on VOI definition, noise and image resolution [42]. A false $SUV$-response was observed when only tumour size changed, but lesion activity was constant, which was caused by partial volume effects [42]. An almost linear correlation amongst $SUV$s obtained with different VOI types was found [61]. Not superiority of any of the evaluated VOIs could be concluded, since no comparison to a gold standard was made.

Krak et al. [44] evaluated chemotherapy in 16 breast carcinoma patients by $SUV$ using different VOIs (manual placement, fixed dimensions (15 mm), threshold based (50% and 70%) and maximum voxel value). They found significant lower responses by manually drawn VOIs as compared to the fixed dimensions VOI. The relative changes
4.4. Discussion

Figure 4.3: Kaplan-Meier survival curves for overall survival in CRC patients (top panels) and NSCLC (lower panels) for VOI-fixed volumes (left panels) and VOI-fixed thresholds (right panels). Populations are dichotomised at the $\Delta M_{R_{glc}}$ value for which >90% of the metabolic responders show one-year overall survival. CRC: colorectal carcinoma; $^{18}$F-FDG PET: $^{18}$fluoro-2-deoxy-D-glucose positron emission tomography; HR: hazard ratio (95%-confidence interval); $\Delta M_{R_{glc}}$: relative change in glucose metabolic rate between baseline and follow-up $^{18}$F-FDG PET; NSCLC: non-small cell lung carcinoma.

<table>
<thead>
<tr>
<th></th>
<th>0yr</th>
<th>1yr</th>
<th>2yr</th>
<th>3yr</th>
<th>4yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-responders</td>
<td>24</td>
<td>16</td>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Responders</td>
<td>25</td>
<td>24</td>
<td>13</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Log rank: p=0.030; HR: 0.48 (0.25-0.93); Median Ratio: 1.6 (1.1-2.1)

<table>
<thead>
<tr>
<th></th>
<th>0yr</th>
<th>1yr</th>
<th>2yr</th>
<th>3yr</th>
<th>4yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-responders</td>
<td>28</td>
<td>20</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Responders</td>
<td>21</td>
<td>20</td>
<td>11</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Log rank: p=0.038; HR: 0.51 (0.27-0.96); Median Ratio: 1.4 (0.89-2.0)

<table>
<thead>
<tr>
<th></th>
<th>0yr</th>
<th>1yr</th>
<th>2yr</th>
<th>3yr</th>
<th>4yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-responders</td>
<td>40</td>
<td>21</td>
<td>11</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Responders</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Log rank: p<0.001; HR: 0.29 (0.15-0.55); Median Ratio: 4.8 (4.4-5.2)

<table>
<thead>
<tr>
<th></th>
<th>0yr</th>
<th>1yr</th>
<th>2yr</th>
<th>3yr</th>
<th>4yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-responders</td>
<td>41</td>
<td>22</td>
<td>13</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Responders</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Log rank: p=0.012; HR: 0.42 (0.22-0.83); Median Ratio: 4.5 (4.1-4.9)

We are not the first to use change in tumour volume combined with changes in metabolic activity in response evaluation. Guillem et al. [227] use a technique previously described by Larson et al. [145] to calculate the change in total lesion glyco-

measured by threshold based or maximum $SUV$ were similar. They concluded that VOI definition has a clear effect on measured change in tumour metabolism and, therefore, that $SUV$s obtained from different VOIs cannot be compared to each other. They, too, were not able to conclude superiority of any of the evaluated VOIs since they did not compare them to a gold standard or outcome measure.
4. VOI definitions for Metabolic Response Evaluation

lysis (δTLG) by multiplication of the $^{18}$F-FDG activity concentration ($SUV$) to the metabolic tumour volume in 15 patients with primary rectal carcinoma treated with preoperative chemoradiotherapy, identifying 100% of responses. They found complete concordance between pathology and the δTLG parameter in six cases, in four the response was overestimated and in five it was underestimated, which was slightly better than usage of $SUV_{max}$ or $SUV_{mean}$ alone. More recently, Benz et al. [146] investigated 20 patients with locally advanced high-grade soft tissue sarcoma that underwent neoadjuvant treatment and compared $SUV_{max}$, $SUV_{mean}$, $TLG_{max}$ and $TLG_{mean}$ (using volumes determined on CT) for response evaluation. They conclude that TLG was less accurate in predicting tumour response than were measurements of the intratumoural $^{18}$F-FDG concentration ($SUV$). This technique has been evaluated for treatment of primary rectal carcinoma [227, 229, 235], NSCLC [272], malignant mesothelioma [273], melanoma [274] and breast carcinoma [275]. The TLG has the disadvantage that it is highly dependent on the threshold level used. When lesion response to treatment causes its metabolic rate to be in the same range as the background, large-volume VOIs will be derived, which will artefactually lead to underestimation of treatment effect. The VOI fixed volumes technique has the advantage that in substantial metabolic response the remaining $MR_{glc}$ can be calculated.

In this study, in CRC and NSCLC, both $MR_{glc}$ estimations showed a significant decrease between baseline and follow-up $^{18}$F-FDG PET as a sign of therapy response. Significant differences in the degree of relative $MR_{glc}$ decrease were seen, which was expected since VOI fixed volumes determination also takes into account tissues with <50% of the maximum activity in the lesion on follow-up $^{18}$F-FDG PET. Thus, the volume-based average will be lower than the value determined by using a threshold for VOI definition.

A significant relation between $\Delta MR_{glc}$, VOI fixed volumes and both $OS$ and $PFS$ was shown, which was still present after correcting for known confounders. $\Delta MR_{glc}$, VOI fixed thresholds did not show significant prediction of $PFS$ in CRC and of in NSCLC after multivariate correction. Other cut-off values may be necessary when performing therapy response evaluation by VOI fixed volumes measurements, since $MR_{glc}$, VOI fixed volumes showed higher decreases than $MR_{glc}$, VOI fixed thresholds as explained before. Using the median to dichotomise $\Delta MR_{glc}$ showed higher separation of median survival times between VOI fixed volumes and VOI fixed thresholds in CRC and highly similar separation in NSCLC.

Threshold methods are relatively simple and user-independent and recovered counts are relatively independent of lesion size and of changes in geometry. Fixed dimension (15 mm) methods are relatively simple, semiautomatic and presumably less sensitive to partial volume effects when tumour size changes during therapy. Our presented VOI fixed volumes method benefits from the advantages of both and also takes into account treatment induced changes in tumour size.
4.5 Conclusions

Metabolic response has a high predictive value for treatment outcome in CRC as well as NSCLC patients. For determination of the VOI in therapy response evaluation studies the VOI\textsubscript{fixed volumes} method proved to be superior to the VOI\textsubscript{fixed thresholds} method. The use of one standardised method for VOI definition is of the utmost importance in future multicentre trials, in order to avoid institution-based variations in evaluation of metabolic response.

***
A Curve-Fitting Approach to Estimate the Arterial Plasma Input Function for the Assessment of Glucose Metabolic Rate and Response to Treatment

*J Nucl Med, 2009 Dec;50(12):1933-9*
(adapted and reprinted with permission)

D. Vriens, L.F. de Geus-Oei, W.J.G. Oyen and E.P. Visser

1 dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands
Abstract

**Purpose** For the quantification of 4D dynamic 2-\(^{18}\text{F}\)fluoro-2-deoxy-D-glucose positron emission tomography (\(^{18}\text{F}\)-FDG PET) studies, the arterial plasma time-activity concentration curve (APTAC) needs to be available. This can be obtained using serial sampling of arterial blood or an image-derived input function (IDIF). Arterial sampling is invasive and often not feasible in practice; IDIFs are biased because of partial volume effects and cannot be used when no large arterial blood pool is in the field-of-view. We propose a mathematical equation to describe the APTAC. The proposed function was compared with serial arterial sampling and the IDIF.

**Methods** To determine the free parameters of the function, plasma time-activity concentration curves based on arterial samples in 80 patients were fitted after normalisation for administered activity (\(AA\)) and initial distribution volume (\(iDV\)) of \(^{18}\text{F}\)-FDG. The medians of these free parameters were used for the model. In 40 other patients (20 baseline and 20 follow-up dynamic \(^{18}\text{F}\)-FDG PET scans), this model was validated. The population-based curve, individually calibrated by \(AA\) and \(iDV\) (APTAC\(_{AA/iDV}\)), by a single late arterial sample (APTAC\(_{1\text{sample}}\)) and by the individual IDIF (APTAC\(_{IDIF}\)), was compared with the gold standard of serial arterial sampling (APTAC\(_{sampled}\)) using the area under the curve (\(AUC\)). Additionally, these three methods of APTAC determination were evaluated with Patlak glucose metabolic rate (\(MR_{glc}\)) estimation and with \(\Delta MR_{glc}\) for therapy effects.

**Results** Excellent individual fits to the function were derived with significantly different decay constants (\(p < 0.001\)). Correlations between \(AUC\) from APTAC\(_{AA/iDV}\), APTAC\(_{1\text{sample}}\) and APTAC\(_{IDIF}\) with APTAC\(_{sampled}\) were 0.880, 0.994 and 0.856, respectively. For \(MR_{glc}\), these correlations were 0.963, 0.994 and 0.966, respectively. In therapy response evaluation, these correlations were 0.947, 0.982 and 0.949, respectively. Additional scaling by a single late arterial sample showed a significant improvement (\(p < 0.001\)).

**Conclusions** The fitted input function calibrated for \(AA\) and \(iDV\) performed similarly to IDIF. Performance improved significantly using a single late arterial sample. The proposed model can be used when an IDIF is not available or serial arterial sampling is not feasible.

**Keywords** Chemotherapy · 2-\(^{18}\text{F}\)fluoro-2-deoxy-D-glucose · Input Function · Least-Squares Analysis · Pharmacokinetics · Positron Emission Tomography · Theoretical Models · Therapy Response Evaluation · Tissue Distribution
5.1 Introduction

Pharmacokinetic analysis of $^{18}$F-FDG by 4D dynamic PET requires both the arterial plasma time-activity concentration curve (APTAC, $C_{\text{plasma}}(t)$) or input function and the tissue time-activity concentration curve measured by PET (TTAC or $C_{\text{tissue}}(t)$) to be available to create a Gjedde-Patlak plot $C_{\text{PET}}(t) \cdot C_{\text{plasma}}(t)^{-1}$ vs $\int_0^t C_{\text{plasma}}(\tau) \, d\tau \cdot C_{\text{plasma}}(t)^{-1}$, from which the glucose metabolic rate ($MR_{\text{glc}}$) in the tissue of interest can be derived. Special interest in this type of quantitative analysis has risen for oncologic patients for prognostic stratification and response evaluation of disease.

The gold standard to obtain the APTAC is by measuring decay-corrected activity concentrations in plasma obtained by serial arterial sampling [135]. This is an invasive and potentially harmful procedure for the patient and exposes the personnel to radiation. Complications attributed to radial artery cannulation include temporary occlusion (19.7%) and hematoma formation (14.4%), local infection (0.72%), sepsis (0.13%), permanent occlusion (0.09%) and pseudoaneurysm (0.09%) [136], some of which require medical intervention. Therefore, different methods have been developed to reduce these drawbacks. The major four alternatives to serial arterial sampling are (arterialised) venous blood sampling [137,276], image-derived input function (IDIF) estimation [83,103,138,277–280], population-based input function modelling [139–141,281,282] and whole-blood input function extraction using sophisticated mathematical image segmentation methods (e.g., cluster analysis [144] and independent component analysis [142,283,284]).

A drawback of venous sampling is the time-dependent ratio of $^{18}$F-FDG in venous to arterial blood (a ratio of 0.61-0.88 at 5-50 min after injection) [137,276]. Therefore, shunting of arterial blood to the venous system (the heated hand procedure) is used to improve these ratios to 0.92-1.05 [137]. This arterialised venous sampling technique still requires cannulation of an extra vein and exposes personnel to radiation.

IDIFs require a large blood pool (aorta, left ventricle) within the field-of-view of the PET image. IDIFs show systematic error, because activity concentration is measured in whole-blood and activity concentration is known to be lower in whole-blood than in plasma (ratio $\sim$0.925-0.95) [103,138,277]. Moreover, partial volume effects cause inaccuracy by spillover of activity from or to the surrounding tissues (e.g., myocardium) [83,103]. Finally, additional noise is introduced because of the limited number of counts in the short early time frames. Because underestimation due to spill-out of the IDIF activity concentration leads to overestimation of $MR_{\text{glc}}$ and overestimation due to spill-in of surrounding IDIF activity concentration (in later time frames) leads to underestimation of $MR_{\text{glc}}$ [83,103], correction is necessary [278,280].

A population-based APTAC is based on the averaging of normalised sampled blood data of multiple patients. Several corrections were introduced, based on administered activity (AA), bodyweight and blood transit time [281] or AA and body surface area [141,282]. The latter provided a reliable estimation of $MR_{\text{glc}}$ but yielded less accurate parameter values in pharmacokinetic analysis. Another method is the fitting of the sampled data to an equation, assuming the $^{18}$F-FDG distribution in the vascular system as a compartment model on its own [128,139,140].
Here, we describe an APTAC model that is based on the fitting of a large series of arterially sampled oncologic patients to a mathematical equation (16,17,22). The model APTAC was calibrated to individual patients of a separate patient population by either an estimation of the distribution volume of $^{18}$F-FDG or a single late arterial blood sample and compared with the gold standard of arterial sampling and the IDIF. The performance of both models was assessed by comparison of the APTACs themselves and by their influence on Patlak $MR_{glc}$ and therapy response ($\Delta MR_{glc}$) evaluation.

5.2 Material & methods

Patient population, arterial sampling procedure

Data of 120 dynamic $^{18}$F-FDG PET scans with serial arterial sampling data were reanalysed. Scans were randomly distributed over two groups (table 5.1, page 90): a parameter-identification group ($n = 80$) and a parameter-validation group ($n = 40$). For the latter group, both a pretreatment and a follow-up scan, after 2-3 courses of chemotherapy, were included. The details of $^{18}$F-FDG PET and arterial plasma data acquisition are described elsewhere [83], with the only difference being that they were reconstructed using ordered-subsets expectation maximisation (OSEM) with 4 iterations and 16 subsets with a Gaussian filter of 5 mm full-width at half-maximum (FWHM) in all directions. In short, normoglycemic patients who had fasted were injected with $^{18}$F-FDG by an automated standardised infusion protocol. Directly thereafter, 17 arterial blood samples were taken at set time points from which plasma was obtained by centrifugation to provide a sampled arterial plasma time-activity concentration curve (APTAC$_{sampled}$). Simultaneously, 4D dynamic PET, consisting of 16 time frames of variable length, was performed to provide both the tissue (TTAC) and the image-derived blood time-activity concentration curves (APTAC$_{IDIF}$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Identification</th>
<th>Validation</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{18}$F-FDG PET scans (individual patients)</td>
<td>80 (49)</td>
<td>40 (20)</td>
<td>0.327*</td>
</tr>
<tr>
<td>Male sex [% of scans]</td>
<td>60</td>
<td>70</td>
<td>0.320*</td>
</tr>
<tr>
<td>Median age (range) [y]</td>
<td>60.7 (44.8-78.9)</td>
<td>60.1 (44.7-71.9)</td>
<td>0.225†</td>
</tr>
<tr>
<td>Mean AA ± SD [MBq]</td>
<td>207 ± 40.1</td>
<td>200 ± 35.8</td>
<td>0.314‡</td>
</tr>
<tr>
<td>Median plasma glucose concentration (range) [mmol·l$^{-1}$]</td>
<td>5.3 (4.2-10.0)</td>
<td>5.3 (4.2-8.3)</td>
<td>0.125†</td>
</tr>
<tr>
<td>Median BMI (range) [kg·m$^{-2}$]</td>
<td>25.8 (22.7-27.4)</td>
<td>25.9 (19.8-33.7)</td>
<td>0.508†</td>
</tr>
<tr>
<td>Cancer localisation [% of scans]</td>
<td></td>
<td></td>
<td>0.618*</td>
</tr>
<tr>
<td>Non-small cell lung carcinoma</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Colorectal carcinoma</td>
<td>45</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Comparison between parameter-identification and parameter-validation groups, with statistical significance of group difference. AA: administered activity; BMI: body mass index; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; SD: standard deviation. *$\chi^2$ test. †Mann-Whitney U test. ‡Independent-samples t test.
Parameter-identification study \((n = 80)\)

**Normalisation of the \(\text{APTAC}_{\text{sampled}}\)**

The initial plasma activity concentration of \(^{18}\text{F-FDG}\) \((C^*_\text{plasma}(0), \text{[MBq·l}^{-1}]\)) was used to normalise the \(\text{APTAC}_{\text{sampled}}\). It was defined as the expected \(^{18}\text{F-FDG}\) concentration directly after tracer injection, assuming instant homogenisation and is dependent on the AA [MBq] and the (apparent) initial distribution volume \((iDV [L])\) [141,282]. To avoid confusion, an “∗” is added, because the true sampled activity concentration at \(t = 0\) is 0 MBq·l\(^{-1}\).

In the period between 5 and 30 min after injection, the plasma and extravascular extracellular \(^{18}\text{F-FDG}\) pool of the whole-body were assumed to be in equilibrium [130]. Before this period, the tracer is being distributed over the body and in the period thereafter, the tracer is mainly being metabolised and excreted. Within this interval, we sampled four times (7.5, 12.5, 17.5 and 25 min after injection) [83] to obtain \(C^*_\text{plasma}(0)\); semilogarithmic recordings of these four points were linearly extrapolated back to \(t = 0\) (y-intercept) [130,141].

**Estimation of \(iDV\) by bodyweight and height**

By definition, the \(iDV\) represents the (virtual) volume of the plasma and extravascular extracellular \(^{18}\text{F-FDG}\) pool. The \(iDV\) can be estimated by [141,282]:

\[
iDV = \frac{\text{AA}}{C^*_\text{plasma}(0)} = c \cdot H^h \cdot W^w
\]

(5.1)

where \(H\) is patient height [m] and \(W\) is patient bodyweight [kg]. Iteratively \(h\), \(w\) and \(c\) were derived for which the coefficient of variation of \(c\) \((CV_c = \frac{\text{SD}_c}{\text{mean}_c})\) is smallest in the parameter-identification data.

**Fitting of normalised \(\text{APTAC}_{\text{sampled}}\)**

The normalised \(\text{APTAC}_{\text{sampled}}\) can be approximated using the three-compartment model for the blood pool as proposed by Feng *et al.* [140], simplified by Eberl *et al.* [139], as the following:

\[
\frac{\text{APTAC}_{\text{sampled}}(t)}{C^*_\text{plasma}(0)} = \begin{cases} 
0 & : t < -\frac{b}{a} \\
a \cdot t + b & : -\frac{b}{a} \leq t < \tau \\
\sum_{i=1}^{3} A_i \cdot e^{-\lambda_i \cdot (t-\tau)} & : t \geq \tau 
\end{cases}
\]

(5.2)

where \(\tau\) is the time to peak activity concentration. The normalised sampled plasma curves were fitted by linear curve fitting \((t < \tau)\) and by nonlinear least squares \((t \geq \tau)\) to obtain the eight free parameters in every patient of the parameter-identification study. Because chemotherapy might influence \(^{18}\text{F-FDG}\) distribution and clearance, the parameter values were compared between scans made in patients who did and did not receive chemotherapy.
Parameter-validation study ($n = 40$)

**IDIF & tumour time-activity concentration curves**

The APTAC\textsubscript{IDIF} was determined in manually placed volumes-of-interest (VOIs) over the ascending aorta (thoracic images) or descending aorta (abdominal images) known to correlate best with the gold standard \cite{83} on summed images of the period 30-90 s after injection. The TTAC was obtained semi-automatically by placing volumes-of-interest in the summed images of the period 20-50 min after injection over the largest lesion, fully present in the field-of-view, using a threshold of 50\% of its maximum voxel value. All images were analysed using the Inveon Research Workplace (IRW) version 2.2 (Siemens Healthcare, Erlangen, Germany).

**Comparison of calibrated plasma time-activity concentration curves & glucose metabolic rates**

Median values of APTAC parameters derived from the parameter-identification study were used for the population-based APTAC, which was calibrated to each individual patient, using two methods: by either multiplication by $AA$ [MBq] divided by $iDV$ [l] (estimated using \textit{equation 5.1} (page 91), further mentioned as APTAC_{AA/iDV}) or multiplication by the plasma activity concentration of a single late arterial sample (APTAC_{1sample}).

The performance of the three curves (APTAC_{AA/iDV}, APTAC_{1sample} and APTAC_{IDIF}) was compared with that of the gold standard (APTAC_{sampled}) by their area under the curve (AUC), determined by trapezoid integration.

$MR_{glc}$ was determined using all four plasma input curves by Gjedde-Patlak graphical analysis \cite{33}. In the Patlak approximation, the lumped constant, accounting for the difference in glucose and $^{18}$F-FDG affinity, was set to 1 and the tumour blood volume fraction was set to 0 cm$^3$·cm$^{-3}$. Therefore, the slope of the Gjedde-Patlak plot was $K_i$ and the intercept was $\frac{K_1 k_2}{(k_2 + k_3)^2}$. $MR_{glc}$ can then be determined by multiplying the linear regression (5-50 min after injection) by the plasma glucose concentration.

**Statistical analysis**

All parameters were assessed for normality (Shapiro-Wilk, skewness and kurtosis) and are either displayed as mean ($\pm$SD) or median (interquartile range, IQR). Comparison between two independent groups was performed by the $t$ test, the Mann-Whitney $U$ test or the $\chi^2$ test. $\lambda_1$-$\lambda_3$ were compared using Friedman ANOVA. Correlations between the AUC, $MR_{glc}$ and $\Delta MR_{glc}$ of the different curves compared with the gold standard were assessed by Pearson $R$ or Spearman $\rho$ and linear regression and are displayed as Bland-Altman plots. Confidence intervals for correlation coefficients were compared after Fisher $z$ transformation \cite{285}. Analysis was performed by the Statistical Package for Social Sciences (SPSS) version 16.0.2 (SPSS Inc., Chicago, IL, USA). Two-sided significance was set at the 0.050 level.
5.3 Results

Parameter-identification study

A median $iDV$ of 12.7 l (corresponding to 0.1683 l·kg$^{-1}$ bodyweight) was calculated. Iterative determination of the function to estimate the $iDV$ led to $h = 1.257$, $w = 0.582$ and $c = 0.533$ l·m$^{-1.257}$·kg$^{-0.582}$ (minimum CV$_c = 0.171$, adjusted $R^2 = 0.3181$).

Fitting of the 80 APTAC$_{sampled}$ by equation 5.3 (page 91) led to a minimum adjusted $R^2$ of 0.8533; 80% of fits had an adjusted $R^2 > 0.9500$. Resulting parameters for the normalised APTAC$_{sampled}$ are provided in table 5.2 (page 93). No significant difference in $^{18}$F-FDG clearance could be detected ($p > 0.326$) between patients who did and did not receive chemotherapy. The three decay constants ($\lambda_1$-$\lambda_3$) were significantly different from each other ($p < 0.001$; figure 5.1, page 94).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All scans</th>
<th>Pretherapy scans</th>
<th>Follow-up scans</th>
<th>$p^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$ [min]</td>
<td>1.25 (1.25-1.50)</td>
<td>1.25 (1.25-1.25)</td>
<td>1.25 (1.25-1.50)</td>
<td>0.077</td>
</tr>
<tr>
<td>$a$ [min$^{-1}$]</td>
<td>4.911 (4.188-5.709)</td>
<td>4.914 (4.154-5.660)</td>
<td>4.911 (4.203-5.771)</td>
<td>0.672</td>
</tr>
<tr>
<td>$b$</td>
<td>-2.077 (-2.817-1.321)</td>
<td>-1.195 (-2.805-1.096)</td>
<td>-2.225 (-2.961-1.503)</td>
<td>0.168</td>
</tr>
<tr>
<td>$A_1$</td>
<td>2.785 (2.302-3.212)</td>
<td>2.718 (2.241-3.152)</td>
<td>2.937 (2.321-3.300)</td>
<td>0.326</td>
</tr>
<tr>
<td>$\lambda_1$ [min$^{-1}$]</td>
<td>6.133 (4.646-8.114)</td>
<td>6.034 (4.121-8.064)</td>
<td>6.133 (4.833-8.322)</td>
<td>0.622</td>
</tr>
<tr>
<td>$A_2$</td>
<td>0.7085 (0.6492-0.8011)</td>
<td>0.6966 (0.6340-0.8434)</td>
<td>0.7229 (0.6694-0.8130)</td>
<td>0.589</td>
</tr>
<tr>
<td>$\lambda_2$ [min$^{-1}$]</td>
<td>0.2541 (0.2075-0.3071)</td>
<td>0.2400 (0.1881-0.3423)</td>
<td>0.2592 (0.2285-0.2847)</td>
<td>0.805</td>
</tr>
<tr>
<td>$A_3$</td>
<td>0.7721 (0.7177-0.8070)</td>
<td>0.7699 (0.6861-0.8286)</td>
<td>0.7741 (0.7385-0.7970)</td>
<td>0.920</td>
</tr>
<tr>
<td>$\lambda_3$ [min$^{-1}$]</td>
<td>0.01443 (0.01248-0.01595)</td>
<td>0.01446 (0.01206-0.01656)</td>
<td>0.01434 (0.01260-0.01590)</td>
<td>0.562</td>
</tr>
</tbody>
</table>

Table 5.2: Results of fitting of normalised plasma data of parameter-identification group and subgroup analysis between pretherapy and follow-up $^{18}$F-FDG PET scans. There is no significant difference in parameters between APTAC$_{sampled}$ of chemotherapy-naive patients and patients that previously underwent chemotherapy. Median values are presented with IQR between parentheses. APTAC$_{sampled}$: arterial plasma time activity concentration curve derived by sampling of the radial artery; $^{18}$F-FDG PET: 2-((18$^F$)fluoro-2-deoxy-d-glucose positron emission tomography; IQR: interquartile range. $^*$Mann-Whitney U test.

Parameter-validation study

The correlation between the $AUC$ of the APTAC$_{AA/iDV}$ and the APTAC$_{sampled}$ was 0.880. This correlation improved to 0.994 ($p < 0.001$) using a single late arterial sample for calibration. The mean $AUC$ of APTAC$_{AA/iDV}$ (462 ± 119 MBq·l$^{-1}$·min) was similar to and that of APTAC$_{sampled}$ (468 ± 107 MBq·l$^{-1}$·min) slightly lower than the gold standard (APTAC$_{sampled}$, 475 ± 108 MBq·l$^{-1}$·min; $p = 0.156$ and $p < 0.001$, respectively). The APTAC$_{IDIF}$ showed much lower $AUC$s than APTAC$_{sampled}$ (392 ± 99 MBq·l$^{-1}$·min; $p < 0.001$), with a correlation coefficient between these parameters of 0.856.
Comparison between $MR_{glc}$ is shown in figure 5.2 (page 95, left panels). Adding a single late arterial sample (25 min after injection) improved the correlation between $MR_{glc}$, determined by APTAC$_{AA/iDV}$ and APTAC$_{sampled}$, from $\rho = 0.963$ to $\rho = 0.994$ ($p < 0.001$). $MR_{glc}$ determined by IDIF was significantly higher than the gold standard; the correlation was similar to that of APTAC$_{AA/iDV}$ ($\rho = 0.966$).

Comparison between therapy effects ($\Delta MR_{glc}$) is shown in figure 5.2 (page 95, right panels). Adding a single late arterial sample (25 min after injection) improved the correlation between $\Delta MR_{glc}$ determined by APTAC$_{AA/iDV}$ and APTAC$_{sampled}$ from $\rho = 0.947$ to $\rho = 0.982$ ($p = 0.012$). $\Delta MR_{glc}$ determined by APTAC$_{IDIF}$ had a correlation to the gold standard similar to that determined by APTAC$_{AA/iDV}$ ($\rho = 0.949$).

Any arterial sample taken at 7.5 min after injection showed similar high correlations when comparing $MR_{glc}$ ($\rho \geq 0.9916$) or $\Delta MR_{glc}$ ($\rho \geq 0.9759$) measured by APTAC$_{sampled}$ and APTAC$_{1sample}$. 
5.3. Results

Figure 5.2: Bland-Altman plots of comparison between any of three evaluated APTACs and the gold standard. In Bland-Altman plots, mean difference is displayed by solid line and the 95%-confidence interval by dotted lines. Unstandardised regression coefficients are displayed with corresponding 95%-confidence interval. Comparisons are either given for $MR_{glc}$ (left panels) or $\Delta MR_{glc}$ (right panels). From top to bottom panels: APTAC$_{AA/iDV}$, APTAC$_{1sample}$ and APTAC$_{IDIF}$ compared to the gold standard of APTAC$_{sampled}$. 1sample: calibrated using a single late arterial sample at 25 min after injection (APTAC$_{1sample}$); AA: administered activity; AA/iDV: calibrated by AA and iDV (APTAC$_{AA/iDV}$); APTAC: arterial plasma time activity concentration curve; IDIF: based on the image-derived input function (APTAC$_{IDIF}$); iDV: initial distribution volume; $\Delta MR_{glc}$: (change in) glucose metabolic rate.
5.4 Discussion

This study describes the performance of a population-based APTAC based on fitting patient data to a mathematical equation on Patlak determination of $MR_{gIc}$ and on chemotherapy response evaluation. Its accuracy is comparable to that of an IDIF, explaining 93% of variance in $MR_{gIc}$ (90% in response evaluation by $\Delta MR_{gIc}$), but the addition of a single late arterial sample improves this to 99% (96% in response evaluation by $\Delta MR_{gIc}$). Where the IDIF is fully individual, the proposed method is a population average, calibrated to individual patient parameters without the need for serial arterial sampling. The method can be used when scanning body regions where blood pools are unavailable, such as the extremities. The model can be further improved by including a single late arterial sample. It can be expected that a late arterialized venous sample might adequately replace the arterial sample, because more than approximately 30 min after $^{18}$F-FDG injection, activity concentrations of arterial and arterialized venous blood are highly similar\cite{137,276}. This sample might also be used to obtain an accurate measurement of the plasma glucose level, as long as radiation safety regulations do not prohibit this analysis in the clinical chemistry laboratory.

As shown, $MR_{gIc}$ was significantly overestimated using an IDIF, compared with the gold standard. Both the partial volume effect and the fact that whole-blood activity concentrations are lower than in plasma cause underestimation of the (integral of the) APTAC. From the Patlak equation (equation 1.10 page 22) it can be derived that an underestimation of the APTAC and its integral cause overestimation of the $MR_{gIc}$. The quality of the IDIF can be improved by correction for the partial volume and spillover effects\cite{286}. Moreover, the activity in whole-blood can be corrected to the activity concentration in plasma using hematocrit\cite{142} or modelled erythrocyte uptake\cite{287}. In addition, the noise in the short early time frames contributes to the inaccuracy of the activity concentration. Recently, eight methods for the estimation of the carotid IDIF in human brain studies were compared with the reference input function (arterially sampled) with respect to cerebral $MR_{gIc}$ and individual rate constants in a study consisting of phantoms and healthy volunteers. The authors concluded that blood-sample-free methods provided less reliable results than those obtained using the methods that require blood samples, even when limited to a single sample\cite{288}. Therefore, calibration of our IDIFs by a single late venous sample would probably have improved the accuracy of these IDIFs. The advantage of a more patient-specific rather than a generic input function is especially important for pharmacokinetic analysis, because the exact shape of the APTAC in the artery feeding the tumour is vital and is usually different from a remote artery (because of delay and dispersion by the intra-individual variation in the impulse-response characteristics of the vascular system).

One method of calibration of the population-based curve was by the approach of Sadato et al.\cite{130} ($AA/iDV$). They reported a mean $iDV$ scaled to bodyweight ($iDV_{BW}$) of 0.1627 l·kg$^{-1}$, similar to this study (median $iDV_{BW}$ of 0.1683 l·kg$^{-1}$). Around 50-60% of the bodyweight of an average adult is water, of which 25-45% l·kg$^{-1}$ is extracellular fluid (the volume of water in which $^{18}$F-FDG dissolves), stressing both the concordance and plausibility of this result\cite{289}. We used the function of Shiozaki et al.\cite{141} to fit the $iDV$. They reported an $iDV$ [ml] that equalled
5.4. Discussion

Table 5.3: Literature review of publications on population-based APTACs and IDIFs. AA/iDV: administered activity/initial distribution volume of \(^{18}\)F-FDG; AUC: correlation or mean % difference between AUC (area under the curve) of population-based curve and serial arterial sampling; APTAC: arterial plasma time activity concentration curve; AVS: arterialised venous samples; \(^{18}\)F-FDG: 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose; IDIF: image-derived input function; \(MR_{glc}\): correlation or mean % difference between glucose metabolic rate of population-based curve and serial arterial sampling; \(n\): no. of patients (to derive curve/to validate curve).

<table>
<thead>
<tr>
<th>Reference</th>
<th>(n)</th>
<th>Method</th>
<th>Calibration</th>
<th>AUC</th>
<th>(MR_{glc})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takikawa et al.</td>
<td>10/24</td>
<td>Arithmetic mean</td>
<td>2 arterial</td>
<td>0.998</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 AVS</td>
<td>0.998</td>
<td>0.989</td>
</tr>
<tr>
<td>Tsuchida et al.</td>
<td>44/10</td>
<td>Arithmetic mean</td>
<td>–</td>
<td>+3.5%</td>
<td>± +2.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 AVS</td>
<td>2.2%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Shiozaki et al.</td>
<td>101/192</td>
<td>Arithmetic mean</td>
<td>AA/iDV</td>
<td>+7.2%</td>
<td>± +8.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.7%</td>
<td>7.3%</td>
<td></td>
</tr>
<tr>
<td>Bentourkia et al.</td>
<td>20/same20</td>
<td>Fit to equation</td>
<td>–</td>
<td>–</td>
<td>0.998</td>
</tr>
<tr>
<td>Eberl et al.</td>
<td>139</td>
<td>Fit to equation</td>
<td>2 AVS</td>
<td>–</td>
<td>0.995</td>
</tr>
<tr>
<td>de Geus-Oei et al.</td>
<td>83</td>
<td>IDIF</td>
<td>–</td>
<td>–</td>
<td>0.98</td>
</tr>
<tr>
<td>This study</td>
<td>80/40</td>
<td>Fit to equation</td>
<td>AA/iDV</td>
<td>0.880</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 arterial</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDIF</td>
<td>0.856</td>
<td>0.966</td>
</tr>
</tbody>
</table>

39.0 \times H[^{cm}]^{0.80} \times W[^{kg}]^{0.35}. This result appears different from the present study \((iDV[^{[l]}]= 0.533 \times H[^{[m]}]^{1.257} \times W[^{[kg]}]^{0.582})\), but they were only slightly off our optimum \((CV_c = 0.18 \text{ vs } CV_c = 0.17)\). A dissimilar distribution of body habitus in their (Asian) population might be the cause for this difference.

As shown in figure 5.2 (page 93), the difference between \(MR_{glc}\) calculated by APTAC\(_{\text{sample}}\) and APTAC\(_{\text{sampled}}\) was small; therefore, the net influx constant determined by Patlak analysis was accurate. For determining the individual rate constants, however, the inaccuracy will probably be much higher. For Patlak analysis, only the integral of the APTAC and the late (>5 min) activity concentrations of the arterial plasma are relevant, but for the accurate estimation of the individual rate constants of glucose pharmacokinetics, the exact shape of the APTAC is vital, because of both the function (including a convolution operation) and the method for parameter estimation (nonlinear least squares). The delay and dispersion between the APTAC in the artery feeding the tissue of interest and a remote artery are of major importance for the accurate determination of the fast rate constants \((K_1\text{ and } k_2)\) and blood volume fraction \((V_B)\). Moreover, the use of a generic input function for pharmacokinetic analysis should be considered because of the large intra-individual variation of the slow clearance rate constants of \(^{18}\)F-FDG \((\lambda_3)\) and to a lesser extent in \(\lambda_2\) in figure 5.1 (page 94) and table 5.2 (page 93). The assumption of a single clearance profile for all patients might therefore lead to the inaccurate estimation of the slower rate constants \((k_3\text{ and, if it exists, } k_4)\) of individual patients. The proposed model, therefore, should be evaluated further for estimating microparameters before implementation. This further evaluation, however, was out of the scope of this study.

None of the previously published articles on population-based input curves used an IDIF for comparison and most studies consisted of a small series of non-oncologic patients, but reproducibility was high (table 5.3 page 97). The studies that used triexponential clearance of \(^{18}\)F-FDG as a function for the APTAC \([128,139,140]\) found a decay similar to that found in this study (figure 5.3 page 98). The results shown
5. Population-Based Input Function

Figure 5.3: Comparison of triexponential $^{18}$F-FDG clearance in our study and three previous publications [128,139,140]. Functions were normalised to total AUC. AUC: area under the curve; Dif.: absolute differences of our clearance curve compared with three found in literature.

In figure 5.1 (page 94) suggest a high (linear) correlation between $A_3$ and $\lambda_3$. This correlation suggests that the model might be simplified by omitting one parameter; we did not verify the possibility of simplifying the model in this study.

5.5 Conclusions

The current model of the APTAC of $^{18}$F-FDG, calibrated by a single late arterial sample, shows high accuracy for Patlak $MR_{gic}$ calculation and therapy response evaluation. The model allows dynamic scanning of areas without large vessels in the field-of-view, such as the extremities; is less invasive; and causes less radiation exposure to personnel than does arterial sampling. Even without a calibrating sample available, this model has the same accuracy as an IDIF, without its inherent disadvantages.

***
Part III

Tissue Heterogeneity & Therapy Response Evaluation
Quantitative Assessment of Heterogeneity in Tumour Metabolism using $^{18}$F-FDG PET

*Int J Radiat Oncol Biol Phys, 2012 Apr;82(5):e725-31*

(adapted and reprinted with permission)

D. Vriens$^1$*,
J.A. Disselhorst$^1$*,
W.J.G. Oyen$^1$,
L.F. de Geus-Oei$^1$ and
E.P. Visser$^1$

$^1$dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands

*Both authors contributed equally to this article*
Abstract

Purpose 2-(18F)fluoro-2-deoxy-d-glucose positron emission tomography (18F-FDG PET) images are usually quantitatively analysed in “whole-tumour” volumes-of-interest. Also parameters determined with 4D dynamic PET acquisitions, such as the Patlak glucose metabolic rate ($MR_{glc}$) and pharmacokinetic rate constants of two-tissue compartment modelling, are most often derived per lesion. We propose segmentation of tumours to determine tumour heterogeneity, potentially useful for dose-painting in radiotherapy and elucidating mechanisms of 18F-FDG uptake.

Methods In 41 patients with 104 lesions, 4D dynamic 18F-FDG PET was performed. On parametric $MR_{glc}$ images, tumours were segmented in quartiles of background subtracted maximum $MR_{glc}$ (0-25%, 25-50%, 50-75% and 75-100%). Pharmacokinetic analysis was performed using an irreversible two-tissue compartment model in the three segments with highest $MR_{glc}$ to determine the rate constants of 18F-FDG metabolism.

Results From the highest to the lowest quartile, significant decreases of uptake ($K_1$), washout ($k_2$) and phosphorylation ($k_3$) rate constants were seen with significant increases in tissue blood volume fraction ($V_B$).

Conclusions Tumour regions with highest $MR_{glc}$ are characterised by high cellular uptake and phosphorylation rate constants with relatively low blood volume fractions. In regions with less metabolic activity, the blood volume fraction increases and cellular uptake, washout and phosphorylation rate constants decrease. These results support the hypothesis that regional tumour glucose phosphorylation rate is not dependent on the transport of nutrients (i.e., 18F-FDG) to the tumour.

Keywords 2-(18F)fluoro-2-deoxy-d-glucose · Heterogeneity · Patlak Analysis · Pharmacokinetics · Positron Emission Tomography · Tissue Distribution
6.1 Introduction

Positron emission tomography (PET) is a molecular imaging technique to quantitatively assess various tissue properties with an appropriate radiotracer such as $^{18}$F-fluoromisonidazole for hypoxia [291] or $^{18}$F-fluorothymidine for proliferation [292]. However, 2-($^{18}$F)fluoro-2-deoxy-D-glucose ($^{18}$F-FDG), which visualises glucose metabolic processes in tissues, remains the most commonly used radiotracer with PET. $^{18}$F-FDG PET has obtained a clear role in tumour staging [27] and is used to assess therapy response and predict outcome [293]. In most cases the standardised uptake value ($SUV$) is used to quantify whole-tumour $^{18}$F-FDG uptake. Most tumours, however, display heterogeneous uptake in $^{18}$F-FDG PET images, which may reflect different biologic behaviour in the different regions within these lesions. Identifying this metabolic heterogeneity would not only be useful for understanding tumour biology, but might be of prognostic significance or of use for radiotherapy planning and dose-painting by intensity-modulated radiation therapy (IMRT) as well [294].

To examine processes underlying the uptake of $^{18}$F-FDG in a tumour, sophisticated quantification methods such as determination of pharmacokinetic rate constants of two-tissue compartment models can be used [34,35], requiring 4D dynamic PET images. With these pharmacokinetic analyses, the properties of the $^{18}$F-FDG metabolism can be elucidated. Delivery of $^{18}$F-FDG into and out of the tumour cells corresponds to the rate constants $K_1$ and $k_2$, respectively. Once intracellular, phosphorylation of $^{18}$F-FDG is represented by $k_3$, dephosphorylation by $k_4$ and finally $V_B$ indicates the fraction of blood within the volume-of-interest (VOI). The two-tissue compartment model of $^{18}$F-FDG metabolism is shown in figure 1.2 (page 19). Pharmacokinetic rate constants and other model-based parameters are usually derived in a VOI, mostly representing the whole tumour. These parameters therefore correspond to averaged values of tumour glucose metabolic activity. To assess intratumoural heterogeneity, the parameters could be obtained for every individual voxel within the tumour (voxel-wise modelling). The time-activity concentration curves for single voxels, however, are relatively noisy. This prevents accurate determination of tumour parameters using nonlinear least squares (NLLS) methods [295]. Moreover, tumour movement (e.g., as a result of breathing) has a relatively large influence in these small volumes because in this case each voxel does not necessarily represent a single volume of tissue.

As a trade-off, less detailed parameters (such as the metabolic rate of glucose: $MR_{glt}$) can be used, for which voxel-wise quantification is feasible because they can be based on linearisation methods that are far less sensitive to noise (e.g., using the Patlak method [33]). Another method to overcome high noise levels is to decrease spatial image resolution by reconstruction at smaller matrix sizes. Postreconstruction smoothing could also be used to increase signal-to-noise ratios at the cost of spatial resolution.

Similarly, higher signal-to-noise ratios can be obtained by tumour segmentation and taking the mean value in each segment, which could be used to determine tumour regional variation in two-tissue compartment model rate constants of $^{18}$F-FDG metabolism ($K_1-k_4$) and blood volume fraction ($V_B$).
Heterogeneity in Tumour Glucose Metabolism

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (range) [y]</td>
<td>60.8 (44.7-77.7)</td>
</tr>
<tr>
<td>Proportion male [%]</td>
<td>68.3</td>
</tr>
<tr>
<td>Median plasma glucose level (range) [mmol·l⁻¹]</td>
<td>5.2 (4.2-8.3)</td>
</tr>
<tr>
<td>Median AA per unit body mass (range) [MBq·kg⁻¹]</td>
<td>2.7 (1.6-5.9)</td>
</tr>
</tbody>
</table>

**Origin of primary tumour [% of patients] / [% of lesions]**
- Non-small cell lung carcinoma: 71 / 82
- Colorectal carcinoma: 24 / 16
- Breast carcinoma: 5 / 2

**Proposed treatment [%]**
- Induction chemotherapy: 22
- First line curative chemotherapy: 76
- Palliative care: 2

**Tumour differentiation [%]**
- Very poor: 20
- Poor: 15
- Intermediate: 27
- Mucinous: 2
- Unspecified: 37

Table 6.1: Patient characteristics (n = 41). AA: administered activity of 2-(¹⁸F)fluoro-2-deoxy-D-glucose.

We applied the latter method and segmented the tumours based on the $MR_{glc}$ images, taking advantage of the higher tumour-background ratio as compared with standard uptake images. Furthermore, we wanted to verify our hypothesis that tumour regions with the highest $MR_{glc}$ are characterised by high $¹⁸F$-FDG extraction and phosphorylation rates.

### 6.2 Material & methods

#### Patient population & data acquisition procedure

From an existing database of 4D dynamic $¹⁸F$-FDG PET scans of oncological patients acquired on an ECAT EXACT47 (Siemens Healthcare, Erlangen, Germany), 41 patients with 104 tumours of different origin (table 6.1, page 106) were selected. Selection criteria were previously untreated patients in whom a fully arterially sampled input function is available. The study was approved by the Institutional Review Board of the Radboud University Medical Centre, Nijmegen, the Netherlands and written informed consent was obtained from each patient. Details of the study have been described previously, with the only difference being that the present images were reconstructed in $128 \times 128 \times 47$ matrices using 2D ordered subsets expectation maximisation (OSEM) with 4 iterations and 16 subsets and a 5-mm full-width at half-maximum (FWHM) 3D Gaussian filter. In brief, fasted patients were injected with $¹⁸F$-FDG by an automated standardised infusion protocol. Immediately thereafter, 17 arterial blood samples were taken at set time points from which plasma was obtained to provide a sampled arterial plasma time-activity concentration curve ($C_{\text{plasma}}(t)$). Simultaneously, a 4D dynamic PET acquisition consisting of 16 time frames of parameter duration was obtained to provide the tissue time-activity concentration curve. Voxel volumes in the reconstructed images were 39.75 mm$^3$ (transaxial voxel size 3.432-3.432 mm, axial voxel size 3.375 mm).
6.2. Material & methods

Data analysis

Data analysis was performed using Inveon Research Workplace (IRW) version 2.2 (Siemens Healthcare, Erlangen, Germany). First, voxel-wise Patlak linear regression of acquired data 10-50 min postinjection (the last five time frames) was used to obtain the influx constant ($K_i$) of each voxel according to equation 1.10 (page 22). The slope of the Gjedde-Patlak plot equals $K_i$ and $MR_{glc}$ was calculated by multiplication with the measured venous plasma glucose concentration, thereby implicitly assuming a lumped constant ($LC_{FDG}$) of 1 (equation 1.9, page 21). The real value of $LC_{FDG}$ is time- and tissue-dependent and therefore unknown [113].

Because pharmacokinetic analysis of dynamic data is sensitive to time-delay ($\tau$; i.e., the difference in time of arrival of the $^{18}$F-FDG bolus in the tumour and in the sampled artery), manual time-offset synchronisation of $C_{plasma}(t)$ and $C_{tumour}(t)$ of VOI$_{tumour}$ was performed by visually shifting the ascending limbs of both curves until they overlapped.

On images of $MR_{glc}$, one representative background VOI (VOI$_{BG}$) and three tumour VOIs were determined per lesion. The tumours were segmented in quartiles of background-corrected $MR_{glc}$ and the upper three quartiles were analysed (VOI$_{low}$, VOI$_{medium}$ and VOI$_{high}$, summing to VOI$_{tumour}$). The quartile with lowest metabolic rate was not included because the edges of the metabolic volume may contain nontumour tissues and the noise levels in this segment is higher because of a lower $^{18}$F-FDG uptake. We did not attempt any partial volume correction strategies. VOI$_{BG}$ was defined as an ellipsoidal volume of at least the size of the corresponding tumour and placed in a representative volume of tissue with normal $^{18}$F-FDG uptake (e.g., contralateral lung in case of a lung metastasis).

IRW uses the iterative Levenberg-Marquardt algorithm for NLLS curve fitting, minimising the weighted least squares deviation to obtain all rate constants and $V_B$ simultaneously in each tumour VOI separately. The weighting function was chosen as the square root of the frame duration divided by the measured nondecay-corrected activity concentration in each time frame (i.e., Poisson weighting). For initialisation of the NLLS algorithm, multiple (=99), randomly selected starting parameters within a defined range (0.0-2.0 for the rate constants, 0.0-1.0 for $V_B$) were used.

Data from the literature suggest that the rate of dephosphorylation of $^{18}$F-FDG by glucose-6-phosphatase activity ($k_4$ in the Phelps 4K model [35]) is very low in mammalian tissues, except in liver tissue [115]. Moreover, from simulation studies it has been warned that a $k_4$ might result from tissue heterogeneity rather than real dephosphorylation [119]. All pharmacokinetic analyses were performed using the Sokoloff 3K ($K_1-k_3$) model [34], as it showed a better fit than the Phelps 4K ($K_1-k_4$) model in a majority of lesions (data on file).

The resulting values for each parameter were classified as being biologically plausible or not. Values within three standard deviations (99.7%-confidence interval) of the mean described elsewhere [298,299] were considered biologically plausible by definition. This criterion was set to label unrealistic tumour parameter values, which could be the result of the instability of the nonlinear optimisation of extremely noisy time-activity concentration curves. For $K_1$ this meant the interval 0-1.05 ml·g$^{-1}$·min$^{-1}$, for $k_2$ 0-1.455 min$^{-1}$, for $k_3$ 0-0.511 min$^{-1}$ and for $V_B$ 0.001-0.335 cm$^3$·cm$^{-3}$. The subgroup of lesions considered having biologically plausible results was analysed separately, to
validate that the conclusions drawn based on the whole group were not due to (implausible) outliers. Next, to evaluate the influence of blurring caused by the partial volume effect and motion, the subgroup of lesions with a volume of at least 14.1 cm$^3$ were analysed because this volume represents a lesion with a spherical diameter of $\sim$3 cm, which is five times the resolution of the scanner (i.e., $\sim$0.6 cm full-width at half-maximum).

Finally, these parameters were interpreted in view of their reliability (relative standard error) and mutual independence (correlation matrix).

### Statistical analysis

(Log$_e$)-normally distributed parameters are described by mean and 95%-confidence interval (95% – CI: mean ± 1.96-standard deviation). Parameters not obeying the (log$_e$)-normal distribution are described by median and interquartile range (IQR). For normally distributed values, the (paired) $t$ test and the squared Pearson’s product-moment correlation coefficient ($R^2$) were used to compare means or express correlation. In case of non-normality, Spearman’s $\rho$ was used. Correlation was qualified based on $R^2$: very high ($R^2 \geq 0.9$), high ($0.7 \leq R^2 < 0.9$), intermediate ($0.5 \leq R^2 < 0.7$) or low ($R^2 < 0.5$).

Comparison of multiple groups was performed by non-parametric analysis of variance (Friedman’s analysis of variance). To correct for multiple comparisons, a posthoc analysis with Wilcoxon signed-rank test was conducted with a Dunn’s (or Bonferroni’s) correction.

Relations between tumour parameters on the one hand and biological plausibility on the other hand were determined using uniparameter logistic regression. The following continuous parameters were candidate predictors: volume, $K_i$, $MR_{glc}$ and $SUV$ of the smallest VOI (segment). Their discriminative ability was evaluated by determination of the area under the receiver operating characteristics curves (AUC). Difference between the AUC and 0.500 (i.e., AUC for an indiscriminative test) was assessed by asymptotic distribution assumption for significance. All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 16.0.2 (SPSS Inc., Chicago, IL, USA). The cut-off point for statistical significance was set at 0.05.

### 6.3 Results

Median lesion volume based on VOI$_{tumour}$ was 11.3 cm$^3$ (corresponding to 283.5 voxels), ranging from 1.7 to 504 cm$^3$ (42 to 12,668 voxels). VOI characteristics of tumour segments are displayed in table 6.2 (page 109).

Correlation of the VOI$_{tumour}$ parameters showed significant but weak correlation between $MR_{glc}$ and $K_1$ ($\rho = 0.286$; $p = 0.003$), $MR_{glc}$ and $k_3$ ($\rho = 0.488$; $p < 0.001$) but not between $MR_{glc}$ and $k_2$ ($\rho = -0.088$; $p = 0.372$) and $MR_{glc}$ and $V_B$ ($\rho = 0.189$; $p = 0.055$). There was also no significant correlation between $k_3$ and $V_B$ ($\rho = 0.045$; $p = 0.652$). The subgroups of non-small cell lung carcinoma and colorectal carcinoma showed similar results, as summarised in table 6.3 (page 109).
6.3. Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>VOI_tumour</th>
<th>VOI_high</th>
<th>VOI_medium</th>
<th>VOI_low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median volume* [cm³] (range)</td>
<td>11.3 (1.67-504)</td>
<td>0.556 (0.0795-8.11)</td>
<td>1.83 (0.278-92.2)</td>
<td>7.85 (1.07-406)</td>
</tr>
<tr>
<td>Mean $MR_{glc}$ [mmol.ml⁻¹.min⁻¹] (95% – CI)</td>
<td>97.7 (36.4-274)</td>
<td>187 (69.8-503)</td>
<td>135 (50.3-361)</td>
<td>82.7 (30.9-221)</td>
</tr>
<tr>
<td>Median $SUV$ [g.cm⁻³] (IQR)</td>
<td>4.1 (2.8-5.5)</td>
<td>6.8 (4.8-8.7)</td>
<td>5.3 (3.7-6.8)</td>
<td>3.6 (2.5-4.8)</td>
</tr>
<tr>
<td>Median $K_1$ [ml.g⁻¹.min⁻¹] (IQR)</td>
<td>0.11 (0.077-0.16)</td>
<td>0.18 (0.081-0.37)</td>
<td>0.15 (0.092-0.23)</td>
<td>0.11 (0.072-0.15)</td>
</tr>
<tr>
<td>Median $k_2$ [min⁻¹] (IQR)</td>
<td>0.44 (0.25-0.89)</td>
<td>0.99 (0.31-2.9)</td>
<td>0.55 (0.29-1.1)</td>
<td>0.40 (0.22-0.75)</td>
</tr>
<tr>
<td>Median $k_3$ [min⁻¹] (IQR)</td>
<td>0.089 (0.058-0.12)</td>
<td>0.19 (0.090-0.39)</td>
<td>0.12 (0.070-0.17)</td>
<td>0.073 (0.051-0.099)</td>
</tr>
<tr>
<td>Median $V_B$ [cm³.cm⁻³] (IQR)</td>
<td>0.089 (0.057-0.14)</td>
<td>0.036 (0.00063-0.086)</td>
<td>0.071 (0.037-0.11)</td>
<td>0.095 (0.056-0.15)</td>
</tr>
</tbody>
</table>

Table 6.2: VOI characteristics. 95% – CI: 95%-confidence interval; IQR: interquartile range; $K_1$-$k_3$: rate constants of the two-compartment model of glucose metabolism; $MR_{glc}$: mean glucose metabolic rate computed assuming a lumped constant of 1 and a $V_B$ of 0 cm³.cm⁻³; $SUV$: standardised uptake value normalised for bodyweight, measured 40-50 min postinjection; $V_B$: blood volume fraction; VOI: volume-of-interest. *One voxel equals a volume of 39.75 mm³. †Calculated after loge-transformation.

<table>
<thead>
<tr>
<th>Histology</th>
<th>$MR_{glc}$ vs $K_1$</th>
<th>$MR_{glc}$ vs $k_2$</th>
<th>$MR_{glc}$ vs $k_3$</th>
<th>$MR_{glc}$ vs $V_B$</th>
<th>$k_3$ vs $V_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n = 104)</td>
<td>0.286 (0.003)*</td>
<td>-0.088 (0.372)</td>
<td>0.488 (&lt;0.001)*</td>
<td>0.189 (0.055)</td>
<td>0.045 (0.652)</td>
</tr>
<tr>
<td>NSCLC (n = 85)</td>
<td>0.255 (0.018)*</td>
<td>-0.114 (0.298)</td>
<td>0.507 (&lt;0.001)*</td>
<td>0.141 (0.198)</td>
<td>0.102 (0.352)</td>
</tr>
<tr>
<td>CRC (n = 17)</td>
<td>0.235 (0.363)</td>
<td>-0.039 (0.881)</td>
<td>0.363 (0.152)</td>
<td>0.365 (0.149)</td>
<td>-0.159 (0.541)</td>
</tr>
</tbody>
</table>

Table 6.3: Correlation of VOI_tumour parameters. Correlations are obtained on a whole-lesion basis and are expressed as Spearman’s $\rho$ with statistical significance between parentheses. CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma; VOI: volume-of-interest. *Statistically significant.

An example of lesion segmentation with corresponding time-activity concentration curves is displayed in figure 6.1 (page 110). As demonstrated in figure 6.2 (page 111), a clear trend in reduction of $K_1$, $k_2$, and increase in $V_B$ was observed from VOI_high toward VOI_low, which was significant for all four investigated parameters. The within-group differences were significant between all segments, except for VOI_high vs VOI_medium for both $K_1$ and $k_2$. These trends were seen in significantly more than the expected 25% (if divided equally between the four possibilities) of the analysed lesions: 42% (95% – CI: 33-52%) for $K_1$, 40% (95% – CI: 31-50%) for $k_2$, 79% (95% – CI: 70-86%) for $k_3$ and 93% (95% – CI: 87-97%) for $V_B$. Similar trends were observed when non-small cell lung carcinoma and colorectal carcinoma lesions were considered separately. Forty lesions had a volume larger than 14.1 cm³. In this subgroup, the trends were observed in a higher percentage of the lesions, as displayed in table 6.4 (page 110).

In 90% of the lesions, all four VOI_tumour parameters were considered biologically plausible. In the ten lesions with one or more implausible parameters, $K_1$ was not within the defined plausible range in 20%, $k_2$ in 70%, $k_3$ in 40% and $V_B$ in 50% of these lesions. Multiparameter logistic regression showed only $K_1$ to be independently related to plausibility.
6. Heterogeneity in Tumour Glucose Metabolism

Figure 6.1: Representative example of a patient with a cT3N2M0 non-small cell lung carcinoma. The lesion’s largest diameter was 58 mm. (panel A) A surface rendered image of the volumes-of-interest (VOIs) clearly demonstrating the heterogeneity in 2-(18F)fluoro-2-deoxy-d-glucose metabolism. (panel B) The same lesion “opened up”. (panel C) Graph showing the time activity concentration curves in the four tumour VOIs (Ctumour of VOIlow, VOImedium, VOIhigh and VOItumour, left axis) and in the sampled arterial plasma (Cplasma, right axis). In panel A and panel B the edge of VOIhigh is indicated with an arrowhead and the edge of VOImedium with an arrow.

When looking at the four parameters in the segments VOIhigh, VOImedium and VOIlow separately in 45 lesions (43% of the analysed tumours) all 12 values were within the range of biologic plausibility. In the other 59 lesions (57%) with at least

<table>
<thead>
<tr>
<th>Histology</th>
<th>K1</th>
<th>k2</th>
<th>k3</th>
<th>Vb</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n = 104)</td>
<td>28.1 (&lt;=0.001)*</td>
<td>26.8 (&lt;=0.001)*</td>
<td>120.5 (&lt;=0.001)*</td>
<td>71.6 (&lt;=0.001)*</td>
</tr>
<tr>
<td>NSCLC (n = 85)</td>
<td>26.6 (&lt;=0.001)*</td>
<td>16.3 (&lt;=0.001)*</td>
<td>99.7 (&lt;=0.001)*</td>
<td>54.5 (&lt;=0.001)*</td>
</tr>
<tr>
<td>CRC (n = 17)</td>
<td>2.5 (0.312)</td>
<td>12.1 (0.002)*</td>
<td>17.3 (&lt;=0.001)*</td>
<td>22.2 (&lt;=0.001)*</td>
</tr>
<tr>
<td>Plausible* (n = 45)</td>
<td>9.0 (0.011)*</td>
<td>2.8 (0.247)</td>
<td>37.9 (&lt;=0.001)*</td>
<td>14.4 (0.001)*</td>
</tr>
<tr>
<td>Volume&gt;14.1 cm³ (n = 40)</td>
<td>25.6 (&lt;=0.001)*</td>
<td>43.6 (&lt;=0.001)*</td>
<td>61.4 (&lt;=0.001)*</td>
<td>52.9 (&lt;=0.001)*</td>
</tr>
</tbody>
</table>

Table 6.4: Comparisons between tumour segments. Comparison between tumour segments is expressed as Friedman’s χ² with statistical significance between parentheses. A trend between the segments is expressed as a percentage of lesions (between brackets), for K1-k3 this trend is VOIhigh>VOImedium>VOIlow and for Vb this trend is VOIhigh<VOImedium<VOIlow. CRC: colorectal carcinoma; K1-k3: rate constants of tumour 2-(18F)fluoro-2-deoxy-d-glucose metabolism; NSCLC: non-small cell lung carcinoma; Vb: tumour blood volume fraction. *Statistically significant. †Lesions with all 12 parameters within the defined range of biological plausibility (see text).
6.3. Results

Figure 6.2: Differences in rate constants and blood volume fraction in different tumour VOIs.
Friedman: Friedman’s nonparametric analysis of variance overall test; $K_1$-$k_3$: rate constants of tumour $2-\text{[18F]}\text{fluoro-2-deoxy-d-glucose metabolism}$; NS: not significant; $V_B$: tumour blood volume fraction; VOI: volume-of-interest. *, ** Significant with Wilcoxon signed-rank test posthoc analysis with Dunn’s correction applied at $^*p < 0.01$ or at $^{**}p < 0.001$.

one or more resulting implausible parameters in one or more VOIs, $K_1$ was not within the plausible range in 15%, $k_2$ in 81%, $k_3$ in 34% and $V_B$ in 61% of these lesion. Only the smallest $K_i$ was related to biologically plausible results, which had moderate discriminative ability ($AUC = 0.680$, $95\% - CI$: 0.578-0.783).

When looking only at the 45 lesions with all 12 values within the defined range of biological plausibility, we saw similar trends between VOI$_{\text{high}}$ and VOI$_{\text{low}}$ for $K_1$, $k_3$ and $V_B$. However this did not reach statistical significance for $k_2$ (Friedman=2.80; $p = 0.257$, downslope trend for VOI$_{\text{high}}$-VOI$_{\text{low}}$ in 33%; table 6.4, page 110).

For the 312 NLLS operations (three tumour-segment VOIs in each of the 104 lesions), 10% of $K_1$, 10% of $k_2$, 12% of $k_3$ and 20% of $V_B$ showed relative standard errors larger than 5%. In the 104 (larger VOI) whole-tumour parameters this was only the case in 4%, 4%, 5% and 5%, respectively. Correlation matrices of the 312
6. HETEROGENEITY IN TUMOUR GLUCOSE METABOLISM

parameters, showed very high mutual dependence between $K_1$ and $k_2$ (median $R = 0.96$, IQR: 0.96-0.97), intermediate mutual dependence between $K_1$ and $V_B$ (median $R = -0.73$, IQR: -0.83- -0.68) and low mutual dependence between $K_1$ and $k_3$ (median $R = 0.06$, IQR: -0.38-0.43), between $k_2$ and $k_3$ (median $R = 0.36$, IQR: -0.13-0.68), between $k_2$ and $V_B$ (median $R = -0.60$, IQR: -0.70- -0.54) and between $k_3$ and $V_B$ (median $R = 0.34$, IQR: 0.03-0.66).

6.4 Discussion

When comparing tumour regions with stepwise decreasing $MR_{glc}$, there is decreasing phosphorylation rate ($k_3$), but increasing fractional blood volume ($V_B$). This cannot be explained by mutual dependence of these parameters in the fitting process, because the fit-correlation was low. We could not find a significant correlation between whole-tumour $V_B$ and $k_3$ between lesions. Glucose phosphorylation rate therefore is not correlated with lesion blood volume fraction. Miles and Williams [300] have reviewed the relation between tumour vascularisation and tumour metabolism. Although the total uptake of $^{18}$F-FDG was considered, rather than $k_3$, they showed that literature on this subject is highly parameter. It was reported that the relationship appeared to depend on tumour type, tumour size and tumour grade. In the present study, no significant correlations between $V_B$ and $k_3$ could be found, not even in the subgroups per tumour type or the subgroup of large tumours (table 6.3, page 109). The weak correlations between $MR_{glc}$ and $k_3$ (significant) and $MR_{glc}$ and $V_B$ (not significant) might indicate that using $MR_{glc}$ as tumour segmentation parameter is responsible for the within-patient inverse relation between the $k_3$ and $V_B$ trends over the segments, but the absence of between-lesion correlation between $V_B$ and $k_3$ suggests another underlying mechanism. Multiple groups reported a correlation between $^{18}$F-FDG uptake during the first few minutes after injection and tumour blood flow measured by $H_2( ^{15}O)$ in various tumour types [301,302]. Because a large $V_B$ is mainly responsible for the tumour $^{18}$F-FDG concentration in the (very) early time frames, at which blood $^{18}$F-FDG activity concentrations are still very high, our results suggest that $V_B$ (hence regional tumour perfusion) is inversely related with phosphorylation rate ($k_3$). Our results suggest that the “metabolic centre” of the tumour can maintain high phosphorylation rates ($k_3$) at relatively low blood volume fractions ($V_B$). This interesting result may support the hypothesis that as the tumour grows, angiogenesis becomes unable to maintain an adequate blood supply, contributing to the uncoupling of glucose metabolism and blood flow, a theory also known as the Warburg effect [10]. It should be noted that $V_B$ as obtained with pharmacokinetic modelling may not reflect the complete blood volume fraction in the tumour. The venous blood time-activity concentration curve may have a shape very different from the arterial blood time-activity concentration curve as used for the compartment analysis. This could result in an underestimation of $V_B$.

Tumour hypoxia is disadvantageous property for several cancer treatments, including radiation therapy [303]. Numerous methods to detect or visualise hypoxia in vivo are available that include hypoxia-specific PET tracers such as ($^{18}$F)fluoromisonidazole. Some suggest employing $^{18}$F-FDG PET for dose-painting because of the disadvantages of hypoxia tracers (e.g., a low tumour-background ra-
6.5 Conclusions

Regions of tumours with highest $MR_{gly}$ are characterised by high cellular uptake and phosphorylation rate constants with relatively low blood volume fractions. In regions with less metabolic activity, the blood volume fraction gradually increases and cellular uptake, washout and phosphorylation rate constants decrease. These results are not due to covariance of the regression coefficients and might be relevant for understanding tumour biology and for dose-painting in radiotherapy.

***
Chemotherapy Response Evaluation of Colorectal Liver Metastases by DCE-MRI Perfusion Parameters and 4D Dynamic $^{18}$F-FDG PET Metabolic Rate

*J Nucl Med, 2009 Nov;50(11):1777-84*
(adapted and reprinted with permission)

D. Vriens$^1$, H.W.M. van Laarhoven$^2$, J.J.A. van Asten$^3$, P.F.M. Krabbe$^4$, E.P. Visser$^1$, A. Heerschap$^3$, C.J.A. Punt$^2$, L.F. de Geus-Oei$^1$ and W.J.G. Oyen$^1$

dep. of $^1$Nuclear Medicine, $^2$Medical Oncology, $^3$Radiology and $^4$Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands
Abstract

Purpose In this study, we examined the in vivo relationship between functional tumour vasculature, determined by dynamic contrast-enhanced (DCE-)MRI and tumour metabolism, determined by 4D dynamic $^{18}$F-FDG PET, during cytotoxic treatment of patients with colorectal liver metastases.

Methods Twenty-three patients underwent DCE-MRI (using gadopentetate dimeglumine) and 4D dynamic $^{18}$F-FDG PET at baseline and after three treatment cycles, unless treatment was terminated because of toxicity. Parameters for vasculature (rate constant between extravascular extracellular space and blood plasma ($k_{ep}$) and volume transfer constant ($K_{trans}$)), extracellular space ($v_e$), tumour size (the maximum axial diameter of each included lesion $MAD$) and metabolism (glucose metabolic rates ($MR_{glc}$)) were derived and changes during treatment were correlated. Overall survival ($OS$) and progression-free survival ($PFS$) served as outcome measures for the predictive abilities of pretreatment parameters and of treatment-related parameter changes.

Results Pretreatment $MR_{glc}$ and $MAD$ were individually predictive for $OS$ and $PFS$. During treatment, $K_{trans}$ increased significantly, but this increase could not be confirmed in a lesion-by-lesion analysis. $MR_{glc}$ decreased significantly ($p < 0.001$). No correlations were found for changes in DCE-MRI parameters and $\Delta MR_{glc}$. No relationship was found between changes in DCE-MRI parameters and $OS$ or $PFS$. $\Delta MR_{glc}$ was able to predict $OS$ ($p = 0.008$) after correction for confounders.

Conclusions The efficacy of cytotoxic chemotherapy assessed by reduction in tumour metabolism does not depend on pretreatment properties of the tumour vasculature determined by DCE-MRI. Cytotoxic chemotherapy does not alter DCE-MRI-derived properties of tumour vasculature but decreases glucose consumption of tumour cells.

Keywords Chemotherapy · Colorectal Carcinoma · Dynamic Contrast-Enhanced MRI · 2-(18F)fluoro-2-deoxy-d-glucose · Liver Metastases · Positron Emission Tomography · Survival · Therapy Response Evaluation
7.1 Introduction

Chemotherapy is usually the treatment of choice in patients with advanced colorectal carcinoma (CRC). Advances in cytotoxic treatment have improved the median survival from eight months to more than 20 months [153]. Unfortunately, chemotherapy is only effective in a subset of patients. Early response prediction would enable individualised treatment and prevent side effects and costs due to futile treatment of nonresponders.

The efficacy of chemotherapy depends on the delivery of cytotoxic drugs by the tumour vasculature, uptake and retention of the drug in tumour cells, metabolic activation of prodrugs, intrinsic chemosensitivity of tumour cells and catabolism and excretion of drugs. Capillary perfusion and permeability of the vessel wall can be measured in vivo by 4D dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) using gadopentetate dimeglumine (Gd-DTPA) as the contrast agent [306]. Glucose metabolic activity can be assessed in vivo by positron emission tomography (PET) using the radionuclide-labelled glucose analogue 2-\(^{18}\)F)fluoro-2-deoxy-d-glucose (\(^{18}\)F-FDG) [33].

Pharmacokinetic analysis of DCE-MRI data yields parameters for perfused capillaries including blood flow, permeability and the total surface area. Assuming that vascularity, as reflected by these parameters, is also relevant for the delivery of relatively small cytotoxics to the tumour, it may be hypothesised that these parameters can be used for the prediction of therapy response. The value of baseline DCE-MRI parameters to predict treatment outcome has been shown for several tumour types including CRC [307,308] and its significance in the early evaluation of therapy response has been shown for rectal [308] and breast [309,311] carcinoma.

\(^{18}\)F-FDG uptake is increased in malignant tumours. An adequate vascular supply and the presence of several membrane-bound glucose transport proteins are necessary for the delivery of glucose to tumour cells and intracellular hexokinase is necessary for subsequent phosphorylation. \(^{18}\)F-FDG uptake can be quantified by glucose metabolic rates (\(MR_{glc}\)) derived from dynamic \(^{18}\)F-FDG PET data. The value of baseline \(MR_{glc}\) as a predictive parameter for treatment outcome has been shown for several tumour types, including non-small cell lung carcinoma (NSCLC) and CRC [15] and its predictive value using changes in \(MR_{glc}\) for the early evaluation of response during chemotherapy has been shown in many tumour types, including NSCLC [21] and CRC [22].

Because DCE-MRI and \(^{18}\)F-FDG PET assess two different determinants of chemotherapy efficacy, their combination could aid in the unravelling of the principles of chemosensitivity. In this prospective study, we investigated the predictive value of pretreatment pharmacokinetic parameters and the value of (early) cytotoxic therapy-induced changes in pharmacokinetic parameters of Gd-DTPA and \(^{18}\)F-FDG with respect to overall survival (OS) and progression-free survival (PFS).
7.2 Material & methods

Patients

Patients with liver metastases of histologically proven CRC who underwent a diagnostic work-up before the start of cytotoxic chemotherapy between June 2002 and September 2005 were eligible. Liver metastases were established during routine staging or during follow-up by abdominal CT (n = 21), ultrasonography (n = 1) or MRI (n = 1). Follow-up included a three monthly ultrasonography or CT scan. Patients with diabetes mellitus, severe claustrophobia or implanted electrical devices and with (multiple) small (<1 cm) lesions, in whom the limited spatial resolution of the PET scanner would pose technical difficulties for quantification, were excluded. The study was approved by the institutional review board of the Radboud University Medical Centre, Nijmegen, the Netherlands and all patients gave written informed consent.

Thirty-three patients with liver metastases of CRC were included in this prospective study. For five patients, DCE-MRI data were not complete because of technical problems (n = 2) or the unavailability of a pretherapy scan (n = 1) or a follow-up scan (n = 2); for four patients, 18F-FDG PET data were not complete because of technical problems during the follow-up dynamic acquisition (n = 1) or the unavailability of a follow-up scan (n = 3). For one patient, both DCE-MRI and 18F-FDG PET at follow-up were inaccessible. Therefore, complete datasets of two DCE-MRI and two 18F-FDG PET scans were available for 23 patients for analysis of therapy response. Patient characteristics are listed in table 7.1 (page 119). Thirty-one lesions in total could be matched for lesion-by-lesion analysis on MRI (34 lesions visible) and 18F-FDG PET (56 lesions visible because of the larger axial field-of-view (FoV) of PET than MRI) by two experienced observers. In four patients, metastases visible on DCE-MRI could not be identified as separate lesions on 18F-FDG PET. Therefore, the MRglc of the combined lesion was used.

Nine patients received chemotherapy in the first, eleven in the second, two in the third and one in fourth line. The median interval between the last day of the previous-line chemotherapy and the baseline scan for patients treated in the second or higher line was 39 days (interquartile range (IQR): 23-150 days). None of these 14 patients had been treated with anti-angiogenic agents before inclusion in the study. The treatment regimens were irinotecan (n = 7), capecitabine (n = 5), capecitabine/irinotecan (n = 4), capecitabine/oxaliplatin (n = 3), 5-folinic acid(FA)/fluorouracil (5-FU)/oxaliplatin (FOLFOX; n = 3) and 5-FU/FA (n = 1).

No patients were lost to follow-up. Median OS was 1.5 y (1-, 2- and 3-y proportions, 70%, 26% and 9%, respectively) and median PFS was 5.3 months. At the close-out date, two patients were alive (follow-up: 4.3 and 5.6 years, respectively). All patients showed progression of disease.

DCE-MRI and 18F-FDG PET were performed before the start and after three cycles of chemotherapy (cycle duration, 21 days, except for FOLFOX, which lasted 14 days). When treatment was terminated before three cycles (n = 6; range: 0.5-2.0 cycles), the follow-up scan was performed earlier. The median duration of treatment was 50 days (IQR: 30-56 days). The median interval between the last treatment day and the follow-up scan was six days (IQR: 2.5-15.5 days).
7.2. Material & methods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>23</td>
</tr>
</tbody>
</table>

**Demographics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (range) [y]</td>
<td>61.5 (44.8-78.9)</td>
</tr>
<tr>
<td>Number of men [%]</td>
<td>17 (74%)</td>
</tr>
</tbody>
</table>

**Matched lesions per patient**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (range)</td>
<td>1.3 (1-5)</td>
</tr>
<tr>
<td>Patients &gt;1 matched lesion [%]</td>
<td>5 (22%)</td>
</tr>
<tr>
<td>Median lesion MAD (IQR) [mm]</td>
<td>56 (38-75)</td>
</tr>
</tbody>
</table>

**Histology [%]**

<table>
<thead>
<tr>
<th>Histology</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>22 (96%)</td>
</tr>
<tr>
<td>Mucinous adenocarcinoma</td>
<td>1 (4%)</td>
</tr>
</tbody>
</table>

**Location of primary tumour [%]**

<table>
<thead>
<tr>
<th>Location</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigmoid</td>
<td>9 (39%)</td>
</tr>
<tr>
<td>Rectum</td>
<td>6 (26%)</td>
</tr>
<tr>
<td>Colon</td>
<td>5 (2%)</td>
</tr>
<tr>
<td>Colon &amp; rectum</td>
<td>3 (13%)</td>
</tr>
</tbody>
</table>

**Stage at presentation [%]**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>II (metachronous CRLM)</td>
<td>3 (13%)</td>
</tr>
<tr>
<td>III (metachronous CRLM)</td>
<td>4 (17%)</td>
</tr>
<tr>
<td>IV (synchronous CRLM)</td>
<td>16 (70%)</td>
</tr>
</tbody>
</table>

Table 7.1: Patient characteristics. CRLM: colorectal liver metastases; IQR: interquartile range; MAD: maximum axial diameter measured on T1-weighted magnetic resonance image.

**DCE-MRI**

**Quantitative DCE-MRI data acquisition & reconstruction**

Measurements were performed on a 1.5 T MRI scanner (Siemens Healthcare, Erlangen, Germany), using a body phased-array coil. After conventional T1- and T2-weighted imaging, 15 ml of 0.5 M Gd-DTPA (Magnevist; Bayer Schering Pharma AG, Leverkusen, Germany) was administered intravenously at an injection rate of 2.5 ml·s$^{-1}$ by a Spectris MR injection system (Medrad Inc.; Bayer Schering Pharma AG, Leverkusen, Germany). An axial T1-weighted fast low-angle shot sequence was performed multislice; that is, 4-6 slices were acquired simultaneously within the $TR$, realised by short $TE$ and short pulse duration. The acquisition matrix was 256-125 pixels per slice. Application of the partial Fourier technique (factor, 6/8) reduced the number of phase-encoding steps to 94, which in turn led to a net temporal resolution of $\sim$21 ms. The acquisition scheme was performed multislice; that is, 4-6 slices were acquired simultaneously within the $TR$, realised by short $TE$ and short pulse duration. The acquisition matrix was 256-125 pixels per slice. Application of the partial Fourier technique (factor, 6/8) reduced the number of phase-encoding steps to 94, which in turn led to a net temporal resolution of $\sim$21 ms. If the 4-6 slices in the axial $FoV$ did not fully cover the tumour in the axial direction, slices were positioned in such a way that the largest diameter of the tumour (on the coronal view) was covered. During the acquisition, patients were instructed to breathe quietly and shallowly.
Immediately before the Gd-DTPA injection, proton density-weighted images were recorded with the same sequence parameters as those for the DCE-MRI, except for a flip angle of 10° and a TR of 250 ms. Data from these images were combined with the DCE-MRI data to calculate the concentration of Gd-DTPA in arbitrary units, using the methods described by Hittmair et al. \[312\].

DCE-MRI data analysis

For the analysis of DCE-MRI, a vascular normalisation function was obtained from pixels in the spleen using an algorithm based on the concentration of Gd-DTPA (high in blood vessels) and time to bolus passage (short in arteries) as described earlier \[313\]. Using a physiologic pharmacokinetic model \[314\], we analysed the Gd-DTPA concentration vs time curves of the pixels in all MRI slices (256·125 pixels) containing tumour tissue and the $k_{ep}$ [s$^{-1}$] of Gd-DTPA uptake was calculated according to the equation $c_t(t) = K^{trans} \cdot e^{-k_{ep} \cdot t} \otimes c_p(t)$, in which $c_t$ is the tissue concentration of Gd-DTPA, $k_{ep}$ is the rate constant of contrast agent exchange between the extracellular extravascular space and the plasma compartment, $K^{trans}$ is the volume transfer constant between these compartments [s$^{-1}$], $c_p$ is the concentration of contrast agent in plasma of a capillary and $\otimes$ denotes a convolution operation \[306\]. In Larsson’s model \[314\], the Gd-DTPA uptake rate constant ($k_{ep}$) is directly related to tumour blood flow, the product of the permeability and the total surface area of perfused capillaries, according to $k_{ep} = \left(1 - e^{-P \cdot S \cdot TBF^{-1}} \right) \cdot TBF \cdot v_e^{-1}$, in which $v_e$ is the volume of extravascular extracellular space per unit volume of tissue [cm$^3$·cm$^{-3}$], $P$ is the permeability of capillaries [cm·s$^{-1}$], $S$ is the total surface area of the vessels [cm$^2$] and $TBF$ is the tumour blood flow [ml·s$^{-1}$]. Previous reports have confirmed that there is a moderately strong positive correlation between $k_{ep}$ and microvessel density in liver metastases ($R = 0.458; p = 0.037$) \[315\].

The spatial distribution of the pharmacokinetic parameters was represented as a map. On a T1-weighted MR image recorded directly before Gd-DTPA injection, a 2D volume-of-interest (VOI) was drawn that comprised the metastases. Only large lesions (i.e., >15 mm) that were totally covered by the FoV of the acquired images and were not disturbed by artefacts of inflow of Gd-DTPA in the abdominal aorta (i.e., lesions directly ventral to the aorta) were included. Six lesions were smaller than 30 mm. This VOI was applied to the map of pharmacokinetic parameters to select the single values of $k_{ep}$, $K^{trans}$ and $v_e$ for all tumour pixels. Whole-tumour values were calculated after log transformation of all voxels within the VOI, thereby excluding zero voxels values for $k_{ep}$, $K^{trans}$ and $v_e$, which are assumed to represent necrotic tissue or fit artefacts. Mean tumour values and 95%-confidence intervals were obtained by backtransformation \[313\]. To obtain a whole-patient value, a mean value, weighted by lesion volume, of all lesions within the FoV was determined by:

$$\text{parameter}_{\text{whole-patient}} = \sum_{i=1}^{n} \left( \frac{\text{parameter}_{\text{lesion},i} \cdot \text{volume}_{\text{lesion},i}}{\sum_{i=1}^{n} \text{volume}_{\text{lesion},i}} \right)$$

(7.1)
7.2. Material & methods

The coronal T1-weighted images were used to measure the maximum axial diameter of each included lesion (MAD, [mm]) to evaluate morphologic therapy response. Patient-based MADs were calculated by the sum of all included lesions.

**18F-FDG PET**

Quantitative 4D dynamic 18F-FDG PET data acquisition & reconstruction

4D Dynamic PET was performed on an ECAT-EXACT47 dedicated PET scanner (Siemens Healthcare, Erlangen, Germany) using the ECAT 7.2.1 software for 2D reconstruction. Patients fasted for at least 6 h before imaging. Intake of sugar-free liquids was permitted. Blood glucose levels (hexokinase method) (Aeroset; Abbott Diagnostics) were determined. The median fasting glucose level was 5.3 mmol·l⁻¹ (maximum, 9.2 mmol·l⁻¹). The location for the dynamic acquisition in the axial FoV (162 mm in 47 planes) of the scanner was based on whole-body 18F-FDG PET and CT scans obtained for routine clinical work-up, including as many measurable tumour lesions as possible. A 20-min transmission scan was made, using the internal ⁶⁸Ge/⁶⁸Ga sources, to correct for photon attenuation. Approximately 200 MBq (mean±SD: 202±40 MBq) of ¹⁸F-FDG (Coviden, Petten, the Netherlands) was injected intravenously using constant infusion by a remote-controlled pump (Mead Inc.; Bayer Schering Pharma AG, Leverkusen, Germany). The dynamic data acquisition, performed in 2D mode, was started simultaneously with the injection of ¹⁸F-FDG and consisted of 16 time frames with parameter duration (10·30, 3·300 and 3·600 s) for a total time of 50 min. During the acquisition, patients were instructed to breathe quietly and shallowly. Correction for decay, randoms and scatter was performed. Attenuation-corrected images were reconstructed in 128-128 matrices using filtered backprojection (FBP) with a Gaussian filter of 4 mm full-width at half-maximum (FWHM). This resulted in forty-seven 3.375 mm slices for each time frame, with voxel dimensions of 3.432·3.432·3.375 mm and a spatial resolution of 6 mm FWHM in the reconstructed images.

**18F-FDG PET data analysis**

¹⁸F-FDG PET data were analysed as described before [83]. In brief, a plasma time-activity concentration curve was obtained by serial arterial sampling. When arterial sampling was not feasible or was contraindicated, an image-derived input function (IDIF) of the abdominal aorta was used (48% of forty-six ¹⁸F-FDG PET scans). Tumour time-activity concentration curves were obtained by determination of volume-weighted mean activity concentration within VOIs. These VOIs were placed semi-automatically over the metastases using 50% of the maximum voxel value within the lesion on the summed images of frames 14-16 (20-50 min after injection). For therapy response evaluation, the lesion-specific VOI with the largest volume was copied to the other scan [316]. Using Gjedde-Patlak graphical analysis [33,112], we determined MR̅⁰ gluc, using a lumped constant (LCFDG) of 1 and a fractional blood volume (VB) of 0 cm³·cm⁻³. A volume-weighted mean value for all lesions corresponding with DCE-MRI was obtained as patient-based data using the same equation as the one used for DCE-MRI data.
Clinical follow-up

Follow-up was performed according to a stringent protocol for three years in accordance with standard clinical care. Tumour response for clinical decision making was evaluated by experienced radiologists, without knowledge of DCE-MRI or dynamic \(^{18}\)F-FDG PET results, according to Response Evaluation Criteria in Solid Tumours (RECIST) \[224\]. Changes in the DCE-MRI parameters \((k_{ep}, K_{trans}, v_c), MAD\) and \(^{18}\)F-FDG PET parameter \(MR_{glc}\) during treatment were calculated as:

\[
\Delta \text{parameter} = \frac{\text{parameter}_{\text{follow-up}} - \text{parameter}_{\text{baseline}}}{\text{parameter}_{\text{baseline}}} \cdot 100\% \quad (7.2)
\]

The date of progression was defined as the earliest date at which disease progression was confirmed. Survival and progression were measured from the date of baseline \(^{18}\)F-FDG PET or DCE-MRI (whichever was performed first) to the date of, respectively, disease-related death or progression. For patients who were alive \((n = 2)\) at the close-out date (May 2009), survival was censored.

The median interval between all DCE-MRI and \(^{18}\)F-FDG PET scans was 0 d \((IQR: 0-1 d)\). The mean interval between the baseline and the first follow-up PET scan was 59.1±11.3 d.

Statistical analysis

Both patient-based analysis and lesion-by-lesion analysis were performed. DCE-MRI, T1-weighted MRI and \(^{18}\)F-FDG PET data were analysed separately and the results were masked. Parameters were assessed for normality by Shapiro-Wilk statistics. Means (±SD) for normally distributed data or medians \((IQR)\) are presented. Differences were assessed for significance by the Mann-Whitney \(U\) test and by the paired \(t\) test (normally distributed) or Wilcoxon signed-rank test for paired data. Correlations were determined by the non-parametric Spearman’s \(\rho\).

Cancer-related \(OS\) and \(PFS\) were calculated using Kaplan-Meier estimates. Both univariate and multivariate analyses were performed by the Cox proportional hazards model. The relationship between DCE-MRI and \(^{18}\)F-FDG PET parameters and \(OS\) and \(PFS\) was assessed for both pretreatment values (predictive value of tumour parameters) and changes in parameter values during treatment (predictive value of early therapy response). Separate (univariate) analysis was performed for the relationship between \(OS\) and \(PFS\) and the following covariates: number of lesions, patient age, TNM classification, tumour differentiation, histology or localisation, chemotherapy line (first line vs second or higher lines) and regimen and the number of chemotherapy cycles before the follow-up scan. Hazard ratios \((HRs)\) are presented with their 95%-confidence intervals (Wald’s \(\chi^2\) test). Multivariate analysis was performed using imaging parameters and significant covariates in a backwardly designed conditional Cox proportional hazards model, removing parameters when \(p > 0.100\).

All statistical analyses were performed with the Statistical Package for Social Sciences (SPSS) version 16.0.2 for Mac (SPSS Inc., Chicago, IL, USA). Statistical tests were based on a 2-sided significance level set at \(p = 0.050\) for all tests.
7.3 Results

Predictive value of pretreatment DCE-MRI & $^{18}$F-FDG PET parameters

An example of DCE-MRI and $^{18}$F-FDG PET data is displayed in figure 7.1 (page 123). No statistically different pretreatment parameters for tumour size and vascular parameters were found for patients who were treated in the first line ($n = 9$) vs patients who had been treated with cytotoxic therapy previously ($n = 14$). $MR_{gic}$ was slightly lower in the patients who were treated in the first line. This, however, did not reach statistical significance (median $MR_{gic}$: 0.1095 $\mu$mol·min$^{-1}$·cm$^{-3}$ vs 0.1373 $\mu$mol·min$^{-1}$·cm$^{-3}$; $p = 0.072$).
Patient-based analysis of the pretreatment scans showed a significant positive correlation between the tumour size (MAD) and the fraction of extravascular extracellular space ($v_e$) ($\rho = 0.426; p = 0.043$). Neither the patient-based analysis nor the lesion-by-lesion analysis yielded significant correlations between pretreatment DCE-MRI and $^{18}$F-FDG PET parameters. The correlation between baseline $k_{ep}$ and $MR_{glc}$ was -0.028 ($p = 0.880$).

Univariate Cox regression analysis established an OS and PFS benefit in patients with low baseline $MR_{glc}$ values. Baseline MAD showed a minor but significant relationship with OS as well, but chemotherapy line (first line vs higher lines) showed no significant relationship with either OS or PFS in this group of patients. None of the DCE-MRI parameters for vascularity showed a significant relationship with either OS or PFS. When these parameters were used in multivariate analysis, correction for MAD increased the HR of $MR_{glc}$ for OS and PFS, but additional correction for chemotherapy line did not improve the predictive ability of the model as a whole (table 7.2 page 124).

### Lesion-by-lesion analysis of changes in scan parameters

A significant change in median $MR_{glc}$ (baseline: 0.138 $\mu$mol-min$^{-1}$-cm$^{-3}$, follow-up: 0.059 $\mu$mol-min$^{-1}$-cm$^{-3}$; $p < 0.001$) was found at the follow-up, but no significant changes were seen in the other parameters. $K^{trans}$ did not change significantly ($p = 0.088$). No correlations were present between changes in $^{18}$F-FDG PET parameters and changes in DCE-MRI parameters during treatment.

### Predictive value of changes in DCE-MRI & $^{18}$F-FDG PET parameters

For response to chemotherapy, no statistical differences were seen between the changes in tumour size, DCE-MRI vascularity parameters and $MR_{glc}$, between the group that was treated with systemic treatment in the first line ($n = 9$) and the group that was treated in higher lines ($n = 14$) (all $p > 0.305$).
7.3. Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Baseline Median</th>
<th>Baseline IQR</th>
<th>Follow-up Median</th>
<th>Follow-up IQR</th>
<th>(p^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCE-MRI (k_{ep}[s^{-1}])</td>
<td>0.014 (0.005-0.034)</td>
<td>0.022 (0.012-0.049)</td>
<td>0.056 (\dagger)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K_{trans}[s^{-1}])</td>
<td>0.009 (0.003-0.020)</td>
<td>0.016 (0.008-0.033)</td>
<td>0.035 (\dagger)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(v_e[cm^3\cdot cm^{-3}])</td>
<td>0.638 (0.516-0.698)</td>
<td>0.614 (0.566-0.744)</td>
<td>0.893 (\dagger)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| \(18F-FDG PET\)  
MR\(_{glc}\) \(\mu mol\cdot min^{-1} \cdot cm^{-3}\) | 0.128 (0.108-0.160) | 0.054 (0.041-0.122)   | \(<0.001\) \(\dagger\) |
| T1-weighted MRI \(MAD[mm]\) | 56 (50-92) | 54 (47-109) | 0.268 \(\dagger\) |

Table 7.3: Baseline and follow-up values of vascular and metabolic parameters. DCE-MRI: 4D dynamic gadopentetate dimeglumine contrast-enhanced magnetic resonance imaging; \(18F-FDG PET\): 2-\(^{18}\)F-fluoro-2-deoxy-\(\alpha\)-glucose positron emission tomography; IQR: interquartile range; \(k_{ep}\): gadopentetate dimeglumine uptake rate constant; \(K_{trans}\): gadopentetate dimeglumine volume transfer constant; \(MAD\): maximum axial diameter measured on T1-weighted magnetic resonance image; \(MR_{glc}\): glucose metabolic rate; \(v_e\): volume of extravascular extracellular space per unit volume of tissue. \(\dagger\): Wilcoxon signed-rank test for paired samples. \(\dagger\): Significant at the \(p = 0.05\) level.

Results of scan parameters before and after the start of treatment are provided in table 7.3 (page 125). There was a significant change in both \(MR_{glc}\) (median baseline: 0.128 \(\mu mol\cdot min^{-1} \cdot cm^{-3}\); median follow-up: 0.054 \(\mu mol\cdot min^{-1} \cdot cm^{-3}\); \(p < 0.001\)) and \(K_{trans}\) (median baseline: 0.009 \(s^{-1}\); median follow-up: 0.016 \(s^{-1}\); \(p = 0.035\)) during chemotherapy. No significant correlations were found between parameter changes during treatment. The correlation between \(\Delta K_{trans}\) and \(\Delta MR_{glc}\) was \(\rho = -0.172\) (\(p = 0.433\)).

To assess whether pre-existing vasculature as assessed by DCE-MRI influenced therapy response by delivery of cytotoxic drugs, \(k_{ep}\) before chemotherapy was correlated to metabolic and anatomic response, but no significant relationship could be established between \(k_{ep}\) and \(\Delta MAD\) (\(\rho = -0.209\); \(p = 0.340\)) or \(\Delta MR_{glc}\) (\(\rho = 0.257\); \(p = 0.237\)). The same applied for \(K_{trans}\) and \(v_e\). In lesion-by-lesion analysis, \(k_{ep}\) before chemotherapy was not correlated with \(\Delta MAD\) (\(\rho = -0.197\); \(p = 0.289\)) or \(\Delta MR_{glc}\) (\(\rho = 0.293\); \(p = 0.109\)).

Univariate Cox regression analysis showed no relationship of any change in DCE-MRI or \(^{18}\)F-FDG PET parameters with respect to \(OS\). When all lesions inside the larger FoV of the PET scanner were quantified (instead of only the matching lesions) to determine a patient-based \(MR_{glc}\), \(\Delta MR_{glc}\) was predictive for \(OS\) (\(HR = 1.15\); \(p = 0.041\)). Only \(\Delta MAD\) was related to both \(OS\) (\(HR = 1.40\); \(p = 0.023\)) and \(PFS\) (\(HR = 1.34\); \(p = 0.026\)). Chemotherapy line was not a significant confounder for \(OS\) or \(PFS\). Multivariate Cox regression modelling showed both \(\Delta MR_{glc}\) (\(HR = 1.22\); \(p = 0.008\)) and \(\Delta k_{ep}\) (\(HR = 0.99\); \(p = 0.100\)) as predictors for \(OS\), but the latter was irrelevantly small. For \(PFS\), both \(\Delta MAD\) (\(HR = 1.48\); \(p = 0.010\)) and chemotherapy line (\(HR = 3.15\); \(p = 0.023\)) were of predictive relevance. Results are shown in table 7.4 (page 126).
7. DCE-MRI & Dynamic FDG PET in Therapy Response Evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\Delta MR_{glc}$ (all in FoV)*</th>
<th>$\Delta MAD^*$</th>
<th>Chemotherapy line$^\dagger$</th>
<th>$\Delta k_{ep}^*$</th>
<th>$\Delta MAD^*$</th>
<th>Chemotherapy line$^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.15 (1.01-1.32)</td>
<td>1.40 (1.06-1.85)</td>
<td></td>
<td>0.99 (0.98-1.00)</td>
<td>0.99 (0.98-1.00)</td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041§</td>
<td>0.023§</td>
<td></td>
<td>0.008§</td>
<td>1.48 (1.10-1.99)</td>
<td>0.010§</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.15 (1.17-8.49)</td>
<td>0.023§</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.4: Predictive value of therapy-induced parameter changes for early response evaluation, assessed by Cox regression analysis. 95% – CI: 95%-confidence interval; HR: hazard ratio; $\Delta MR_{glc}$: relative change in glucose metabolic rate including all lesion in the (larger) field-of-view of the positron emission tomography scanner; $\Delta k_{ep}$: relative change in gadopentetate dimeglumine uptake rate constant; $\Delta MAD$: relative change in maximum axial diameter measured on T1-weighted magnetic resonance image; Os: overall survival; PFS: progression-free survival. $^\ast$ Per 10% change. $^\dagger$ First line vs higher lines. $^\ddagger$ Wald’s $\chi^2$ test. $^§$ Significant at the $p = 0.05$ level.

7.4 Discussion

Pretreatment parameters for vascularity & metabolism

No association was seen between vascularity, assessed by DCE-MRI and glucose metabolism, assessed by $^{18}$F-FDG PET. Our results confirm those of Brix et al. [317], who found no correlation between $k_{ep}$ and $^{18}$F-FDG uptake (standardised uptake value ($SUV$) in breast carcinoma. Semple et al. [318], however, found a positive correlation between $k_{ep}$ and $SUV$ ($\rho = 0.5$) and a positive, though non-significant, correlation between $K_{trans}$ and $SUV$ in breast carcinoma patients before the commencement of treatment. They suggested that $^{18}$F-FDG delivery was restricted by the blood flow dynamics of the tumour. In a previous study [315], we found a negative correlation between $k_{ep}$ and $^{18}$F-FDG uptake (tumour-nontumour ratios) in CRC.

The different relationships between vascular and metabolic tumour parameters described in the literature could imply that tumour vasculature is not related to $MR_{glc}$. However, a more complex relationship between vasculature and $MR_{glc}$, mediated by acute and chronic tumour hypoxia, could also play a role: chronic hypoxia and diminished delivery of glucose could be caused by a low blood flow, a low permeability or a small surface area of tumour blood vessels resulting in low values for $k_{ep}$. All these vascular parameters may result in a decreased supply of nutrients such as glucose and oxygen to the tumour, similar to the decreased delivery of Gd-DTPA, which would lead to decreased cell proliferation as an energy-saving method [319] or to necrosis or apoptosis. This situation would result in a positive correlation between $k_{ep}$ and $MR_{glc}$. Conversely, (transient) hypoxia due to poor vascular function (as measured by a low value for $k_{ep}$) might induce higher glucose uptake in the tumour for anaerobic glycolysis [315,320,321]. The latter explanation would result in a negative correlation between $k_{ep}$ and $MR_{glc}$. Therefore, the opposite, combined, complex effects of acute and chronic hypoxia and nutrient supply could explain the lack of correlation between $k_{ep}$ and $MR_{glc}$. Another reason for this lack of correlation could be non-specificity of $K_{trans}$ and $k_{ep}$, which may represent flow, vessel permeability or surface area or...
a combination of these. Finally, imprecision in the determination of the DCE-MRI or PET parameters combined with the limited study size might explain the lack of correlation.

Both pretreatment tumour metabolism ($MR_{glc}$) and size ($MAD$) were associated with higher hazards for death and progression. This confirms previous data [16], which showed that a one unit increase in $SUV$ results in a 17% increase in the risk of death. Multivariate analysis showed no influence of chemotherapy line (first- vs higher-line treatment) for the predictive abilities of $MR_{glc}$ and $MAD$.

Previously, Semple et al. [322] had observed a significant correlation ($p < 0.05$) between pretreatment $k_{ep}$ and $\Delta SUV$ during chemotherapy of 17 breast carcinoma patients and concluded that the reduction of measured metabolism may be partly attributable to pretherapy vascular delivery ($k_{ep}$). Because Gd-DTPA and phenylacetate (which has a size similar to 5-FU) are similarly distributed in the interstitium of tumour tissue [323], a restriction of Gd-DTPA delivery to the interstitium reflects a restriction of 5-FU delivery to the immediate neighborhood of tumour cells. We could not confirm these findings in patients with CRC.

### Early changes in metabolic & vascularity parameters during treatment

We have found no significant treatment-induced changes in $k_{ep}$, $v_e$ and $MAD$, whereas $MR_{glc}$ and $K^{trans}$, respectively, significantly decreased and increased on a patient level. On a lesion level, the increase in $K^{trans}$ could not be reproduced. Our results suggest that the observed reduction in $MR_{glc}$ during chemotherapy cannot be explained by changes in tumour vasculature ($\Delta k_{ep}$). The antivascular effect of cytotoxic drugs is small and marginally influences cell metabolism and patient survival. It seems that direct cytotoxic effects leading to necrosis and apoptosis cause disease response and improved survival and that possible effects of chemotherapeutic agents on nutrient delivery play a minor role.

The reduction in $k_{ep}$ during therapy has been described in breast carcinoma [324, 325] and reduction in $K^{trans}$ during treatment has been described in rectal [308] and breast [309, 310, 326] carcinoma. Some authors explained these changes by the direct antivascular effect of the cytotoxic drugs [310, 326] or by the loss of immature tumour vessels [308]. Our results, however, do not suggest an effect of cytotoxic drugs on tumour vasculature in CRC.

We found no relationship between vascular parameters and clinical outcome. This confirms the results of our previous study showing no evident relevance of pretreatment $k_{ep}$, $K^{trans}$ or $v_e$ for $OS$ and $PFS$ in colorectal liver metastases or any change in these parameters during first-line chemotherapy [327]. In the present study, we observed a positive relationship between $\Delta MR_{glc}$ and hazards for death but not for progression, which was mainly predicted by chemotherapy line and $\Delta MAD$.

Because DCE-MRI and $^{18}$F-FDG PET both address different aspects of tumour physiology, they might be used complementarily in treatment-response evaluation. The choice for a specific imaging modality should depend on the treatment regimen. The cytotoxic drugs given to our patients interfere with DNA synthesis and stabilisation, eventually leading to cell death. Therefore, the direct interaction with tumour vascularity is limited, as is suggested by our data, showing no change in DCE-MRI
parameters during cytotoxic treatment. When anti-angiogenic drugs are used, response might be predicted by DCE-MRI, as shown by two studies \([328,329]\) in clear cell renal cell carcinoma patients treated with sorafenib. Therapy response might be evaluated by DCE-MRI, showing a decrease of \(k_e\) [330], but results are still contradictory [329]. The effect of cytotoxic drugs on cell metabolism can be evaluated by \(^{18}\)F-FDG PET. Tumour metabolic response has predictive value [22], which is also shown by our data. Early disease-related deaths did occur in the group of metabolic responders. Therefore, a reduction in metabolism during therapy does not guarantee a long survival and more effects presumably play a role.

**Study limitations**

Patients were treated in different lines of chemotherapy using different chemotherapy regimens; thus, the included population was heterogeneous. Furthermore, currently combinations of cytotoxic treatment with anti-angiogenic treatment are standard in first-line treatment. DCE-MRI may have a role as a tool for response evaluation in those patients who are being treated with anti-angiogenics. Previous chemotherapy might already have influenced both metabolism and vascularity. However, subgroup analysis of the nine patients treated with first-line chemotherapy did not change our conclusions.

Only a selection of lesions could be analysed because of the limited FoV of the acquired images and the limited spatial resolution of (especially) the PET system, which might have caused some bias during selection.

The parameters for vascularity as derived from DCE-MRI were not verified by histologic quantification of micro-vessel density, because biopsies of stage IV patients are taken only in exceptional circumstances. However, we have previously described a relationship between both histology and DCE-MRI parameters [315]. An advantage of DCE-MRI over histology is that DCE-MRI measures functional vasculature only, whereas the quantification of histologic staining of endothelial cells also includes non-perfused vessels.

**7.5 Conclusions**

4D Dynamic Gd-DTPA-enhanced MRI parameters of tumour vasculature showed no relationship to tumour metabolic response on dynamic \(^{18}\)F-FDG PET or to patient survival during cytotoxic chemotherapy either before or during treatment. Therefore, a decrease in metabolic activity and an increase in \(OS\) and \(PFS\) during chemotherapy cannot be attributed to changes in tumour vascularity, resulting in the altered delivery of drugs or nutrients of the same size as Gd-DTPA. The present study underlines the potential of \(^{18}\)F-FDG PET for therapy response evaluation. The main conclusions of the article are valid for patients not receiving vascular-targeted therapy and further study is required in the subset of patients who do receive additional anti-angiogenic agents.

***
Vascular and Metabolic Response to Bevacizumab-Containing Regimens in Two Patients with Colorectal Liver Metastases Measured by DCE-MRI and 4D Dynamic $^{18}$F-FDG PET

(adapted and reprinted with permission)

D. Vriens\textsuperscript{1},
L.F. de Geus-Oei\textsuperscript{1},
A. Heerschap\textsuperscript{2},
H.W.M. van Laarhoven\textsuperscript{3} and
W.J.G. Oyen\textsuperscript{1}

dep. of \textsuperscript{1}Nuclear Medicine,
\textsuperscript{2}Radiology and \textsuperscript{3}Medical Oncology, Radboud University Medical Centre, Nijmegen, the Netherlands
Abstract

Purpose Early evaluation of response to therapy is one of the cornerstones of personalised treatment. As new and often expensive targeted therapies, which are tumourstatic rather than tumouricidal, become available, new demands are posed on response assessment. Bevacizumab, an anti-angiogenic agent causing normalisation of the tumour microvasculature, potentiates the effect of cytotoxic agents on colorectal liver metastases. It is known that assessment of glucose metabolism by (4D dynamic) positron emission tomography using 2-(^{18}F)fluoro-2-deoxy-d-glucose (^{18}F-FDG PET) can be used as an early surrogate endpoint to determine treatment efficacy. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) can be used to quantify functional tumour vasculature (permeability, vascular surface area).

Methods Here, we describe the response of colorectal liver metastases to cytotoxic regimens including bevacizumab using both ^{18}F-FDG PET and DCE-MRI in two cases.

Results In both cases, a large reduction in glucose metabolic rate and functional tumour vasculature are observed after three treatment cycles.

Conclusions Metabolic and vascular effects to combination therapy can be evaluated using DCE-MRI and ^{18}F-FDG PET.

Keywords Angiogenic Inhibitors · Chemotherapy · Colorectal Carcinoma · DCE-MRI · Therapy Response Evaluation
Progress in cytotoxic treatment of advanced colorectal carcinoma (CRC) has improved median survival from eight months to more than 20 months [153]. For many tumour types it is known that they induce neoangiogenesis, leading to an intratumoural microcirculation that is characterised by tortuous microvessels with chaotic architecture and irregular, sluggish blood flow with unstable rheology. Angiogenesis inhibitors lead to normalisation of tumour vasculature, which may improve chemotherapy delivery [331]. In fact, the addition of bevacizumab to the standard first-line chemotherapy has improved response rate (34.8% to 44.8%), progression-free survival ($PFS$: 6.2 months to 10.6 months) and overall survival ($OS$: 15.6 months to 20.3 months) [332–335] and this humanised monoclonal antibody against vascular endothelial growth factor A (VEGF-A) was approved for the treatment of metastatic CRC. Unfortunately, not all patients respond to systemic treatment. Early response prediction would help to identify patients who are early in treatment who most likely benefit from the addition of this monoclonal antibody therapy, increasing the beneficial effects and preventing side effects and costs because of ineffective treatments.

Capillary perfusion and permeability of the vessel wall can be measured in vivo by 4D dynamic gadopentetate dimeglumine (Gd-DTPA), contrast-enhanced magnetic resonance imaging (DCE-MRI) [306]. Pharmacokinetic analysis of DCE-MRI data using a two-compartment model [314,336] yields parameters for perfused capillaries including blood flow, permeability and the total vessel surface area ($K_{trans}$: volume transfer constant, $k_{ep}$: compartmental rate constant and $v_e$: volume of contrast extravascular extracellular space per unit volume of tissue). Assuming that vascularity, as reflected by these parameters, is modified by treatment by bevacizumab, it may be hypothesised that these parameters can be used for prediction of therapy response. The value of early evaluation of therapy response by DCE-MRI has been shown for rectal [308] and breast [309,311] carcinoma. Glucose metabolic activity can be assessed in vivo by positron emission tomography (PET) using the radionuclide labelled glucose-analogue 2-($^{18}$F)fluoro-2-deoxy-D-glucose ($^{18}$F-FDG) [33]. $^{18}$F-FDG uptake is increased in malignant tumours. An adequate vascular supply is necessary for delivery of glucose to tumour cells as well as the presence of several membrane-bound glucose transport proteins and intracellular hexokinase for subsequent phosphorylation. $^{18}$F-FDG uptake can be quantified by pharmacokinetic compartment modelling [33,34] of 4D dynamic $^{18}$F-FDG PET data resulting in glucose metabolic rates ($MR_{glc}$). The predictive value for early evaluation of therapy response, using changes in $MR_{glc}$ during (chemo)therapy, has been shown in many tumour types including non-small cell lung carcinoma (NSCLC) [21] and CRC [22].

Previously, we have shown that the vascularity parameter $k_{ep}$, reflecting the transfer rate of Gd-DTPA over the capillary membrane, was inversely correlated to $^{18}$F-FDG uptake [315]. A positive correlation was observed between $k_{ep}$ and histopathological vascular density and in surgically removed colorectal liver metastases (CRLM). More recently, we described the influence of cytotoxic treatment on changes in vascular perfusion and permeability (DCE-MRI) and cell metabolism (dynamic $^{18}$F-FDG PET) showing no changes in Gd-DTPA transfer rates but high decreases in $MR_{glc}$,
which was correlated to patient survival (chapter 7) [327,337]. Here, we present findings of two patients with CRLM imaged by both DCE-MRI and $^{18}$F-FDG PET before and following three months of first-line chemotherapy including bevacizumab.

### 8.2 Case Report

#### Patient A

In patient A, a 57-year-old man (Karnofsky performance status score: 90) with a history of myocardial infarction, a pT$_3$N$_2$M$_x$ moderately differentiated adenocarcinoma of the sigmoid colon was resected. Computed tomography at follow-up showed two liver metastases, in Couinaud segment II/III (left) and VI (right), considered eligible for resection. However, during surgery it was concluded that these metastases were unresectable. The patient was treated by first-line FOLFOX4 therapy (bimonthly two-hour infusion of oxaliplatin (85 mg·m$^{-2}$) combined with weekly two-hour infusion of folinic acid (200 mg·m$^{-2}$) and 24-hour continuous infusion of 5-fluorouracil (2,600 mg·m$^{-2}$) combined with bevacizumab (10 mg·kg$^{-1}$)) 30-90 minute infusion every three weeks [338]. Initial assessment after three treatment cycles by MRI showed stabilisation of disease (28% reduction according to the Response Evaluation Criteria in Solid Tumours (RECIST) v1.0 [224]). Compared with baseline, the 4D dynamic $^{18}$F-FDG PET (performed three days after the last therapy day of the third cycle as previously described [337]) showed a 80% reduction in MR$_{glc}$ of the lesions corresponding to a reduction in standardised uptake value ($SUV$, the activity concentration in the tumour normalised for administered activity and patient bodyweight) on the static images of 23%. The DCE-MRI performed the same day showed a clear decrease in $K_{trans}$ and $k_{ep}$ of 40% and 58%, respectively (table 8.1 page 135 figure 8.1 page 136).

Because of polyneuropathy and therapy-resistant diarrhea, oxaliplatin had to be stopped after nine cycles. After 12 cycles in total, stable disease (SD) on MRI persisted according to the RECIST criteria and treatment was stopped. Six days after the last therapeutic day, the MR$_{glc}$ and $SUV$ remained at similarly reduced levels compared with baseline (reduction of 89% and 20%, respectively). The $k_{ep}$, however, was only slightly lower than baseline and $K_{trans}$ had increased.

At the follow-up three months after the end of the previous treatment scheme, progressive disease was observed, which did not respond to second-line therapy with irinotecan. The patient died two years after the start of chemotherapy, which is somewhat longer than the median $OS$ known for this treatment scheme [332].

#### Patient B

Patient B, a 58-year-old man (Karnofsky performance status score: 90) with a history of prostate carcinoma treated by laparoscopic prostatectomy 55 months earlier was diagnosed with a pT$_3$N$_2$M$_1$ poorly differentiated adenocarcinoma in the sigmoid. A low-anterior resection with end colostomy was performed. Palliative chemotherapy was started for multiple synchronous CRLM with first-line capecitabine orally (1,000 mg·m$^{-2}$ twice daily on days 1-14), combined with infusion of oxaliplatin (130 mg·m$^{-2}$ on day 1) and bevacizumab (7.5 mg·kg$^{-1}$ on day 1) plus cetuximab (400 mg·m$^{-2}$ in
8.2. Case Report

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Patient A</th>
<th>Patient B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{18}$F-FDG PET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MR_{glc}$</td>
<td>Baseline: 91.7</td>
<td>Baseline: 80.4</td>
</tr>
<tr>
<td></td>
<td>Early: 18.4</td>
<td>Early: 7.8</td>
</tr>
<tr>
<td></td>
<td>Late: 10.1</td>
<td>Late: 3.2</td>
</tr>
<tr>
<td></td>
<td>(-80%)</td>
<td>(-90%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-96%)</td>
</tr>
<tr>
<td>$SUV$</td>
<td>Baseline: 5.3</td>
<td>Baseline: 4.3</td>
</tr>
<tr>
<td></td>
<td>Early: 4.1</td>
<td>Early: 3.2</td>
</tr>
<tr>
<td></td>
<td>Late: 4.2</td>
<td>Late: 3.4</td>
</tr>
<tr>
<td></td>
<td>(-23%)</td>
<td>(-24%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-22%)</td>
</tr>
<tr>
<td>$DCE$-MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{trans}$</td>
<td>Baseline: 15.6</td>
<td>Baseline: 89.5</td>
</tr>
<tr>
<td></td>
<td>Early: 9.4</td>
<td>Early: 6.4</td>
</tr>
<tr>
<td></td>
<td>Late: 18.3</td>
<td>Late: –</td>
</tr>
<tr>
<td></td>
<td>(-40%)</td>
<td>(-93%)</td>
</tr>
<tr>
<td>$k_{ep}$</td>
<td>Baseline: 30.4</td>
<td>Baseline: 122</td>
</tr>
<tr>
<td></td>
<td>Early: 12.7</td>
<td>Early: 8.6</td>
</tr>
<tr>
<td></td>
<td>Late: 29.5</td>
<td>Late: –</td>
</tr>
<tr>
<td></td>
<td>(-58%)</td>
<td>(-93%)</td>
</tr>
</tbody>
</table>

Survival

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Patient A</th>
<th>Patient B</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFS [months]</td>
<td>9.1</td>
<td>13.8</td>
</tr>
<tr>
<td>OS [months]</td>
<td>26.8</td>
<td>23.2</td>
</tr>
</tbody>
</table>

Table 8.1: Responses in imaging parameters of the two patients described between baseline, after three treatment cycles and a late acquisition (12 cycles for patient A, nine cycles for patient B). Relative changes compared with the baseline scan are displayed in parentheses. DCE-MRI: 4D dynamic gadopentetate dimeglumine contrast-enhanced magnetic resonance imaging; $^{18}$F-FDG PET: $^{18}$F-2-fluoro-2-deoxy-d-glucose positron emission tomography; $k_{ep}$: gadopentetate dimeglumine uptake rate constant; $MR_{glc}$: glucose metabolic rate; OS: overall survival; PFS: progression-free survival; $SUV$: standardised uptake value (the $^{18}$F-FDG activity concentration in the tumour normalised for administered activity and patient bodyweight).

week 1 of the first treatment cycle and 250 mg·m$^{-2}$ weekly thereafter [339]. All cycles were administered every three weeks. To prevent serious peripheral sensory neurotoxicity, oxaliplatin was administered during a maximum of six cycles, after which the capcitabine dose was increased to 1,250 mg·m$^{-2}$ orally twice daily. Initial assessment after three cycles by MRI showed a partial response (82% reduction according to RECIST). Compared with baseline, the dynamic $^{18}$F-FDG PET (performed the day after the last therapy day of the third cycle) a 90% reduction in $MR_{glc}$ of the lesions were seen corresponding to a reduction in $SUV$ on the static images of 24%. The DCE-MRI performed the same day showed a clear decrease in both $K_{trans}$ and $k_{ep}$ of 93% (table 8.1 page 135 figure 8.2 page 137). Follow-up after nine cycles of chemotherapy by MRI showed a complete response according to RECIST. The dynamic $^{18}$F-FDG PET (performed the day after the last therapy day of the ninth cycle) showed a further reduction of $MR_{glc}$ with 96% compared with baseline, the $SUV$ remained stable at 22% below baseline (table 8.1 page 135). Unfortunately, the DCE-MRI performed that day could not be analysed because of a human mistake in saving the images.

After 17 cycles this treatment regimen was discontinued because of the development of lung metastases and cetuximab-related skin toxicity. Less than 30% of patients on this treatment scheme have a PFS of this duration [339]. After the fifth cycle of second-line treatment with three-weekly irinotecan, treatment was switched to oxaliplatin combined with capcitabine because of a peritonitis carcinomatosis. Despite treatment, the disease remained progressive and the patient died 21 months after the start of chemotherapy. His OS was comparable to the median survival of this treatment scheme (19.4 months) [339].
8. DCE-MRI & Dynamic FDG PET in Bevacizumab Response Evaluation

Figure 8.1: Patient A. Static $^{18}$F-FDG PET acquired 40-50 minutes after injection of the tracer (left panels). Enhanced T1 (middle panels) and parametric $k_{ep}$ (right panels) DCE-MRI images. From top row (baseline) via middle row (after three treatment cycles) to lower row (after 12 cycles of therapy) it can be clearly seen that glucose metabolism and $k_{ep}$ decrease in the lesion displayed (Couinaud segment VII (arrows)). DCE-MRI: 4D dynamic gadopentetate dimeglumine contrast-enhanced magnetic resonance imaging; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography; $k_{ep}$: gadopentetate dimeglumine uptake rate constant

8.3 Discussion

In this case series of two patients treated with bevacizumab-containing therapeutic regimens we saw a strong decrease in vessel perfusion and permeability as assessed by DCE-MRI. These results suggest that DCE-MRI can be used to assess tumour response to vascular targeting therapy. In contrast, in 23 other patients with CRLM treated with cytotoxic (non-targeting) chemotherapeutic agents, we saw similar responses in glucose metabolism, which were correlated with patient OS and PFS (chapter 7) [337]. However in that study, no changes in DCE-MRI parameters were observed.

To the best of our knowledge, this is the first report of clinical response evaluation of anti-angiogenic treatment of CRLM with bevacizumab by DCE-MRI, excluding therapy with tyrosine kinase inhibitors. Wedam et al. [330] showed in 20 patients with inflammatory and locally advanced breast carcinoma that treatment with bevacizumab for one cycle followed by bevacizumab/doxorubicin/docetaxel for six cycles reduced vascular parameters quantified by DCE-MRI with Gd-DTPA. From baseline to the end of the first cycle (bevacizumab alone) they saw a decrease in $k_{ep}$, $K_{trans}$
8.3. Discussion

Figure 8.2: Patient B. Static $^{18}$F-FDG PET acquired 40-50 minutes after injection of the tracer (left panels). Enhanced T1 (middle panels) and parametric $k_{ep}$ (right panels) DCE-MRI images. From top row (baseline) via middle row (after three treatment cycles) to lower row (after nine cycles of therapy) it can be clearly seen that glucose metabolism and $k_{ep}$ decrease in the two lesions displayed (Couinaud segment IVa (arrowheads) and VII (arrows)). DCE-MRI: 4D dynamic gadopentetate dimeglumine contrast-enhanced magnetic resonance imaging; $^{18}$F-FDG PET: 2-($^{18}$F)fluoro-2-deoxy-d-glucose positron emission tomography; $k_{ep}$: gadopentetate dimeglumine uptake rate constant.

and $v_e$ of 58.0%, 49.4% and 20.4%, respectively. After seven cycles in total, these decreased further, being 75.5%, 59.1% and 20.2% compared with baseline. However, they could not find a significant difference in any of the DCE-MRI parameters comparing clinical responders with nonresponders. Thukral et al. [31] followed 19 patients with inflammatory or locally advanced breast carcinoma treated with one cycle of bevacizumab alone followed by six cycles of bevacizumab added to chemotherapy using DCE-MRI with Gd-DTPA. The median values of $K^{trans}$, $k_{ep}$ and the integrated (180 s) area under the concentration-time curve ($iAUC_{180}$) showed significant decreases from baseline to cycle 1 of 34% ($p = 0.003$), 15% ($p < 0.001$) and 23% ($p = 0.009$), respectively. Similarly, Baar et al. [34] recently showed that bevacizumab decreases tumour perfusion as measured by DCE-MRI. Forty-nine patients with inoperable breast carcinoma were randomised to either two cycles of neoadjuvant docetaxel weekly for six weeks or docetaxel/bevacizumab every other week for 16 weeks. DCE-MRI using Gd-DTPA was performed at baseline and after the first and second treatment cycle. All patients showed a decrease in $iAUC_{90}$ that was larger for patients using docetaxel/bevacizumab than patients treated with docetaxel alone ($p = 0.024$). This decrease was more profound after two treatment cycles than after one treatment cycle.
Therefore, therapy response assessment using DCE-MRI shows parameter results, depending on the treatment schedule used, contrast medium administered and analysis performed (heuristic or compartment analysis) and may even differ between different tumours. As these two cases are illustrative of therapy response assessment by DCE-MRI, larger trials should be performed to assess its true value and to determine the most useful parameter and optimal time-point for evaluation.

8.4 Conclusions

Therapy response evaluation using anti-angiogenic combination therapy using both DCE-MRI and $^{18}$F-FDG PET seems feasible. Further studies are needed to support incorporation of molecular imaging in therapy response evaluation of targeted therapies.

***
Part IV

Clinical Impact of Molecular Imaging
The Role of $^{18}$F-FDG PET in Thyroid Nodules with Indeterminate Fine-Needle Aspiration Cytology: Systematic Review and Meta-Analysis of the Literature

Cancer, 2011 Oct;117(20):4582-94 (adapted and reprinted with permission)

D. Vriens\textsuperscript{1},
J.H.W. de Wilt\textsuperscript{2},
G.J. van der Wilt\textsuperscript{3},
R.T. Netea-Maier\textsuperscript{4},
W.J.G. Oyen\textsuperscript{1} and
L.F. de Geus-Oei\textsuperscript{1}

dep. of \textsuperscript{1}Nuclear Medicine,
\textsuperscript{2}Surgery, \textsuperscript{3}Health Evidence and
\textsuperscript{4}Internal Medicine, Section
Endocrinology, Radboud
University Medical Centre,
Nijmegen, the Netherlands
Abstract

Purpose Indeterminate results at fine-needle aspiration cytology (FNAC) of thyroid nodules pose a clinical dilemma, because only 20% to 30% of patients suffer from malignancy. Previous studies suggested that the false-negative ratio of 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose positron emission tomography (\(^{18}\)F-FDG PET) is very low; therefore, it may help identify patients who would benefit from (hemi)thyroidectomy.

Methods A systematic literature search was performed in five databases. After assessment, the identified studies were analysed for heterogeneity and the extracted data of test characteristics were pooled using a random-effects model. Threshold effects were examined and publication bias was assessed.

Results The query resulted in 239 records, of which six studies met predefined inclusion criteria. Data from 225 of the 241 described patients could be extracted. There was mild to moderate heterogeneity in study results (inconsistency index (\(I^2\)): 0.390-0.867). The pooled prevalence of malignancy was 26%. Pooled sensitivity, specificity, positive predictive value, negative predictive value and accuracy were 95% (95%-confidence interval (95% – CI): 86-99%), 48% (95% – CI: 40-56%), 39% (95% – CI: 31-47%), 96% (95% – CI: 90-99%) and 60% (95% – CI: 53-67%), respectively. Sensitivity increased to 100% for the 164 lesions that measured >15 mm in greatest dimension. There was no evidence of threshold effects or publication bias.

Conclusions A negative \(^{18}\)F-FDG PET scan in patients who had thyroid nodules >15 mm with indeterminate FNAC results excluded thyroid carcinoma in a pooled population of 225 patients. Conversely, a positive \(^{18}\)F-FDG PET result did not identify cancer, because approximately 50% of these patients had benign nodules. It was concluded that the incorporation of \(^{18}\)F-FDG PET into the initial work-up of such patients before surgery deserves further investigation.

Keywords 2-(\(^{18}\)F)fluoro-2-deoxy-d-glucose · Meta-Analysis · Positron Emission Tomography · Systematic Review · Thyroidectomy · Thyroid Nodules
9.1 Introduction

Thyroid carcinoma is the most common endocrine malignancy. It represents approximately 1% of all cancers, corresponding to an incidence of up to 10.2 per 100,000 individuals per year in the United States \[341,342\] with increasing incidence over the last decades \[342,343\]. The most common clinical presentation of thyroid carcinoma is a thyroid nodule (TN), either solitary or (dominant) within a multinodular goitre. Approximately 5% to 10% of adults with thyroid carcinoma have palpable TNs and 17% to 45% have nodules identified by ultrasonography \[344,345\]. The majority of these nodules are benign, but approximately 5% to 15% of all palpable TNs are malignant \[345,346\]. Fine-needle aspiration cytology (FNAC) is the most important diagnostic test in the initial evaluation of a patient with a TN and has high diagnostic accuracy (70-97% in experienced centres). Approximately 70% of FNAC results are classified as benign, 4% are classified as malignant, 2% to 10% are classified as insufficient material and the remaining are classified as either indeterminate or suspicious (16-24%). Because of similar cytologic features, it is particularly challenging to distinguish between different types of thyroid neoplasms of the follicular type (i.e., follicular thyroid adenoma, follicular thyroid carcinoma (FTC) and follicular type of papillary carcinoma). Therefore, patients with indeterminate or suspicious FNAC results have to undergo diagnostic hemithyroidectomy to exclude malignancy \[347\]. Because only 20% to 30% of these nodules are malignant \[347\], most patients are undergoing unnecessary thyroid surgery with the potential risk of irreversible complications.

Several promising markers (e.g., thyroid peroxidase, galectin-3, telomerase, rearranged in transformation/papillary thyroid carcinoma (RET/PTC), p53) have been studied in patients with TNs to improve the accuracy of FNAC. However, none of these markers have reached routine clinical use, because they have been documented in only a subset of tumours \[348\]. Both ultrasonographic and scintigraphic features of TNs have been investigated in the past for their diagnostic value in the preoperative diagnostic work-up of patients with TNs, but none of these technique could accurately distinguish between benign and malignant nodules.

The characterisation of tissue using the glucose analogue 2-\((^{18}\text{F})\)fluoro-2-deoxy-D-glucose \((^{18}\text{F}-\text{FDG})\) together with positron emission tomography (PET) has proven beneficial in diagnostics and follow-up of many malignancies \[349\]. The use of \(^{18}\text{F}-\text{FDG PET/computed tomography (CT)}\) in the management of thyroid disease has been limited primarily to the postoperative surveillance of patients with known differentiated thyroid carcinoma (postoperative staging for remnant disease and in therapy response assessment \[345\]). There is a special role for \(^{18}\text{F}-\text{FDG PET/CT}\) in postoperative surveillance if a patient has an elevated thyroglobulin level but a negative whole-body \(^{131}\text{I}\) scintigraphy study \[345\]. Finally, \(^{18}\text{F}-\text{FDG PET/CT}\) thyroid incidentalomas are identified on approximately 1% to 2% of \(^{18}\text{F}-\text{FDG PET/CT}\) studies and harbor a 14% to 47% chance of being confirmed as malignant \[350,352\], warranting further investigations \[345,353\]. Currently, there is no routine place for \(^{18}\text{F}-\text{FDG PET/CT}\) in the work-up of a TN \[345\].

The objective of this systematic review was to provide an up-to-date summary of the value of \(^{18}\text{F}-\text{FDG PET/CT}\) for the preoperative evaluation of patients with TNs and either indeterminate or repeatedly insufficient FNAC results. We used a
systematic literature search and meta-analysis to quantify the false-negative rate of 18F-FDG PET/CT and to investigate whether a negative 18F-FDG PET/CT scan can select patients who have a low suspicion of malignancy and, thus, can avoid surgery.

9.2 Material & methods

Literature search & study selection

A systematic search was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (figure 9.1, page 147). The main research question was defined using the target Population (including previous tests), Index test, Comparator test, Target condition and Study design (PICTS) strategy, which was formulated into a search query containing a combination of Medical Subject Headings (MeSH) or keywords and truncated synonyms (boolean operators). Then, a search using this query was performed on November 21, 2010, with the following five search engines, which use partially overlapping databases: PubMed, Scopus, the library of the Cochrane Collaborations, OvidSP MEDLINE, OvidSP MEDLINE In-Process and other non-indexed citations. Review articles and letters to the editor, articles that included <5 patients and articles that were written in languages other than English, French, German or Dutch were excluded.

Quality assessment & data extraction

Quality appraisal of retrieved full-text articles were all graded independently by two investigators for quality and applicability according to the Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS). This widely used tool consists of 14 items that cover patient spectrum, reference standard, disease progression bias, verification and review bias, clinical review bias, incorporation bias, test execution, study withdrawals and intermediate results. Disagreements were resolved by consensus after a re-evaluation of the references.

Individual patient data were extracted from the approved published articles for all patients with indeterminate (follicular or Hurthle cell (i.e., oxyphilic, oncocytic) proliferation) or (repeatedly) inconclusive FNAC in whom both an 18F-FDG PET study and surgery were performed and a final histopathologic diagnosis was available. Heterogeneity in PET acquisition and quantification of 18F-FDG uptake led to the conclusion that between-study comparison of quantitative parameters was impossible and, thus, was not attempted. Therefore, a positive test result was based solely on visual assessment of the PET scan and was defined as “any focal increased uptake in the region of the TN above background”. The final histopathologic diagnosis served as the gold standard outcome parameter. When final histopathology revealed microcarcinoma (defined as <1 cm in greatest dimension), then it was considered a coincidental finding in benign disease. The maximum lesion dimension measured on histopathology was noted for all patients. Because the greatest dimensions of the lesions previously described by our group were not published, they were retrieved...
9.2. Material & methods

![Flowchart of the article-selection process used in the current study.](image)

from the original data. In case of multiple nodules, bias caused by FNAC sampling error or spatial mismatch between the nodule and on the $^{18}$F-FDG was neglected. Therefore, a patient-based rather than a lesion-by-lesion analysis was performed.

**Data synthesis & statistical analysis**

Agreement in the per-item QUADAS score per reference between the two reviewers was expressed using the Cohen $\kappa$-coefficient and was interpreted according to the suggestions of Landis and Koch [363]. The intraclass correlation coefficient (ICC) using a two-way random-effects model with a definition of absolute agreement was used to quantify agreement for the overall score.
9. Diagnostic Performance of FDG PET in Thyroid Nodules

Heterogeneity of the study populations was assessed by comparing the distribution of possible confounders in the included references. Proportions (sex, FNAC results, prevalence of malignancy, histology results) were compared between references using either the $\chi^2$ statistic or the Fisher exact test. The continuous potential confounders (patient age, greatest tumour dimension) were assessed for (log-)normality (histograms, skewness, kurtosis, Shapiro-Wilk). In case of (log-)normality, mean values ($\pm$standard deviation (SD) and, in all other cases, median values (with interquartile range (IQR)) are presented. Comparison of (log-)normal distributed parameters between references was performed by using one-way analysis of variance (ANOVA) followed by a Tukey honestly significant difference (HSD) posthoc test. In case of violation of normality, the Kruskal-Wallis $H$ statistic was used as non-parametric equivalent for comparison between multiple independent groups and the Mann-Whitney $U$ test was used for comparisons between two independent groups.

Extracted data were ordered into 2·2 contingency tables from which disease prevalence and diagnostic test characteristics (sensitivity ($Se$), specificity ($Sp$), positive predictive value ($PPV$), negative predictive value ($NPV$), positive likelihood ratio ($LR$), negative $LR$, diagnostic odds ratio ($dOR$) and accuracy ($Acc$)) could be calculated using the classic equations. To avoid calculation problems by having zero values, 0.5 was added to each cell of the respective contingency table, which is a common method \[364\]. Ninety-five percent confidence intervals (95\% − CIs) for the proportions were calculated using the $\beta$-distribution (the exact Clopper-Pearson interval), because the commonly used, asymptotic, normal approximation only holds true for observed frequencies $>5$ individual patients \[156\].

Because sensitivity and specificity often are related inversely because of the threshold effect, study heterogeneity in these diagnostic test characteristics was observed using a summary receiver operating characteristic (sROC) curve for which the area under the curve ($AUC$) was calculated \[365366\]. Other causes of between-study heterogeneity in diagnostic test characteristics were assessed using $\chi^2$ statistics (heterogeneity was defined as $p < 0.10$) and were quantified by using the inconsistency index ($I^2$) (i.e., the amount of variability in the results attributable to between-study variation \[367\]). A funnel plot of the $dOR$ of each study was constructed to provide insight into publication bias \[368\]. Because distribution of the $dOR$ is approximately log-normal, correlation between the log($dOR$) and the standard error of the study effect was determined by the Kendall $\tau_b$ rank-correlation coefficient \[369\] and the Egger statistic \[370\], because a significant correlation suggests publication bias.

Pooled sensitivity and specificity were computed based on individual patient data (i.e., using the sum of the true-positive, true-negative, false-positive and false-negative individuals). Ratios (positive $LR$, negative $LR$ and $dOR$) were pooled using three strategies: 1) pooling individual patient data, which assumes negligible heterogeneity; 2) weighted (Mantel-Haenzel) averaging of per-study data, which assumes a fixed-effect and weights studies by their precision and, thus, is less sensitive to (small) inaccurate studies with extreme effects; and 3) weighted (DerSimonian-Laird) averaging of per-study data, which allows a random-effects and, thus, is less sensitive to study heterogeneity. Data are presented in forest plots and $\tau^2$ is used as a value for the between-study variance of the random-effects model.
All analyses were performed using the Statistical Package for Social Sciences (SPSS) version 16.0.2 for Mac (SPSS Inc., Chicago, IL, USA). A 2-tailed \( p < 0.05 \) was considered significant. All meta-analyses (pooling and sROC analysis) were performed using Meta-DiSc version 1.4 (Unit of Clinical Biostatistics, Ramon y Cajal Hospital, Madrid, Spain) [371].

9.3 Results

Literature search & study selection

After the removal of duplicates, the query resulted in 239 articles and, after we discarded references that fulfilled the exclusion criteria, 29 references remained (figure 9.1, page 147). Of these, 20 articles dealt with the prevalence of malignancy in thyroid incidentalomas identified on \(^{18}\)F-FDG PET/CT studies and nine articles dealt with the main search question [362,372–379].

Quality assessment

There was substantial agreement between both reviewers, who concurred on 104 of 126 QUADAS items (\( \kappa = 0.693; \ p < 0.001 \)). The correlation of absolute overall scores between both reviewers was moderate (\( ICC = 0.721; \ p = 0.010 \)). Three references were unsuitable for meta-analysis because of limitation to lesion-by-lesion analysis [373], lack of definition of \(^{18}\)F-FDG PET-positive results, lack of a final histopathologic diagnosis in each patient [377] or lack of clarity with respect to which and how many patients had indeterminate FNAC results [378]. None of the reviewed articles interpreted the \(^{18}\)F-FDG PET and histology data in combination with other clinical data that would be available in practice (QUADAS item 12). Only one article [373] described the methodology of \(^{18}\)F-FDG PET in sufficient detail to permit replication (QUADAS item 9) and only one article [379] mentioned blinding of the pathologist to the \(^{18}\)F-FDG PET findings (QUADAS item 11).

Study heterogeneity

The selected studies were performed in Austria (in an endemic goitre area) [372], the Netherlands [362], South Korea [374], Brazil [375] and the United States [376,379]. Four of the studies were performed in university hospitals [374,376,379], one was performed in a general hospital [372] and one was a multicentre trial that was performed in a university hospital and a general hospital [362]. In all but one study [376], all patients had thyroid-stimulating hormone levels within the normal range. The portion of patients with a single TN varied from 50% to 100%. The populations, inclusion criteria and exclusion criteria from these six studies are described in table 9.1 (page 150). Inclusion criteria varied somewhat among studies, mainly concerning nodule size and the preselection of patients who had nodules with suspicious ultrasonographic or scintigraphic features [372].
## Table 9.1: Characteristics of the patient populations included in the studies selected for the current meta-analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>City, Country</th>
<th>Hospital Inclusion criteria</th>
<th>Exclusion criteria</th>
<th>TSH</th>
<th>STN: No./total</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kresnik (2003)</td>
<td>Klagenfurt, Austria</td>
<td>G All patients with TN Hypoechogenic/no uptake on scintigraphy FNAC: follicular or Hürthle cell proliferation Scheduled for surgery</td>
<td>Autonomous goitre</td>
<td></td>
<td>24/43</td>
<td>Endemic goitre area</td>
</tr>
<tr>
<td>Kim (2007)</td>
<td>Seoul, South-Korea</td>
<td>U TN &gt;1 cm FNAC: follicular proliferation</td>
<td>–</td>
<td>Normal in all</td>
<td>32/46</td>
<td>–</td>
</tr>
<tr>
<td>Traugott (2010)</td>
<td>St. Louis, MO, USA</td>
<td>U Adults with TN or dominant TN Palpable or &gt;1 cm on US Scheduled for surgery</td>
<td>Previous neck surgery Previous radiotherapy</td>
<td>Normal in all</td>
<td>51/51</td>
<td>Interim analysis</td>
</tr>
</tbody>
</table>

DM: diabetes mellitus; FNAC: fine-needle aspiration cytology; G: general hospital; H&N: head and neck; (S)TN: (single) thyroid nodule; TSH: thyroid-stimulating hormone (thyrotropin); U: university hospital; US: ultrasonography.
<table>
<thead>
<tr>
<th>Reference</th>
<th>No. of patients</th>
<th>Women [%]</th>
<th>Mean age±SD [y]</th>
<th>Median Greatest histological lesion dimension (range) [mm]</th>
<th>FNAC results [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Follicular proliferation</td>
<td>Hürthle cell proliferation</td>
</tr>
<tr>
<td>Kresnik (2003) [372]*</td>
<td>37</td>
<td>78</td>
<td>55.1±13.8</td>
<td>16 (6-80)</td>
<td>65</td>
</tr>
<tr>
<td>de Geus-Oei (2006) [362]</td>
<td>44</td>
<td>93</td>
<td>48.5±13.8</td>
<td>30 (3-55)†</td>
<td>75</td>
</tr>
<tr>
<td>Kim (2007) [374]*</td>
<td>36</td>
<td>86</td>
<td>44.2±13.1</td>
<td>25 (10-90)</td>
<td>100</td>
</tr>
<tr>
<td>Sebastianes (2007) [375]</td>
<td>42</td>
<td>90</td>
<td>45.3±16.3</td>
<td>28 (4-85)</td>
<td>–</td>
</tr>
<tr>
<td>Hales (2008) [376]</td>
<td>15</td>
<td>93</td>
<td>47.5±14.9</td>
<td>25 (1-60)</td>
<td>80</td>
</tr>
<tr>
<td>Traugott (2010) [379]</td>
<td>51</td>
<td>80</td>
<td>49.6±10.6</td>
<td>15 (5-50)‡</td>
<td>69</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>225</strong></td>
<td><strong>86</strong></td>
<td><strong>48.5±13.9</strong></td>
<td><strong>20 (15-35)</strong></td>
<td><strong>62</strong></td>
</tr>
</tbody>
</table>

Table 9.2: Clinical characteristics of the patients included in the meta-analysis (extracted data). FNAC: fine-needle aspiration cytology; NOS: not otherwise specified; SD: standard deviation. *Presented data vary from published data: not all patients who were included in publications could be included in the current meta-analysis. †Lesion size was unavailable for nine lesions. ‡Lesion size was unavailable for one lesion. §Computed using the Fisher exact test. #Computed using a one-way analysis of variance; the difference was caused by older patients in the study by Kresnik et al. [372] compared with the patients in the studies by Kim et al. [374] and Sebastianes et al. [375] (Tukey honestly significant difference test; p = 0.010 and p = 0.020, respectively). ††Computed using the Kruskal-Wallis test.

\[ \text{Table 9.2: Clinical characteristics of the patients included in the meta-analysis (extracted data).} \]

\[ \text{FNAC: fine-needle aspiration cytology; NOS: not otherwise specified; SD: standard deviation.} \]

\[ ^* \text{Presented data vary from published data: not all patients who were included in publications could be included in the current meta-analysis.} \]

\[ ^\dagger \text{Lesion size was unavailable for nine lesions.} \]

\[ ^\ddagger \text{Lesion size was unavailable for one lesion.} \]

\[ ^\S \text{Computed using the Fisher exact test.} \]

\[ ^\# \text{Computed using a one-way analysis of variance; the difference was caused by older patients in the study by Kresnik et al. [372] compared with the patients in the studies by Kim et al. [374] and Sebastianes et al. [375] (Tukey honestly significant difference test; } p = 0.010 \text{ and } p = 0.020, \text{ respectively).} \]

\[ ^\†† \text{Computed using the Kruskal-Wallis test.} \]
In two studies, not all patients could be used for the current meta-analysis: six of 43 patients in one study [372] had papillary thyroid carcinoma (PTC) at FNAC and were reported as positive controls, ten of 46 patients in another study [374] refused surgery and, thus, no histopathologic diagnosis was available. Consequently, the data from 225 of 241 individual patients were available for pooling. Results of the US investigation were described for 35% of those patients.

Data on age, sex, FNAC results, $^{18}$F-FDG PET results and final histopathologic diagnosis were known on the individual patient level. Patient age, the maximum histologically measured nodule dimension and FNAC results, all of which are potential confounders that may cause heterogeneity in study results, differed significantly between the six included studies. The proportion of women was not significantly different between the six included studies (table 9.2, page 151).

$^{18}$F-FDG PET studies were obtained from all patients at least 17 days after FNAC [362,374,375], although three studies did not mention this interval [372,376,379]. The studies were obtained 60 to 70 minutes after the injection of 188 to 555 MBq of $^{18}$F-FDG and visually interpreted by one or two (blinded) experienced observers. A positive $^{18}$F-FDG PET result was defined as any $^{18}$F-FDG uptake greater than the background thyroid bed in all but one study [376], which added a further restriction that the maximum standardised uptake value ($SU_{max}$) of the lesion had to be greater than 2.0 g·cm$^{-3}$ to be considered $^{18}$F-FDG PET positive [376]. This quantitative restriction was not considered in the meta-analysis for the purpose of homogenizing the definition of a positive test. Therefore, the results from patients #5, #13 and #15 in the study by Hales et al. [376] were reassessed as PET-positive according to our definition. Another study [379] described four patients who had “incidental” PTC (0.3-17 mm in greatest dimension). Because those lesions were located distant from the nodules of interest, as observed on ultrasonography studies and because the nodules of interest were caused by another (benign) thyroid disease, the nodules were considered benign.

The prevalence of malignancy in the pooled population was 25.8% (range: 13.6-41.7%; Fisher exact test: 10.7; $p = 0.055$) (table 9.3, page 153). Of the benign disorders, multinodular goitre was the cause of TNs in 44% (range: 15-71%), follicular adenoma was the cause of TNs in 33% (range: 0-52%), Hürthle cell adenoma was the cause of TNs in 9% (range: 0-37%), (lymphocytic) thyroiditis was the cause of TNs in 4% (range: 0-33%) and a combination of benign disorders was the cause of TNs in 10% (range: 0-33%). The distribution of benign causes of TNs differed statistically between studies (Fisher exact test: 70.7; $p < 0.001$). Malignant disorders were caused mainly by FTC (35%; range: 0-73%), PTC (26%; range: 0-50%), Hürthle cell carcinoma (5%; range: 0-13%) and anaplastic carcinoma (3%; range: 0-20%). The remaining 31% of malignancies (range: 0-60%) were either a combination of malignant histologies, variants of FTC or PTC or malignancy not otherwise specified by the authors. The distribution of malignant causes of TNs differed statistically between studies (Fisher exact test: 38.2; $p < 0.001$).
### Table 9.3: Test characteristics of $^{18}$F-FDG PET for the detection of malignancy in thyroid nodules with indeterminate fine-needle aspiration cytology results according to the published literature.

<table>
<thead>
<tr>
<th>Reference</th>
<th>$^{18}$F-FDG PET</th>
<th>Cancer prevalence</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>NPV</th>
<th>PPV</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>TP</td>
<td>TN</td>
<td>FP</td>
<td>FN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kresnik (2003)</td>
<td>10</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>27.0 (13.8-44.1)</td>
<td>100 (69.2-100)</td>
<td>55.6 (35.3-74.5)</td>
</tr>
<tr>
<td>de Geus-Oei (2006)</td>
<td>6</td>
<td>25</td>
<td>13</td>
<td>0</td>
<td>13.6 (5.2-27.4)</td>
<td>100 (54.1-100)</td>
<td>65.8 (48.6-80.4)</td>
</tr>
<tr>
<td>Kim (2007)</td>
<td>15</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>41.7 (25.5-59.2)</td>
<td>100 (78.2-100)</td>
<td>0 (0-16.1)</td>
</tr>
<tr>
<td>Sebastianes (2007)</td>
<td>11</td>
<td>12</td>
<td>19</td>
<td>0</td>
<td>26.2 (13.9-42.0)</td>
<td>100 (71.5-100)</td>
<td>38.7 (21.8-57.8)</td>
</tr>
<tr>
<td>Hales (2008)</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>40.0 (16.3-67.7)</td>
<td>83.3 (35.9-99.6)</td>
<td>33.3 (7.5-70.1)</td>
</tr>
<tr>
<td>Traugott (2010)</td>
<td>8</td>
<td>25</td>
<td>16</td>
<td>2</td>
<td>19.6 (9.8-33.1)</td>
<td>80.0 (44.4-97.5)</td>
<td>61.0 (44.5-75.8)</td>
</tr>
<tr>
<td>Pooled</td>
<td>55</td>
<td>80</td>
<td>87</td>
<td>3</td>
<td>25.8 (20.2-32.0)</td>
<td>94.8 (85.6-98.9)</td>
<td>47.9 (40.1-55.8)</td>
</tr>
</tbody>
</table>

$^{18}$F-FDG PET: $2^{(18)}$fluoro-2-deoxy-D-glucose positron emission tomography; FN: false-negative; FNAC: fine-needle aspiration cytology; FP: false-positive; NaN: not a number; NPV: negative predictive value; PPV: positive predictive value; TN: true-negative; TP: true-positive. *Presented data vary from published data: not all patients who were included in publications could be included in the current meta-analysis.
Pooled data

The pooled sensitivity of $^{18}$F-FDG PET for the detection of cancer was 94.8% (95% − CI: 85.6-98.9%), but there was moderate though non-significant inconsistency among studies ($I^2 = 0.390$; $\chi^2$ test with 5 degrees of freedom ($df$): 8.2; $p = 0.146$). The pooled specificity was 47.9% (95% − CI: 40.1-55.8%), but there was high and significant inconsistency between studies ($I^2 = 0.867$; $\chi^2$ test with 5 $df$: 37.6; $p<0.001$), mainly because a specificity of 0% was reported in one study [374] in which all patients had positive $^{18}$F-FDG PET scans (i.e., there were no negative results). We could not identify a (methodological) cause and, thus, did not exclude that study from further analysis. The pooled NPV was 96.4% (95% − CI: 89.8-99.2%) and the pooled PPV was 38.7% (95% − CI: 30.7-47.3%) and the overall accuracy of $^{18}$F-FDG PET for the determination of TN malignancy was 60% (95% − CI: 53.3-66.5%) (Table 3).

On the basis of the pooled analysis of individual patient data, the positive LR was 1.82 (95% − CI: 1.56-2.13). By using fixed-effects, the positive LR was 1.56 (95% − CI: 1.36-1.80; high heterogeneity: $\chi^2$ test with 5 $df$: 76.0; $p<0.001$; $I^2 = 0.934$); however, considering the heterogeneity, random-effects modelling seemed more appropriate. With random-effects, the positive LR was 1.67 (95% − CI: 0.98-2.84; $\tau^2 = 0.39$). The negative LR could not be computed for one study [374], because there were no $^{18}$F-FDG PET-negative results. Therefore, pooling was based on the remaining five studies. By using fixed-effects, the negative LR was 0.19 (95% − CI: 0.076-0.46; low heterogeneity: $\chi^2$ test with 4 $df$: 2.33; $p = 0.676$; $I^2 < 0.001$). With random-effects, the negative LR was 0.24 (95%−CI: 0.10-0.59; $\tau^2 < 0.01$) (figure 9.2 page 155). Therefore, the pre-$^{18}$F-FDG PET probability of malignancy (prevalence) rose from 25.8% to 38.7% for a positive $^{18}$F-FDG PET (PPV) and decreased to 3.6% if the $^{18}$F-FDG PET results were negative (1-NPV).

The sROC curve revealed no “shoulder-arm” plot, suggesting no threshold effect (figure 9.3 page 156 and the AUC was $0.84\pm 0.079$. The symmetrical funnel plot of the dOR produced no evidence of publication bias (figure 9.4 page 157), which was confirmed by a non-significant correlation (Kendall $\tau_b = 0.067; p = 0.851$; Egger statistic: $t = -0.054; p = 0.959$).

In total, three of 58 patients who had differentiated thyroid carcinoma had false-negative $^{18}$F-FDG PET scans. Patient #1 in the study by Hales et al. [376] was diagnosed with a 14 mm PTC that had diffuse, moderate uptake ($SUV_{max} = 1.5$ g·cm$^{-3}$), similar to the thyroid background $^{18}$F-FDG uptake. Patients #1 and #12 in the study by Traugott et al. [379] had follicular variant PTCs that measured 7 mm and 9 mm, respectively, but no focal increase of $^{18}$F-FDG uptake was observed in the thyroid gland. These three false-negative lesions were significantly smaller in greatest dimension than other malignant lesions (median greatest dimension: 26 mm; $IQR$: 18-50 mm; Mann-Whitney $U$ test: 15.0; $p = 0.011$) as well as all other lesions (median greatest dimension: 21 mm; $IQR$: 15-35 mm; Mann-Whitney $U$ test: 72.0; $p = 0.014$). Thus, evaluation of the 164 TNs that measured >15 mm led to a sensitivity of $^{18}$F-FDG PET for the detection of thyroid carcinoma of 100% (95%−CI: 92.5-100%), whereas specificity remained similar (46.6%; 95%−CI: 37.4-56%) (figure 9.5 page 158).
9.4 Discussion

In this meta-analysis, $^{18}$F-FDG PET results were considered positive in 142 of 225 patients (63%) who had TNs with either indeterminate or repeatedly inconclusive FNAC cytology results. When patients consequently underwent surgery, almost 40% of these TNs were confirmed at final histopathology as carcinoma. A negative $^{18}$F-FDG PET result demonstrated malignancy in only three of 83 (3.6%) individual patients without focal $^{18}$F-FDG uptake, all of which had a greatest histologic dimension $<$1.5 cm. In the clinical work-up of patients with larger nodules, $^{18}$F-FDG PET is considered a useful tool and a negative $^{18}$F-FDG PET result can reliably exclude T2 malignancies.

We computed that it would require 72 patients with thyroid malignancy and all a positive $^{18}$F-FDG PET scans (i.e. measured sensitivity = 100%) to be reasonably confident that the true sensitivity is $>$95%, i.e., the lower level of the 2-tailed Clopper-Pearson 95% – CI or the solution to the following equation for n:

\[ 1 - \beta_1 - \frac{\alpha}{2} (n - x + 1, x) \geq 0.95 \]  
(9.1)
Figure 9.3: Summary receiver operating characteristic (sROC) curve. Sensitivity vs (1-specificity) are illustrated for 2-\(^{18}\text{F}\)fluoro-2-deoxy-D-glucose positron emission tomography in the detection of malignancy among patients who had thyroid nodules with indeterminate fine-needle aspiration cytology results along with corresponding boundaries of the 95%-confidence interval (represented by the outer curved lines). There are no signs of threshold effects. The area under the curve is \(0.84 \pm 0.079\).

where \(n = x\) and \(\alpha = 0.05\). Because, in the data presented, the prevalence of malignancy in TN was 25.8%, at least 276 patients who had TN with indeterminate FNAC results should have been included. To date, published studies have had sample sizes much smaller than this number; therefore, we undertook the current, comprehensive meta-analysis.

Because meta-analyses are prone to error caused by factors like as low study quality, study inhomogeneity and publication bias, these factors were minimised by careful selection and quality appraisal of references and descriptions of potential causes of heterogeneity and publication bias. This resulted in a pooled negative \(LR\) of 0.19 (fixed-effects) to 0.24 (random-effects), indicating that the pre-\(^{18}\text{F}\)-FDG PET probability of malignancy in these patients decreases from 26% (the prevalence) to 6.2% (fixed-effects) or 7.7% (random-effects) after a negative \(^{18}\text{F}\)-FDG PET scan. In other
9.4. Discussion

Figure 9.4: Funnel plot of the diagnostic odds ratios (dOR) of articles that were included in current analyses [362,372,374–376,379]. The solid vertical line denotes the pooled dOR and the dashed vertical lines indicate 95%-confidence intervals (95%–CI). The standard errors (SE) of the individual and pooled dORs were calculated using Mantel-Haenzels weighing (fixed-effects (FE)). The symmetric distribution of references can be observed. No correlation could be detected between loge(dOR) and SE(loge(dOR)) as reflected by Kendall τb = 0.067 (p = 0.851) and Egger statistic t = −0.054 (p = 0.959); therefore, there is no evidence for publication bias.

In words, the incorporation of 18F-FDG PET into the work-up of a TN with indeterminate FNAC results correctly saves 27 patients from undergoing diagnostic surgery at the cost of one patient who has surgical treatment unjustifiably delayed (80 of 225 patients correctly would not undergo surgery when 18F-FDG PET is incorporated into the work-up of these nodules (true-negative results); however, three of 225 patients would have surgical treatment unjustifiably delayed because of a false-negative result (80/3 ≈ 27)). The number of false-negative results can be decreased by only taking lesions that measure >15 mm in greatest dimension into consideration. This may be explained as a consequence of the partial volume effect because of the limited spatial resolution (approximately >6 mm full-width at half-maximum (FWHM)) of the PET scanners in use.

Currently, the total costs for 18F-FDG PET in the Netherlands is approximately €1,400 (in the United States, for comparison, the current reimbursement rate is approximately $1,050) and the treatment costs are driven mainly by the costs of surgery and hospitalisation, with mean costs per patient amounting to €3,311 for benign cases and €5,228 in malignant cases, without considering additional treatment costs, economic costs or indirect costs [380,381]. In our pooled population for the current meta-analysis, surgery in all patients would have resulted in an average cost per patients of €3,805 without the use of 18F-FDG PET and €3,958 with the
use of $^{18}$F-FDG PET in the Netherlands (approximately €3,608 in based on the reimbursement rate in the United States). Thus, both scenarios generate similar direct costs, whereas the cost of additional treatments, economic costs or indirect costs of futile surgery are not even considered. For this computation, we did not consider extra costs for the 1.3% false-negative $^{18}$F-FDG PET scans.

However, there remain other reasons to consider surgery despite this cost-effectiveness, including mechanical or cosmetic concerns and mere reassurance. Nonetheless, as in larger lesions, the complication rates can be higher (because of extension toward the large vessels, trachea or recurrent laryngeal nerve) and these patients may benefit most from reassurance of a true-negative PET. However, in patients with small lesions (<15 mm), the value of $^{18}$F-FDG PET should be weighed against the disadvantage of the risk of a false-negative result. When renouncing surgery in these patients, follow-up remains warranted.

A limitation of our meta-analysis is that different studies used different definitions for a positive $^{18}$F-FDG PET scan. E.g., Hales et al. [376] required a threshold $SUV_{max}$ of 2.0 g·cm$^{-3}$ for a positive test. We tried to use exactly the same definition for $^{18}$F-FDG PET positivity for each reference. However, because the $^{18}$F-FDG PET scans were not examined centrally, some inhomogeneity remained because of inter-observer variability in the interpretation of these images. Another limitation was

---

Figure 9.5: Graph depicting the size dependence of sensitivity ($Se$) and specificity ($Sp$). 95%-confidence intervals are provided with whiskers and between brackets. Test characteristics are of $2\times^{(18)}$F-fluoro-2-deoxy-D-glucose positron emission tomography for detecting malignancy in patients who had thyroid nodules with indeterminate fine-needle aspiration cytology results. $n$: number of patients for whom the diagnostic test characteristics were computed.
the definition of malignancy: Some studies excluded incidentally identified papillary microcarcinomas from participation in the study [374,375], others even considered malignant lesions that were located distant from the (benign) index nodule [379]. Of the four lesions in the latter category (0.3 mm, 4 mm, 8 mm and 17 mm PTCs), two were negative on $^{18}$F-FDG PET studies and, thus, also could be considered false-negative rather than true-negative. This would have increased the total number of false-negative results to five and, thus, would have decreased the pooled sensitivity from 95% (95% – CI: 86-99%) to 92% (95% – CI: 82-97%). A final limitation concerns the heterogeneity of the population, particularly the vast variation in the prevalence of malignancy (range: 14-42%) in different parts of the world with both endemic goitre areas and iodine-sufficient areas.

9.5 Conclusions

In conclusion, this comprehensive, systematic review and meta-analysis of the literature revealed that, in patients with TN who have indeterminate FNAC results, a negative $^{18}$F-FDG PET scan improves diagnostic accuracy, particularly for patients who have lesions >15 mm. All false-negative $^{18}$F-FDG PET results were from lesions <15 mm (i.e., T$_1$ tumours). A positive $^{18}$F-FDG PET scan increases the chance of malignancy from 25.8% to 38.7% in these patients. Further prospective series are ongoing and ultimately will reveal the value of $^{18}$F-FDG PET in the diagnostic evaluation of TNs and confirm these findings.

***
Cost-Effectiveness of $^{18}\text{F-FDG}$ PET/CT for Cytologically Indeterminate Thyroid Nodules: a Decision Analytic Approach

*J Clin Endocrinol Metab, 2014 Sep;99(9):3263-74*
(adapted and reprinted with permission)

D. Vriens¹, E.M.M. Adang², R.T. Netea-Maier³, J.W.A. Smit⁴, J.H.W. de Wilt⁵, W.J.G. Oyen⁴ and L.F. de Geus-Oei¹

dep. of ¹Nuclear Medicine, ²Health Evidence, ³Internal Medicine, Section Endocrinology, ⁴Internal Medicine and ⁵Surgery, Radboud University Medical Centre, Nijmegen, the Netherlands
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

Abstract

**Purpose** Patients with thyroid nodules of indeterminate cytology undergo diagnostic surgery according to current guidelines. In 75% of patients, the nodule is benign. In these patients, surgery was unnecessary and unbefitting because complications may occur. Preoperative $^{18}$F-FDG PET/CT was found to have a very high negative predictive value (96%) and might therefore avoid futile surgery, complications and costs. In the United States, two molecular tests of cytology material are routinely used for this purpose. Five-year cost-effectiveness for routine implementation of $^{18}$F-FDG PET/CT was evaluated in adult patients with indeterminate fine-needle aspiration cytology and compared with surgery in all patients and both molecular tests.

**Methods** A Markov decision model was developed to synthesise the evidence on cost-effectiveness about the four alternative strategies. The model was probabilistically analysed. One-way sensitivity analyses of deterministic input parameters likely to influence outcome were performed. The model was representative for adult patients with cytologically indeterminate thyroid nodules. The discounted incremental net monetary benefit ($iNMB$), the efficiency decision rule containing outcomes as quality-adjusted life-years and (direct) medical cost, of implementation of $^{18}$F-FDG PET/CT is displayed.

**Results** Full implementation of $^{18}$F-FDG PET/CT resulted in 40% surgery for benign nodules, compared with 75% in the conventional approach, without a difference in recurrence free and overall survival. The $^{18}$F-FDG PET/CT modality is the more efficient technology, with a mean $iNMB$ of €3,684 compared with surgery in all. Also, compared with a gene expression classifier test and a mutation marker panel, the mean $iNMB$ of $^{18}$F-FDG PET/CT was €1,030 and €3,851, respectively and consequently the more efficient alternative.

**Conclusions** Full implementation of preoperative $^{18}$F-FDG PET/CT in patients with indeterminate thyroid nodules could prevent up to 47% of current unnecessary surgery leading to lower costs and a modest increase of health-related quality of life. Compared with an approach with diagnostic surgery in all patients and both molecular tests, it is the least expensive alternative with similar effectiveness as the gene-expression classifier.

**Keywords** Cost-Effectiveness · Economic Analysis · 2-($^{18}$F)fluoro-2-deoxy-D-glucose · Positron Emission Tomography · Thyroidectomy · Thyroid Nodules
10.1 Introduction

Thyroid nodules are common and as many as 3-8% of European adults have palpable nodules, but the risk of differentiated thyroid carcinoma in these nodules is less than 5%. In healthy adults, a screening ultrasonography can detect asymptomatic thyroid nodules in up to 68% of volunteers [382]. Due to the increasing use of ultrasonography and other imaging techniques, more and more asymptomatic thyroid nodules are discovered, most of which have no clinical relevance. Once a nodule is established, screening for cancer is warranted because most of the thyroid carcinomas present as thyroid nodules [3]. In particular, in localised (∼68%) and regional (∼25%) stage at diagnosis, the prognosis of differentiated thyroid carcinoma is favourable because the five-year relative survival in these patients is greater than 97% [3].

In case of unsuppressed TSH, the recommended initial diagnostic test of a thyroid nodule according to current guidelines is ultrasonography-guided fine-needle aspiration cytology (FNAC) [345, 383]. Aspirates are classified into six diagnostic categories according to the Bethesda System for Reporting Thyroid Cytopathology [384]. In approximately 75% of patients, this will lead to a definite diagnosis and treatment, for benign, suspicious for malignancy or definite malignant disease. However, in the remaining cases, repetitive FNAC cannot determine whether the lesion is benign or malignant due to cellular atypia, follicular neoplasia or repetitive nondiagnostic or unsatisfactory specimens. Without further classification, in 69-88% of these patients, the nodule is found to be benign at diagnostic hemithyroidectomy (lobectomy) [385]. In most malignant nodules, secondary surgery with adjuvant treatment including radioactive ($^{131}$I)iodine thyroid remnant ablation (RRA) and TSH-suppression therapy is recommended. Only in the case of subcentimeter (pT$_{1a}$), indolent, unifocal papillary microcarcinoma, additional treatment is considered unnecessary [345].

The use of one of two molecular tests as an adjunct to diagnosis in FNAC-indeterminate thyroid nodules is standard of care in the United States. One, a 142-gene expression classifier (GEC), is used to minimise unnecessary diagnostic thyroid surgery and another one, a mutation marker panel (MMP), is used to select patients for initial total thyroidectomy, thereby saving on the two-step surgery. The GEC [386, 388] showed a positive (PPV) and negative predictive value (NPV) of 47% and 93%, respectively and was found to be cost effective [389]. Another molecular test [390] includes a MMP for mutations in BRAF and RAS and rearrangements in RET/papillary thyroid carcinoma (RET/PTC) and paired box transcription factor-8/peroxisomal proliferator-activated receptor-γ (PAX8/PPAR$\gamma$). It showed a PPV and NPV of 87% and 90%, respectively. Its limited NPV made the authors suggest an up-front total thyroidectomy after a positive test result and lobectomy otherwise. By saving on two-stage surgery, they showed a moderate increase in costs of nodule evaluation (+18% or US$ 104 per patient overall costs) [391]. Currently both these tests are unavailable in Europe or Asia.

Recently we summarised the data of 225 individual patients with indeterminate thyroid nodules from our own series [362] and five other published prospective studies (chapter 9) [385]. In all patients a 2-($^{18}$F)fluoro-2-deoxy-d-glucose ($^{18}$F-FDG) positron emission tomography (PET) was performed on previous-generation PET scanners (most without computed tomography (CT) capabilities and none with time of flight technology) prior to the scheduled surgery and therefore, gold standard his-
Cost-Effectiveness of FDG PET in Thyroid Nodules

10.0 Cost-Effectiveness of FDG PET in Thyroid Nodules

tology was available. We described a PPV and NPV of 39% and 96%, respectively. These data were recently confirmed by two prospective series of 55 and 46 patients, respectively [392, 393], concluding that $^{18}$F-FDG PET/CT could reduce the number of diagnostic (hemi)thyroidectomies by 13-25% [393]. Even though none of the studies summarised in our published meta-analysis [385] adopted the Bethesda criteria (five of six were published before its establishment [384]), confirmation of the performance of $^{18}$F-FDG PET/CT in a Bethesda classified population [392, 393] supports its predictive value in this population.

Based on the high NPV of $^{18}$F-FDG PET/CT to exclude malignancy in case of cellular atypia or follicular neoplasia in asymptomatic thyroid nodules, we hypothesise that its incorporation could reduce futile surgery from 74% to 39%. This would lead to fewer symptoms and cosmetic complaints of a neck scar. Also, fewer patients would need life-long daily thyroid hormone supplementation because up to one third of lobectomised patients have functional insufficiency of the remaining thyroid tissue [394]. Although rare, surgical complications may be severe (haemorrhage, infection, permanent hoarseness) [394, 396] and could be decreased using the proposed strategy.

Because surgery, hospitalisation, follow-up, $^{18}$F-FDG PET/CT and both molecular tests entail significant costs, current health economic evaluation was undertaken to model the potential impact of implementation of each one of these tests separately in the work-up of FNAC-indeterminate thyroid nodules on direct health care costs and patients’ health-related quality of life (HRQoL). We determined the cost-effectiveness of an $^{18}$F-FDG PET/CT driven approach compared with either a surgical approach (being the standard of care in Europe/Asia) or one of both molecular tests (USA standard).

10.2 Material & methods

Decision model

An eight (health)state Markov decision model, with yearly cycle length, was developed in accordance with the 2009 American Thyroid Association (ATA) guidelines for the management of patients with thyroid nodules [345] and the strategies proposed by the developers of both molecular tests [389, 391]. Treatment for adult patients with thyroid nodules that are scheduled for surgery based on indeterminate FNAC (Bethesda categories III and IV) was simulated, being driven by diagnostic thyroid surgery (surgery), a molecular test aiming at the prevention of unnecessary surgery (GEC), a molecular test aiming at the prevention of two-step surgery (MMP) and routine $^{18}$F-FDG PET/CT. Branches were developed to represent patient care after an indeterminate FNAC result (decision tree; figure 10.1 page 165), leading to one of eight potential health states. These health states include surveillance (after a negative $^{18}$F-FDG PET/CT or GEC), surveillance after thyroid surgery, permanent complications due to thyroid surgery, recurrence after thyroid surgery or death.
10.2. Material & methods

Figure 10.1: Decision tree. Simulated patients with FNAC-indeterminate TNs will either be treated based on diagnostic thyroid surgery, based on one of two molecular test or based on the result of $^{18}$F-FDG PET/CT. They will enter the Markov model (figure 10.2, page 166) in one of eight (health)states based on this decision tree. Diamonds are decision nodes and decision are based on probabilities. Boxes are interventions and cost money. (c)TT: (completion) total thyroidectomy; $^{18}$F-FDG PET/CT: 2-$^{18}$F)fluoro-2-deoxy-0-glucose positron emission tomography/computed tomography; FNAC: fine-needle aspiration cytology; GEC: gene-expression classifier; HT: hemithyroidectomy; MMP: mutation marker panel; MT: molecular test; PA: histopathology; RRA: radioactive ($^{131}$I)iodine thyroid remnant ablation; TN: thyroid nodule; UPM: unifocal papillary microcarcinoma. *(The initial diagnostic surgical approach (either HT or TT) may depend on e.g., the nodule size, location, ultrasonography features of the contralateral thyroid lobe. This decision node also is the entry point of the ‘Jump to “Type of Surgery”’ branches.

$^{18}$F-FDG PET/CT & molecular tests

Diagnostic performance of $^{18}$F-FDG PET/CT is based on the six studies summarised in our meta-analysis [385]. Diagnostic performance of the GEC is based on Li et al. [389] and for the MMP on Yip et al. [391]. In contrast with Yip et al. [391], we chose not to incorporate a repeated FNAC in any of the four study arms to homogenise the simulated clinical course in all patients.
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

Figure 10.2: Markov tree. Simulated patients with FNAC-indeterminate TNs will end up in any of these eight (health)states (ellipses), based on the decision tree (figure 10.1, page 165). After each cycle duration (one year), transitions to other health states may occur (arrows, transition probabilities). In the case of recurrence after HT, intervention (box) can take place, which has a certain decision (diamond). During surveillance after a negative $^{18}$F-FDG PET/CT at some point suspicion for malignancy might arise and the patient will undergo (diagnostic) thyroid surgery after all. Decision nodes are based on probabilities, interventions cost money and health states cost money and have a certain HRQoL. (c)TT: (completion) total thyroidectomy; $^{18}$F-FDG PET/CT: 2-($^{18}$F)fluoro-2-deoxy-d-glucose positron emission tomography/computed tomography; FNAC: fine-needle aspiration cytology; HRQoL: health-related quality of life; HT: hemithyroidectomy; RRA: radioactive ($^{131}$I)iodine thyroid remnant ablation; TN: thyroid nodule.

Because the different tests were originally benchmarked on different populations, with individual study cancer prevalence ranging from 20% [391] to 32% [389], we computed PPVs and NPVs based on an uniform a priori risk of malignancy of 25% (i.e., the weighted mean of all three study populations [385,389,391]) and the test sensitivities and specificities as stated in the original references.

Risk & probability estimation

The duration of each Markov cycle was considered to be one year; therefore, the transition between health states reflect annual probabilities governed by factors such as a priori probability of malignancy, surgical complication rates, recurrence rates and age- and sex-specific mortality rates (Markov tree: figure 10.2, page 166). Stochastic transition probabilities were collected from a variety of international literature sources including several other decision analyses on the diagnostic approach of an FNAC-indeterminate thyroid nodule (table 10.1, page 167). Missing parameter values or those that varied highly among literature were elicited from a panel consisting of six medical, surgical and imaging thyroid experts from the Radboud University Medical Centre, Nijmegen, the Netherlands and one health economist.
### 10.2. Material & methods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distr.</th>
<th>Expected Value</th>
<th>Source</th>
<th>Range for SA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discount rate of costs</strong></td>
<td>C</td>
<td>0.040</td>
<td>389, 397</td>
<td>0.030-0.050</td>
</tr>
<tr>
<td><strong>Discount rate of utilities</strong></td>
<td>C</td>
<td>0.015</td>
<td>389, 397</td>
<td>0.010-0.050</td>
</tr>
<tr>
<td><strong>Fraction of female patients</strong></td>
<td>β</td>
<td>0.86 (0.81-0.90)</td>
<td>362, 372, 374, 376, 379</td>
<td>0.78-0.93</td>
</tr>
<tr>
<td><strong>Age of female patient when diagnosed [year]</strong></td>
<td>γ</td>
<td>47.3 (21.0-73.6)</td>
<td>362, 372, 374, 376, 379</td>
<td>–</td>
</tr>
<tr>
<td><strong>Age of male patient when diagnosed [year]</strong></td>
<td>γ</td>
<td>55.6 (26.1-85.0)</td>
<td>362, 372, 374, 376, 379</td>
<td>–</td>
</tr>
<tr>
<td><strong>Incidence of DTC in healthy females</strong></td>
<td>β</td>
<td>3.1·10^{-6} (2.1-4.3·10^{-6})</td>
<td>398</td>
<td>–</td>
</tr>
<tr>
<td><strong>Incidence of DTC in healthy males</strong></td>
<td>β</td>
<td>1.3·10^{-6} (0.69-2.2·10^{-6})</td>
<td>383</td>
<td>–</td>
</tr>
<tr>
<td><strong>Yearly probability of surgery after surveillance</strong></td>
<td>C</td>
<td>Age/sex dependent life table</td>
<td>398</td>
<td>–</td>
</tr>
<tr>
<td><strong>Fraction HT of all surgery</strong></td>
<td>β</td>
<td>0.95 (0.90-0.98)</td>
<td>EO</td>
<td>0.50-0.99</td>
</tr>
<tr>
<td><strong>Fraction of UPM in indeterminate nodules</strong></td>
<td>β</td>
<td>0.023 (0.0076-0.047)</td>
<td>362, 372, 374, 376, 379</td>
<td>0.01-0.10</td>
</tr>
<tr>
<td><strong>Prevalence of cancer in indeterminate nodules</strong></td>
<td>Dir</td>
<td>0.25 (0.22-0.28)</td>
<td>362, 372, 374, 376, 379</td>
<td>0.15-0.35</td>
</tr>
<tr>
<td><strong>18F-FDG PET/CT sensitivity</strong></td>
<td>Dir</td>
<td>0.95 (0.88-0.99)</td>
<td>385, 392, 393</td>
<td>0.70-0.99</td>
</tr>
<tr>
<td><strong>18F-FDG PET/CT specificity</strong></td>
<td>Dir</td>
<td>0.48 (0.40-0.55)</td>
<td>385, 392, 393</td>
<td>0.35-0.70</td>
</tr>
<tr>
<td><strong>GEC sensitivity</strong></td>
<td>Dir</td>
<td>0.92 (0.85-0.97)</td>
<td>385, 392, 393</td>
<td>0.65-0.99</td>
</tr>
<tr>
<td><strong>GEC specificity</strong></td>
<td>Dir</td>
<td>0.52 (0.44-0.59)</td>
<td>385, 392, 393</td>
<td>0.40-0.75</td>
</tr>
<tr>
<td><strong>MMP sensitivity</strong></td>
<td>Dir</td>
<td>0.59 (0.49-0.69)</td>
<td>390, 391</td>
<td>0.35-0.70</td>
</tr>
<tr>
<td><strong>MMP specificity</strong></td>
<td>Dir</td>
<td>0.98 (0.96-0.99)</td>
<td>390, 391</td>
<td>0.75-0.99</td>
</tr>
<tr>
<td><strong>Yearly probability of recurrence after HT for UPM</strong></td>
<td>β</td>
<td>0.0070 (0.0014-0.021)</td>
<td>computed (= 1 – NPV^{-0.2})</td>
<td>0.00-0.05</td>
</tr>
<tr>
<td><strong>Yearly probability of recurrence after (c)TT</strong></td>
<td>β</td>
<td>0.0019 (0.0009-0.0032)</td>
<td>computed (= 1 – NPV^{-0.2})</td>
<td>0.00-0.01</td>
</tr>
<tr>
<td><strong>Yearly probability of cTT after recurrence after HT</strong></td>
<td>β</td>
<td>0.0047 (0.0002-0.016)</td>
<td>407</td>
<td>0.001-0.025</td>
</tr>
<tr>
<td><strong>Yearly probability of recurrence after (c)TT</strong></td>
<td>β</td>
<td>0.017 (0.009-0.023)</td>
<td>389, 408</td>
<td>0.90-1.00</td>
</tr>
<tr>
<td><strong>Yearly probability of death due to cancer</strong></td>
<td>β</td>
<td>0.0051 (0.0020-0.0095)</td>
<td>389, 408</td>
<td>0.00-0.01</td>
</tr>
</tbody>
</table>

Table 10.1: Accountability of base-case parameter values and stochastic distributions (distr.) for base parameters and transition probabilities. 95% – CI: 95%-confidence interval; C: constant; (c)TT: (completion) total thyroidectomy; Dir: Dirichlet; DTC: differentiated thyroid carcinoma; EO: expert opinion; 18F-FDG PET/CT: 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; GEC: gene-expression classifier; HT: hemithyroidectomy; MMP: mutation marker panel; NPV: negative predictive value; SA: one-way sensitivity analysis; UPM: unifocal papillary microcarcinoma.
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

Table 10.2: Accountability of base case parameter values for costs and utilities. All cost parameters were assumed to be $\gamma$-distributions and all utility parameters $\beta$-distributions. 95% – CI: 95% confidence interval; (c)TT: (completion) total thyroidectomy; DOT: the system of imbursement of the NZa; EO: expert opinion; $^{18}$F-FDG: 2-($^{18}$F)fluoro-2-deoxy-D-glucose; GEC: gene-expression classifier;

Cost & utility estimation

The Markov state information contained costs and utilities with a time frame of one year.

The model considers stochastic direct medical costs data (table 10.2, page 168). These were derived from 2012 reimbursement rates of the Dutch system of Diagnosis-Treatment Combinations and published in the international literature. All prices were indexed to January 2013 euros, using country-specific consumer price indexes [416–419] and up-to-date exchange rates [420] (January 2013: €1.00000 \equiv \text{US$} 1.31139 \equiv \text{CAN$} 1.32909). These prices include reimbursement tariffs for the molecular test, $^{18}$F-FDG PET/CT, to physicians, anaesthesia, pathology, laboratory investigations, ultrasonography procedures, thyroid surgery, RRA, medication, hospital facilities and all other costs incurred during inpatient and outpatient treatment. The costs of both transient and permanent complications were based on estimates from literature; its wide distribution reflects the variety of the severity of these complications.

Utilities for each cycle in a particular health state were derived from literature (table 10.2, page 168). Quality-adjusted life-years (QALYs) were calculated by the discounted sum of utilities over the five-year evaluation period. Utility values from the literature were used where available or elicited from a previously mentioned expert panel based on time-trade-off weighting.
10.2. Material & methods

All costs and utilities were exponentially discounted at a constant rate of 4.0% and 1.5% per year, respectively [397].

Base case cost-effectiveness analysis & sensitivity analyses

For the base-case scenario, the model has been run in a probabilistic fashion, with microsimulation of 100,000 first-order trials (patients) for 10,000 second-order parameter samples over five cycles. A five-year evaluation period was chosen because most costs (and HRQoL losses) are made in the first years, the ATA guidelines [345] have difficulty in providing recommendations after the first five years, greater than 50% of recurrences occur in the first two years [383,421] and only limited data exist supporting the fact that probabilities, costs and effects after the initial first five years differ between scenarios. Half-cycle correction was applied. Results are displayed in a cost-effectiveness plane [422].

One-way sensitivity analyses were performed to explore the variation of base-case model parameters on their range of extremes (10,000 hypothetical patients, 1,000 second-order parameter samples). One way-sensitivity analyses for transition probabilities, costs and utilities were performed over a wide range of values identified from the literature (tables 10.1-10.2, pages 167-168). Among the parameters examined are parameters connected to the procedure and follow-up after hemithyroidectomy,
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

the procedure and the follow-up after $^{18}$F-FDG PET/CT, the performance of molecular tests and $^{18}$F-FDG PET/CT (sensitivity, specificity) and the demographics of the population under review (prevalence of malignancy in thyroid nodules).

The mean costs and utilities acquired during this five-year period for each scenario were used to compute the incremental net monetary benefit ($iNMB$ in €):

$$iNMB = \lambda \cdot (E_2 - E_1) - (C_2 - C_1)$$  \hspace{1cm} (10.1)

where $\lambda$ is the willingness-to-pay threshold, $E$ is the effects (utilities) and $C$ is the costs of both scenarios under comparison. The subscript 1 denotes the comparator (surgery, GEC or MMP) and 2 denotes $^{18}$F-FDG PET/CT-driven treatment [423]. From the $iNMB$, the decision rule for cost-effectiveness can be inferred: $iNMB > €0$.

The Dutch Council for Public Health and Health Care recommends a willingness-to-pay threshold of €80,000·QALY$^{-1}$ for conditions with a maximal disease burden [424] and this is used throughout this study. However the cost-effectiveness acceptability curve, defined as the probability of $iNMB > €0$ for a wide willingness-to-pay range, is displayed.

modelling and Monte-Carlo analysis were performed using TreeAgePro Suite version 2011 (TreeAge Software Inc., Williamstown, MA, USA). Data analyses were performed using MATLAB version R2013a (The MathWorks Inc., Natick, MA, USA).

10.3 Results

Base case cost-effectiveness analysis

After five years of treatment for and follow-up after an FNAC-indeterminate thyroid nodule, mean discounted costs were €8,804 (surgery), €9,341 (GEC), €8,913 (MMP) and €7,983 ($^{18}$F-FDG PET/CT).

Their mean discounted utilities were 4.52, 4.56, 4.52 and 4.55 QALY, respectively. Therefore, $^{18}$F-FDG PET/CT driven surgery proved to be the more efficient alternative, being on average €822 less expensive per patient with moderately higher HR-QoL of 0.036 QALY over five years compared with surgery in all patients. Compared with GEC and MMP, it was €1,358 and €930 less expensive with slight differences in HRQoL over five years. The mean $iNMB$ was €3,684 compared with surgery, €1,030 compared with GEC and €3,851 compared with MMP (table 10.3, page 171). The robustness of these findings is displayed in the cost-effectiveness plane in figure 10.3 (page 172): all of the 10,000 projections actually show a reduction of costs of $^{18}$F-FDG PET/CT compared with the other three strategies. None of these 10,000 simulations indicated that $^{18}$F-FDG PET/CT would be more costly and less effective, less costly and less effective or more costly and more effective except in comparison with the GEC, in which PET showed a lower HRQoL of 0.0040 QALY (i.e., 1.5 quality-adjusted life days). This makes a convincing case that the $^{18}$F-FDG PET/CT modality is the most efficient approach. For the willingness-to-pay range of €0–€80,000·QALY$^{-1}$, the probability of a positive $iNMB$ equals 1 for $^{18}$F-FDG PET vs any of the other three strategies (figure 10.4, page 173).

The fraction of futile surgery of histologically benign, FNAC-indeterminate thyroid nodules was 75% (surgery), 38% (GEC), 75% (MMP) and 40% ($^{18}$F-FDG PET/CT), respectively. Therefore, unbeneﬁcial surgery could potentially be decreased by up to
37% and 35% by the full implementation of GEC and $^{18}$F-FDG PET/CT, respectively. This would lead to a reduction of surgery-related (permanent) complications (including surgery related death) from 7.7% (surgery or MMP) to 4.4% (GEC) or 4.6% ($^{18}$F-FDG PET/CT), i.e., almost halving unbeneficial surgery and surgery-related complications. Mean five-year overall and recurrence-free survival in this population was similar in all four strategies, being 96.5% and 97.2% respectively.

One-way sensitivity analyses

The most influential parameter (under assumptions of independency) was found to be the utility attributed to watchful surveillance (after a negative $^{18}$F-FDG PET/CT scan or GEC). At the minimum evaluated value (0.90), a worse quality of life was found for $^{18}$F-FDG PET/CT-driven treatment vs either thyroid surgery in all patients or MMP (in both mean incremental utility: -0.10 QALY), leading to a mean iNMB of €-7,418 (vs surgery) and €-7,264 (vs MMP). At a value for the utility attributed to watchful surveillance of 0.953 (vs surgery) or 0.952 (vs MMP), the mean iNMB equals €0. For comparison, the utility attributed to the health state after uncomplicated hemithyroidectomy is set at 0.99.

Other parameters that proved influential in affecting cost-effectiveness included the utility of surveillance and permanent complications after hemithyroidectomy, the probability of hemithyroidectomy-induced (transient and permanent) complications, the probability of performing hemithyroidectomy as the primary method for thyroid surgery and surgical mortality as well as the costs of a hemithyroidectomy procedure. In comparison with the GEC, which has a similar place in the work-up as $^{18}$F-FDG PET/CT, the crucial parameters leading to a preference of GEC over $^{18}$F-FDG PET/CT were the test specificity of both (sensitivity and specificity), the cost-price of the GEC, the test sensitivity of the GEC and the yearly probability that sur-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Surgery</th>
<th>GEC</th>
<th>MMP</th>
<th>$^{18}$F-FDG PET/CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute values, mean (95% − CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costs [€]</td>
<td>8,804 (8,774-8,835)</td>
<td>9,341 (9,300-9,383)</td>
<td>8,913 (8,884-8,942)</td>
<td>7,983 (7,941-8,025)</td>
</tr>
<tr>
<td>futile surgery [%]</td>
<td>75.0% (74.8-75.3%)</td>
<td>38.2% (37.8-38.5%)</td>
<td>75.0% (74.8-75.3%)</td>
<td>40.3% (39.9-40.7%)</td>
</tr>
</tbody>
</table>

| Incremental values of $^{18}$F-FDG PET/CT vs alternative strategy, mean (95% − CI) |
| incremental costs [€] | -822 (-871- -772) | -1,358 (-1,377- -1,340) | -930 (-970- -980) |
| incremental utilities [QALY] | +0.036 (+0.031+0.041) | -0.0040 (-0.0050- -0.0030) | +0.037 (+0.033+0.041) |
| iNMB [€] | +3.684 (+3.278+4.094) | +1.030 (+0.916+1.142) | +3.528 (+3.528+4.170) |
| incremental futile surgery [%] | -34.7% (-35.2- -34.3%) | +2.1% (+2.0+2.3%) | -34.7% (-35.2- -34.3%) |

Table 10.3: Base case main results over a duration of 5 years. 95% − CI: 95%-confidence interval; $^{18}$F-FDG PET/CT: $^{18}$F-fluorodeoxyglucose positron emission tomography/computed tomography; GEC: gene-expression classifier; iNMB: incremental net monetary benefit (using a willingness-to-pay threshold (λ) of €80,000-QALY$^{-1}$); MMP: mutation marker panel; QALY: quality-adjusted life year.
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

Figure 10.3: Incremental costs–incremental utility plot (cost-effectiveness plane) comparing $^{18}$F-FDG PET/CT-driven treatment with current practice. Each of the 10,000 dots represents the mean value of 100,000 simulated patients. The left upper quadrant represents situations in which the novel strategy is less effective but more costly than the conventional treatment; the right upper quadrant represents more effective and more costly treatment; left lower quadrant represents less effective and less costly treatment; the right lower quadrant represents more effective but less costly treatment. The oblique lines represent a willingness-to-pay threshold of €20,000-QALY$^{-1}$ and €80,000-QALY$^{-1}$, respectively. 95%-confidence ellipses are drawn. $^{18}$F-FDG PET/CT: 2-(18F)fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; GEC: gene-expression classifier; HRQoL: health-related quality of life; MMP: mutation marker panel; QALY: quality-adjusted life year.

gery has to be performed after a (false-negative) $^{18}$F-FDG PET/CT. For the range of the prevalence of thyroid carcinoma tested (15-35%), $^{18}$F-FDG PET/CT was the preferred modality over the GEC (figures 10.3, 10.6, pages 174-177).

10.4 Discussion

We presented an economic decision analytical model, forecasting that implementation of $^{18}$F-FDG PET/CT in the work-up of FNAC-indeterminate thyroid nodules could lead to a substantial reduction in direct medical costs and, compared with two of the three alternatives, modestly improvement of patients HRQoL over the duration of five years.
Figure 10.4: Cost-effectiveness acceptability curves. A plot of the probability of a positive $iNMB$ ($p(iNMB > 0)$) for a range of values for the willingness-to-pay threshold ($\lambda$). The dotted line is at a willingness-to-pay threshold of $\mathcal{E}80,000$ per QALY. $^{18}$F-FDG PET/CT: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; GEC: gene-expression classifier; $iNMB$: incremental net monetary benefit; MMP: mutation marker panel; QALY: quality-adjusted life year.

Avoidance of (complications of) unnecessary thyroid surgery to provide a definite histopathological diagnosis is the principal cause of cost-reduction. The fraction of surgeries performed for a benign thyroid nodule could almost be halved when fully implementing $^{18}$F-FDG PET/CT compared with thyroid surgery in all FNAC-indeterminate thyroid nodules (40.3% and 75.0%, respectively). Because it is estimated that 60,220 men and women are diagnosed with differentiated thyroid carcinoma (DTC) in the United States in 2013 [8] and about half are found after surgery for FNAC-indeterminate nodules [425, 426], it can be approximately estimated that 120,000 patients undergo thyroid surgery for a FNAC-indeterminate thyroid nodule, of whom 90,000 undergo surgery for a benign disease. Full implementation of $^{18}$F-FDG PET/CT could save up to 42,000 unnecessary surgeries annually, $\mathcal{E}$99 million and 4.3 thousand QALYs in the United States only, assuming FNAC indeterminacy was the sole reason for thyroid surgery. Compared with those in the USA current practice of GEC, a change from full implementation of GEC to $^{18}$F-FDG PET/CT could potentially result in an annual cost-reduction of $\mathcal{E}$164 million. On the draw-
Figure 10.5: Tornado plots showing the results of one-way sensitivity analyses of the top ten inputs of the model. The results on the \(iNMB\) of\(^{18}\)F-FDG PET/CT vs one of the other three strategies (upper panel: surgery; middle panel: GEC; lower panel: MMP), for a willingness-to-pay threshold (\(\lambda\)) of \(€80,000\cdot QALY^{-1}\), the whiskers represent the limits of the 95%-confidence interval (95% – CI); the ranges of tested values tested are between parentheses. The vertical dotted line is set at the mean \(iNMB\) of the base-case scenario. The vertical line at \(€0\) represents the break-even situation at a willingness-to-pay threshold of \(€80,000\cdot QALY^{-1}\). (c)TT: (completion) total thyroidectomy; \(^{18}\)F-FDG PET/CT: 2-(\(^{18}\)F)fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; GEC: gene-expression classifier; HT: hemithyroidectomy; \(iNMB\): incremental net monetary benefit; MMP: mutation marker panel; QALY: quality-adjusted life year.

back, the somewhat lower specificity of \(^{18}\)F-FDG PET/CT compared with the GEC might lead to a modestly higher fraction of surgery for benign nodules of 2.1%, responsible for a negligible (but negative) effect on HRQoL (table 10.3 (page 171) and figure 10.3 (page 172)).

We found a higher mean five-year discounted costs of \(€8,913\) (MMP) compared with \(€8,804\) (surgery). This is similar to the published economic analysis [391], which describes an additional US$ 104 to the overall cost of nodule evaluation only. The numerical difference can be explained by the fact that Yip et al. [391] allowed a second FNAC in case of a negative MMP, which is able to revoke FNAC indeterminacy and thus futile surgery.

Compared with the economic analysis of the GEC [389], we found higher mean five-year discounted costs of \(€9,341\) (GEC) compared with \(€8,804\) (surgery), whereas these authors describe a lower economic burden when adopting the GEC (US$ 10,719 compared with US$ 12,172). The main reason explaining our contrary conclusion is
that we attribute lower values for surgery and surgery-related costs than they do. E.g., in our model uncomplicated hemithyroidectomy plus five-year follow-up would cost €5,499, but in their model this would be US$ 10,319 (≡ €8,311, indexed to January 2013). Because we adopted the same cost-price of the GEC, this example shows that in our model the prevention of one uncomplicated surgery by the GEC equals the costs of two diagnostic tests only, whereas in their model it saves enough to pay for more than three GECs. This is further supported by the fact that the costs attributed to the GEC was one of the most influential determinants in one-way sensitivity analysis (figure 10.5, page 174, middle panel).

Modest improvement of HRQoL was found as long as estimated HRQoL of surveillance after a negative $^{18}$F-FDG PET/CT was higher than 0.95, this parameter was found to be the sole parameter that could lead to a situation in which an $^{18}$F-FDG PET/CT-driven approach did not dominate current European practice of surgery in all patients and even a decremental net monetary benefit. To the best of our knowledge, currently there have been no prospective studies published that investigate the HRQoL of a wait-and-see policy in benign thyroid nodules. To further substantiate this parameter and our results, a prospective study should be undertaken to investigate the consequences of implementation in daily practice with respect to (in)direct costs, measured HRQoL and other measures of effectiveness.

The HRQoL attributed to surveillance after a negative $^{18}$F-FDG PET/CT could be depreciated due to factors related to the thyroid nodule itself or to the fear of a false-negative $^{18}$F-FDG PET/CT result (1.3% of all $^{18}$F-FDG PET/CT scans performed were false-negative [385]). The former can be prevented by not offering $^{18}$F-FDG PET/CT in case thyroid surgery is considered for other than mere diagnostic purposes only. A false-negative $^{18}$F-FDG PET/CT scan could delay treatment for thyroid malignancy. Our model assumes that on average most of these are treated during a five-year follow-up period. Outcome with respect to progression-free and overall survival, costs and HRQoL is not known for delayed treatment; therefore, no additional costs or detrimental effects are incorporated into the model. However, the oncological-, economical- and HRQoL-related consequences are considered to be minimal due to the relative indolent course of this disease. Furthermore, there is limited impact on survival on the transition from localised to regional disease (five-year relative overall survival: 99.9% and 97.4%, respectively [3]), all with good treatment options. Finally, the false-negative ratio is based on the sensitivity of $^{18}$F-FDG PET/CT, which was found to be highly dependent on the scanners’ resolution (5-8 mm full-width at half-maximum (FWHM) for the PET scanners used in the meta-analysis). With state-of-the-art time-of-flight technology (3-4 mm FWHM), it is likely that sensitivity, and thus $NPV$, are higher and that the false-negative cases that occur are the smallest DTCs.

General weaknesses of any model are oversimplification of daily practice and the accuracy of the definition of each parameter. However, the current model was designed closely adhering to the ATA guidelines. By using data from a variety of sources including international literature, government publications, guidelines and expert estimates and allowing a stochastic uncertainty in these estimates, we substantiated the generalisability of the model.
When the available literature showed heterogeneous parameter values, we elicited these from an expert panel because we expected that this variation was both based on study heterogeneity and threshold effect due to unclear definitions. E.g., parameter values for probability, costs and utility of complications highly depend on what the authors define as complication: if a minor bleeding is included in the definition of a transient complication, the probability of having a transient complication will increase, the average costs will decrease and the average utility will probably increase. By adopting a higher scale parameter, determining the statistical dispersion of the distribution, we tried to cover these higher uncertainties.

It is likely that the value of QALYs rise over time and because this rise is not taken into account by other means in an economic evaluation, it is suggested to discount utilities with a lower rate than costs. Therefore, we adopted a non-uniform discount rate for costs and utilities. Because non-uniform discounting is still uncommon in the international literature, we repeated the analyses of the base-case scenario with a uniform discount rate of 3\%-year\(^{-1}\) for both costs and utilities, showing no different conclusion.

One-way sensitivity analyses over a plausible but wide range of parameter estimates showed that the outcome of the simulations were most critically influenced by the utility of surveillance after a negative \(^{18}\)F-FDG PET/CT or hemithyroidectomy, costs of hemithyroidectomy, fractions and utilities of hemithyroidectomy-induced complications (including death), distribution of the initial type of surgery and \(^{18}\)F-FDG PET/CT sensitivity and specificity. Furthermore, only direct costs for a five-year duration were computed. One could argue that indirect costs (e.g., sick leave days, decreased productivity and money spent on care outside the medical setting), would further support the inclusion of \(^{18}\)F-FDG PET/CT in the diagnostic algorithm.

A limitation of the sensitivity analyses is the assumption of independency. The parameters in the model are clearly related due to threshold effects. Because these relationships are complex and because it is impossible to accurately substantiate any assumption as to the quantitative relationship between these parameters, this was not attempted, and a wide range value for the sensitivity analyses was chosen.

Due to the limited specificity and positive predictive value, still 40% of patients undergo thyroid surgery for a benign thyroid nodule. The only independent predictive factor for \(^{18}\)F-FDG uptake in literature was cellular atypia (present in both benign and malignant nodules). The current literature mainly focuses on \(^{18}\)F-FDG uptake in known thyroid carcinoma or (in vitro) in thyroid cells; therefore, the limited specificity of \(^{18}\)F-FDG PET/CT for (FNAC-indeterminate) thyroid nodules is still poorly understood.

Test characteristics of \(^{18}\)F-FDG PET/CT are based on populations with a heterogeneous fraction of people suffering from multinodular disease (15-71% \[385\]-\[392\]-\[393\]), which might influence results for two reasons: 1) from a methodological point of view, the nodule under investigation by FNAC, \(^{18}\)F-FDG PET/CT and histopathology might not be the same and 2) the result of a negative \(^{18}\)F-FDG PET/CT might not modify surgical treatment decision because reasons other than merely indeterminate FNAC might be the reason for surgery. In practice the former issue is being by most studies by only including patients with a clear, dominant nodule. The latter can be overcome by offering \(^{18}\)F-FDG PET/CT to patients that are scheduled for surgery only for reason of indeterminate FNAC. Although this further selected population is
different from that we obtained the negative and positive predictive value of $^{18}$F-FDG PET/CT, we believe that the robustness of our main conclusions shown by one-way sensitivity analysis is still valid for a wide range of values. The global impact might be overestimated because not all patients with a FNAC-indeterminate thyroid nodule and a negative $^{18}$F-FDG PET/CT might wish to refrain from surgery.

10.5 Conclusions

In conclusion, our cost-utility analysis demonstrates that full implementation of $^{18}$F-FDG PET/CT in the work-up of adult patients with thyroid nodules scheduled for surgery for FNAC-indeterminacy (i.e., cellular atypia and follicular neoplasia) could lead to a decrease in costs and a moderate increase in HRQoL compared with diagnostic surgery in all patients according to current European practice and is competitive to the current USA standard of the GEC. These results are primarily based on a decrease in costs and complications of surgery in patients with benign thyroid nodules that are not resected for being symptomatic. Sensitivity analyses showed the robustness of these data. Prospective studies are needed to further support cost-effectiveness and implementability and to gain insight in the false positivity of $^{18}$F-FDG PET/CT. A prospective head-to-head comparison with alternative strategies or combinations of strategies should be considered.

***

Figure 10.6: Tornado plots showing the results of one-way sensitivity analysis of all inputs of the model on the $iNMB$ vs one of the other three strategies (next pages). Panel a: surgery, Panel b: GEC, Panel c: MMP. The willingness-to-pay threshold ($\lambda$) was €80,000-QALY$^{-1}$. The whiskers represent the limits of the 95%-confidence interval; the ranges of tested values tested are between parentheses. The vertical dotted line is set at the mean $iNMB$ of the base-case scenario. The vertical line at €0 represents the break-even situation at a willingness-to-pay threshold of €80,000-QALY$^{-1}$. (c)TT: (completion) total thyroidectomy; $^{18}$F-FDG PET/CT: 2-($^{18}$F)fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; GEC: gene-expression classifier; HT: hemithyroidectomy; iNMB: incremental net monetary benefit; MMP: mutation marker panel; QALY: quality-adjusted life year; RRA: radioactive ($^{131}$I)iodine thyroid remnant ablation; UPM: unifocal papillary microcarcinoma.
10. Cost-Effectiveness of FDG PET in Thyroid Nodules
10. Cost-Effectiveness of FDG PET in Thyroid Nodules

FDG-PET/CT vs. MMP

Incremental Net Monetary Benefit (NMB) [thousand €] (Willingness-to-Pay threshold (λ): € 80,000/QALY)

Variable (range tested)

- Utility of surveillance after FDG-PET/CT or GEC (0.90 - 0.99)
- Utility of surveillance after HT (0.90 - 0.99)
- Fraction of permanent complications due to HT (0.01 - 0.25)
- Utility of permanent complication due to HT (0.01 - 0.25)
- Fraction of transient complications due to HT (0.01 - 0.25)
- FDG-PET/CT specificity (0.36 - 0.70)
- Costs of TT (€ 3,433 - € 20,796)
- Costs of HT (€ 2,994 - € 16,679)
- Fraction of permanent complications due to c(T)T (0.01 - 0.25)
- FDG-PET/CT sensitivity (0.70 - 0.99)
- Utility of surveillance after c(T)T (0.90 - 0.99)
- Fraction of HT of all surgery (0.50 - 0.99)
- Prevalence of cancer in indeterminate nodules (0.15 - 0.35)
- MMP specificity (0.75 - 0.99)
- Costs of c(T)T (€ 3,892 - € 16,878)
- Fraction of UPM in indeterminate nodules (0.01 - 0.10)
- Yearly probability of surgery after surveillance after negative FDG-PET/CT (0.00 - 0.05)
- Fraction of death due to any type of surgery (0.00 - 0.01)
- Utility of permanent complication due to c(T)T (0.21 - 0.97)
- Costs of surveillance after HT - 2nd - 5th year (€ 0 - € 725)
- Costs of surveillance after FDG-PET/CT or GEC - 2nd - 5th year (€ 0 - € 493)
- Fraction of transient complications due to c(T)T (0.01 - 0.65)
- MMP sensitivity (0.36 - 0.70)
- Costs of FDG-PET/CT (€ 800 - € 1,200)
- Utility of transient complication due to c(T)T (0.90 - 0.99)
- Utility of transient complication due to HT (0.90 - 0.99)
- Discount rate of utilities (0.010 - 0.050)
- Costs of transient complication due to c(T)T (€ 189 - € 5,154)
- Costs of MMP (€ 400 - € 820)
- Discount rate of costs (0.030 - 0.050)
- Costs of transient complication due to HT (€ 100 - € 5,200)
- Costs of surveillance after HT - 1st year (€ 317 - € 1,208)
- Costs of surveillance after FDG-PET/CT or GEC - 1st year (€ 228 - € 889)
- Costs of surveillance after c(T)T - 2nd - 5th year (€ 180 - € 954)
- Utility of recurrence after c(T)T (0.54 - 0.98)
- Yearly probability of recurrence after c(T)T (0.01 - 0.07)
- Yearly probability of recurrence after HT for UPM (0.001 - 0.025)
- Costs of surveillance after c(T)T - 1st year (€ 274 - € 1,727)
- Costs of permanent complication due to HT - 2nd - 5th year (€ 95 - € 606)
- Costs of permanent complication due to HT - 1st year (€ 3,123 - € 4,903)
- Utility of recurrence after HT (0.54 - 0.98)
- Fraction of female patients (0.70 - 0.99)
- Costs of RRA (€ 1,277 - € 2,692)
- Costs of permanent complication due to c(T)T - 1st year (€ 5,327 - € 9,025)
- Costs of permanent complication due to c(T)T - 2nd-5th year (€ 517 - € 1,772)
- Costs of recurrence after c(T)T (€ 326 - € 2,184)
- Yearly probability of c(T)T after recurrence after HT (0.90 - 1.00)
- GEC sensitivity (0.62 - 0.99)
- Costs of recurrence after HT (€ 325 - € 2,013)
- GEC specificity (0.44 - 0.75)
- Yearly probability of death due to cancer (0.00 - 0.01)
- Costs of GEC (€ 1,611 - € 4,026)
- Yearly probability of surgery after surveillance after negative GEC (0.00 - 0.05)

MMP more cost-effective

FDG-PET/CT more cost-effective

←

10. Cost-Effectiveness of FDG PET in Thyroid Nodules
Epilogue
General Discussion & Future Prospects

Unpublished

D. Vriens

1dept. of Nuclear Medicine,
Radboud University Medical Centre, Nijmegen, the Netherlands
Introduction

The observation that cancer cells favour inefficient anaerobic over aerobic glycolysis under normoxic conditions and therefore generally have higher glycolytic rates is one of the main pathophysiological features responsible for the high interest in quantification of tumour glucose metabolism [10]. Positron Emission Tomography using 2-(18)F-fluoro-2-deoxy-d-glucose (18F-FDG PET) currently is the only widely available technology that allows absolute whole-body quantification of glucose metabolism in vivo using a non-invasive approach. Other approaches, such as studies using magnetic resonance spectroscopy (MRS) chemical shift imaging of 19F-FDG are under (pre-clinical) development but still require very long acquisition times and high doses of tracers (typically 500 mg·kg⁻¹ bodyweight) [433].

The aim of this thesis was to address three diagnostic challenges for 18F-FDG PET in solid tumours: (i) the influence of different methods for data-analysis on measurement of the metabolic response to therapy and their relation with clinical outcome, (ii) metabolic and vascular characterisation of solid tumours and (iii) clinical and economical impact of the use of 18F-FDG PET/CT for tissue characterisation.

Methodological considerations & clinical relevance

Part I of this thesis reviewed both methodological considerations for 18F-FDG PET and its clinical utility in patients with colorectal carcinoma (CRC), the third most frequent type of cancer in adults.

Reliable and reproducible quantification of tumour glucose metabolism using 18F-FDG PET is challenged by many biological, technical and analytical factors, which have led major nuclear medicine associations and societies to develop procedure guidelines [30,31,46,47,434]. To further harmonise the interpretation of 18F-FDG PET/CT, guidelines with respect to therapy response evaluation for both malignant lymphoma [435–443] and solid tumours [4–6] have been developed. Chapter 1 reviewed the many factors that influence quantification of glucose metabolism by 18F-FDG PET. The negative influence of high patient blood glucose levels on 18F-FDG uptake and therefore the accuracy of 18F-FDG PET in detection and quantification of tissues with high metabolic rates, has led to strict preparation and inclusion criteria in many trials. This selection of patients might compromise general applicability of published research, as in clinical practice a substantial number of cancer patients also suffer from diabetes mellitus due to an increase of prevalence of both with age. The effect of reaching normoglycemia prior to injection of 18F-FDG by administering glucose-lowering pharmaceuticals is still under debate [52,53,444]. Since 18F-FDG uptake is a dynamic process, the timing of single time-point static acquisition has major impact on the resulting numeric value of the uptake value. For clinical applicability, 18F-FDG uptake periods of 60 ± 5 minutes are used in most situations. Uniform uptake periods are especially important for studies that require repetitive measurements, such as therapy response evaluation, but also for comparison of results between patients. Delayed or dual phase imaging can increase specificity for a limited number of indications [445,446]. Respiratory gated acquisition has come available to clinical scanners, reducing motion artefacts which results in a decrease in metabolic
volume and increase in $SUV_{mean}$ in pulmonary nodules situated peripherally in the middle and lower lobes \cite{447,448}. Brown adipose tissue, previously considered as an interfering finding only, lowering the diagnostic accuracy of $^{18}$F-FDG PET and therefore requiring patient preparations preventing brown adipose tissue $^{18}$F-FDG uptake, has transformed to an independent imaging research field \cite{449}.

Apart from these biological factors needing special attention during patient preparation, the chapter also discussed acquisition settings, such as the relation between image quality (signal-to-noise ratio), patient bodyweight and acquisition duration. Current investigations point to a power law based rather than a linear dosage regime to keep image quality similar between patient of different body size \cite{31,450,451}. Briefly, the influence of reconstruction algorithms and settings, algorithms for correction of tissue attenuation, scatter and random coincidences and filtering are mentioned. The chapter concludes with a review of methods of image analysis, including tumour segmentation and different quantification parameters, which is the main subject of the studies described in part II of this thesis.

**Chapter 2** described the established added value of $^{18}$F-FDG PET/CT in the management of patients with CRC. Most clinical guidelines concur that there is a role for $^{18}$F-FDG PET/CT in this disease for tissue characterisation of liver lesions detected by morphological image modalities, to exclude extra-hepatic disease before local treatment of colorectal liver metastases is commenced. Recently, in a randomised trial in 404 patients with potentially resectable hepatic metastases of colorectal adenocarcinoma, the use of $^{18}$F-FDG PET/CT compared to CT alone did not result in frequent change in surgical management nor in patient survival. The authors concluded that these findings raise questions about the value of PET/CT scans in this setting. However, about half of their patients received chemotherapy within 12 weeks of surgery, which could have resulted in false-negative PET/CT imaging \cite{452}. Other established indications for $^{18}$F-FDG PET in CRC are: an unexplained rise of the serum tumour marker carcinoembryonic antigen (CEA), detection of local recurrence after radiofrequency ablation or cryoablation of a liver metastasis and distinguishing scar from vital tumour after radiotherapy for rectal carcinoma. Current clinical studies also address the potential role of $^{18}$F-FDG PET in detecting peritoneal disease and exclusion of extra-abdominal metastases before hyperthermic intraperitoneal chemotherapy (HIPEC) for peritoneal metastases \cite{453}, in biological target volume delineation for radiotherapy planning \cite{454} and in early response evaluation during neoadjuvant treatment \cite{455}. So far no interventional trials on these subjects have been published.

**Optimisation of quantification methodology**

Several parameters for quantification of $^{18}$F-FDG uptake have been reviewed in chapter 1, varying from simple semiquantitative estimates to increasingly sophisticated fully quantitative parameters. Semiquantitative estimates of glucose metabolism include the tumour-nontumour ratio, using a patient’s presumably healthy reference tissue as internal control and the standardised uptake value ($SUV$), which is the $^{18}$F-FDG activity concentrations in tissue normalised to the administered activity remaining at the time of the acquisition and its volume of distribution ($V_D$).
Fully quantitative parameters require the use of 4D dynamic PET and include glucose metabolic rate ($MR_{glc}$) and the pharmacokinetic rate constants of cellular glucose uptake and phosphorylation. Each parameter has its own assumptions and therefore its own limitations. Analysis of 4D dynamic $^{18}$F-FDG PET is extensively used in this thesis and key assumptions of the tracer kinetic modelling theory, such as the tracer principle and assumptions of steady-state, tissue homogeneity, instantaneous mixing, linearity, tracer dynamics and low plasma extraction are often difficult to substantiate. The compartment model assumed to represent $^{18}$F-FDG metabolism is a simplification of tissue metabolic processes. Irreversible trapping of the radiopharmaceutical, a main assumption in determination of the $MR_{glc}$ using Gjedde-Patlak analysis, is disputed in some types of tissues. These issues have been extensively discussed in chapter 1.

Part II of this thesis focused on three aspects of $^{18}$F-FDG image analysis: normalisation of $SUV_{mean}$ by different parameters for $V_D$ of $^{18}$F-FDG in studies of therapy response evaluation, the use of different segmentation methods for the determination of therapy response using $MR_{glc}$ and the use of different arterial plasma time-activity concentration curves in the determination of $MR_{glc}$. This part thereby focused on the first objective of this thesis: the influence of different methods for data-analysis on measurement of the metabolic response to therapy and their relation with clinical outcome.

The assumption of first-order kinetics (i.e., linearity principle) dictates a linear relation between tissue activity concentration and administered activity per unit $V_D$, reflected by the $SUV$ parameter. However, the optimal method to normalise for $V_D$ is disputed. As fat has a much lower uptake of $^{18}$F-FDG than other tissues, other metrics for $V_D$ than simply bodyweight have been advocated, such as ideal bodyweight, lean body mass and body surface area. Normalising for bodyweight is recommended by most guidelines given its simplicity, but since this method of normalisation results in blood and tumour $SUV$ values still positively correlated to patient bodyweight ($R^2 = 0.50$), others recommended different normalisation factors. E.g., PET response criteria in solid tumours (PERCIST) advocates the determination of an $SUV$ normalised for lean body mass.

Four different normalisation methods for $V_D$ were evaluated in the analysis of a cohort of 97 patients with CRC or non-small cell lung carcinoma (NSCLC) in chapter 3. These included bodyweight, lean body mass, body surface area and a combination of body mass with fasting plasma glucose level. The change in mean bodyweight in this cohort was limited: a mean decrease of 1.2 kg in the CRC subgroup and no significant change in the patients with NSCLC. However, individually, the bodyweight ranged from -15 kg to +10 kg between baseline and follow-up scan. Similarly, there was a moderately lower mean fasting serum glucose level at the follow-up scan of the NSCLC patients ($0.3 \text{ mmol}\cdot\text{l}^{-1}$). Therefore, it is not very surprising that the mean patient change in $SUV$ ($\Delta SUV$) did not differ between the normalisation methods used. The fact that none of the normalisation procedures showed a stronger relation between $\Delta SUV$ and clinical outcome can be explained by the observed minimal influence of the chemotherapy provided on body habitus and plasma glucose. However, when cut-offs of $\Delta SUV$ are used for treatment decisions, such as in the PERCIST criteria, the exact and accurate computation of metabolic response is important.
up to ten of 97 individual patients, the use of a different normalisation factor resulted in a $\Delta SUV$ that was considered higher than its reproducibility limit, where it was considered within the reproducibility limit (i.e., $|\Delta SUV| \leq 20\%$) using bodyweight as $V_D$.

Interestingly, the equation to compute lean body mass used by many authors [108, 110, 131, 456] and which was also provided in chapter 1 and used in chapter 3 of this thesis, in fact is the Hamwi “ideal” bodyweight equation [457]. This is a function dependent on sex and length only, based on estimates rather than measurements on a defined population and results in values higher than those of the real fat-free lean body mass. As shown by Sugawara et al. [456], $SUV$s corrected for bodyweight and ideal bodyweight both remain positively correlated to bodyweight. When these authors applied correction for either body surface area or the accepted equation for lean body mass [458], the weight-dependency of the $SUV$ was lost.

Apart from the metric used to quantify glucose metabolic rate in tumours, the tissue sample from which it is determined, i.e., the volume-of-interest (VOI), must be taken into account. The observation that $^{18}$F-FDG uptake is positively correlated with cellular proliferation in many cancer types [459] but also to other factors related with an adverse outcome (e.g., invasiveness, dedifferentiation) is one of the reasons that many studies quantify the maximum uptake within a lesion (i.e., $SUV_{max}$), as this value is supposed to represent the most malignant region of the tumour. The $SUV_{max}$ shows perfect interrater reproducibility, but is positively biased by image noise and therefore dependent on technical factors such as administered activity, scan speed, scanner sensitivity, reconstruction settings including filter, matrix and zoom. This knowledge has led other investigators to use the mean lesion $SUV$ ($SUV_{mean}$). $SUV_{mean}$ is far less sensitive to image noise, but requires the definition of a VOI, for which again many different techniques are being used. Segmentation methods often used include manual contouring on either PET or CT images, threshold-based methods that only incorporate voxels surpassing a predefined absolute or relative tumour uptake value or semi-automatic segmentation methods (e.g., fuzzy locally adaptive Bayesian delineation). Some of these methods are more robust in situations of low tumour-background contrast [32, 460] and have high interrater reproducibility. The volume of tissue incorporated by such a VOI, often called metabolic tumour volume ($MTV$), is found to be an independent prognostic parameter in many solid tumours [461] and holds important therapeutic potential as it can be used for determination of the target volume in radiotherapy dose distribution [462]. From a statistical point of view, the $SUV_{mean}$ determined from this volume might not accurately describe the uptake in a lesion, as tumour $SUV$-volume histograms are positively skewed [463] and therefore the arithmetic mean is (often positively) biased.

Recently, a parameter that combines the advantages of both $SUV_{max}$ (perfect interrater reproducibility, representing the most malignant cluster) and $SUV_{mean}$ (less dependent on image quality) is being used more frequently. This parameter is computed as the mean voxel value of a VOI defined as the cluster with highest $^{18}$F-FDG uptake: $SUV_{peak}$. Since the exact definition of the peak VOI must be described,
Optimisation of quantification methodology

e.g., “a 3D 1.2 cm diameter (≈1.0 ml) spherical VOI centred on area with maximum uptake” [464], again different definitions are being used. The recommendation to use $SUV_{peak}$ is increasingly mentioned in guidelines [6,30,31].

The abovementioned asymmetric positively skewed $SUV$-volume histograms and thereby the dependence of the arithmetic mean $SUV$ on the size of the lesion, could have an effect in therapy response evaluation studies. A mere treatment-induced reduction in tumour volume without any effect on overall lesion $^{18}$F-FDG uptake (distribution) could theoretically result in a decrease in $MTV$ but an unchanged or potentially an increased $SUV_{mean}$. To account for both metabolic and morphometric changes during therapy, chapter 4 compared the influence of analysing the exact same VOI before and after treatment (fixed volumes) to a method in which the tumour VOI was redefined on each scan separately (fixed thresholds). This was performed in the same patients as chapter 3, but using Gjedde-Patlak $MR_{glice}$ as a metric for glucose metabolism. A clear numeric difference in the change in $MR_{glice}$ between baseline and follow-up 4D dynamic $^{18}$F-FDG PET ($\Delta MR_{glice}$) between both approaches was observed: median fixed volume-based reductions in treatment-induced $MR_{glice}$ exceeded fixed threshold-based reductions both in patients with CRC and NSCLC. This can be explained by the fact that tumour size reduction leads to incorporation of peritumoural tissue in the fixed-sized VOI. This tissue only has limited $MR_{glice}$ compared to vital tumour tissue. The fixed volume methodology was more strongly associated with overall and progression-free survival than the fixed thresholds technique, thereby showing the additional prognostic effect of morphometric changes in tumour volume.

The proposed methodology has some limitations related to inter- and intraobserver reproducibility and for lesions near organs with high $^{18}$F-FDG uptake, such as brain, myocardium, liver, ureters and bladder. Placement of the fixed-sized VOI on the follow-up scan was performed visually by 3D translation of the VOI, centred on the lesion, without any rotation, scaling or deformation of the VOI. In case the original lesion could not be identified at follow-up, the VOI was placed approximately in the region of the original lesion. The accuracy and reproducibility of placement could have been improved by fusion of both PET-scans prior to VOI-placement. Furthermore, it could be argued that non-rigid transformations are not sufficient.

Other authors have described a different solution to incorporate both the morphometric changes and the metabolic changes during treatment: the use of the Larson-Ginsberg Index, also known as relative change in total lesion glycolysis ($\delta TLG$). The $TLG$, having the same unit as the parameter used for $V_D$ (mass in most of the cases), has different definitions, but generally it is calculated by the sum of the product of each lesion’s $MTV$ and $SUV_{mean}$ [145]. Again, different definitions for the VOI to determine the $MTV$ (often a relative 41% $SUV_{max}$ threshold or an absolute 2.5 g·cm$^{-3}$ $SUV$ threshold) and different normalisation factors for the $SUV$ are used. Recently, a more sophisticated approach was described: the area under the $SUV$-volume histogram corrected for background [465]. The last few years an increasing number of publications indicated that ($\delta$) $TLG$ to be an independent prognostic marker in many different tumour types [466-468] but not in all [469]. It should be noted however, that morphometric changes are equally weighed as metabolic changes by the $TLG$. 

191
Most of the techniques to perform pharmacokinetic analysis of 4D dynamic $^{18}$F-FDG PET/CT scans described in the first chapter require an input function, \emph{i.e.}, the time-dependent activity concentration of $^{18}$F-FDG in the plasma of the artery feeding the individual lesion. It is practically impossible to sample this artery, therefore other approaches are being used. These approaches can be divided into three main categories: (i) serial sampling via an intravascular device in a peripheral vessel and determine the activity concentration in plasma after centrifuging the whole-blood samples, (ii) obtaining an image-derived whole-blood time-activity curve (image-derived input function, IDIF) from a large vessel or extracting this data from reference tissue in a subject imaged by 4D dynamic $^{18}$F-FDG PET or (iii) using a generalised population-based curve individualised to the subject under investigation. Hybrid variants have been described as well. Arterial blood sampling is rather invasive and assumes similar dynamics in the sampled vessel and the location of the tumour, the flow dynamics in the vascular tree however might lead to delay or dispersion of the $^{18}$F-FDG bolus compared to the sampled artery. The time-dependent plasma-to-whole-blood ratio, decreasing in time, can be overcome using the supernatant plasma after centrifuging. If venous samples are taken, the time-dependent arteriovenous gradient of $^{18}$F-FDG activity concentration should be taken into account. Physiologically, this gradient decreases in time, but can be rendered near to unity by using arterialised venous blood, by decreasing the extraction fraction using the heated-hand procedure \cite{137}. The second option of image-derived determination of the input function requires a large vessel in the field-of-view or extraction of this data from reference tissue, however, given the scanners’ spatial resolution, an IDIF often results in a noisy whole-blood curve. Conversely, population-based curves lack individualisation and depend strongly on the $^{18}$F-FDG infusion protocol used. In \textit{chapter 5} of this thesis, a population-based input function for our infusion protocol was determined. The use of either an individualised variant of this generic function or an IDIF for determination of Gjedde-Patlak $MR_{glc}$ were compared to the gold standard of serial peripheral arterial sampling before and during treatment of breast carcinoma, CRC and NSCLC patients. The chapter concluded that for $MR_{glc}$ determination, an individualised population-based curve outperforms an image-derived input function when using a single late venous plasma sample for individualisation. This method can be helpful in case arterial sampling is not desirable and an IDIF is not available, such as in lesion located outside the torso or in 4D dynamic PET of small animals. From the Gjedde-Patlak equation it can be derived that only the area under the input function and the late (>5 min) activity concentrations are of importance to determine $MR_{glc}$. The early section of the input curve is likely to be most dependent on the $^{18}$F-FDG infusion protocol and patient haemodynamics and would therefore probably show most interpatient variability. As for determination of the pharmacokinetic rate constants of glucose metabolism, especially transmembranous $^{18}$F-FDG flux ($K_1, k_2$) and blood volume fraction ($V_B$), the exact shape of this first part of the curve is also of importance, an individualised population-based function would likely not work in these types of analyses.

A general critical remark on the $MR_{glc}$-values provided in this thesis refers to the lumped constant of $^{18}$F-FDG ($L_F^{FDG}$): the steady-state ratio of the net extraction of $^{18}$F-FDG to that of glucose at constant plasma levels of both substrates. As there
Tissue heterogeneity & therapy response evaluation

Improvement of technology, including higher spatial resolution and detection sensitivity, has shifted the interest from whole-tumour metabolism ($SUV$, $MR_{glc}$) towards other metabolic parameters including $MTV$ [167, 168, 170, 171], intra-tumour heterogeneity and in vivo assessment of transmembranous glucose transporter (GLUT) and hexokinase activity. Furthermore, parameters such as tumour perfusion are being explored, as this is an important feature for treatment with angiogenesis-inhibitors.

Part III of this thesis investigated some of these developments relating to the second objective of this thesis: metabolic and vascular characterisation of solid tumours.

Image-derived blood pool and tumour time activity concentration curves are noisy, due to the limited amount of photon pairs that originate in the VOIs. This is mainly due to the limited activity present in these lesions at time of the acquisition: for a tumour $SUV_{mean}$ of 5.0 and 15 g·cm$^{-3}$, 60 minutes after injection of 3.45 MBq·kg$^{-1}$ bodyweight, the expected activity concentration would be 12 and 35 kBq·cm$^{-3}$, respectively. Another reason is the limited acquisition time, required to obtain a high temporal resolution, as sampling frequencies of 0.2 min$^{-1}$ to 1 s$^{-1}$ are used. To improve statistical accuracy of the curves, larger tissue volumes are analysed simultaneously. Therefore, mean time-activity concentration curves of a whole tumour VOI are obtained to determine whole-tumour pharmacokinetic parameters such as $MR_{glc}$, $^{18}$F-FDG influx constant ($K_i$), transmembranous $^{18}$F-FDG flux ($K_1, k_2$), $^{18}$F-FDG phosphorylation rate ($k_3$), dephosphorylation rate ($k_4$) and tumour blood volume fraction ($V_B$). With improved photon-pair detection sensitivity of newer PET-scanners and regression analysis methods less sensitive to statistical noise, voxel-by-voxel modelling will become feasible at acceptable levels of administered activity [296]. This remains problematic as a spherical lesion with a diameter of 2 cm (4.2 cm$^3$) typically contains approximately 100 voxels (isotropic 3.4 mm voxels). In order to describe regional differences in tumour metabolism, chapter 6 investigated a tumour segmentation method in which three independent tumour regions are analysed instead of whole-tumour or voxel-based analysis. Regions with a higher mean $MR_{glc}$ showed higher values for $K_1$, $k_2$ and $k_3$ but lower values for $V_B$. It was concluded that increased phosphorylation rates in tumour regions with low blood content support the Warburg hypothesis. Unfortunately, no histological tissue analyses were available to support these findings. Regional differences in metabolic characteristics might be of interest to determine the optimal biopsy site [172] or for radiotherapy dose painting [173]. Although this study demonstrates the feasibility of identifying regional differences in tumour glucose handling within tumours, so far no studies have confirmed these results. No significant correlation could be established between tumour $V_B$ and parameters for glucose metabolism ($MR_{glc}, k_3$) as assessed by 4D dynamic $^{18}$F-FDG PET in CRC, NSCLC and breast carcinoma. Similarly, tumour blood volume measured by DCE-CT could not be related to $SUV_{max}$ or $SUV_{mean}$ in rectal carcinoma by Fisc-
her et al. [474] or to $SUV_{\text{max}}$ or TLG in CRC by Tixier et al. [475]. Blood volume measured by DCE-MRI could not be related to $MR_{\text{glc}}$ in CRC (chapter 7). Furthermore, Zhang et al. found a negative relation between tumour blood volume assessed by DCE-MRI and $SUV_{\text{mean}}$ or $SUV_{\text{max}}$ in adenocarcinoma but not in squamous cell carcinoma in NSCLC [476]. The paper by Miles and Williams [300] discusses the complex relation between tumour blood flow (and volume) and metabolism: on the one hand a positive relation is expected as tumour vascular supply is required for glucose (and $^{18}$F-FDG) to reach its destination. On the other hand, inadequate vascularisation of tumour will result in hypoxia, a factor that stimulates anaerobic metabolism of glucose, which would favour a negative relation between both.

Interestingly, the last few years, more data have been published on intra- and interlesional heterogeneity using textural feature analysis [477–481] as independent prognostic and predictive markers in different tumour types, although these results should be interpreted with care in smaller lesions [482].

Apart from metabolism, vascularity is known to be increased in tumours. This is the basis for the use of intravascular contrast agents in tumour imaging by CT and MRI. Tissue contrast-enhancement is often used as a main criterion for tissue characterisation. The vascular supply of a tumour is a prerequisite for systemic therapy to reach its target. Moreover, it is well known that a tumour requires induction of angiogenesis to grow beyond a size of 1-2 mm$^3$. This latter phenomenon has been a target for tumour therapy since the Food and Drug Administration approved the first anti-angiogenetic drug bevacizumab in 2004. Bevacizumab is a recombinant humanised monoclonal antibody that inhibits angiogenesis by neutralising vascular endothelial growth factor A. This drug causes normalisation of tumour vasculature. Tumour vasculature typically lacks hierarchy and retains the disordered, tortuous and leaky characteristics of incipient vasculature. These vessels have poorly aligned defective endothelial cells with wide fenestrations, lacking a smooth muscle layer or innervation with a wider lumen. Together with the observation that tumour tissues usually lack effective lymphatic drainage, these factors lead to abnormal molecular and fluid transport dynamics, especially for macromolecular drugs. This phenomenon is referred to as the enhanced permeability and retention (EPR) effect of macromolecules and lipids in solid tumours. The EPR effect causes drugs with a high molecular weight, such as antibodies and antibody fragments, to accumulate in tumour tissue even in the absence of the targeted antigen. However, this abnormal fluid transport leads to inefficient delivery of smaller molecules such as most cytotoxic drugs. By itself, anti-angiogenic drugs have a limited effect on tumour growth, but bevacizumab-induced vascular normalisation improves delivery of small-molecule cytotoxic drugs and thereby enhances their effect when given consecutively or in combination. However, bevacizumab counteracts the EPR effect by vascular normalisation and thereby deteriorates the delivery of larger molecular drugs such as other antibodies to their targets in both preclinical [483] and clinical investigations [339]. It is therefore not surprising that much effort has been made to quantify functional tumour vascularity in vivo mostly using 4D dynamic Gd-DTPA contrast-enhanced MRI (DCE-MRI) and more recently also using DCE-CT. Chapter 7 and chapter 8 investigated the use of both quantification of tumour vascularity
Clinical impact of molecular imaging

(by DCE-MRI) and metabolism (by 4D dynamic $^{18}$F-FDG PET) during cytotoxic or concurrent use of cytotoxic and anti-angiogenic agents in a population of patients with liver metastases of CRC (CRLM).

In chapter 7, no changes in functional tumour vascularity parameters were found in 23 patients with CRLM during treatment with cytotoxic drugs: baseline DCE-MRI parameters were not associated with overall and progression-free survival when corrected for tumour size and chemotherapy line. Similarly, Gu et al. describe a small series of four patients that underwent $^{18}$F-FDG PET/CT, diffusion weighted (DW)-MRI and DCE-MRI in the evaluation of therapy response to pre-operative chemoradiation in locally invasive T$_3$/T$_4$ rectal carcinoma [484]. These patients did not show any change in DCE-MRI parameters. In contrast, Oberholzer et al. demonstrated that pretreatment DCE-MRI has a significant impact on prediction of response to neoadjuvant chemoradiation in rectal carcinoma, in relation to T-downstaging [485]. Furthermore, this technique showed a significant predictive value [485] in response prediction to local treatment of head and neck squamous cell carcinoma [486]. DCE-MRI also proved to be predictive in response evaluation of neoadjuvant chemoradiation in locally advanced rectal carcinoma (T-downstaging) [487], in early response evaluation of concurrent chemoradiation in cervical carcinoma (tumour size changes) [488], following two cycles of induction chemotherapy in advanced head and neck carcinoma [489] and after neoadjuvant chemotherapy in breast carcinoma [490]. Accordingly, the study presented in chapter 7 confirmed $MR_{glc}$ as an independent prognostic marker.

Chapter 8 provided a description of two cases of CRLM, which were treated concurrently with cytotoxic drugs and bevacizumab, showing a remarkable decrease in mean $^{18}$F-FDG uptake and DCE-MRI vascularity parameters. A larger study including 19 patients with metastatic CRC treated with cytotoxic chemotherapy plus bevacizumab before surgery has been published since then [491]. This study confirms a decrease in heuristic parameters (area under the time-enhancement curve) and compartment model-based vascularity parameters ($K^{trans}$) of DCE-MRI and $^{18}$F-FDG uptake measured by PET/CT during treatment. A high decrease in $K^{trans}$ and a low $SUV_{max}$ were associated with favourable prognosis. The same group described in a subgroup of 18 patients, that the relative change in $SUV_{max}$ was also related to overall and progression-free survival after five cycles of combination treatment, consisting of neoadjuvant cytotoxic and anti-angiogenic therapy prior to resection CRLM [492].

Clinical impact of molecular imaging

Next to methodological issues in analysis of (4D dynamic) $^{18}$F-FDG PET data and multi-modality therapy response evaluation, this thesis focused on clinically relevant aspects of this molecular imaging modality for tissue characterisation. Part IV focussed on the potential role of $^{18}$F-FDG PET in cytologically indeterminate thyroid nodules and thereby addressed the third objective of this thesis: clinical and economical impact of the use of $^{18}$F-FDG PET/CT for tissue characterisation.
Clinically overt thyroid nodules occur in 3-8% of European adults. The risk of differentiated thyroid carcinoma in these nodules, however, is less than 5%. Asymptomatic nodules occur in up to 68% of adults and due to increasing use of imaging such as US, CT and MRI, more and more of these subclinical nodules are being found. In case of unsuppressed thyrotropin levels, the primary diagnostic approach is cytological classification of a fine-needle aspirate. The Bethesda System covers six diagnostic categories, all with a different risk of malignancy, occurrence and usual management (table 7, page 196) [493]. In case of indeterminate cytology, defined as Bethesda category III or IV, diagnostic surgical thyroid lobectomy (hemithyroidectomy) is performed which is positive for malignancy in 15-26%, meaning that the majority of patients are undergoing unnecessary thyroid surgery with the potential risk of irreversible complications.

Several diagnostic tests are currently under investigation for a better selection of patients before diagnostic thyroid surgery, including cytological tissue markers and imaging tests. The first category includes molecular tissue biomarkers such as galectin-3 [494, 495], but also consists of a currently commercially available 167-gene-expression classifier (GEC) [386] and a mutation marker panel (MMP) for common mutations and rearrangements [390]. The second category includes ultrasonography-elastography, $^{99m}$Tc-methoxyisobutylisonitrile ($^{99m}$Tc-MIBI) SPECT [499, 500] and $^{18}$F-FDG PET.

Chapter 9 systematically reviewed the available literature on the use of $^{18}$F-FDG PET/CT in case of indeterminate thyroid nodule cytology and provided individual patient based meta-analysis, showing a very high negative predictive value but a moderate positive predictive value in a population with a representative prevalence of malignancy of 26%. Full implementation of this technique could potentially reduce unnecessary diagnostic thyroid surgery from $\sim$75% to $\sim$40%. The false-negative fraction of $^{18}$F-FDG PET proved to be very low (3 in 225 patients). Sensitivity proved to be size-dependent, increasing to 100% for lesions $\geq$ 15 mm, which can be explained by the limited resolution of the scanner used in the studies analysed in this meta-analysis. It may be argued that there is a larger chance of

<table>
<thead>
<tr>
<th>Cat.</th>
<th>Description</th>
<th>Risk of malignancy</th>
<th>Prevalence</th>
<th>Usual management*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Nondiagnostic or unsatisfactory</td>
<td>1-4%</td>
<td>13%</td>
<td>Repeat FNA with US guidance</td>
</tr>
<tr>
<td>II</td>
<td>Benign</td>
<td>0-3%</td>
<td>59%</td>
<td>Clinical follow-up</td>
</tr>
<tr>
<td>III</td>
<td>AUS/FLUS</td>
<td>$\sim$5-15%</td>
<td>9.6%</td>
<td>Repeat FNA / Diagnostic surgical lobectomy</td>
</tr>
<tr>
<td>IV</td>
<td>FN/SFN</td>
<td>15-30%</td>
<td>10%</td>
<td>Diagnostic surgical lobectomy</td>
</tr>
<tr>
<td>V</td>
<td>Suspicious for malignancy</td>
<td>60-75%</td>
<td>2.7%</td>
<td>Surgical lobectomy or near-total thyroidectomy†</td>
</tr>
<tr>
<td>VI</td>
<td>Malignant</td>
<td>97-99%</td>
<td>5.5%</td>
<td>Near-total thyroidectomy†</td>
</tr>
</tbody>
</table>

Table 1: The Bethesda system for reporting thyroid cytopathology. Implied risk of malignancy, result prevalence and recommended clinical management. AUS: atypia of undetermined significance; FLUS: follicular lesion of indeterminate significance; (S)FN: (suspicious for a) follicular neoplasm; FNA: fine-needle aspiration; US: ultrasonography. *Actual management may depend on other factors (e.g., clinical, ultrasonographic). †In case of metastatic tumour rather than primary thyroid carcinoma, surgery might not be indicated.
sampling error in larger nodules, which then are classified as cytologically benign and were therefore not included in the reviewed studies. However, a large series showed that thyroid nodule size had no influence on the distribution of cytology aspirates in each Bethesda category \[501\]. This latter study also concluded that greater nodule size influences cancer risk, although the increase in absolute risk between small (1-2 cm) and large (>4 cm) nodules is modest. A threshold effect was detected at approximately 2 cm in nodule diameter. However, larger nodules, if cancerous, were found significantly more likely to be follicular or Hürthle cell carcinomas or other rare malignancies in comparison with smaller nodules. Subgroup re-analysis for indeterminate cytology (Bethesda category III: 381 nodules; Bethesda category IV: 532 nodules), however, showed a significant increase in the number of category IV nodules for the four different size groups (being 6, 8, 11 and 9% for sizes 1-2 cm, 2-3 cm, 3-4 cm and >4 cm). Although no direct conclusion can be drawn from this dissimilar distribution of cytology categories with respect to lesion size, it should be kept in mind that lesions of different sizes represent a different cytological group with a different possibility of being malignant (and thus being \(^{18}\)F-FDG avid).

A review of the three false-negative lesions, did not lead to a definite explanation. Likewise, the reason why 87 of 167 benign lesions showed \(^{18}\)F-FDG uptake cannot be easily explained. Especially follicular lesions show higher uptake. Thyroid \(^{18}\)F-FDG incidentalomas, occurring in about 1.6% of all scans, have a probability of being malignant of \(\sim 25\text{-}35\% \) and a high uptake of \(^{18}\)F-FDG in such an incidentaloma is correlated with disease persistence and progression \[503\,504\]. This supports the value of \(^{18}\)F-FDG PET for detection of malignant thyroid disease. Therefore, it seems contradictory that, in case of thyroid carcinoma, well-differentiated lesions are in general not \(^{18}\)F-FDG avid and dedifferentiation is associated with increase in \(^{18}\)F-FDG uptake as is well known in anaplastic \[505\] and Hürthle cell carcinomas \[506\,507\]. Often, the acquisition of a higher glucose metabolic rate coincidences with a loss of the sodium-iodine symporter. This is also known as the not fully elucidated “flip-flop phenomenon” \[429\,508\,510\]. Thus, the paradox arises that in indeterminate, predominantly benign, thyroid nodules, \(^{18}\)F-FDG uptake is associated with differentiated thyroid carcinoma, but in well-differentiated thyroid carcinomas most lesions are \(^{18}\)F-FDG negative and increased uptake is strongly correlated to dedifferentiation.

Although the accuracy of \(^{18}\)F-FDG PET/CT to characterise thyroid nodules is very similar compared to some newer cytological diagnostics, the use of \(^{18}\)F-FDG PET/CT has disadvantages: it requires the patient to pay an extra hospital visit, exposes her or him to a low dose of ionising radiation, potentially leads to irrelevant incidental findings due to the rather unspecific nature of the radiopharmaceutical and it may not yet be accessible for all patients. The exposure to ionising radiation and the false-positive rate can be minimised by optimising the acquisition protocol only to cover neck and upper thoracic regions. \(^{18}\)F-FDG PET/CT also has advantages: on average performing three PET-scans will save one futile hemithyroidectomy. PET with follow-up therefore can be cost effective if it is less than one third of the price of surgery plus follow-up. With a mean cost price for PET of \(\€ 1,000\) and of diagnostic thyroid lobectomy of over \(\€ 4,000\), it can easily be calculated that full implementation is likely to be cost-effective. As the GEC, current standard clinical practice in the USA, is by itself also is cost-effective \[389\] and has a similar sensitivity
and specificity to $^{18}$F-FDG PET, but costs over €2,500 per sample, it is straightforward to deduct that routine implementation of the imaging test would be more cost-effective than the genetic test on cytological material. Chapter 10 used a more sophisticated approach to prove this reasoning, incorporating uncertainties with respect to patient outcome, cost-price and diagnostic test accuracies, also incorporating estimates of patient-experienced health-related quality of life and evaluates different scenarios using a broad range of values for all parameters. This Monte-Carlo analysis also concluded that $^{18}$F-FDG PET/CT is the least costly alternative. Recent publications showed a similar result for $^{99m}$Tc-MIBI imaging \[496\] or a multi-step approach using the gene-expression classifier and gene mutation panel consecutively on cytological material \[511\].

Especially in an era in which medical costs show an exponential increase and available resources are limited, not only efficacy but also cost-effectiveness is of importance. Investigational techniques or treatments should not only be safe (phase I-II clinical research) and effective (phase III) but also cost-effective (phase IV). Oncological $^{18}$F-FDG PET/CT cost-effectiveness studies have been published since 1996 \[512\] and are slowly increasing, but still only cover little over 1% of published literature on oncological $^{18}$F-FDG PET since 1982 \[513\].

Main conclusions & directions for future research

Quantitative $^{18}$F-FDG PET & other molecular imaging

Many favourable characteristics of the radiopharmaceutical $^{18}$F-FDG make it a very good tracer in oncology: the physical half-life of $^{18}$F (slightly less than 110 min) making centralised cyclotron production, transportation and time for biodistribution after injection possible, $^{18}$F’s low positron range ($E_{\text{max}} = 0.635$ MeV, $R_{\text{max}} = 2.4$ mm, $R_{\text{mean}} = 0.6$ mm), its favourable dosimetry \(19 \, \mu\text{Sv} \cdot \text{MBq}^{-1}\) in a healthy 70-kg adult, its high tissue contrast, differentiating disease from surrounding healthy tissue in most organs, its disease unspecificity (many different cancer types can be imaged), its predictive and prognostic ability and its cost-effectiveness. Thus >95% of all clinical PET/CT procedures are performed with $^{18}$F-FDG. Its limited specificity has sparked the interest in more specific (PET-)radiopharmaceuticals for specific disease categories such as prostate carcinoma (choline-derivatives and prostate-specific membrane antigen-ligands), breast carcinoma (oestrogen and progesterone hormone receptor imaging, HER2-receptor imaging), neuroendocrine tumours (somatostatin analogues, neurotransmitter precursors or analogues) or for specific treatments (hypoxia or proliferation imaging for radiotherapy, receptor imaging as diagnostic companion in targeted therapy). Independent of the radiopharmaceutical used, methods for acquisition and analysis need to be standardised to benefit comparability of study results. Many of the observations and conclusions of the investigations performed in this thesis can be used for other tracers as well.

For most clinical studies, standardised acquisition of whole-body $^{18}$F-FDG PET/CT would suffice. Quantification can be performed using $SUV$ normalised for lean body mass or bodyweight. For assessment of malignant potential of tissue, $SUV_{\text{peak}}$ is recommended, as it is less dependent to image noise than $SUV_{\text{max}}$. 
Main conclusions & directions for future research

Prospective studies need to be undertaken to support the ability of (early) response assessment using quantitative $^{18}$F-FDG PET for PET-guided treatment individualisation, which has already been performed to some extent for malignant lymphoma and adenocarcinoma of the oesophagogastric junction. This could be used to intensify treatment in metabolic nonresponders in order to improve survival or by reducing toxicity without adversely affecting cure rates by de-escalating therapy in metabolic responders \[514\]. Although in solid tumours, model studies have been published as early as 2007 \[20\], not many have followed since. This lack of evidence, therefore, limits the clinical use of the prognostic and predictive ability of $^{18}$F-FDG PET/CT for many solid tumours.

Tumour burden can be assessed using MTV and TLG. The use of MTV will likely influence the planning target volume or boost-volume during radiotherapy in the near future. Especially tumour parameters including both the amount of tumour tissue and its metabolic characteristics, such as the TLG, seem promising parameters for therapy response evaluation. The use of user and image-quality independent segmentation methods should be explored. These features should also be made available for clinicians and for use in clinical trials by incorporation in software used in radiology, nuclear medicine and radiotherapy.

Tissue heterogeneity is a relatively new area of investigation for quantitative molecular imaging. Many quantitative parameters can be derived to describe the spatial and temporal distribution of tissue tracer uptake. A major problem in describing heterogeneity in image tracer distribution is that it is a multifactorial parameter reflecting both real, biological heterogeneity superimposed by stochastic image noise. Description of tissue heterogeneity can only be performed on a multi-millimetre scale, as current PET-scanners’ spatial resolution is limited. Comparison of measured heterogeneity with biological surrogates (e.g., spatial distribution of cellular expression of GLUT, hexokinase, proliferation et cetera) using histopathology is very challenging due to difference in orientation, resolution and timing of the methods used. Heterogeneity might not only be another parameter reflecting the non-uniform nature of tumours but this parameter might also be used in intensity-modulated radiation therapy dose painting.

Biological correlates for $^{18}$F-FDG uptake are well-known \[515\], however many questions remain to be answered. Why are some tumours with high proliferative indexes, such as pancreatic carcinoma, lacking $^{18}$F-FDG uptake? What is the influence of inflammatory tumour micro-environment, such as tumour-associated macrophages? Why are many benign (thyroid) lesions also $^{18}$F-FDG avid?

Currently, there is only limited evidence that simple, heuristic parameters for dynamic PET but also for DCE-MRI are of less value than fully quantitative parameters, such as those based on compartment models. Although some claim that $MR_{glc}$ provides a more accurate evaluation of glucose consumption than the $SUV$ \[516\], many studies do not find a clear advantage of 4D dynamic PET parameters of glucose metabolism. This fact, together with the sophisticated and lengthy dynamic acquisition protocol and a reproducibility of metrics that is less than of “simple” static whole-body $^{18}$F-FDG PET \[52\] has caused 4D dynamic PET to be replaced by 3D static PET in daily clinical practice.
Even though the technique of 4D dynamic PET shows limited additional value, there are specific roles for it: it can be used to quantify target expression and tracer/target binding, it can be used to further ascertain tracer (or labelled drug) pharmacokinetics and to explore potential metabolic pathways (compartment models). In the specific case of targeted radionuclide radiotherapy, 4D dynamic PET can be used for dosimetry purposes. 4D dynamic PET should be considered the gold standard for quantitative molecular imaging of radiopharmaceuticals and should therefore be used as a benchmark for simplified methods before these become routine practice. Likewise, 1-hour $SUV$ measurement in clinical oncological $^{18}$F-FDG PET has been established as a simplified measure for determination of $MR_{glo}$ thereby replacing the need for 4D dynamic PET acquisition.

Many other imaging modalities address *in vivo* tissue characterisation. Especially nuclear magnetic resonance techniques addressing vascular contrast enhancement (i.e., DCE-MRI), quantification of water diffusivity (diffusion weighted imaging), but also biochemical information (MRS) or MRI using hyperpolarised tracers. The information from each of these techniques is likely to be of additional value and therefore, multimodal multiparametric assessment of tumour characteristics is likely to provide even more specific information to improve and individualise treatment. In prostate carcinoma, the use of multiparametric MRI is already being advised in guidelines [518]. The combination of imaging techniques with other techniques, such as gene-expression profiling, resulting in a multidimensional fingerprint of a tumour, could provide breakthroughs in understanding tumour biology, drug development and treatment tailoring. However, as most of these techniques are relatively expensive, cost management is of high importance. Therefore, after a technique has shown to be effective, the relation of additional effectiveness and additional costs of a technique needs to be described, in order for society to judge routine implementation. It is the responsibility of researchers in the field to describe this cost-effectiveness relationship.

**Molecular imaging in drug research & development**

Research and development of new drugs is very costly, mainly because most experimental compounds will never reach phase I clinical trials and many drugs that do progress to the very lengthy and expensive phase III clinical trials will never be approved [519]. Large numbers of patients need to be included in pivotal trials to secure enough events of the outcome measure under investigation to find statistical differences. During drug discovery, biodistribution and pharmacokinetics of the new drug can be evaluated using molecular imaging by labelling the drug or labelling the target ([table 2, page 201] [520] [521]). Timing and optimal dose finding of compounds can be aided by the use of molecular imaging. Both during preclinical testing and during clinical phases of drug development, more accurate exclusion of animals or patients that are unlikely to benefit of a targeted treatment (sample enrichment), would lead to trials with smaller and more homogeneous patient cohorts. Besides, the use of molecular imaging as intermediate or surrogate endpoint would lead to shorter follow-up duration necessary. The non-invasive character of molecular imaging allows it to be used in longitudinal studies. These advantages could therefore lead to
Main conclusions & directions for future research

<table>
<thead>
<tr>
<th>Topic</th>
<th>Phase</th>
<th>Potential roles of molecular imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics &amp; proteomics</td>
<td>• Assessment of target expression</td>
<td>• Target identification and localisation (reporter gene, receptor)</td>
</tr>
<tr>
<td></td>
<td>• Compound screening</td>
<td>• Target quantification</td>
</tr>
<tr>
<td>Drug discovery</td>
<td>• Preclinical testing of lead compound</td>
<td>• Drug effects (presence of target, metabolism, perfusion, apoptosis, hypoxia, et cetera)</td>
</tr>
<tr>
<td></td>
<td>(optimisation and efficacy)</td>
<td>• Drug pharmacokinetics/microdosing (radiolabelled compound or target)</td>
</tr>
<tr>
<td></td>
<td>• Phase 0, I, II trials</td>
<td>• Patient selection</td>
</tr>
<tr>
<td></td>
<td>• Phase III trials</td>
<td>• Surrogate endpoints (therapy response evaluation)</td>
</tr>
<tr>
<td>Drug development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Use</td>
<td>• Sales / postmarketing surveillance</td>
<td>• Patient selection</td>
</tr>
<tr>
<td></td>
<td>(phase IV)</td>
<td>• Early identification of (non)responders</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Efficacy for new indications</td>
</tr>
</tbody>
</table>

Table 2: Potential roles of molecular imaging in the drug research and development pipeline. EMEA: European Medicines Agency; FDA: Food and Drug Administration.

more rapid and cheaper drug development. Finally, once approved, molecular imaging might be used in these new drugs to carefully select patients in order to withhold these medicines in those who will most likely not benefit from it by screening for their molecular target or early response evaluation. More investigations should explore the use of molecular imaging for these roles and would probably make these kinds of imaging more affordable. Some pharmaceutical companies are currently developing therapeutic compounds in combination with companion diagnostics, e.g., Amgen and Illumina market a test based on multigene next-generation sequencing to establish RAS-mutation status of patients who would be appropriate to receive panitumumab, a fully human anti-EGFR monoclonal antibody\[523\].

Randomised controlled clinical trials using molecular imaging

The previously mentioned use of $^{18}$F-FDG PET and other molecular imaging techniques as an imaging biomarker for patient selection or (early) evaluation of therapy response, could not only be used as surrogate endpoint in clinical drug development studies, but could also be applied in clinical trials to individualise patient treatment. Currently, treatment strategies in most solid tumours are based on patient factors, such as operability, age, general health and tumour factors, such as histology and staging, i.e., localised treatment such as surgery or radiotherapy for localised disease, systemic treatment in case of metastasised disease. Newer, targeted therapies also take into account receptor expression, such as trastuzumab in HER2-positive breast carcinoma, hormonal therapy in oestrogen/progesterone-receptor positive breast carcinoma, rituximab in CD20 positive malignant lymphomas or leukaemias and cetuximab in Epidermal Growth Factor Receptor positive K-RAS wild-type metastatic CRC, NSCLC and head and neck carcinomas. Using histopathological biomarkers, such as immunohistochemistry, target expression on a tumour sample can be deter-
General Discussion & Future Prospects

ted. However, molecular imaging has the potential advantage that target expression of all known tumour lesions can be measured in vivo, therefore removing the risk of tissue sampling error or heterogeneity.

In groups of patients with similar patient and tumour characteristics, the problem of biological heterogeneity in tumour response to (targeted) therapy remains. Currently, no clear biomarkers have been found that can select patients who will or will not respond to therapy, but it has been shown that $^{18}$F-FDG PET/CT guided treatment decisions improve survival in malignant lymphoma and some solid tumours [514,524]. In solid tumours, most studies focus on staging in NSCLC [525,526], including prevention of futile thoracotomies [527–529]. Others focus on the follow-up in NSCLC [530], prevention of futile direct laryngoscopy in larynx carcinoma [531], recurrence detection in CRC [217], prevention of futile laparotomies in CRLM [18] and treatment planning in cervical carcinoma [532]. In PET-response adapted treatment, most trials focus on neoadjuvant chemotherapy in adenocarcinoma of the oesophagogastric junction [20,533,534] and breast carcinoma (ClinicalTrials.gov Identifier: NCT01641406). For nonresponders to neoadjuvant chemotherapy in locally advanced oesophagogastric tumours, the prognosis is inferior to that of patients primarily treated with surgery. This may be due to tumour progression during ineffective chemotherapy, but also due to toxicity. The Metabolic response evalUatioN for Individualisation of neoadjuvant Chemotherapy in oesOphageal and oesophagogastric adeNocarcinoma phase I (MUNICON)-I trial evaluated the clinical outcome of a cohort of 110 patients with locally advanced $^{18}$F-FDG avid oesophagogastric tumours in which PET-response after two weeks of neoadjuvant induction chemotherapy using platinum and fluorouracil based treatment. The response was dichotomised at a cut-off $\Delta$SUV of -35%, previously found to be a relevant cut-off value to differentiate metabolic responders and nonresponders in this disease. The metabolic responders completed 12 weeks of chemotherapy and then proceeded to surgery and metabolic nonresponders immediately proceeded to surgery. Compared with cohorts from previous studies, one can conclude that the outcome of metabolic nonresponders was not at all compromised by the early discontinuation of chemotherapy [20].

Instead of stopping unbeneficial cancer treatment in metabolic-nonresponders, intensification of therapy is also investigated; in the similar MUNICON-II trial, the metabolic nonresponders were switched to salvage neoadjuvant chemoradiation in order to improve histopathological response, showing an increased histopathological response compared to MUNICON-I (26% vs 0%). However, the primary endpoint of the study to increase the $R_0$ resection rate from 74% to 94% was not met. No benefit in prognosis was observed in the subgroup of PET-nonresponders and almost 50% of the metabolic nonresponders showed distant metastases shortly after chemoradiation, indicating the unfavourable tumour biology that could not be reversed by radiation (total dose of 32 Gy) plus concurrent chemotherapy (cisplatin or 5-fluorouracil) [533].

Despite the promising results of these elegantly designed studies, so far no randomised controlled trials have been published that use treatment-adaptation based on molecular imaging, but according to online trial registers these studies are currently running and will hopefully be published in the second half of this decade.
Final remarks

Positron Emission Tomography, which once emerged as a scientific tool, is currently highly integrated in the clinical care of patients. Many developments in both tracers and equipment are driven by their clinical application. Trial designs in which therapy response-adaptation based on molecular imaging are emerging; mostly in malignant lymphoma but some, non-randomised, uncontrolled trials have been published in solid tumours. These studies would not only take into account the value of $^{18}$F-FDG PET in tissue characterisation, disease detection and staging, but would also exploit its prognostic and predictive value. The proliferation of many variants in methodology, however, hampers not only comparability of different studies or multicentre study initiation, but suboptimal methods of PET acquisition and analysis potentially lead to underpowered or false-negative studies, which would have a negative effect on the potential of this technology. Efforts should be made towards the harmonisation of different technologies, not only in clinical practice but also in preclinical and clinical research. This, together with optimisation of research design strategies, surrogate outcome measures and the obligation to provide evidence for the cost-benefit ratio should become the responsibility of researcher and clinicians in the field of molecular imaging in order to keep medical research and clinical medicine affordable.

***
Summary & Conclusions

Unpublished

D. Vriens

1dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands
Introduction

The aim of this thesis was to address three diagnostic challenges in 2-(\(^{18}\text{F}\))fluoro-2-deoxy-d-glucose positron emission tomography (\(^{18}\text{F}\)-FDG PET) in solid tumours: (i) the influence of different methods for data-analysis on measured metabolic therapy response and their effect on its relation with clinical outcome, (ii) metabolic and vascular characterisation of solid tumours and (iii) clinical and economical impact of the use of \(^{18}\text{F}\)-FDG PET/CT for tissue characterisation. The thesis started with a general introduction to the methodology of quantitative \(^{18}\text{F}\)-FDG PET including its aims and scope.

Methodological considerations & clinical relevance

Part I of this thesis provided a general introduction to quantitative \(^{18}\text{F}\)-FDG PET/CT, reviewing methodological considerations from patient selection and preparation to analysis of the acquired data and a clinical example of the use of \(^{18}\text{F}\)-FDG PET in current and near-future practice. Chapter 1 reviewed all major factors influencing quantitative results of \(^{18}\text{F}\)-FDG PET in oncological applications. Biological factors that influence quantification of glucose metabolism are important for study design and execution. The impact of fasting blood glucose on the resulting quantitative parameters of tumour metabolism is important for proper patient selection (e.g., excluding patient with diabetes mellitus) and preparation. As uptake time is an important factor for the parameter standardised uptake value (SUV), execution of the investigation must be standardised to this factor. A trade-off between high temporal and spatial resolution with low signal-to-noise ratios in 4D dynamic PET on the one hand and patient safety, keeping the ionising radiation dose as low as reasonable achievable (ALARA) requires optimisation of the acquisition protocol. The important role of the low-dose CT in quantitative PET, stresses the importance to prevent misalignment and the use of intravenous contrast agents. Reconstruction algorithms are vendor-dependent and improve over different generations of PET-scanners. Optimisation of the algorithm settings with respect to quantification instead of visual interpretation is often necessary and performance between different scanners should be addressed in multicentre trials. Even at the time images have been generated, the vast analytical approaches that exist which use different volumes-of-interest and different quantitative parameters may in part be redundant, for another part be very useful for specific questions only. This chapter forms the basis for the research questions and methods used in the rest of this thesis.

Chapter 2 reviewed the role of \(^{18}\text{F}\)-FDG PET from a clinician’s point-of-view in the example of colorectal carcinoma (CRC). The limited role of \(^{18}\text{F}\)-FDG PET staging on management of both rectum carcinoma (16%) and colorectal carcinoma (CRC, 11%) has resulted that this technology has not been advised in current clinical guidelines for staging purposes. However, the evidence for its influence in case of suspected recurrence on individual patient management in case of equivocal radiologic findings (management change in 45-66%) and in unexplained rise of tumour markers (management change in up to 82%) has been formalised in official guidelines.
Especially its value in the detection of (unexpected) distant metastases has led to the routine use of $^{18}$F-FDG PET/CT during restaging of resectable local recurrence (management change up to 59%) or colorectal liver metastases (CRLM, management change in 22-32%) leading to a decrease in futile laparotomies. Molecular imaging using $^{18}$F-FDG PET, however, also has a role in prognostic stratification of patient prior to and during therapy. Many studies support the evidence that a lower $SUV$ before therapy is related to a favourable prognosis and that a higher $SUV$ correlates with a larger morphological response to treatment. The (early) effect of treatment measured by a relative change in $SUV$ between baseline and follow-up $^{18}$F-FDG PET was found to correlate with patient histopathological or morphological response and survival for radiotherapy in rectal carcinoma, for local ablative therapy in CRLM and for chemotherapy in metastasised CRC. No prospective randomised trials have been published that use the prognostic and predictive ability of $^{18}$F-FDG PET, thus the evidence is lacking for routine clinical use for guidance in a tailored therapy or follow-up regime.

Optimisation of quantification methodology

Part II of this thesis provided evidence for the use of different analytical methods from a clinical point of view, using hard clinical endpoints such as overall and progression-free survival. It also questioned the use of the most invasive part of 4D dynamic $^{18}$F-FDG PET: serial arterial sampling. This part thereby focussed on the first aim of this thesis: the influence of different methods for data-analysis on measured metabolic therapy response and their effect on its relation with clinical outcome.

Chapter 3 addressed different normalisation factors used for the $SUV$. The activity concentration is normalised for physical decay corrected administered activity per unit distribution volume. Additionally, normalisation can be used for fasted serum plasma glucose. Measures for distribution volume used are bodyweight, ideal bodyweight, lean body mass or body surface area. As $SUV$s normalised for (ideal) bodyweight are still strongly positively correlated with on bodyweight itself, some guidelines such those of the European Organisation of Research and Treatment of Cancer, recommend lean body mass as normalisation parameter of distribution volume for $SUV$ computation. In chemotherapy response evaluation using $^{18}$F-FDG PET, relative changes in $SUV$ ($\Delta S U V$) of a PET-scan performed before and during the course of chemotherapy were used. These parameters may all be influenced by a change in body composition during treatment. In a prospectively collected dataset of 48 patients with CRC and 49 patients with non-small cell lung carcinoma (NSCLC), the prognostic ability of $\Delta S U V$ with respect to overall and progression-free survival were evaluated. Despite moderate changes in individual patient bodyweight (from -15 kg to +10 kg), the prognostic ability of four differently normalised $\Delta S U V$s was very similar in both cancer types. It was concluded that with respect to relative changes in $SUV$, no method showed statistical advantage over the other.

Chapter 4 investigated another factor of analysis of $^{18}$F-FDG PET in therapy response evaluation. At different evaluation time points, the volume-of-interest (VOI) is often redefined. Therefore, in the hypothetical case a tumour shows a morphological response to treatment without any change in the metabolism of the remaining
malignant tissue, repetitive definition of the VOI would lead to the same value of the parameter of glucose metabolism. An alternative VOI-segmentation method was introduced, where the same volume of tissue is evaluated twice, both before and during the course of chemotherapy. In a prospectively collected dataset of 49 patients with CRC and 50 patients with NSCLC, metabolic response to chemotherapy was evaluated using relative changes in Gjedde-Patlak glucose metabolic rate (\(\Delta MR_{glc}\)) using either redefinition of the metabolic tumour volume (fixed threshold) or applying the original VOI to the follow-up scan (fixed volume). As expected, mean \(\Delta MR_{glc}\) values of a fixed volume VOI were significantly higher compared to a fixed threshold VOI as also tumour shrinkage was taken into account (CRC: -52% vs -30%, NSCLC: -50% vs -37%). However, when dichotomised at their individual median values, \(\Delta MR_{glc}\) determined using a fixed volume VOI showed a better separation between responders and nonresponders in overall and progression-free survival than \(\Delta MR_{glc}\) determined in a fixed threshold VOI in both cancer types. Although this method is more user dependent, this chapter shows that a method both incorporating changes in tumour metabolism and metabolic tumour volume has better prognostic value than only incorporating changes in tumour metabolism in these two types of cancer.

Chapter 5 studied three different alternatives to a fully peripherally sampled arterial plasma time activity concentration curve (APTAC). For pharmacokinetic analysis of a 4D dynamic \(^{18}\)F-FDG PET using a compartment model, this ‘input function’ is needed for computation of tumour metabolic parameters. A population-based input function was derived and tailored to the individual patient either non-invasively by administered activity and initial volume of distribution of \(^{18}\)F-FDG (iDV) or by a single late peripheral blood sample. The third approach is to use the 4D dynamic PET images to obtain an image-derived input function (IDIF). From a set of 120 studies of patients with NSCLC, CRC or breast carcinoma, 80 datasets including fully arterially sampled APTACs were randomly selected. After normalisation for administered activity and iDV, a population-based function was derived based on the mean fit parameters of a three-compartment model. In a validation set consisting of 40 other, paired base-line and follow-up other dynamic \(^{18}\)F-FDG PET-scans, individual study \(MR_{glc}\) and paired study \(\Delta MR_{glc}\) were computed using either an individualised version of the population-based function (either by individual patient administered activity and iDV or one late arterial plasma sample) or IDIF and were compared to the values obtained using an individual fully sampled APTAC. The \(MR_{glc}\) and \(\Delta MR_{glc}\) values obtained, using non-invasive approaches such as an individualised population-based input function or IDIF, were very similar to the gold standard of a fully sampled input function (correlation coefficient: 0.95-0.96). This could be further improved by using only one late arterial plasma sample (correlation coefficient: >0.99). These alternative methods can be used when an IDIF is not available and when serial arterial sampling is not desirable.

**Tissue heterogeneity & therapy response evaluation**

**Part III** of this thesis focused on tissue characterisation: regional rather than whole-tumour metabolic parameters of glucose metabolism are explored as a parameter for tissue heterogeneity. Also, multimodality multiparametric approaches are used to
describe the metabolic and vascular response of tumours to therapy. This part thereby focussed on the second aim of this thesis: metabolic and vascular characterisation of solid tumours.

Chapter 6 focused on a detailed description of regional pharmacokinetic parameters of glucose metabolism using 4D dynamic $^{18}$F-FDG PET. Most studies describe whole-tumour parameters of transmembranous glucose uptake ($K_1$), washout ($k_2$), phosphorylation rate ($k_3$) and tumour blood volume fraction ($V_B$). In this study, 104 lesions of 41 previously untreated patients with NSCLC, CRC or breast carcinoma were imaged using 4D dynamic $^{18}$F-FDG PET. Parametric $MR_{gls}$ datasets were generated using voxelwise modelling and tumours were segmented in four quartiles of background subtracted lesion maximum $MR_{gls}$. Pharmacokinetic analysis was performed using an irreversible two-tissue compartment model in the three segments with highest $MR_{gls}$ to determine the rate constants ($K_1$-$k_3$) of glucose metabolism and $V_B$. From the highest to the lowest quartile, significant decreases in uptake, washout and phosphorylation rate constants were seen with significant increases in tissue blood volume fraction. It was concluded that tumour regions with highest $MR_{gls}$ are characterised by high cellular uptake and phosphorylation rate constants with relatively low blood volume fractions. In regions with less metabolic activity, the blood volume fraction increases and cellular uptake, washout and phosphorylation rate constant decrease. These results support the hypothesis that regional tumour glucose phosphorylation rate is not dependent on the transport of nutrients such as glucose and $^{18}$F-FDG to the cells.

Chapter 7 and 8 described response in morphologic, metabolic and vascularity parameters in patients with CRLM between baseline studies and studies acquired after three cycles of chemotherapy. 4D dynamic gadopentetate dimeglumine-contrast enhanced MRI (DCE-MRI) was used to obtain tumour maximal axial diameter and pharmacokinetic rate constants of contrast agent exchange between the extracellular extravascular space and the intravascular plasma compartment ($k_{ep}$), the volume transfer constant between these compartments ($K^{trans}$) and the volume fraction of the extracellular extravascular space of the lesions. Likewise, whole tumour $MR_{gls}$ values were computed using Gjedde-Patlak analysis of 4D dynamic $^{18}$F-FDG PET data. In chapter 7, twenty-three patients with CRLM were evaluated after three cycles of only cytotoxic chemotherapy. In chapter 8, two patients were evaluated after three cycles of anti-angiogenic treatment in combination with cytotoxic therapy. Without the anti-angiogenic drug bevacizumab, pretreatment $MR_{gls}$ and tumour diameter were individually predictive of overall and progression-free survival. During treatment, $K^{trans}$ increased and $MR_{gls}$ decreased significantly. There were no correlations between relative changes in vascularity and relative changes in metabolism. $\Delta MR_{gls}$ was significantly correlated with overall survival, but the changes in vascularity parameters were not associated with a change in survival. In the two patients in whom bevacizumab was administered, large reductions both in parameters of functional tumour vascularity and in metabolic parameters were observed, suggesting an additional role of DCE-MRI when anti-angiogenic treatment is given.
Clinical impact of molecular imaging

Part IV of this thesis returned to clinical application of $^{18}$F-FDG PET in a potential new indication of tissue characterisation of thyroid nodules. It used methods of systematic review, meta-analysis and Monte-Carlo decision analysis on published data to support safety and cost-effectiveness of this indication of $^{18}$F-FDG PET. This last part thereby focussed on the third aim of this thesis: clinical and economical impact of the use of $^{18}$F-FDG PET/CT for tissue characterisation.

Chapter 9 systematically reviewed six studies in 225 individual patients on the use of $^{18}$F-FDG PET in tissue characterisation of cytologically indeterminate thyroid nodules. Current practice advises diagnostic thyroid lobectomy in all these patients, however, a malignancy is established in only 25%. There was mild to moderate heterogeneity in the published studies. A pooled sensitivity and specificity of $^{18}$F-FDG PET of 95% and 48% was found, respectively. With a prevalence of malignancy of 26%, a negative and positive predictive value of 96% and 60% was found, respectively. No evidence of threshold effects or publication bias could be established. However, a relation between $^{18}$F-FDG PET sensitivity and nodule diameter was found, as in nodules larger than 15 mm this was 100%, potentially reflecting the effect of the limited resolution of PET-scanners used in the included studies.

Chapter 10 further explored the benefit of the use of $^{18}$F-FDG PET in cytologically indeterminate thyroid nodules. The high predictive value of the 37% negative $^{18}$F-FDG PET-scans, found in chapter 9, led to the hypothesis that full incorporation of this technology in the work-up of these patients could be cost-effective by saving on surgery, hospital admission and complications. Currently, advanced analyses on cytological material are marketed, either with a highly sensitive gene-expression classification (GEC) or highly specific mutation marker panel (MMP). Highly sensitive markers such as $^{18}$F-FDG PET and GEC could save on unbeneficial surgery and highly specific markers such as the MMP on two-stage surgery by directly indicating total thyroidectomy as treatment of choice. The decision analysis presented in this chapter computed individual patient direct medical costs and health-related quality-adjusted life expectancy using a Monte-Carlo approach on all four strategies (PET, GEC, MMP and current standard) over a duration of five years. $^{18}$F-FDG PET could potentially prevent 47% of current unnecessary surgery procedures, leading to lower costs and a modest increase in health-related quality of life. Compared to the other approaches it is the least expensive alternative, on average saving €822 per patient over five years compared to diagnostic thyroid lobectomy in all patients. Its effectiveness was similar to the GEC, but this latter approach was found to be on average €1,358 more expensive per patient over five years than the use of molecular imaging.

Final remarks

The thesis concluded with a general discussion including future perspectives. Multimodality multiparametric imaging is described using a combination of $^{18}$F-FDG PET and DCE-MRI. A multidimensional approach describing different aspects of individual tumour lesions simultaneously holds many advantages for future individualisation of treatments, which will become more and more important when treatments
become targeted and more expensive. The clinical example used in the last part of this thesis is an illustration that clinical studies should not only aim at a mere description of accuracy and optimisation of a diagnostic test but also must compare the biomarker under investigation to alternative biomarkers with respect to accuracy, cost-effectiveness and implementability. At last: any measurement should be able to influence the phenomenon (i.e., disease) under investigation (free interpretation of the quantum theory’s observer effect).

***
Samenvatting & Conclusies in het Nederlands voor niet-ingewijden (Layman’s Summary & Conclusions in Dutch)

Unpublished

D. Vriens

1dept. of Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands
Kwantitatieve karakterisering van solide tumoren met behulp van \( ^{18} \text{F-FDG PET} \)

- What’s in a Number?

Het in beeld brengen en kwantificeren van de stofwisseling van tumoren is één van de belangrijkste manieren om gedetailleerde informatie over het gedrag van verschillende soorten kanker te verkrijgen en deze in de tijd te vervolgen. In de dagelijkse praktijk wordt hiervoor een merkstof (tracer) gebruikt welke wordt afgebeeld met positronen emissie tomografie (PET), vaak in combinatie met computer tomografie (CT), zodat zowel de functie als de anatomie of structuur van de tumor wordt afgebeeld.

In dit proefschrift werden drie diagnostische problemen met betrekking tot \( ^{18} \text{F-FDG PET} \) onderzocht, namelijk: (1) de invloed van verschillende manieren om \( ^{18} \text{F-FDG PET} \)-scans te analyseren tijdens de behandeling van verschillende typen kanker, (2) het bepalen van zowel de stofwisseling als de bloedvatvoorziening in tumoren en (3) de relevantie van het gebruik van deze technologie in de dagelijkse praktijk, zowel op diagnostisch als economisch vlak.

Introductie

In de westerse wereld is kanker sinds 2008 de voornaamste doodsoorzaak: in 2012 waren in Nederland 27% van alle sterfgevallen het gevolg van hartvaatziekten en bijna 32% het gevolg van kanker. Bij ruim 40% van de bevolking wordt gedurende het leven kanker gediagnostiseerd, waarvan de helft voor het 66e levensjaar. De geschatte jaarlijkse kosten van deze ziekte in de Europese Unie is € 126 miljard. De meeste kankers worden veroorzaakt door zogenaamde solide tumoren (tumoren die niet afkomstig zijn van de bloedcellen, zoals leukemie en het lymfomen). De vijf meest voorkomende solide tumoren zijn: kanker van de prostaat, de borst, de long, de dikke darm en endeldarm en huidkanker (melanoom). Long- en darmkanker veroorzaken het hoogste aantal sterfgevallen.

Bij het vaststellen van de uitbreiding van de ziekte (lokaal, regionaal of uitgezaaid) wordt vaak gebruik gemaakt van morfologische beeldvormende onderzoeken of moleculaire beeldvormingstechnieken. Morfologische beeldvormingstechnieken (zoals endoscopie, echografie, röntgen-opnamen, CT, kernspinresonantie-tomografie (MRI)) richten zich op zichtbare afwijkingen in de structuur van weefsels. Deze onderzoeken worden meestal gebruikt voor een bepaalde lichaamsregio en zijn dus met name behulpzaam voor het bepalen van lokale en regionale uitbreiding van de ziekte. Molecuulaire beeldvorming (zoals PET of Enkel Foton Emissietomografie (SPECT)) brengt de functie van de verschillende weefsels van het gehele lichaam in beeld en wordt daarom voornamelijk gebruikt voor het uitsluiten van uitzettingen. Bij kanker worden over het algemeen vroegtijdiger functionele afwijkingen van de weefsels gezien en volgen structurele afwijkingen pas later. Daardoor is morfologische beeldvorming veelal pas later in het ziekteproces afwijkend dan moleculaire beeldvorming. Bovendien is het vaak onduidelijk of een afwijking in de structuur het gevolg is van de kanker of een andere (goedaardige) oorzaak heeft. Molecuulaire beeldvorming richt zich voornamelijk
op ofwel de stofwisseling ofwel het aantonen van moleculaire oppervlaktekenmerken van een tumor (bv. receptoren) ofwel het bepalen van de chemische samenstelling van een weefsel. Hiervoor wordt een zogenaamde tracer gebruikt, die zichtbaar wordt gemaakt door het aan een radioactief isotoop te binden. Een radioactief isotoop kan afgebeeld worden middels PET of SPECT en hiermee dus ook de locatie van de aan deze isotoop verbonden tracer. Deze beeldvormende technieken zijn beiden erg gevoelig voor het detecteren van zeer kleine hoeveelheden tracer in het lichaam. De voornaamste nadelen van deze technieken zijn de beperkte beeldscherpte (oplossend vermogen, resolutie), het gebruik van potentieel schadelijke ioniserende straling en de relatief hoge kosten van beide beeldvormingstechnieken. PET heeft als voordeel boven SPECT dat het een betere beeldscherpte heeft dan SPECT \((\pm 3 \text{ mm} \ vs \ \pm 10 \text{ mm})\) en ook een hogere gevoeligheid heeft dan SPECT, waardoor nog kleinere hoeveelheden tracer kunnen worden gedetecteerd. Daarnaast is het mogelijk de opname van de tracer op de PET beelden in getal uit te drukken, terwijl dit bij SPECT veel minder eenvoudig is.

Voor moleculaire beeldvorming van de stofwisseling van tumoren door middel van PET, wordt de tracer 2-(\(^{18}\text{F}\))fluoro-2-deoxy-d-glucose (\(^{18}\text{F-FDG}\)) het meest gebruikt. \(^{18}\text{F-FDG}\) is een stof die veel overeenkomsten heeft met glucose (suiker), maar die – anders dan glucose – aangepast is zodat hij niet wordt verbrand door de cel, maar opstapelt in weefsel. De aan glucose verwante moleculen kunnen in het lichaam gelokaliseerd worden doordat deze gekoppeld zijn aan de radioactieve isotoop \(^{18}\text{F}\), welke gedetecteerd kan worden met PET. Na injectie in de bloedbaan, wordt de tracer actief door levend weefsel opgenomen, afhankelijk van diens glucosebehoefte. Vervolgens wordt de \(^{18}\text{F-FDG}\) gebonden tot \(^{18}\text{F-FDG-6-fosfaat}\) en kan het niet verder worden afgebroken noch worden vrijgemaakt uit dit weefsel: de hoeveelheid zal in de loop van de tijd dus toenemen. Kwaadaardige weefsels, dat wil zeggen kanker, hebben over het algemeen meer glucose nodig dan goedaardige weefsels omdat ze enerzijds meer energie verbruiken en anderzijds een minder efficiënte energieproductie hebben. Kwaadaardig weefsel stapelt na injectie van \(^{18}\text{F-FDG}\) dus sneller en meer van deze tracer dan gezonde weefsel. De hoeveelheid opgenomen \(^{18}\text{F-FDG}\) in weefsel is dus een afspiegeling van diens energiebehoefte. Helaas is dit proces niet specifiek voor kwaadaardigheid: ook ontstekingsprocessen en infecties verbruiken veel energie en vertonen daardoor ook een hogere stapeling van deze tracer.

In de hedendaagse oncologie (kankerkunde) is \(^{18}\text{F-FDG PET/CT}\) het meest gebruikte moleculaire beeldvormingsonderzoek voor het onderscheiden van goed- en kwaadaardig weefsel en het bepalen van de uitbreiding van de ziekte. Naast het vaststellen van de uitbreiding van de ziekte, kan het ook bijdragen bij het vinden van de oorspronkelijke tumor als er alleen nog uitzaaingen bekend zijn of helpt het vaststellen van de beste locatie voor een biopsie (waarbij een representatief stukje weefsel verwijderd om onderzocht te worden). De mate van stapeling van \(^{18}\text{F-FDG}\) in een tumor is vaak direct gerelateerd aan de delingssnelheid. Daarom voorspelt de mate van tracerstapeling ook de prognose van patiënt, waarbij een hogere opname (en dus delingssnelheid) meestal is gerelateerd aan een slechtere prognose.
Ook kan de effectiviteit van een behandeling voorspeld worden aan de hand van het gemeten energieverbruik: snel delende weefsels zijn weefsels die meer $^{18}$F-FDG opnemen en blijken vaak gevoeliger voor chemotherapie en bestraling. Tenslotte wordt een $^{18}$F-FDG PET-scan ook tijdens bestraling gebruikt, om het te bestralen gebied af te bakenen.

$^{18}$F-FDG PET kan eveneens ingezet worden om kankerpatiënten te vervolgen tijdens of na een behandeling tegen kanker. Omdat veranderingen van de energiestofwisseling vlak nadat een therapie is gestart, meestal gerelateerd zijn aan de prognose van de patiënt, zou $^{18}$F-FDG PET gebruikt kunnen worden om patiënten, die niet reageren op een ingestelde therapie in een vroeg stadium, ofwel een andere, ofwel een intensievere therapie te geven. Ook komt het voor dat tijdens of na de behandeling van kanker gedacht wordt aan het achterblijven van een deel van de tumor of aan terugkeer van de kwaadaardigheid, bijvoorbeeld omdat een afwijking gezien wordt bij morphologische beeldvorming waarvan niet bekend is of dit bestaat uit littekenweefsel of levend tumorweefsel. Ook kan naar aanleiding van een bepaling van specifieke tumormerkstoffen in het bloed, resterend tumorweefsel of terugkeer van de kanker veroord worden, maar weet men nog niet waar in het lichaam zich dit bevindt. In dit geval is $^{18}$F-FDG PET vaak in staat de oorzaak aan te wijzen. Dit proefschrift richtte zich op het bepalen van de prognose van de patiënt vóór start van de behandeling, het bepalen van de effectiviteit van de behandeling door vroeg tijdens de behandeling de verandering in stofwisseling te bepalen en het karakteriseren van weefsel om het onderscheid tussen goed- en kwaadaardig weefsel te maken.

In dit proefschrift werd naast 3D statische PET ook gebruikt gemaakt van 4D dynamische PET. Bij deze laatste methode wordt niet alleen gekeken waar de tracer zich na 1 uur bevindt, maar ook hoe die daar in de tijd gekomen is. Bij 4D dynamisch PET wordt vanaf het moment van injectie van de tracer de stapeling van de stof in verschillende delen van de tumor gedurende de tijd bepaald. Voor analyse van deze beelden, die te vergelijken zijn met een video, wordt gebruik gemaakt van zogenaamde “compartiment-modellen”. Dit zijn beschrijvingen van de wiskundige relatie tussen de hoeveelheid tracer in de tumor en de tijd. Hierbij zijn de reactiesnelheden van de verschillende deelprocessen te bepalen die uiteindelijk ervoor zorgen dat de tracer als $^{18}$F-FDG-6-fosfaat in het weefsel stapelt. Deze deelprocessen zijn: doorbloeding van het weefsel, opname van de tracer in het weefsel en binding van de tracer tot $^{18}$F-FDG-6-fosfaat in dit weefsel. Door deze 4D metingen kan niet alleen inzicht verkregen worden hoeveel $^{18}$F-FDG in een bepaalde tijd wordt opgenomen, zoals bij regulier 3D statische PET, maar ook wat de stofwisselings snelheid van het weefsel is en wat het aandeel van de eerder genoemde deelprocessen is aan deze stofwisselings snelheid is. Er kan zo dus een zeer gedetailleerd beeld verkregen worden van “de processen achter het 3D statische PET plaatje”. Echter, omdat in eerder wetenschappelijk onderzoek aangetoond is dat de minder complexe en patiëntvriendelijker methode van 3D statische PET voor de meeste oncologische vragen 4D dynamische PET kan vervangen, wordt deze laatste techniek alleen voor wetenschappelijke doeleinden ingezet. Mede door de uitgebreide inzetbaarheid van $^{18}$F-FDG PET werden in Europa in 2011 ruim 900,000 PET-scans per jaar vervaardigd. Meer dan 95% van alle PET-scans wordt gemaakt met $^{18}$F-FDG, hoewel inmiddels honderden andere PET-tracers bestaan.
Het feit dat de opname van de tracer, bepaald met PET, uitgedrukt kan worden in een getal (d.w.z. gekwantificeerd kan worden) biedt veel voordelen voor wetenschappelijk onderzoek. Echter, de manier waarop de getallen bepaald worden, varieert enorm. Vele factoren, zowel biologisch, technisch, als analytisch, zijn van invloed op het uiteindelijke numerieke resultaat van kwantitatieve PET. Al deze variaties leiden er toe dat de resultaten van verschillende studies moeilijk te interpreteren zijn en lastig met elkaar kunnen worden vergeleken. Zeker wanneer patiënten voor eenzelfde studie in meerdere centra in kaart worden gebracht, is uniformiteit en optimalisatie van het studieprotocol zeer wenselijk. Dit proefschrift testte een aantal verschillende methodologische aspecten die aan bod moeten komen in het studieprotocol en doet uiteindelijk aanbevelingen voor het gebruik hiervan in toekomstige studies.

Methodologische overwegingen & klinische relevantie

In deel I van dit proefschrift werd uitgebreid stilgestaan bij de methodologische overwegingen die belangrijk zijn voor de selectie van geschikte patiënten voor (kwantitatieve) $^{18}$F-FDG PET tot en met de uitwerking van scans. Tevens werd als voorbeeld de huidige en toekomstige rol van deze technologie bij darmkanker beschreven.

In hoofdstuk 1 werden de belangrijkste factoren benoemd die de resultaten van kwantitatief $^{18}$F-FDG PET beïnvloeden. Het bepalen van de invloed van biologische factoren die de kwantificatie van glucosestofwisseling beïnvloeden, zoals het nuchter bloedglucose van de patiënt en het interval tussen toediening van de tracer en het kwantificeren van tracer-opname, zijn belangrijk voor het ontwerpen en uitvoeren van wetenschappelijke studies. Technische factoren die de kwantificatie van glucosestofwisseling beïnvloeden, waaronder de hoeveelheid toegediende radioactiviteit, maar ook de "reconstructie-algoritmen" die tot de uiteindelijke beelden leiden, dienen in aanbevelingen te worden opgenomen, met als doel betrouwbare en herhaalbare kwantificatie, ook wanneer met meerdere PET-scanners gewerkt wordt. Tenslotte zijn er talrijke manieren van analyse van de beelden, waarin verschillende interessegebieden en verschillende parameters voor glucosestofwisseling gebruikt worden. Waarschijnlijk zijn deze variaties grotendeels overbodig, maar deels mogelijk juist erg zinvol voor specifieke doeleinden. Dit hoofdstuk vormde de basis van de onderzoeksvragen die in de rest van dit proefschrift werden uitgewerkt.

Hoofdstuk 2 richtte zich op de rol van $^{18}$F-FDG PET voor kanker van de dikke- en endeldarm vanuit het gezichtspunt van de clinicus. Er werd een samenvatting gegeven van gepubliceerde patiënten onderzoek om te onderbouwen wanneer het gebruik van $^{18}$F-FDG PET slechts beperkt zinvol is, juist erg zinvol is en waarvoor deze technologie in de toekomst mogelijk nog meer gebruikt kan worden. Zo is de aanvullende waarde van $^{18}$F-FDG PET voor het bepalen van de uitbreiding van die ziekte beperkt als reeds structurele beeldvorming van de romp heeft plaatsgevonden. Het verrichten van een $^{18}$F-FDG PET is echter wel van waarde bij patiënten waarbij vermoed wordt dat er sprake is van terugkeer van de kanker, bijvoorbeeld op basis van verdachte afwijkingen tijdens structurele beeldvorming of doordat er een toename is in tumormerkstoffen in het bloed. Ook is het in de dagelijkse zorg zinvol $^{18}$F-FDG PET te gebruiken om uitzaaiingen buiten het geplande operatiegebied uit te sluiten om vergeefse operaties te voorkomen. Aangezien het is aangetoond dat de mate van
Optimalisatie van kwantificatie-methodologie

Optimalisatie van kwantificatie-methodologie

Deel II van dit proefschrift richtte zich op het gebruik van verschillende analyse methoden vanuit een klinisch perspectief: door het gebruik van ‘harde klinische uitkomstmaten’ zoals de levensverwachting van de patiënten. Dit deel besprak tevens de noodzaak van het voor de patiënt meest belastende deel van 4D dynamisch $^{18}$F-FDG PET, namelijk het repetitief bepalen van het gehalte $^{18}$F-FDG in het slagaderlijk bloed. Deel II richtte zich dan ook op de eerste doelstelling van dit proefschrift: de invloed van verschillende manieren om $^{18}$F-FDG PET-scans te analyseren tijdens de behandeling van verschillende typen kanker.

Hoofdstuk 3 richtte zich op verschillende methoden om de “gestandaardiseerde $^{18}$F-FDG opname-waarde” (SUV) te berekenen. De SUV is de meest gebruikte maat voor glucosestofwisseling. Er worden verschillende methoden gebruikt om de SUV te berekenen, waaronder verschillende parameters voor de grootte van de patiënt (lichaamsgewicht, ideaal lichaamsgewicht, vetvrij lichaamsgewicht of lichaamsoppervlakte) en nuchter bloed glucose gehalte. Wanneer het behandelresultaat van chemotherapie bepaald wordt middels $^{18}$F-FDG PET, wordt vaak de relatieve verandering in de SUV ($\Delta$SUV) tussen een PET-scan vóór en tijdens therapie gebruikt. Deze waarden worden dus ook beïnvloed door een verandering in lichaamssamenstelling tijdens de behandeling, zoals bijvoorbeeld het frequent voorkomen van afvallen of verlies van vet- of spiermassa. In een groep patiënten met darmkanker of longkanker behandeld met chemotherapie werd het voorspellende vermogen van verschillende varianten van het berekenen van de $\Delta$SUV op totale en progressievrije levensverwachting bepaald. Ondanks dat er veranderingen optraden in lichaamsgewicht van deze patiënten (individueel variërend van -15 kg tot +10 kg), was het voorspellend vermogen van alle vier de onderzochte $\Delta$SUV varianten bij beide kankertypen vergelijkbaar.

Hoofdstuk 4 onderzocht een andere onderdeel van de analyse van $^{18}$F-FDG PET voor evaluatie van de effectiviteit van de ingezette behandeling. Tijdens de verschillende scan-momenten wordt meestal het interessegebied (het weefsel waarvan de stofwisseling moet worden geanalyseerd) opnieuw bepaald. In het geval dat een tumor tijdens behandeling wel kleiner wordt, maar de stofwisseling in het resterende tumorweefsel niet verandert, zou het deze analysemethode leiden tot een onjuiste waarde voor de glucosestofwisseling. Als men naar het getal kijkt, lijkt het dan dus alsof er geen verandering is opgetreden door de therapie. Er werd in dit hoofdstuk een alternatieve methode voorgesteld, waarbij hetzelfde interessegebied gehanteerd werd zowel vóór als tijdens chemotherapie. Als beide methoden voor het meten van tumorstofwisseling gebruikt werden bij patiënten met darmkanker of longkanker die chemotherapie kregen, bleek de door ons voorgestelde methode beter in staat patiënten met een goede en een slechte levensverwachting te scheiden dan de methode.
waarbij het interessegebied opnieuw wordt bepaald. Om onderscheid te maken tussen patiënten met een goede of een slechte prognose moet men dus naast met veranderingen in stofwisseling ook met veranderingen in tumorgrootte rekening houden.

In hoofdstuk 5 werden drie alternatieve methoden voor het bemonsteren van de $^{18}$F-FDG concentratie in het bloed onderzocht. Zoals boven beschreven moet voor analyse van 4D dynamisch PET het gehalte $^{18}$F-FDG in het bloedplasma van de slagader die de tumor voedt bekend zijn. Aangezien deze vaak onbereikbaar is omdat de tumor dieper in het lichaam gelegen is, wordt vaak een beter toegankelijke slagader zoals die in de pols bemonsterd. Ook in de pols is dit een ingrijpende, pijnlijke meting welke ook kans op complicaties kent. In dit hoofdstuk werd gezocht naar een alternatief voor deze methode, door het bepalen van een $^{18}$F-FDG bloedcurve die gold voor al de patiënten (generieke populatie-curve) of door het meten van de $^{18}$F-FDG concentratie in het bloed in de grote lichaamsslagader op de PET-beelden. Er werd gezien dat deze niet-ingrijpenden methoden vrijwel gelijkwaardig zijn aan het repetitief bemonsteren van de polsslagader. Met name wanneer de generieke populatie curve geïndividualiseerd werd door gebruikt te maken van een enkel bloedmonster kwam deze nagenoeg overeen met de metingen uit de polsslagader. Individualisering van de populatie-curve door gebruik te maken van lengte en gewicht van de patiënt was even goed als het bepalen van deze curve uit de PET-beelden. Er werd geconcludeerd dat een geïndividualiseerde generieke populatie-curve gebruikt kan worden wanneer er geen grote slagader is afgebeeld (zoals bij het afbeelden van hoofd of leden) en wanneer herhaaldelijke slagaderlijke bemonstering niet wenselijk is.

Weefselheterogeniteit & evaluatie van effectiviteit van behandeling

Deel III van dit proefschrift richtte zich op de meerwaarde van $^{18}$F-FDG PET bij de beschrijving van verschillende tumorweefsel eigenschappen, zowel wat betreft de regionale verschillen (heterogeniteit) in glucosestofwisseling binnen tumoren als hun functionele bloedvatvoorziening. Deel III richtte zich op de tweede doelstelling van dit proefschrift: het bepalen van zowel de stofwisseling als de bloedvatvoorziening in tumoren.

Hoofdstuk 6 beschreef de regionale verschillen in de deelprocessen van de glucosestofwisseling binnen in tumoren. De meeste gepubliceerde studies beschrijven de waarden van de gehele tumor van deze parameters. In dit hoofdstuk werden tumoren van onbehandelde patiënten met longkanker, darmkanker en borstkanker afgebeeld door middel van 4D dynamisch $^{18}$F-FDG PET. Elke tumor werd opgedeeld in drie niveaus van aflopende waarden van glucosestofwisseling. Hierdoor ontstonden drie schillen, variërend van de “metabole kern” tot de buitenste, minst actieve “schil”. In deze drie tumorregio’s werden de verschillende deelprocessen betrokken bij de opname van $^{18}$F-FDG gekwantificeerd (de snelheid van glucose-opname en verlies over de tumorcelmembrana, de fosforyleringssnelheid in de tumorcel en het relatieve volume van bloed in de tumor). Wanneer vanuit de “metabole tumorkern” naar regio’s met lagere glucosestofwisseling werd gekend werden significante afnames van transport van glucose over de tumorcelmembrana en van fosforyleringssnelheid in de tumorcel-
Klinische invloed van moleculaire beeldvorming

Deel IV van dit proefschrift keerde terug naar de klinische toepassing van $^{18}$F-FDG PET voor de potentiële nieuwe rol voor weefseltypeering van schildklierknobbeltjes. Dit deel van het proefschrift maakte middels geavanceerde rekenkundige methoden gebruik van de beschikbare patiëntengegevens uit de literatuur bij het behalen van de laatste doelstelling van dit proefschrift: de relevantie van het gebruik van deze technologie in de dagelijkse praktijk, zowel op diagnostisch als economisch vlak.

Hoofdstuk 9 maakte gebruik van gepubliceerde studies van patiënten met een schildklierknobbeltje waarbij de patholoog op basis van een punctie geen onderscheid tussen goed- of kwaadaardig kon maken. Huidige richtlijnen adviseren bij deze patiënten de halve schildklier met het knobbeltje te verwijderen. Er wordt echter bij slechts 25% van deze patiënten kanker vastgesteld, welke vervolgens nog een tweede keer geopereerd worden om de andere helft van de schildklier weg te halen. De overige 75% zijn achteraf bezien voor niets geopereerd, met alle risico’s van dien. In deze studie werd gekeken of het maken van een $^{18}$F-FDG PET-scan deze onnodige operaties
kon voorkomen. Wanneer alle patiënten in de literatuur systematisch geanalyseerd werden, bleek de kans op kanker na een negatieve $^{18}$F-FDG PET-scan slechts 4% en de kans op kanker na een positieve $^{18}$F-FDG PET-scan 60%. Het lijkt er dus op dat onnutte operaties, opnamedagen en complicaties voorkomen kunnen worden door het maken van een $^{18}$F-FDG PET-scan vóór het uitvoeren van de operatie.

Naar aanleiding van deze resultaten werd in hoofdstuk 10 verder gekeken naar de veiligheid en de kosten-effectiviteit van het maken van een $^{18}$F-FDG PET-scan voor deze schildklierknobbelzjes. Naast de operatie waarbij de halve schildklier verwijderd wordt en de $^{18}$F-FDG PET-scan worden ook twee andere methoden gebaseerd op genexpressie (GEC) en een test gericht op mutaties (MMP) gebruikt om de kans op kanker in deze knobbelzjes te schatten. Alle vier de strategieën (PET, GEC, MMP en de huidige standaardzorg van operatie bij allen) werden met elkaar vergeleken door middel van zogenaamde Monte-Carlo simulaties, waarbij de medische kosten voor een patiënt maar ook diens kwaliteit van leven berekend. Deze analyses toonden aan dat $^{18}$F-FDG PET in staat is om bijna de helft van de huidige onnutte operaties te voorkomen, hetgeen kan leiden tot lagere kosten en een bescheiden toename in kwaliteit van leven. In vergelijking met de drie andere strategieën is $^{18}$F-FDG PET de goedkoopste optie, zonder dat het afbreuk deed aan veiligheid.

**Slotopmerkingen**

Het proefschrift eindigde met een algemene discussie waarin ook toekomstperspectieven werden besproken.

Alle in dit proefschrift beschreven studies hebben bijgedragen aan de kennis van numerieke beeldvorming middels $^{18}$F-FDG PET bij verschillende frequent voorkomende solide tumoren. Deze kennis kan gebruikt worden voor het optimaliseren van toekomstige studieprotocollen, waarbij $^{18}$F-FDG PET gebruikt wordt voor het bepalen van de prognose en het optimale behandelprotocol voor de patiënt. Daarnaast werd beeldvorming middels een combinatie van technologieën met het oog op het beschrijven van meerdere tumoreigenschappen uitgevoerd. Een benadering waarbij gebruik gemaakt wordt van verschillende beeldvormende technieken om verschillende aspecten van tumoren te beschrijven heeft grote potentie voor het aanpassen van de behandeling op de individuele patiënt. Deze individualisering wordt steeds belangrijker omdat behandelingen steeds doelgerichter maar ook duurder worden. Het onderzoek bij patiënten met schildklierknobbelzjes illustreert dat toekomstig klinisch onderzoek niet beperkt moet blijven tot het beschrijven van de voor de arts belangrijke aspecten van een diagnostische test, maar dat ook de voor de patiënt (levenskwaliteit) en de maatschappij (kosteneffectiviteit) belangrijke aspecten en implementeerbaarheid bepaald moeten worden. Dit met als doel om zowel het medisch wetenschappelijk onderzoek, als de gezondheidszorg beter en doelmatiger te maken, nu en in de toekomst. Immers: elke meting moet in staat zijn het onderzochte fenomeen, in dit geval de ziekte, te beïnvloeden (vrij naar de waarnemer-afhankelijke werkelijkheid uit de kwantumtheorie).
Appendices
Glossary
Glossary

A

$\alpha^{2+}$ – particle consisting of two protons and two neutrons bound together into a particle identical to a helium nucleus. Used for treatment.

$\text{AA}$ – administered activity [Bq]

$\text{AC}$ – activity concentration [Bq cm$^{-3}$]

$\text{AC}$ – attenuation correction

$\text{Acc}$ – accuracy of a diagnostic test

$\text{ADP}$ – adenosine diphosphate (molecule)

$\text{AIC}$ – Akaike’s information criterion

$\text{ALARA}$ – as low as reasonable achievable

$\text{ANOVA}$ – analysis of variance

$\text{AOG}$ – adenocarcinoma of the oesophagogastric junction

$\text{APTAC}$ – arterial plasma time-activity concentration curve, see IF and $C_{\text{plasma}}(t)$.

$\text{ATA}$ – American thyroid association

$\text{ATP}$ – adenosine triphosphate (molecule)

$\text{AUC}$ – area under the curve

$\text{AUROC}$ – area under the receiver operating characteristic

B

$\beta$ – family of continuous probability distributions defined on the interval 0-1.

$\beta^-$ – particle identical to an electron. Used for treatment.

$\beta^+$ – particle identical to an anti-electron. Annihilates with an electron to 2-$\gamma$ and is used for PET-imaging.

$\text{Ba}$ – barium (element). $^{137m}\text{Ba}$ is a $\gamma$-emitting isotope, which is for transmission PET imaging for attenuation correction ($t_{1/2} = 2.55$ min). Its mother nuclide is $^{137}\text{Cs}$.

$\text{BAT}$ – brown adipose tissue

$\text{Bq}$ – Becquerel, unit of radio-activity.

$\text{BRAF}$ – member of the Raf kinase family of growth signal transduction protein kinases, playing a role in regulating the MAP kinase/ERKs signalling pathway, which affects cell division, differentiation and secretion. Mutations in this gene are associated with PTC.

$\text{BSA}$ – body surface area [m$^2$]

$\text{BSO}$ – bismuth germanium oxide

$\text{BW}$ – bodyweight [kg]

C

$\chi^2$ – Pearson’s $\chi^2$ test

$C$ – carbon (element). $^{11}\text{C}$ is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 20.3$ min).

$\text{C}$ – constant

$C\$ – costs [€]

$\text{CANS}$ – Canadian dollar

$C_{\text{bound}}(t)$ – the intracellular activity concentration of bound $^{18}\text{F}$-FDG-6-phosphate in time.

$C_{\text{CCR}}$ – correct classification rate

$C_{\text{CDW}}$ – conventional diagnostic work-up

$C_{\text{CEA}}$ – carcinoembryonic antigen [µg·l$^{-1}$], tumour marker in CRC.

$C_{\text{free}}(t)$ – the intracellular activity concentration of free $^{18}\text{F}$-FDG in time.

$C_{\text{glc}}$ – plasma concentration of glucose [mmol·l$^{-1}$]

$CI$, 95% – (95%-)confidence interval

$C_{\text{I}}$ – cobalt (element). $^{57}\text{Co}$ decays by electron capture. The emitted $\gamma$’s are used for transmission PET imaging for attenuation correction ($t_{1/2} = 271.746$ d).

$\text{CO}_2$ – carbon dioxide (molecule)

$C_{\text{PET}}(t)$ – time activity concentration curve measured by PET.

$C_{\text{plasma}}(t)$ – arterial plasma time activity concentration curve, see APTAC and IF.

$\text{CR}$ – complete response

$\text{CRC}$ – colorectal carcinoma

$\text{CRLM}$ – colorectal liver metastases
Glossary

**C** (continued)

**CRT** – chemoradiotherapy

Cs – caesium (element). $^{137}$Cs decays to $^{137m}$Ba by $\beta^-$-decay. The latter is used for transmission PET imaging for attenuation correction ($t_{1/2} = 30.17$ yr). Its mother nuclide is $^{137}$Cs.

**CSA** – cryosurgical ablation

CT – computed tomography

$C_{tissue}(t)$ – tissue time activity concentration curve, see TTAC.

$cTT$ – completion total thyroidectomy, see HT and TT.

Cu – copper (element). $^{64}$Cu is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 12.69$ h).

**D**

$\Delta MR_{glc}$ – relative change in $MR_{glc}$ [%]

$\Delta SUV$ – relative change in $SUV$ [%]

DCE-CT – dynamic contrast enhanced CT

DCE-MRI – dynamic Gd-DTPA contrast enhanced MRI

$df$ – degrees of freedom

$DFS$ – disease-free survival

DG – 2-deoxy-d-glucose (molecule)

DG-6P – 2-deoxy-d-glucose-6-phosphate (molecule)

Dir – Dirichlet family of continuous multivariate probability distributions.

Dist. – stochastic distribution of a parameter.

DM – diabetes mellitus

DNA – deoxyribonucleic acid

dOR – diagnostic odds ratio

DOT – the system of imbursement of the Dutch healthcare authority (“diagnosis-treatment combination on their way to transparency”).

DTC – differentiated thyroid carcinoma

DTPA – diethylenetriaminepentacetate, chelator.

DWI – diffusion weighted imaging, NMR technique.

**E**

$E$ – effect (utility) [QALY]

EANM – European association of nuclear medicine

EC – electron capture

EES – extravascular extracellular space

EGFR – epidermal growth factor receptor

EMEA – European medicines agency

EO – expert opinion

EORTC – European organisation for research and treatment of cancer

EPR – enhanced permeability and retention

ESMO – European society for medical oncology

**F**

$F$ – blood flow

F – fluorine (element). $^{18}$F is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 109.771(20)$ min).

FA – folinic acid

FBP – filtered backprojection, see MLEM and OSEM.

$FD$ – fractal dimensions

FDA – food and drug administration

$^{18}$F-FDG – 2-($^{18}$F)fluoro-2-deoxy-d-glucose (molecule)

$^{18}$F-FDG-6P – 2-($^{18}$F)fluoro-2-deoxy-d-glucose-6-phosphate (molecule)

FE – fixed-effects

FLAB – fuzzy locally adaptive Bayesian

FN – false-negative result of a diagnostic test

FNAC – fine-needle aspiration cytology

FOLFIRI – chemotherapy regimen for treatment of CRC, made up of the drugs folinic acid, 5-fluorouracil and irinotecan.
**F (continued)**

FOLFOX – chemotherapy regimen for treatment of CRC, made up of the drugs folinic acid, 5-fluorouracil and oxaliplatin.

Campo de visão (FoV) – field-of-view

Falso positivo (FP) – false-positive result of a diagnostic

T4 – serum free, unbound 3,3′,5,5′-tetraiodo-L-thyronine, free, unbound thyroxine [pmol·l⁻¹]

FTC – follicular thyroid carcinoma, second most common type of DTC.

5-FU – 5-fluorouracil chemotherapy

FWHM – full-width at half-maximum

**G**

γ – family of continuous probability distributions.

Ga – gallium (element). ⁶⁷Ga is a γ-emitting isotope used for SPECT-imaging ($t_{1/2} = 3.26$ d), ⁶⁸Ga is a β⁺-emitting isotope used for PET-imaging ($t_{1/2} = 67.719$ min).

Gd – gadolinium (element), stable isotopes used for MRI.

Gd-DTPA – gadolinium-diethylenetriaminepentacetate (gadopentetic acid, gadopentetate dimeglumine).

Ge – germanium (element). ⁶⁸Ge decays by electron capture to the β⁺-emitting isotope ⁶⁸Ga and is used in a ⁶⁸Ge/⁶⁸Ga-generator ($t_{1/2} = 270.95$ d).

GEC – gene-expression classifier

glc – d-glucose (molecule)

GLUT – facilitative membrane-bound sodium-independent glucose transporters (13 known isoforms).

Gy – Gray, unit of dose of ionising radiation absorbed by an organ or tissue.

**H**

H – Kruskal-Wallis non-parametric statistic.

H₂O – water (molecule)

HER2 – human epidermal growth factor receptor, also known as erbB-2 or proto-oncogene Neu.

HIPEC – hyperthermic intraperitoneal chemotherapy

H&N – head and neck

Ho – holmium (element). ¹⁶⁶Ho is a β⁺-emitting isotope used for treatment ($t_{1/2} = 26.83(2)$ h).

HR – hazard ratio

HRQoL – health-related quality of life, see QoL.

HSD – honestly significant difference

HT – hemithyroidectomy, thyroid lobectomy, see TT and cTT.

**I**

I – iodine (element). ¹²³I is a γ-emitting isotope used for SPECT-imaging ($t_{1/2} = 13.2$ h), ¹²⁴I is a β⁺ (and other)-emitting isotope used for PET-imaging ($t_{1/2} = 4.18$ d), ¹²⁷I is a stable isotope and ¹³¹I is a β⁻/γ-emitting isotope used for treatment and SPECT-imaging ($t_{1/2} = 8.02$ d).

$I^2$ – inconsistency index

$iAUC_x$ – integrated area under the Gd-DTPA concentration-time curve for the first $x$ seconds

$ICC$ – intraclass correlation coefficient

IDIF – image-derived input function, whole-blood time activity concentration curve

$iDV$ – initial distribution volume [l]

IF – input function, see APTAC and $C_{plasma}(t)$

IFα – interferon-α

IMRT – intensity-modulated radiation therapy

In – indium (element). ¹¹¹In is a γ-emitting isotope used for SPECT-imaging ($t_{1/2} = 67.32$ h).

$iNMB$ – incremental net monetary benefit [€]

$IQR$ – interquartile range

IRW – inveon research workplace

---

233
Glossary

J

κ – Cohen κ-coefficient for agreement.

K

κ₁ – GLUT influx rate constant [ml·g⁻¹·min⁻¹]
κ₂ – GLUT efflux rate constant [min⁻¹]
κ₃ – hexokinase phosphorylation rate constant [min⁻¹]
κ₄ – glucose-6-phosphatase dephosphorylation rate constant [min⁻¹]
κ₅, κ₆ – phosphoglucose isomerase rate constants [min⁻¹]
κₑₚ – rate constant of Gd-DTPA between the EES and plasma compartment measured by DCE-MRI [s⁻¹]
Kᵢ – ¹⁸F-FDG net influx constant [ml·g⁻¹·min⁻¹]
Kₘₐ – Michaelis constant [mol·l⁻¹]
Kr – krypton (element). ⁸¹mKr is a γ-emitting isotope used for SPECT-imaging (t₁/₂ = 13.1 s).
Kᵣₑₚ – volume transfer constant of Gd-DTPA between the EES and plasma compartment measured by DCE-MRI [s⁻¹]

L

λ – decay rate [min⁻¹]
 – ratio of partition coefficients
 – willingness-to-pay threshold [€·QALY⁻¹]
λ₉FDG, λ₉glc – partition coefficient (C_free · C_plasma) of ¹⁸F-FDG or glucose.
LASER – light amplification by stimulated emission of radiation
LBM – lean body mass [kg]
LC, LC₉FDG – lumped constant [no unit]
LITT – LASER induced thermotherapy
LOR – line of response
LR⁻ – negative likelihood ratio of a diagnostic test
LR⁺ – positive likelihood ratio of a diagnostic test
Lu – lutetium (element). ¹⁷⁷Lu is a β⁻-emitting isotope used for treatment (t₁/₂ = 6.6475(20) d).
L(Y)SO – lutetium (yttrium) oxyorthosilicate

M

μ – mean
MAC – measured attenuation correction, see SAC.
MAD – maximum axial diameter on morphological imaging [mm]
MeSH – major subheading
MIP – maximum intensity projection
MLEM – maximum likelihood expectation maximisation, see FBP and OSEM.
MMP – mutation marker panel
MR₉glc – metabolic rate of d-glucose [nmol·min⁻¹·cm⁻³]
MRI – magnetic resonance imaging, NMR technique.
MRS – magnetic resonance spectroscopy, NMR technique.
MT – molecular test
MTV – metabolic tumour volume [cm³]

N

N – nitrogen (element). ¹³N is a β⁺-emitting isotope used for PET-imaging (t₁/₂ = 9.97 min).
N/A – not applicable
NaN – not a number
NCI – national cancer institute
NEC – noise equivalent count rate [cts·s⁻¹]
NEDPAS – Dutch society of nuclear medicine PET standard
NEMA – national electrical manufacturers association
### N (continued)

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH</td>
<td>National Institute of Health</td>
</tr>
<tr>
<td>NIS</td>
<td>Sodium-Iodine Symporter</td>
</tr>
<tr>
<td>NLLS</td>
<td>Nonlinear Least Squares</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NOS</td>
<td>Not Otherwise Specified</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative Predictive Value of a Diagnostic Test</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Non-Small Cell Lung Carcinoma</td>
</tr>
<tr>
<td>NVNG</td>
<td>Dutch Society of Nuclear Medicine</td>
</tr>
<tr>
<td>NZa</td>
<td>Dutch Healthcare Authority</td>
</tr>
</tbody>
</table>

### O

- **O**: oxygen (element). $^{15}\text{O}$ is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 122$ s).
- **O$_2$**: oxygen (molecule)
- **OR**: odds ratio
- **OS**: overall survival
- **OSEM**: ordered subsets expectation maximisation, see FBP and MLEM.

### P

- **$\varphi$**: fraction of glucose-6-phosphate that continues down the Embden-Meyerhof pathway (i.e., regular glycolysis), normally quite close to 1.0.
- **$p$**: level of statistical significance
- **$P$**: permeability of capillaries [cm·s$^{-1}$]
- **P**: chemical product
  - phosphorus (element). $^{32}\text{P}$ is a $\beta^-$-emitting isotope used for treatment ($t_{1/2} = 14.29$ d).
- **PA**: histopathology
- **PAX8/PPAR$\gamma$**: fusion oncogene between transcription factor paired box gene 8 and peroxisome proliferator-activated receptors subtype $\gamma$ associated with FTC.
- **PD**: progressive disease
- **PERCIST**: PET response criteria in solid tumours
- **PET**: positron emission tomography
- **PFS**: progression-free survival
- **PICTS**: target population, index test, comparator test, target condition and study design.
- **PO$_4^{3-}$**: phosphate
- **PPV**: positive predictive value of a diagnostic test
- **PR**: partial remission
- **$P_{rev}$**: prevalence of a disease
- **PRISMA**: preferred reporting items for systematic reviews and meta-analyses.
- **PSMA**: prostate-specific membrane antigen
- **PTC**: papillary thyroid carcinoma, most common type of DTC.
- **PVE**: partial volume effect

### Q

- **QALY**: quality-adjusted life year
- **QoL**: quality of life, see HRQoL
- **QUADAS**: quality assessment tool for diagnostic accuracy studies

### R

- **R**: Pearson product-moment correlation coefficient
- **$\rho$**: Spearman’s $\rho$ non-parametric rank correlation coefficient
- **$\rho$**: volumetric mass density [g·cm$^{-3}$]
- **Ra**: radium (element). $^{223}\text{Ra}$ is an $\alpha$-emitting isotope used for treatment ($t_{1/2} = 11.4$ d).
- **RAS**: “rat sarcoma” tumour suppressor gene, mutations in this gene are associated with DTC.
- **Rb**: rubidium (element). $^{82}\text{Rb}$ is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 76.4$ h).
<table>
<thead>
<tr>
<th><strong>Glossary</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R</strong> (continued)</td>
</tr>
<tr>
<td><strong>RCT</strong></td>
</tr>
<tr>
<td><strong>Re</strong></td>
</tr>
<tr>
<td><strong>RE</strong></td>
</tr>
<tr>
<td><strong>RECIST</strong></td>
</tr>
<tr>
<td><strong>RET/PTC</strong></td>
</tr>
<tr>
<td><strong>RFA</strong></td>
</tr>
<tr>
<td><strong>RH</strong></td>
</tr>
<tr>
<td><strong>ROC</strong></td>
</tr>
<tr>
<td><strong>ROI</strong></td>
</tr>
<tr>
<td><strong>RRA</strong></td>
</tr>
<tr>
<td><strong>RT</strong></td>
</tr>
<tr>
<td><strong>RTL</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>S</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$, SD</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>$S$</td>
</tr>
<tr>
<td>SA</td>
</tr>
<tr>
<td>SAC</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>$SE$</td>
</tr>
<tr>
<td>$SE$</td>
</tr>
<tr>
<td>SKM</td>
</tr>
<tr>
<td>Sm</td>
</tr>
<tr>
<td>SNMMI</td>
</tr>
<tr>
<td>$SNR$</td>
</tr>
<tr>
<td>$Sp$</td>
</tr>
<tr>
<td>SPect</td>
</tr>
<tr>
<td>SPSS</td>
</tr>
<tr>
<td>sROC</td>
</tr>
<tr>
<td>STN</td>
</tr>
<tr>
<td>$SUV$</td>
</tr>
<tr>
<td>$Sv$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>T</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{1/2}$</td>
</tr>
<tr>
<td>$T_3$</td>
</tr>
<tr>
<td>$T_4$</td>
</tr>
<tr>
<td>TAC</td>
</tr>
<tr>
<td>$TBF$</td>
</tr>
<tr>
<td>Tc</td>
</tr>
<tr>
<td>TCA</td>
</tr>
<tr>
<td>TE</td>
</tr>
<tr>
<td>$T_g$</td>
</tr>
<tr>
<td>$T_l$</td>
</tr>
<tr>
<td>$T/L$</td>
</tr>
<tr>
<td>TLE</td>
</tr>
<tr>
<td>$T_LG$</td>
</tr>
<tr>
<td>TN</td>
</tr>
<tr>
<td>TN</td>
</tr>
<tr>
<td>TNM</td>
</tr>
<tr>
<td>$T/NT$</td>
</tr>
<tr>
<td>ToF</td>
</tr>
<tr>
<td>TP</td>
</tr>
<tr>
<td>TR</td>
</tr>
</tbody>
</table>
T (continued)

TREUS – transrectal endo-ultrasonography
TRG – tumour regression grade
TSH – serum thyroid-stimulating hormone, thyrotropin [mU·l⁻¹]
TT – total thyroidectomy, see HT and cTT.
TTAC – tissue time-activity concentration curve, see $C_{tissue}(t)$.

U

UPM – unifocal classic papillary thyroid microcarcinoma (pT₁a cN₀M₀ R₀).
US – ultrasonography
US$ – USA dollar
USA – united states of America

V

$V_B$ – blood volume fraction [cm³·cm⁻³]
$V_D$ – volume of distribution of a tracer [l]
$\nu_e$ – the volume of EES per unit volume of tissue, DCE-MRI parameter [cm³·cm⁻³]
VEGF – vascular endothelial growth factor
$V_{max}$ – Michaelis-Menten maximum reaction rate [s⁻¹]
VOI – (3D) volume-of-interest, see ROI.
VRS – visual response score

W

WHO – world health organisation

X

Xe – xenon (element). $^{133}$Xe is a $\beta^-$- and $\gamma$-emitting isotope used for SPECT-imaging ($t_{1/2} = 5.2475(5)$ d).

Y

Y – yttrium (element). $^{90}$Y is a $\beta^-$-emitting isotope used for treatment ($t_{1/2} = 64.1$ h).

Z

Zr – zirconium (element). $^{89}$Zr is a $\beta^+$-emitting isotope used for PET-imaging ($t_{1/2} = 78.4$ h).

***
References
REFERENCES


REFERENCES


REFERENCES


“German Consumer Price Index,” 2013.


S. De Bruyne, N. Van Damme, P. Smeets, L. Ferdinande, W. Ceelen, J. Mertens, C. Van de Wiele, R. Troisi, L. Lib-


∗∗∗
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>NEMA properties of currently commercially available modern PET/CT scanners.</td>
<td>9</td>
</tr>
<tr>
<td>1.2</td>
<td>Various methods for normalisation of $^{18}$F-FDG uptake for different parameters representing distribution volume.</td>
<td>17</td>
</tr>
<tr>
<td>2.1</td>
<td>Search strategy.</td>
<td>32</td>
</tr>
<tr>
<td>2.2</td>
<td>Management changes and test characteristics in recurrent CRC by $^{18}$F-FDG PET.</td>
<td>41</td>
</tr>
<tr>
<td>2.3</td>
<td>Management changes and test characteristics in recurrent CRC by $^{18}$F-FDG PET (continued).</td>
<td>42</td>
</tr>
<tr>
<td>2.4</td>
<td>Management changes and test characteristics in recurrent CRC by $^{18}$F-FDG PET (continued).</td>
<td>42</td>
</tr>
<tr>
<td>2.5</td>
<td>Prognostic stratification by baseline $^{18}$F-FDG PET.</td>
<td>44</td>
</tr>
<tr>
<td>2.6</td>
<td>Therapy response prediction by baseline $^{18}$F-FDG PET.</td>
<td>45</td>
</tr>
<tr>
<td>2.7</td>
<td>Radiotherapy and multimodality (neoadjuvant) therapy response evaluation in locally advanced rectal carcinoma by $^{18}$F-FDG PET.</td>
<td>48</td>
</tr>
<tr>
<td>2.8</td>
<td>Response evaluation for local ablative treatment for unresectable CRC liver metastases by $^{18}$F-FDG PET.</td>
<td>50</td>
</tr>
<tr>
<td>2.9</td>
<td>Chemotherapy response evaluation by $^{18}$F-FDG PET for CRC metastases.</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>Characteristics of patients with CRC or NSCLC.</td>
<td>63</td>
</tr>
<tr>
<td>3.2</td>
<td>Outcome of follow-up of patients using Kaplan-Meier analysis.</td>
<td>63</td>
</tr>
<tr>
<td>3.3</td>
<td>Results of univariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta SUV$s between the scan at baseline and evaluation.</td>
<td>65</td>
</tr>
<tr>
<td>3.4</td>
<td>Results of univariate Kaplan-Meier analysis for overall and progression-free survival using $\Delta SUV$s between the scan at baseline and evaluation dichotomised at their median for response or nonresponse.</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Characteristics of patients with CRC or NSCLC.</td>
<td>75</td>
</tr>
</tbody>
</table>
List of Tables

4.2 Outcome of follow-up of CRC and NSCLC patients using Kaplan-Meier analysis. ........................................... 76
4.3 Number of colorectal carcinoma patients with decreased, stable or increased metabolism based on both VOI definitions. ............. 80
4.4 Number of non-small cell lung carcinoma patients with decreased, stable or increased metabolism based on both VOI definitions. ............. 80
4.5 Results of univariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta MR_{\text{glc}}$ between the $^{18}$F-FDG PET at baseline and follow-up. ................................. 81
4.6 Results of multivariate Cox proportional hazards regression analysis for overall and progression-free survival using $\Delta MR_{\text{glc}}$ between the $^{18}$F-FDG PET at baseline and follow-up for NSCLC. ................................. 82

5.1 Comparison between parameter-identification and parameter-validation groups, with statistical significance of group difference. ................. 90
5.2 Results of fitting of normalised plasma data of parameter-identification group and subgroup analysis. ........................................... 93
5.3 Literature review of publications on population-based APTACs and IDIFs. ......................................................... 97

6.1 Patient characteristics. ................................................................. 106
6.2 VOI characteristics. ................................................................. 109
6.3 Correlation of VOI$_{\text{tumour}}$ parameters. ........................................ 109
6.4 Comparisons between tumour segments. ........................................ 110

7.1 Patient characteristics. ................................................................. 119
7.2 Predictive value of pretreatment parameters to survival assessed by Cox regression analysis. ........................................... 124
7.3 Baseline and follow-up values of vascular and metabolic parameters. ................................................................. 125
7.4 Predictive value of therapy-induced parameter changes for early response evaluation, assessed by Cox regression analysis. ........................................... 126

8.1 Responses in imaging parameters of the two patients described between baseline, after three treatment cycles and a late acquisition. ......................... 135

9.1 Characteristics of the patient populations included in the studies selected for the current meta-analysis. ........................................... 150
9.2 Clinical characteristics of the patients included in the meta-analysis. ................................................................. 151
9.3 Test characteristics of $^{18}$F-FDG PET for the detection of malignancy in thyroid nodules with indeterminate fine-needle aspiration cytology results according to the published literature. ................................................................. 153

10.1 Accountability of base-case parameter values and stochastic distributions for base parameters and transition probabilities. ................................. 167
10.2 Accountability of base case parameter values for costs and utilities. ................................................................. 168
10.3 Base case main results over a duration of 5 years. ................................................................. 171

1 The Bethesda system for reporting thyroid cytopathology. ................................................................. 196
2 Potential roles of molecular imaging in drug research and development. ................................................................. 201
## List of Figures

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>1.1</td>
</tr>
<tr>
<td>1.2</td>
</tr>
<tr>
<td>1.3</td>
</tr>
<tr>
<td>2.1</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
</tr>
<tr>
<td>Figure</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td>4.1</td>
</tr>
<tr>
<td>4.2</td>
</tr>
<tr>
<td>4.3</td>
</tr>
<tr>
<td>5.1</td>
</tr>
<tr>
<td>5.2</td>
</tr>
<tr>
<td>5.3</td>
</tr>
<tr>
<td>6.1</td>
</tr>
<tr>
<td>6.2</td>
</tr>
<tr>
<td>7.1</td>
</tr>
<tr>
<td>8.1</td>
</tr>
<tr>
<td>8.2</td>
</tr>
<tr>
<td>9.1</td>
</tr>
<tr>
<td>9.2</td>
</tr>
<tr>
<td>9.3</td>
</tr>
<tr>
<td>9.4</td>
</tr>
<tr>
<td>9.5</td>
</tr>
<tr>
<td>10.1</td>
</tr>
<tr>
<td>10.2</td>
</tr>
<tr>
<td>10.3</td>
</tr>
<tr>
<td>10.4</td>
</tr>
<tr>
<td>10.5</td>
</tr>
<tr>
<td>10.6</td>
</tr>
</tbody>
</table>
Curriculum Vitae

D. Vriens
Curriculum Vitae

Dennis Vriens was born on the 15th of April 1981 in Tilburg, the Netherlands. He graduated cum laude in 1999 at Gymnasion “Theresia Lyceum” and was awarded a prize for “best student in exact sciences, Southern Netherlands”. That year, he moved to Nijmegen and started medical school at the Radboud University.

In 2005 he finished the clinical part of his medical training by an internship at the Sengerema Designated District Hospital, Tanzania. In 2006, he graduated medical school with an undergraduate research project entitled “Mortality of severely injured adult patients: evaluation of adherence to the regional triage-protocol” (supervisor: prof. dr. A.B. van Vugt).

After 1.5 year working experience at the emergency medical dept. at the Rijnstate Hospital in Arnhem, he switched to the dept. of nuclear medicine of the Radboud University Medical Centre, Nijmegen, in 2008 for a PhD-project leading to this thesis. In 2010 he was awarded a stipends for Qualified Doctor Training to become a Clinical Researcher (AGIKO) by the Netherlands Organisation for Health Research and Development (ZonMw), allowing him to combine this with a residency in nuclear medicine.

In 2011, he was nominated by the Royal Dutch Academy of Arts and Sciences (KNAW) to attend the 61st Nobel Laureate Meetings on physiology & medicine at Lindau, Germany. End 2013, he left for a 3-month clinical internship on prostate carcinoma imaging and PET/MRI at the dept. of nuclear medicine of the Klinikum Rechts der Isar, Technische Universität München under supervision of prof. dr. M. Schwaiger and prof. dr. K. Scheidhauer.

Begin 2014, a project based on the research presented in the last part of this thesis was awarded a Dutch Cancer Society (KWF) Project Grant, entitled “Efficacy of 18F-FDG PET in the evaluation of cytologically indeterminate thyroid nodules prior to surgery (EffECTS): a nationwide multicentre cost-effectiveness trial” (ClinicalTrials.gov Identifier: NCT02208544).

Dennis is secretary of the Junior Chamber of the Dutch Society of Nuclear Medicine (NVNG). He is member of the committee developing a new resident training program, integrating radiology and nuclear medicine (CORONA). March 2015 he started working as a nuclear medicine physician in the Leiden University Medical Centre.

In 2009 he met his partner Saskia Koene. October 2014, their son Thijn was born. Together they live in Nijmegen and enjoy life most while travelling around the world.

∗∗∗
List of Publications

Chronologically
List of publications

Peer-reviewed publications

Arens AIJ, **Vriens D**, Janssen M, Simon A, Oyen WJG. Anakinra injection site reaction on $^{18}$F-FDG PET/CT. *Clin Nucl Med*. **2015 Jan**; accepted for publication.


* Included in this thesis.
† Both authors contributed equally to this article.


*Included in this thesis.


**Non-peer reviewed publications**


* ***

* Included in this thesis.
† Both authors contributed equally to this article.
Acknowledgements

Dankwoord
Acknowledgements

With this acknowledgement section, I do not only conclude this thesis but also a period of over seven years in which I was given the opportunity to develop myself as a researcher, a medical specialist and as a person at the department of Nuclear Medicine of the Radboud University Medical Centre in Nijmegen, the Netherlands. After a tough career change, this new environment with new people provided me with ample opportunities I needed. These words of thanks are an attempt to express my deepest gratitude to all of you I’ve worked with. There probably are important people that I have forgotten: I apologise, exhaustion plays havoc on memory.

Supervisors & co-supervisor

Professor Oyen, dear Wim, I am extremely grateful to you. When I first sent an email to a, for me unknown, department, I had no idea what you would mean to me for my future in Nuclear Medicine. You’ve spent much effort providing me with opportunities in scientific research and coached me profoundly in my personal development. You are my true teacher and inspiring mentor, adjusting to my needs but always encouraging me to achieve more. With your help, I’ve made the choice to start my residency in Nuclear Medicine. You’ve always informed me of interesting courses, grants or other opportunities. When I started to make my own choices in research, you’ve always supported me. I still remember the night I received confirmation of acceptance of our first coauthored article. You were the first to reply my email, at 2 a.m., to congratulate me and to ask me if I suffered from insomnia. I think I’ve never told you I was at the Nijmegen Summer festival. I sincerely hope that we will find more opportunities to work together in the future.

Professor De Geus-Oei, dear Lioe-Fee, you are my thesis supervisor, my mentor, my trainer and, since recently, my fellow staff member at the LUMC. I hope you’ll find this thesis a worthy successor to yours, insofar size matters, it must be! When I first met you and dr. Visser at the barracks of our former department, there was chemistry between us from the first moments on. You’ve allowed me to develop myself as an independent and critical scientist. Most patients in the studies presented in this thesis have been included by you. You’ve helped me a lot during the many (mostly vain) grant proposal writings. You were personally involved during some difficult periods I had to endure, but also during many happy moments: you learned me to bake carrot-cake, you sent me home when Saskia was in labour and you were the first colleague to pay us a maternity visit, twice!

Doctor Visser, dear Eric, thank you for the interesting discussions on physics and all original ideas on new analyses. You’ve improved all manuscripts and presentations by rephrasing technical terminology to be used most properly. Every time I thought I oversaw all possible alternative analysis options, you were able to think of more. I hope you keep convincing researchers to use 4D dynamic $^{18}$F-FDG PET and that we keep on working together in the future.
ACKNOWLEDGEMENTS

Doctoral thesis committee

I want to thank the members of the doctoral thesis committee, professor Winald Gerritsen, professor Ronald Boellaard and dr. Jan Bussink for the time and effort you’ve spent in judging this thesis and opposing to it during my public defence.

Involved Colleagues

I want to express my gratitude to all technicians and laboratory employees involved in the PET-subdivision of our department, especially Peter Kok, Michel de Groot and Maichel van Riel, that showed unlimited stamina helping me out with the sophisticated dynamic $^{18}$F-FDG PET protocol. Also, I could not have done without the help of Marie Claire Attard, Diana Valks, Jurrian Butter and Merijn Janssen.

Dr. Jonathan Disselhorst, especially during the earlier work presented in this thesis you were my technical twin. I have enjoyed the hours spent together hypothesising on strategies of optimal time framing, code-named “pim-pam-pet”. We have spent hours and hours together programming in MATLAB, using impossible pseudonyms for variables and, of course, the function ‘sendolmail.m’. It was a pleasure to work on our PhD-projects together.

Many studies would have been impossible without the help of involved members of the depts. of Pulmonology and Cardiothoracic Surgery: dr. Erik van der Heijden, dr. Olga Schuurbers-Siebers, dr. Chantal Smits-van der Graaf, dr. Ad Verhagen and last, but certainly not least, Nicolle Peters, who helped me with inclusion of the patients. Also my gratitude to dr. Monika Looijen-Salamon from the department of Pathology for her work on the resection specimens.

Professor Arend Heerschap, dr. Sjaak van Asten from the department of Radiology and prof. Hanneke van Laarhoven from the department of Medical Oncology, thank you for introducing me in the world of advanced MRI. Hanneke, a special ‘thank you’ for inclusion of the CRC patients in our studies.

I want to express my gratitude to the involved members of the department of Endocrinology, especially prof. Jan Smit and dr. Romana Netea-Maier, and surgery, most of all prof. Hans de Wilt, for their clinical vision on our thyroid research.

Professor Gert-Jan van der Wilt, dr. Paul Krabbe and dr. Eddy Adang from the Health Evidence department, thank you for the hours of interesting discussions on biostatistics and health-economical evaluations. It is nice to share a passion on number crunching.

Furthermore I want to thank all other co-authors of my work for their feedback on the numerous versions of the manuscripts I’ve sent to them, especially prof. Winette van der Graaf, prof. Kees Punt and dr. Anja Timmer-Bonte for their critical clinical review of the manuscripts.
Other colleagues & co-workers

I want to thank all employees of the department of Nuclear Medicine (researchers, technicians, secretaries) of the Radboud University Medical Centre. You all created a friendly, joyful environment in which everything was possible.

Professor Boerman, dear Otto, thank you for the highly fertile soil you’ve created for junior researchers to flourish. Special thanks to my colleague researchers, dr. Jonathan Disselhorst, dr. Sandra Heskamp and dr. Maarten Brom; we’ve all started our PhD-projects around the same time, but I am the last to write this chapter. I have to admit, it is true, croquettes taste much better with ketchup than with mustard. A word of gratitude to fellow “believers” in dynamic $^{18}$F-FDG PET: apart from Jonathan Disselhorst, also dr. Linda Heijmen and Anouk van Berkel.

Also thanks to the many other colleagues at our research department, especially ir. Annemarie Eek, ing. Gerben Franssen, Willem Grootjans, Stijn Muselaers, but also all others that I’ve forgotten to mention here.

Thanks to the other medical staff at our department, who were always available in (ethical) needs: prof. Martin Gotthardt, dr. Marcel Janssen and Anne Arens.

Also to the numerous other fellow-residents that I have worked with during my long time at the department, you all had to endure a lot, both due to my personality and my moments of absence.

Also my gratitude to my supervisors during my internal medicine internship: especially prof. Winette van der Graaf, dr. Carla van Herpen, dr. Jolien Tol (Medical Oncology), dr. Chantal Bleeker-Rovers (Infectious Diseases) and prof. Joep Smeets, dr. Arie van Dijk and Sjoerd Westra (Cardiology). Without you these internships would have been not as enjoyable.

During my residency and PhD-project, I was given the special opportunity of personal development in the Da Vinci Challenge. Ellen van der Linden and Anja Schumann have given me the professional guidance and provided me with exceptional experiences needed for personal development. Michiel van den Brand, it was a very learning experience to give a lecture on personal development together with you. Tong Xi, my roommate, in the end we found a way to work together as oral and maxillofacial surgeon and nuclear medicine physicians on our Na$^{18}$F PET/CT project in adolescents. Arjan Bergsma, Rocío Acuña Hidalgo, dr. Peter Makai and Tom Gevers: we’ve shared some very special moments, I will never forget. Finally, Sami Simons and dr. Jeroen de Baaij, thank you for all the long evenings with profound discussions fuelled by lots of wine. These Bacchanalia are indeed the perfect continuation of the Da Vinci feeling, I hope to be able to continue for many more years. By the way, a maternity visit until 00:30 a.m. is quite unusual.

Patients

No clinical research is possible without the altruistic motives of the patients who we have bothered with intravascular catheters and with long scanning times in and sometimes awkward positions in periods of their lives where most likely would rather be at home than in a hospital. Thank all of you.
International colleagues

A short word of thanks to colleagues in the department of Nuclear Medicine of the Klinikum Rechts der Isar of the Technische Universität München, which provided a very interesting internship. Especially prof. Klemens Scheidhauer, prof. Ambros Beer and dr. Matthias Eiber who have made my stay very interesting.

Friends & family

Dear friends and family, I believe this is the only section of this thesis you will be reading. To all my friends and fraternity brothers (too many to list here but you know who you are), thank you for providing the support, friendship and distraction, I needed.

Dear paranymphs, Dirk Ottink and John Storm. Dirk – wie die Nase des Mannes – Ottink, we shared some ups and downs together both during our time at the “Fox street” and thereafter. The most important highlights being the birth of your daughter Emme and our son Thijn, but also our last trips together in Bavaria and Bolivia. I hope we will share endless more memories: “mateloos, redeloos, laveloos, zedeloos” (excessive, irrational, soaked, immoral). Dear John, I met you as Saskia’s best friend Laura’s wife. However, you became a friend and a person I look up to. You dare to follow your dream, you won’t let others drive you crazy, without being ignorant or indifferent. Hopefully your twins Joris and Sophia, will not make your water lukewarm either. Thank you both for being my paranymphomaniacs, my friends and thank you for all welcome distraction.

Dear Laura Storm, thank you for all the love and friendship you have given Saskia and me. We enjoy every moment we have shared with John and you. If it were not for you, Saskia would never have endured me.

Dear Boudevijn and Kitty Koene, I am not only very grateful to you for setting Saskia Koene to this world and accepting me as her partner, but also to the fact that you had the patience to listen to my endless monologues on “three to four letter abbreviations”. Therefore, I dedicate the glossary of this thesis specially to you both. You’ve both helped me through some tough times but also we shared many moments of joy, including our shared passion on travelling and high-speed super-bike driving. Thank you for your endless support. Erik and Nienke Koene, thankfully now I have a little brother and Thijn has an uncle and an aunt. Erik, next Christmas you will eat (yellow) snow, I promise!

Dearest Mom and Dad, I cannot thank you enough for all the support and love you have given me. Where would I’ve been without you (which in fact is more than an existentialistic rhetorical question)? Mom’s perseverance and Dad’s technical insight, together with your perfect upbringing I have had, have made me who I am now. We share passion on science and technology, travelling, SCUBA-diving and motorbiking (at least we did). I still enjoy your support for us and our family every week. Although roles have changed the last years, the amount of happiness I feel with you will never change. This thesis is product of your love and therefore dedicated to you both.

Saskia. You are the love of my life. A buddy, a friend, a partner and now also mother. I am so endlessly happy for having met you. Our passion to backpack to the end of the world and beyond has led to big adventures of extreme heights
(over 5,500 m above sea level in Bolivia’s South West Circuit), extreme depths (31 m below sea level in Cuba’s Bahía de Cochinos), extreme south (Antarctica’s Port Lockroy), extreme cold (ice camping under the open sky on Antarctica) and extreme heat (probably Cambodia’s Siem Reap). We share our passion for exotic foods (frog, guinea pig and some Indian multi-resistant enteroinvasive bacteria only found in raw goat’s meat) but also European specialities, all of which we eat either in the street of prepared on our own kitchen. You also share my passion of research as we are equally determined. Fortunately, you are much more patient than I am, as I would not get along with myself for longer than a day or two. Thank you for all the support, all the tolerance and your acceptance of my shortcomings. Especially the last few months have been the hardest as the apotheosis of this thesis, the implementation of a new multicentre study, a job change and the birth of our first child coalesced. We promise each other every New Year’s Eve that next year will be more peaceful than the last. Up till now we have not managed to keep these New Year’s Resolutions. Nevertheless, I’ve enjoyed, am enjoying and will enjoy every moment with you, even during the sparse arguments. I sincerely hope we can keep travelling together until our final journey. I love you.

Little Thijn, tabula rasa, we see you acquiring new abilities every day. Your mother and I will try to outperform your grandparents’ upbringing, which will be an almost impossible task. We love you endlessly and hope you will also find happiness like we did. “Ragoooh!!”.

***