



UvA-DARE (Digital Academic Repository)

Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer

Konert, T.

Publication date

2020

Document Version

Final published version

License

Other

[Link to publication](#)

Citation for published version (APA):

Konert, T. (2020). *Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

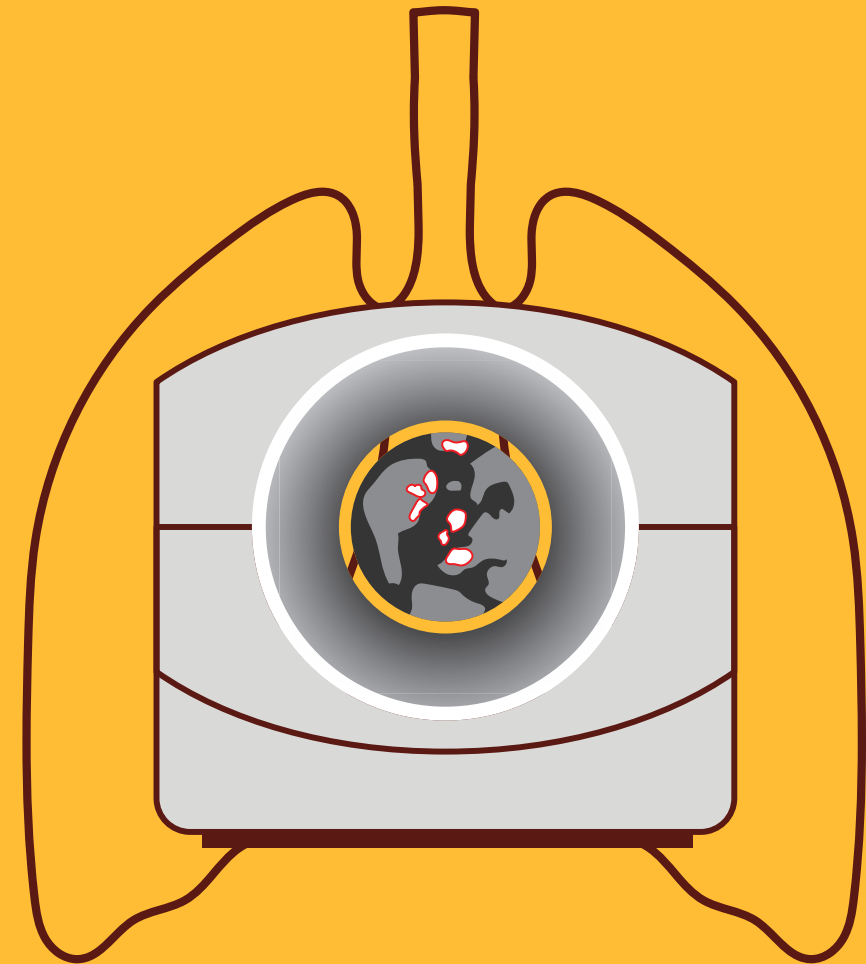
Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer

In low- and middle-income countries, management of patients with locally advanced non-small cell lung cancer (NSCLC) can be improved. Centers that are suitably equipped to provide nuclear medicine and radiation oncology services have minimal experience in the multidisciplinary use of a hybrid imaging technique called PET/CT for concurrent chemoradiotherapy, resulting in many patients receiving suboptimal therapy selection and treatment delivery. The work in this thesis focused on improved patient selection for concurrent chemoradiotherapy, and increased and standardized treatment accuracy using PET/CT applications, with the goal to improve survival in patients with NSCLC. Our work, which was in collaboration with the International Atomic Energy Agency, contributed to broader multidisciplinary use of PET/CT in 9 centers from Brazil, Estonia, India, Jordan, Pakistan, Turkey, and Vietnam. We developed pragmatic standardized guidelines to accurately delineate tumor, provided practical and online training sessions on PET/CT based concurrent chemoradiotherapy over a year, and provided evidence on the effectiveness of PET/CT based concurrent chemoradiotherapy through a multi-center trial. Currently, the choice of treatment is largely dependent on the disease stage and patient condition, and this is not always accurate enough for optimal treatment selection in locally advanced NSCLC. This thesis also contributed to the debate of improving therapy selection in locally advanced NSCLC patients by studying the prognostic accuracy of quantitative imaging features from PET, called PET radiomics features. Future studies should investigate if the prognostic accuracy of current prognostic models can be further improved by combining imaging features from multiple imaging modalities with clinical and genomic data.

Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer

Tom Konert

Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer



Tom Konert

**Towards worldwide use of FDG
PET/CT applications for optimal
treatment of lung cancer**

Tom Konert

Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 13 november 2020, te 14.00 uur

door

Tom Konert
geboren te Seria

Cover by: Guus Gijben | Proefschrift-aio.nl

Printed by: Guus Gijben | Proefschrift-aio.nl

ISBN: 978-94-93184-71-8

The work in this thesis was performed at the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands, and the Peter MacCallum Cancer Center, Melbourne, Australia.

The International Atomic Energy Agency, Netherlands Cancer Institute and Onderzoekschool Oncologie Amsterdam kindly provided financial support.

© 2020 Tom Konert, Amsterdam, the Netherlands

Promotiecommissie:

Promotor:	prof. dr. ir. J.-J. Sonke	AMC-UvA
Copromotores:	dr. W.V. Vogel	NKI-AvL
	dr. M.A.J. Stokkel	NKI-AvL
Overige leden:	prof. dr. R.J. Bennink	AMC-UvA
	prof. dr. C.R.N. Rasch	AMC-UvA
	prof. dr. M.B. van Herk	AMC-UvA
	prof. dr. M.M. van den Heuvel	Radboud Universiteit Nijmegen
	dr. H.M.U. Peulen	Catharina Ziekenhuis
	prof. dr. R. Boellaard	Vrije Universiteit Amsterdam

Faculteit der Geneeskunde

*Voor mijn lieve opa en oma,
van wie ik weet dat ze apetrots zouden zijn*

Table of Contents

Chapter 1

General introduction and outline of this thesis9

Chapter 2

The developing role of FDG PET imaging for prognostication and radiotherapy target volume delineation in non-small cell lung cancer 23

Chapter 3

PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014..... 45

Chapter 4

Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol. 67

Chapter 5

Introducing FDG PET/CT guided chemoradiotherapy for stage III NSCLC in low and middle-income countries: preliminary results from the IAEA PERTAIN trial 87

Chapter 6

Robust, independent, and relevant prognostic ¹⁸F-fluorodeoxyglucose positron emission tomography radiomics features in non-small cell lung cancer: are there any?... 115

Chapter 7

General discussion and future perspectives 167

Appendices

Summary180

Samenvatting..... 184

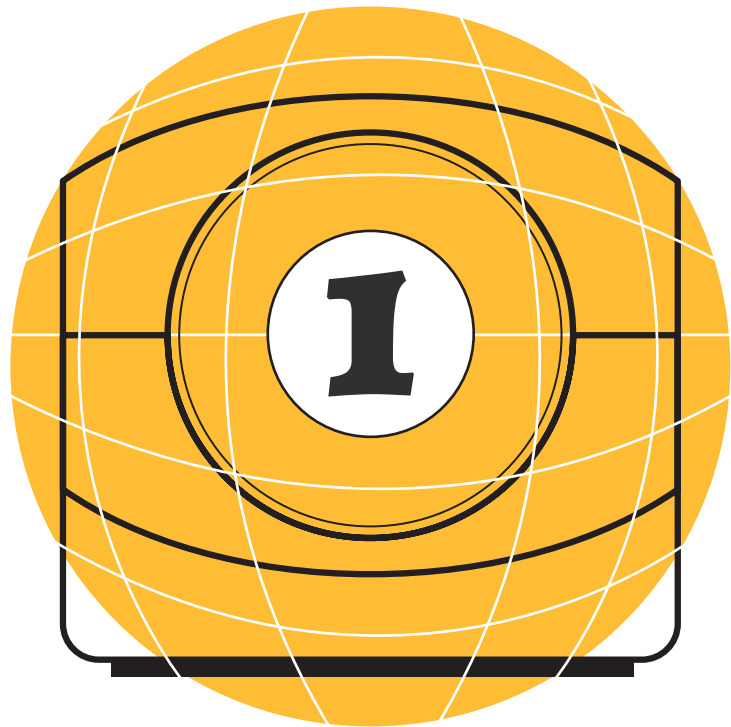
List of publications and author contributions188

Dankwoord.....190

Curriculum Vitae 194

Chapter 1

General introduction and outline of this thesis



Lung cancer: not only a first world problem

Lung cancer is the most common cause of death from cancer worldwide (Figure 1) [1]. The incidence of lung cancer worldwide is expected to further increase in the next years, due to the rise in tobacco use in low and middle-income countries (LMIC) [2]. The expected increase of lung cancer cases is a major concern in LMIC as there are already insufficient treatment facilities to provide the necessary healthcare services [3]. In addition, the lack of expertise and insufficient resources has an impact on many aspects of diagnosis and treatment of lung cancer [4, 5]. As a result, patients may receive ineffective treatments, leading to poor treatment outcome. Hence, management of patients with lung cancer can be improved in LMIC.

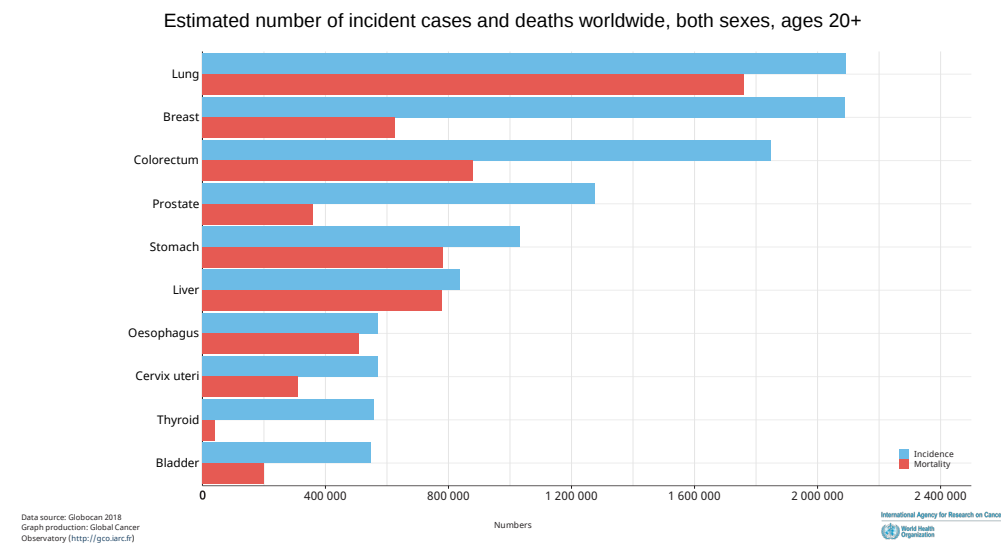


Figure 1. Overview of the incidence and mortality of lung cancer and other cancers worldwide [1].

Diagnosis, treatment and survival

Rapid advancements are made in the development of new medical technology and treatment options for lung cancer in high-income countries (HIC). However, the accessibility of these new techniques and treatments remain poor in LMIC and are only limitedly available in a small number of treatment facilities. Despite these developments, globally, overall survival of patients with lung cancer has improved only gradually in the last decades [6-8]. Worldwide, more than 75% of patients with lung cancer are diagnosed with (locally) advanced lung cancer, *i.e.*, cancer cells have spread to regional lymph nodes or to more distant lymph nodes and to other organs [9,10]. The late diagnosis contributes

to low survival rates in these patients, and, in the case of advanced lung cancer, there is almost no chance of cure [11]. In HIC, preventive measures such as lung cancer screening are proposed to improve early detection and management of lung cancer [12]. However, in LMICs with insufficient treatment facilities these screening programs are less meaningful. Treatment options for lung cancer consists of surgery, radiotherapy (RT), chemotherapy, targeted drug therapy, immunotherapy, or a combination of these, although not all treatment options are available in LMIC.

Locally advanced non-small cell lung cancer

The most common type of lung cancer is non-small cell lung cancer (NSCLC) and accounts for 85% of lung cancer cases [13]. In this thesis, we focus on the management of patients with locally advanced NSCLC. *Locally advanced* means that tumor cells have spread to regional lymph nodes, but not yet to distant lymph nodes or other organs proven on imaging or by clinical examination. The 5-year overall survival rates range between 13% (IIIC) and 36% (IIIA), but these numbers are expected to be lower in LMIC [14, 15]. The current standard treatment with curative intent in HIC for locally advanced NSCLC is concurrent chemoradiotherapy (CRT).

Non-invasive imaging in NSCLC

Non-invasive diagnostic techniques like computed tomography (CT) and positron emission tomography (PET) are used in almost any lung cancer treatment procedure for malignancy detection and accurate staging of NSCLC [16]. CT imaging provides anatomical information, used to localize and delineate target structures like the primary tumor, involved lymph nodes and organs at risk. However, the soft-tissue contrast is limited in CT, especially noticeable in the mediastinum. Hence, functional imaging such as PET is used in conjunction with CT to aid in detection of metastases and distinguishing between healthy tissue and lung cancer.

PET imaging

PET imaging provides information on biochemical and functional processes in which radiolabeled molecular tracers are involved, also known as PET tracers. PET imaging relies on these PET tracers that emit positrons, and visualizes their estimated location within the patient. The most commonly used PET tracer in RT planning is ^{18}F -fluorodeoxyglucose (FDG), an analogue for glucose. FDG is injected intravenously and distributed via the circulatory system throughout the body and accumulates in cells with high rates of metabolism, such as the brain or the heart. High rates of metabolism also occur in cancer cells as a result of proliferation or as an adaptation to hypoxia in cancer. As FDG enters the tumor cells, where it is phosphorylated and trapped intracellularly, positrons are emitted due to radioactive decay of the radionuclide ^{18}F . A positron travels a short distance before it

loses enough energy to annihilate on collision with a nearby electron. Annihilation produces two 511 keV gamma rays, moving in opposite direction, and these may be detected by the PET detector ring surrounding the patient. PET scanners are able to estimate the position of these annihilation events and generate a 3D volume showing the tracer distribution throughout the body.

FDG PET/CT applications in NSCLC

The treatment of locally advanced NSCLC is challenging due to the macroscopic extent of the disease, i.e., (bulky) tumors with extensive lymph node involvement. It is therefore of major importance to identify the exact location of the cancer cells to gain local tumor control and be able to reduce toxicity in healthy tissue. Another challenge in treating locally advanced NSCLC originates from the strong intra and interpatient tumor heterogeneity. Tumor heterogeneity includes differences in cellular morphology, gene expression, metabolism, proliferation, and metastatic potential. For tumors, this intra-patient heterogeneity increases the probability to adapt features such as treatment resistance and other survival mechanisms. The interpatient tumor heterogeneity requires the need for a more personalized treatment approach as patients can respond differently to the same treatment. The first challenge requires accurate and precise tumor and nodal delineation. The second challenge requires a better characterization of the tumor heterogeneity before treatment to avoid futile therapies. FDG PET/CT serves as a valuable tool for accurate and precise tumor delineation and may also play a role in improved prognostic tumor characterization.

Radiotherapy target volume definition

Target volume definition (TVD) starts with contouring of the primary tumor and involved lymph nodes visual on PET/CT imaging, and the resulting contour is named the gross target volume (GTV) [17]. To account for microscopic disease surrounding the tumor that cannot be seen on the scans, a margin should be added, called the clinical target volume (CTV) [18]. Lastly, geometric uncertainties with regards to position and shape of the tumor between treatment planning and treatment as well as during treatment calls for an extra margin: the planning target volume (PTV) [19].

Observer variation

GTV delineation in NSCLC is subject to observer variation [20], hence in conjunction with a poorly defined PTV there is a potential risk of geographic miss of tumor and/or unnecessary inclusion of healthy tissue [21]. The use of CT alone for TVD is associated with poor reproducibility. A significant reduction in interobserver variability (IOV) has been noted when PET was used in conjunction with CT [22-25]. Even with the use of both PET and CT, uncertainty remains in TVD. A disadvantage of FDG PET is that it is not tumor

specific, i.e., FDG uptake can also be seen in inflammatory regions. This can complicate GTV delineation as illustrated in Figure 2. Previously published guidelines on PET/CT based contouring may not be sufficiently detailed for optimal TVD of lung cancer.

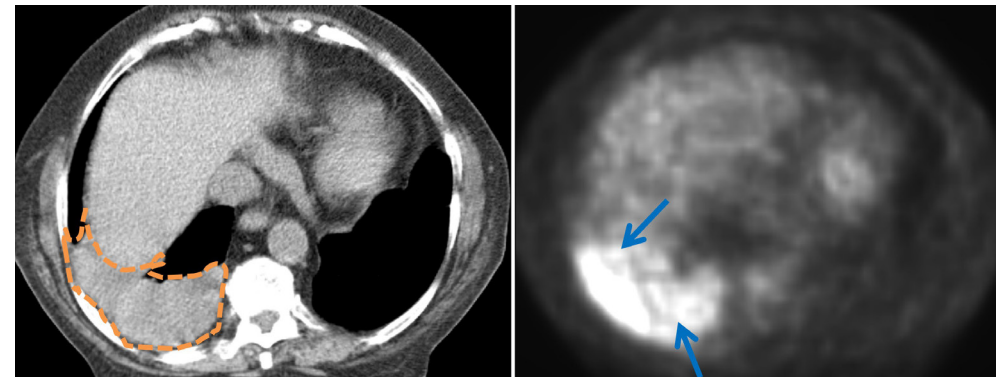


Figure 2. Left image: transversal slice of a CT scan of a lung cancer patient showing in the right lung collapsed lung tissue (orange dashed contour), called atelectasis, hampering identification of tumor boundaries. Right image: transversal slice of the corresponding PET scan showing high FDG uptake in a primary tumor in the right lower lobe; GTV delineation is hampered by the presence of inflammation within atelectasis (blue arrows) and tumor boundaries remain unclear.

PET/CT based prognostic tumor characterization

Information on tumor heterogeneity can be deployed to optimize the management of lung cancer. This information can be retrieved from biopsy based histopathologic examinations on the microscopic and molecular level. However, biopsy based assays comprise only a small area of an otherwise highly heterogeneous bulky tumor. This limits the use of histopathologic examinations, but in contrast gives a huge potential for PET/CT imaging. PET/CT imaging is able to rapidly assess the whole tumor volume. PET/CT could therefore complement histopathologic examination by imaging based characterization of tumor heterogeneity.

Semiquantitative PET

Tumor heterogeneity can be evaluated through visual assessment of FDG uptake patterns on PET images [26] or by quantifying the uptake patterns of radiotracers. The uptake of radiotracers in a voxel or a larger volume of interest (VOI) in a PET image is commonly assessed using a semiquantitative metric, called the standardized uptake value (SUV). The SUV takes into account the weight of the patient and the injected tracer activity. It is called semiquantitative as SUV measurements will vary depending on many biological, technical, and physical factors [27, 28]. As a result of these dependencies, a patient that

is scanned twice within a short timeframe will not have identical PET scans. This adds to the complexities of performing quantitative analysis on PET scans of the same patient and even more when performing quantitative analysis on PET scans acquired from patients in multiple centers. Namely, when PET/CT scans are not acquired and processed in a standardized manner, these variations in scanning protocol can lead to under and overestimation of SUV of $\geq 50\%$ in SUV measurements [29]. It is therefore in our interest to harmonize PET/CT scanning protocols in order to establish SUV interchangeability across centers [30, 31]. SUV interchangeability across centers will open up the possibility to perform quantification studies on tumor heterogeneity in a multicenter setting. Subsequently, this can lead to more insight in prognostic tumor characterization that can possibly aid in treatment decision making.

Prognostic factors to guide therapy in NSCLC

Currently, decisions on treatment strategy still heavily depend on TNM staging [32, 33] and Eastern Cooperative Oncology Group (ECOG) performance status [34, 35]. A number of other prognostic factors are known, however not routinely used to guide clinicians in making treatment decisions, such as weight loss [36], gender [37], histology [38], age [39], serum blood levels [40, 41], mutation status [42], and protein expression levels [43, 44]. In locally advanced NSCLC, treatment selection based on TNM staging and other clinical variables may not be accurate enough to guide therapy with area under the curve (AUC) values up to 0.72 [45, 46]. Therefore, the search for additional prognostic factors is warranted to maximize therapeutic efficacy and to decrease the number of unnecessary treatments to non-responder patients.

Radiomics

A current field of interest is the prognostic value of quantitative image features and its complementary value to well-established clinical prognostic models. The field of research that studies a large number of quantitative image features extracted from medical images using data-characterization algorithms, is termed radiomics. Radiomics focuses on improved quantitative image analysis, capturing additional prognostic information currently not used by physicians. It is hypothesized that underlying tumor biology can be captured via the use of quantitative image features [47], also named radiomics features. Combining radiomics features with other prognostic information could improve treatment decision making.

PET and CT radiomics features in NSCLC

CT radiomics features have been associated with tumor phenotype and genotype, which strengthens the hypothesis of a potential link between radiomics features and tumor biology [48-50]. The assessment of tumor heterogeneity with PET radiomics features may further improve tumor characterization, as a tumor appearing homogeneous on CT can still exhibit heterogeneity on PET (Figure 3). Assessing the relationship between PET radiomics features and tumor biology has been proven difficult and still warrants further investigation [51-53]. Nonetheless, both PET and CT radiomics features have been associated with patient survival in NSCLC [54-57]. These studies are confounded by small sample size, suboptimal study design, and inappropriate statistical methods, and therefore there is no strong evidence yet that PET radiomics features do exhibit complementary value [58].

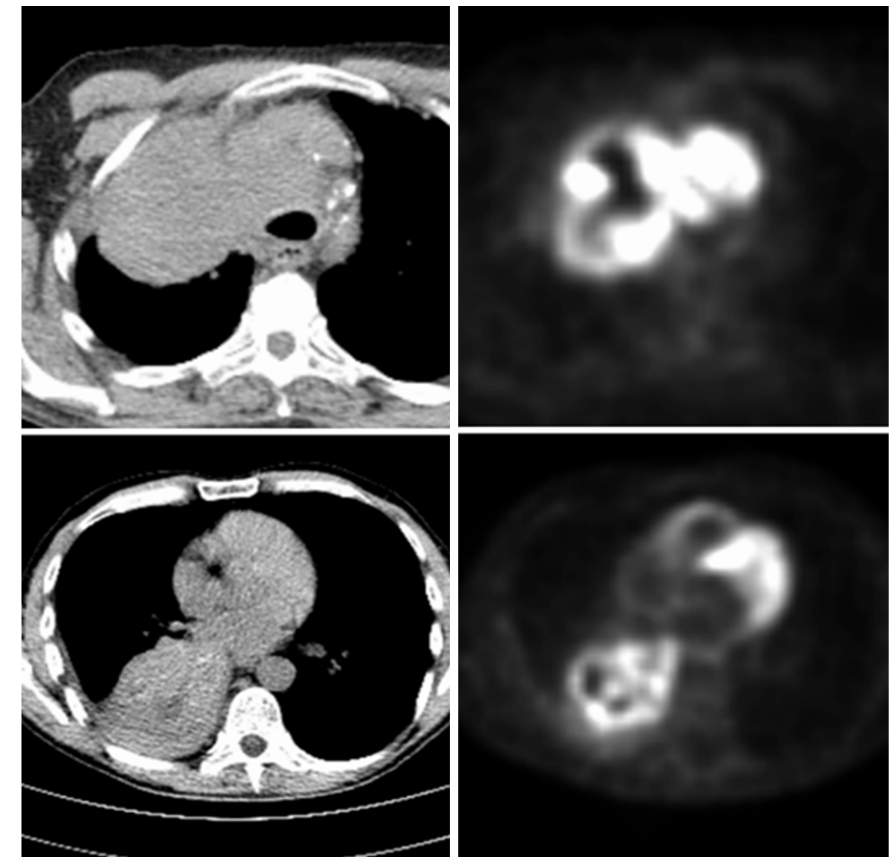


Figure 3. Axial view of two primary lung tumors visualized on registered images from CT (left) and PET (right). On the upper CT slice, the tumor looks like a solid homogeneous mass, but in contrast shows a highly heterogeneous pattern on PET. The CT slice on the bottom does show small photopenic areas centrally located in the tumor, however the extent of this heterogeneity is underestimated in CT when looking at PET.

FDG PET/CT applications for optimized lung cancer management in LMIC

In recent years, a limited number of healthcare facilities in LMIC have established nuclear medicine (NM) services in conjunction with modern radiation oncology (RO) facilities [3, 4]. According to the World Health Organization's Global Atlas of Medical Devices, 20% of LMIC have at least one single PET/CT unit installed compared to 65% of HIC [59]. The International Atomic Energy Agency (IAEA) is committed to helping Member States improve the use of NM applications and RO services in healthcare. The IAEA is an independent intergovernmental, science and technology-based organization, in the United Nations family that has 171 Member States as of February 2019. The work in this thesis is also supported by the IAEA.

The limited number of treatment facilities that are suitably equipped have yet to achieve widespread utilization of their NM and RO services in routine clinical practice. Patients with locally advanced NSCLC would benefit from the multidisciplinary use of PET/CT imaging for concurrent CRT, the current standard treatment in HIC. Concurrent CRT requires accurate treatment delivery as it is more radical than sequential CRT and comes with increased risk of adverse events and longer hospitalization [60, 61]. In LMIC, there is an increased demand for more expertise in RO and NM in these centers to apply the current treatment standards. In addition, investigating complementary prognostic factors to aid in decision making could avoid futile costly medical interventions thus optimizing use of resources.

Objectives of this thesis

Management of patients with lung cancer is suboptimal in LMIC and reasons for this include lack of human resources and limited access to diagnostic and therapeutic techniques. Sites that do have implemented PET/CT in clinic have minimal experience in the multidisciplinary use of PET/CT imaging for concurrent CRT, resulting in many patients receiving suboptimal staging, therapy selection, and treatment. Therefore, the work in this thesis focuses on FDG PET/CT applications that can improve the management of patients with locally advanced NSCLC.

The general aim of this thesis is to achieve accurate treatment delivery and improved prognostic tumor characterization with PET/CT in patients with locally advanced NSCLC treated with CRT. The following objectives are defined:

- To develop pragmatic guidelines for TVD;
- To reduce the interobserver variability in TVD;
- To improve current prognostic models for patients with lung cancer.

Ultimately, the goal is to improve the survival of patients with locally advanced NSCLC.

Outline of this thesis

Chapter 2 is devoted to describing the context of this research by reviewing the relevant literature. This chapter introduces the two main research topics and also describes the newest developments in – and future challenges of PET/CT imaging for radiotherapy TVD and prognostication.

In **Chapter 3** we described detailed guidelines for PET/CT-based TVD, dealing with the uncertainties of PET/CT for management of respiration motion, and introduce the PET/CT-based respiration expanded GTV approach. Overall, the aim of the report was to provide comprehensive guidance on using PET/CT for radiotherapy planning in NSCLC, which had never been provided to this extent.

These guidelines were incorporated in a one-year training program on PET/CT-based TVD in lung cancer that extended on the interface between Nuclear Medicine and Radiation Oncology in LMIC. The aim of this training program was to reduce the IOV and results are discussed in **Chapter 4**.

In **Chapter 5**, the impact of optimized PET/CT-based TVD for radiotherapy planning on overall survival in patients with NSCLC in LMIC is assessed.

The research that is performed in **Chapter 6** generates more insight in the prognostic potential of PET imaging by assessing PET-based radiomics features. We investigated the repeatability of PET radiomics features, and also assessed the relationship with established clinical variables. The rationale was to search for radiomics features derived from pre-treatment PET imaging that are robust, independent, and prognostic, hence, complementary to current clinical prognostic variables and use these to develop an improved prognostic model.

Chapter 7 summarizes the most important findings from this thesis, and discusses the future role of PET/CT on radiotherapy TVD and prognostication.

References

1. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: <https://gco.iarc.fr/today>, accessed 08 Aug 2019.
2. Ferlay J, Soerjomataram I, Dikshit R, et al. *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Int J Cancer 2015; 136(5): 359-86.
3. Datta, Niloy R. et al. *Radiation Therapy Infrastructure and Human Resources in Low- and Middle-Income Countries: Present Status and Projections for 2020*. Int J Radiat Oncol Biol Phys 2014;89(3):448-457.
4. Dondi M, Kashyap R, Paez D, et al. *Trends in nuclear medicine in developing countries*. J Nucl Med. 2011;52(2):16-23.
5. Macbeth FR, Abratt RP, Cho KH, et al. *Lung cancer management in limited resource settings: guidelines for appropriate good care*. International Atomic Energy Agency. Radiother Oncol. 2007;82(2):123-31.
6. Sankaranarayanan R, Swaminathan R. *Cancer survival in Africa, Asia, the Caribbean and Central America*. Lyon: International Agency for Research on Cancer, 2011.
7. Coleman MP, Quaresma M, Berrino F, et al. *Cancer survival in five continents: a worldwide population-based study (CONCORD)*. Lancet Oncol 2008;9:730-56.
8. Antonia SJ, Villegas A, Daniel D, et al.; PACIFIC Investigators. *Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC*. N Engl J Med 2018. [Epub ahead of print]
9. Chansky K, Detterbeck FC, Nicholson AG, et al. *The IASLC Lung Cancer Staging Project: External Validation of the Revision of the TNM Stage Groupings in the Eighth Edition of the TNM Classification of Lung Cancer*. J Thorac Oncol 2017;12(7):1109-1121.
10. Walters S, Maringe C, Coleman MP, et al. *Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007*. Thorax 2013;68(6):551-64.
11. David EA, Clark JM, Cooke DT, et al. *The Role of Thoracic Surgery in the Therapeutic Management of Metastatic Non-Small Cell Lung Cancer*. J Thorac Oncol 2017;12(11):1636-1645.
12. Oudkerk M, Devaraj A, Vliegenthart R, et al. *European position statement on lung cancer screening*. Lancet Oncol 2017;18(12):e754-e766.
13. Molina JR, Yang P, Cassivi SD, et al. *Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship*. Mayo Clin Proc. 2008;83(5):584-594.
14. Aupérin A, Le Péchoux C, Rolland E, et al. *Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small cell lung cancer*. J Clin Oncol 2010; 28: 2181-90.
15. Goldstraw P, Chansky K, Crowley J, et al. *The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer*. J Thorac Oncol 2016;11:39-51.
16. Lardinois D, Weder W, Hany TE, et al. *Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography*. N Engl J Med 2003;348:2500-7.
17. ICRU (1993) ICRU report, vol. 50. Bethesda: International Commission on Radiation Units and Measurements. Prescribing, Recording and Reporting Photon Beam Therapy.
18. ICRU (1999) ICRU report, vol. 62. Bethesda: International Commission on Radiation Units and Measurements. Prescribing, Recording and Reporting Photon Beam Therapy.
19. Grégoire V, Mackie R. ICRU (2010) ICRU report, vol. 83. Bethesda: International Commission on Radiation Units and Measurements. Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT).
20. Steenbakkers RJ, Duppen JC, Fitton I, et al. *Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation*. Radiother Oncol. 2005; 77(2): 182-90.
21. Rasch C, Belderbos J, van Giersbergen A, et al. *The Influence Of A Multi-disciplinary Meeting For Quality Assurance On Target Delineation In Radiotherapy Treatment Preparation*. International Journal of Radiation Oncology. 2009; 75(3): 452-453.

22. Hanna GG, McAleese J, Carson KJ, et al. (18)F-FDG PET-CT simulation for non-small-cell lung cancer: effect in patients already staged by PET-CT. *Int J Radiat Oncol Biol Phys*. 2010; 77(1): 24–30.
23. Caldwell CB, Mah K, Ung YC, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non small-cell lung tumors on CT: The impact of 18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys*. 2001; 51: 923–931.
24. Greco C, Rosenzweig K, Cascini GL, et al. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). *Lung Cancer*. 2007; 57: 125–34.
25. Fox JL, Rengan R, O'Meara W, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys*. 2005; 62: 70–75.
26. Tixier F, Hatt M, Valla C, et al. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *J Nucl Med*. 2014; 55(8):1235–1241
27. Keyes JW Jr. SUV: standard uptake or silly useless value? *J Nucl Med* 1995; 36(10):1836–9.
28. de Jong EEC, van Elmpt W, Hoekstra OS, et al. Quality assessment of positron emission tomography scans: recommendations for future multicenter trials. *Acta Oncol* 2017; 56(11): 1459–1464.
29. Boellaard R, Krak NC, Hoekstra OS, et al. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med* 2004; 45(9):1519–27.
30. Aide N, Lasnon C, Veit-Haibach P, et al. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging* 2017; 44(1):17–31.
31. Boellaard R, Oyen WJG, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging* 2008; 35(12):2320–2333.
32. Detterbeck FC, Boffa DJ, Kim AW, et al. *The Eighth Edition Lung Cancer Stage Classification*. *Chest* 2017; 151(1): 193–203.
33. Rami-Porta R, Bolejack V, Giroux DJ, et al, and International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Advisory Board Members and Participating Institutions. *The IASLC Lung Cancer Staging Project: the new database to inform the eighth edition of the TNM classification of lung cancer*. *J Thorac Oncol* 2014; 9: 1618–1624.
34. Paesmans M. Prognostic and predictive factors for lung cancer. *Breathe*. 2012; 9: 112–121.
35. Berghmans T, Paesmans M, Sculier JP. Prognostic factors in stage III non-small cell lung cancer: a review of conventional, metabolic and new biological variables. *Ther Adv Med Oncol* 2011; 3: 127–138.
36. G Buccheri, D Ferrigno. Importance of weight loss definition in the prognostic evaluation of non-small-cell lung cancer. *Lung Cancer* 2001; 34: 433–440.
37. Nakamura H, Ando K, Shinmyo T, et al. Female gender is an independent prognostic factor in non-small-cell lung cancer: a meta-analysis. *Ann Thorac Cardiovasc Surg* 2011; 17: 469–480.
38. Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016; 7(7): 12474.
39. Pallis AG, Gridelli C. Is age a negative prognostic factor for the treatment of advanced/metastatic non-small-cell lung cancer? *Cancer Treat Rev* 2010; 36(5): 436–41.
40. Yu Z, Zhang G, Yang M, et al. Systematic review of CYFRA 21-1 as a prognostic indicator and its predictive correlation with clinicopathological features in Non-small Cell Lung Cancer: A meta-analysis. *Oncotarget* 2017; 8(3): 4043–4050.
41. Dehing-Oberije C, Aerts H, Yu S, et al. Development and Validation of a Prognostic Model Using Blood Biomarker Information for Prediction of Survival of Non-Small-Cell Lung Cancer Patients Treated With Combined Chemotherapy and Radiation or Radiotherapy Alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys* 2009; 81(2):360–368.
42. Steels E, Paesmans M, Berghmans T, et al. Role of p53 as prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur Respir J* 2001; 18: 705–719.
43. Tong J, Sun X, Cheng H, et al. Expression of p16 in non-small cell lung cancer and its prognostic significance: a meta-analysis of published literatures. *Lung Cancer*. 2011; 74: 155–163.
44. Martin B, Paesmans M, Mascaux C, et al. KI-67 expression and patients survival in lung cancer: systematic review of the literature with meta-analysis. *Br J Cancer*. 2004; 91: 2018–2025.
45. C. Dehing-Oberije, S. Yu, D. De Ruyscher, et al. Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2009; 74:355–362.
46. Liu L, Shi M, Wang Z, et al. A molecular and staging model predicts survival in patients with resected non-small cell lung cancer. *BMC Cancer* 2018; 18(1):966.
47. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48(4):441–6.
48. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 5:4006.
49. Weiss GJ, Ganeshan B, Miles KA, et al. Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. *PLoS One* 2014; 9(7):e100244.
50. Yamamoto S, Korn RL, Oklu R, et al. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. *Radiology* 2014; 272:568–576.
51. Yip SS, Kim J, Coroller TP, Parmar C, et al. Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer. *J Nucl Med* 2017; 58(4):569–576.
52. Del Gobbo A, Pellegrinelli A, Gaudioso G, et al. Analysis of NSCLC tumour heterogeneity, proliferative and 18F-FDG PET indices reveals Ki67 prognostic role in adenocarcinomas. *Histopathology* 2016; 68(5):746–51.
53. van Baardwijk A, Bosmans G, van Suylen RJ, et al. Correlation of intra-tumour heterogeneity on 18F-FDG PET with pathologic features in non-small cell lung cancer: a feasibility study. *Radiother Oncol* 2008; 87:55–58.
54. Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* 2012; 264:387–396.
55. Cook GJR, O'Brien ME, Siddique M, et al. Non-small cell lung cancer treated with erlotinib: heterogeneity of 18F-FDG uptake at PET—association with treatment response and prognosis. *Radiology* 2015; 276:883–893.
56. Depeursinge A, Yanagawa M, Leung AN, et al. Predicting adenocarcinoma recurrence using computational texture models of nodule components in lung CT. *Med Phys* 2015; 42:2054–2063.
57. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 2014; 90:834–842.
58. Chalkidou A, Michael J, O'Doherty MJ, Paul K, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS ONE* 2015; 10(5):e0124165.
59. World Health Organization. Global atlas of medical devices. Geneva: World Health Organization; 2017. Accessed on 07 April 2020.
60. Curran WJ Jr, Paulus R, Langer CJ, et al. Sequential vs. concurrent chemoradiation for stage III non-small cell lung cancer: randomized phase III trial RTOG 9410 [published correction appears in *J Natl Cancer Inst*. 2012 Jan 4; 104(1):79]. *J Natl Cancer Inst*. 2011; 103(19):1452–1460.
61. Vergnenegre A, Combesure C, Fournel P, et al. Cost-minimization analysis of a phase III trial comparing concurrent versus sequential radiochemotherapy for locally advanced non-small-cell lung cancer (GFP-C-GLOT 95-01). *Ann Oncol*. 2006; 17(8):1269–1274.



Chapter 2

The developing role of FDG PET imaging for prognostication and radiotherapy target volume delineation in non-small cell lung cancer

2

Tom Konert
Jeroen B. van de Kamer
Jan-Jakob Sonke
Wouter V. Vogel

Journal of Thoracic Disease 2018; 10(21): 2508–2521

Abstract

Advancements in functional imaging technology have allowed new possibilities in contouring of target volumes, monitoring therapy, and predicting treatment outcome in non-small lung cancer (NSCLC). Consequently, the role of ^{18}F -fluorodeoxyglucose Positron Emission Tomography (FDG PET) has expanded in the last decades from a stand-alone diagnostic tool to a versatile instrument integrated with Computed Tomography (CT), with a prominent role in lung cancer radiotherapy. This review outlines the most recent literature on developments in FDG PET imaging for prognostication and radiotherapy target volume delineation in NSCLC. We also describe the challenges facing the clinical implementation of these developments and present new ideas for future research.

Introduction

Lung cancer is the most common cause of death from cancer worldwide (2). For patients with non-small cell lung cancer (NSCLC) who are being considered for curative intent treatment, ^{18}F -fluorodeoxyglucose Positron Emission Tomography (FDG PET)/ Computed Tomography (CT) imaging has become the standard of care in baseline staging, and has also shown benefit for radiotherapy planning (RTP) (2-4).

The implementation and applications of FDG PET and CT for NSCLC have changed over time. FDG PET was initially acquired as a standalone modality, and was demonstrated to be superior to computed tomography (CT) alone in the staging of lung cancer (5,6). When PET was acquired in conjunction with a CT using an integrated scanner (PET/CT), the combined information has shown to have higher staging accuracy than PET imaging alone (7-12). Evaluation of combined PET/CT images has traditionally been based on visual interpretation, and was predominantly applied for tumor detection, staging and treatment selection. Subsequently, the value of FDG PET/CT in manual target volume delineation (TVD) for radiotherapy was also demonstrated (13). In the more recent years, PET tracers other than FDG have come available for evaluation of different biological tumor characteristics. Advancements in multimodal technology now create possibilities for PET/CT imaging in new applications, including (semi-)automatic definition of target volumes, quantitative response assessment, determining patient prognosis, and predicting treatment outcome in NSCLC (14-20).

Radiomics has been introduced as a sophisticated way to extract and mine a large number of quantitative image features believed to provide a comprehensive picture of tumor phenotypes, for example related to necrosis, angiogenesis and radioresistant cells (21). The assessment of biological processes with PET radiomics features may further improve tumor characterization (22). The relation between CT image texture patterns and tumor biology has provided further insights into tumor phenotype and genotype (19,23,24); such a relationship for PET still warrants extensive investigation (25-27). Meanwhile, studies already indicated that radiomics features contain prognostic information regarding response to therapy or treatment outcome in lung cancer studies (19,28-31). In addition, radiomics features may also contribute to visual or (semi-)automatic definition of gross tumor volume (GTV) (32). Ultimately, proper use of PET/CT imaging could contribute in all steps of the treatment procedure and optimal use may lead to better tumor characterization, treatment decision making, treatment guidance, tumor response assessment, and local tumor control.

Based on these factors, the use and role of PET imaging has expanded from a primarily diagnostic tool to a more central role in the context of personalized medicine. This review discusses the current state of art in applications of FDG PET/CT for prognostication and target volume delineation in NSCLC, and describes the challenges related to the clinical implementation of these new developments.

A literature research was conducted to assess recent advancements in the use of PET/CT imaging for prognostication and radiotherapy TVD in NSCLC. A search query was undertaken at the PubMed database, using a combination of the following keywords: (“PET”) and (“non-small cell lung cancer” or “NSCLC”) and (“target volume delineation” or “segmentation” or “prognostic” or “prognostication” or “radiomics” or “textural features” or “precision medicine”). The search query yielded a total of 410 papers. Only studies written in English, related to PET/CT for prognostication and TVD in the treatment of NSCLC with radiotherapy and of relevance to this overview were included. To illustrate this, papers regarding PET guided patient examination, PET and particle therapy, surgery, PET/CT adapted RT, PET and dosimetric planning, PET and lung toxicity, PET and lung ventilation studies, drug assessment, or economic related were excluded. No limitations were set on the year of publication. However, in the case of review papers, solely the most recent ones for each topic were included. Other reasons for exclusion were inaccessibility, case reports, editorials or conference abstracts, resulting in 77 papers. In addition, references within retrieved articles were analyzed to expand the search. In the end, this led to 124 papers covering the relevant topics, which were studied and incorporated in the descriptive evaluations below.

Prognostic factors in NSCLC

In the last decades, the overall survival (OS) of lung cancer patients has not improved tremendously (33). The selection of treatment strategies for NSCLC patients is mainly based on empirical models. The most important prognostic indicator is the disease stage, which is determined by the extent of the primary tumor, nodal involvement, and the presence of distant metastasis according to the TNM classification (34). Disease staging also plays an important part in guiding therapy (6). In locally advanced NSCLC, however, treatment selection based on TNM staging and other clinical variables may not be accurate enough for survival probability prediction (35,36). As technology improves and more treatment options become available, the search for more accurate prognostic factors is warranted in the context of personalized medicine.

Conventional evaluations

The value of many well-established prognostic factors, such as the distinction between stage IIIA, IIIB, and IIIC, performance status (PS), histology, and other clinical and therapeutic variables, have been confirmed in locally advanced NSCLC (37). Other studies have shown that basic imaging-derived features in pre-treatment and post-treatment scans provide clinically relevant prognostic information for patients with NSCLC of various stages. Examples include tumor size and volume on CT, and standardized uptake value (SUV) based metrics like SUV_{max} , SUV_{peak} , and SUV_{mean} , or metabolic active tumor volume (MTV) on FDG PET (38-44). As a combined parameter, total lesion glycolysis (TLG) was shown to be more promising than MTV, and in combination with other parameters such as shape based features complementary prognostic information could be extracted from PET images (19,45,46). These metrics are all related to tumor burden and metabolic characterization. An overview of studies about image-based prognostication using PET in NSCLC together with their findings is given in table 1.

Table 1. Prognostication with conventional PET image features in NSCLC

Ref	Subject	Features	No. of pts	Results	Conclusion
(36)	Tumor prognosis and response assessment with FDG PET	First order	51	SUV_{max} , PS, and stage were significantly prognostic for disease-specific survival. SUV_{max} and performance were prognostic for OS.	SUV_{max} is an important prognostic factor for survival of inoperable NSCLC and predictive for treatment response.
(37)	Response assessment with FDG PET	First order	40	>20% decrease in SUV_{mean} predicted longer PFS (9.7 versus 2.8 months)	PMR after 3 weeks significant prognostic factor for PFS; OS remained poor
(38)	Response assessment with FLT and FDG PET	First order	51	>15% decrease in SUV_{max} after both 14 days and 56 days associated with longer PFS	PMR significant prognostic factor for PFS; OS was only prognostic after 14 days
(39)	Tumor prognosis with FDG PET	First order	309	MTV and TLG significantly associated with increased risk of death (HR =1.27; HR = 1.22, respectively)	Volume-based PET parameters are significant prognostic factors for OS
(40)	Tumor prognosis with FDG PET	First order	52	Multivariate analysis demonstrated that TLG is significantly associated with OS (HR =1.03) and PFS (HR =1.04)	TLG is a significant independent prognostic factor of PFS and OS

PET, positron emission tomography; NSCLC, non-small cell lung cancer; PS, performance status; SUV, standardized uptake value; PMR, partial metabolic response; PFS, progression free survival; OS, overall survival; MTV, metabolic tumor volume; TLG, total lesion glycolysis; HR, hazard ratio.

Even though basic PET metrics have been proclaimed to be of prognostic or predictive value, there are still contradictions found in the literature where certain SUV metrics do not show any prognostic value, when used with other prognostic factors (47). Paesmans et al. conducted a meta-analysis about the prognostic value of SUV_{max} in NSCLC and concluded that these contradictory findings could be due to influence of disease stage or tumor histology, or due to differences in assessment methods (11). These dependencies may not hold for all PET features, hence PET features still may play a role in prognostication..

Radiomics textural features for improved prognostication

PET radiomics represents the high-throughput mining of quantitative image features from PET imaging to characterize tumor phenotypes. Radiomics features include first order features, that are based on the gray level intensity and its distribution in the image, but do not consider the three dimensional (3D) distribution of gray levels. Examples are the max, mean, standard deviation, skewness, and kurtosis of SUV within a volume of interest. The prognostic value of these ‘simple’ features may be weaker in large well differentiated tumors that are known to exhibit higher hypoxia, necrosis, or anatomic and physiologic complexity, which translates to higher complexity in the spatial distribution of PET tracer uptake (11).

Radiomics features also include second and higher order features, called textural features, which may cover this higher complexity by describing the relationship between the voxel intensity and their position within an image. This relationship can be calculated with various mathematical methods, such as the gray level co-occurrence matrix (GLCM) for pairwise arrangement of voxels (48), the gray level run-length matrix (GLRLM) for alignment of voxels with the same intensity (49), the gray level size-zone matrix (GLSZM) for characteristics of zones with identical voxel values (50), and the neighborhood grey tone difference matrix (NGTDM) for determining changes in neighboring voxel intensities (51). As an example, GLCM entropy measures the variability in neighborhood intensity values and may be useful to characterize necrotic cores, a factor that has been associated with worse prognosis. Higher entropy represents a more heterogeneous FDG PET activity within the tumor, as is depicted in. Another example includes GLSZM features that mostly relate to the size of subregions within a tumor with similar intensities, which hypothetically characterizes cell subpopulations with distinct clonogenic growth. Some of these textural features have shown prognostic value for clinical outcome and tumor response (17,29,52-54). An overview of studies about the prognostic value of textural features in NSCLC together with their findings is given in table 2.

Table 2. The prognostic value of PET textural features in NSCLC

Ref	Subject	Features	No. of pts	Results	Conclusion
(17)	Tumor prognosis with FDG PET	First order and textural features	101	Entropy, MTV, and stage were significant prognostic factors for OS. The HR was 3.81 between a low and high risk group based on the 3 features above	Entropy and MTV contain complimentary information next to TNM staging
(29)	Tumor prognosis and response assessment to Erlotinib with FDG PET	First order and textural features	47	Contrast at 6 weeks (HR = 1.81) and % change in first-order entropy (HR = 1.14) were significantly prognostic for OS. Percentage change in first-order entropy was also associated with treatment response (OR = 0.30)	Contrast and % change in first-order entropy are significantly prognostic for OS and the latter also associated with treatment response following RECIST
(52)	Tumor prognosis and response assessment with FDG PET	First order and textural features	53	ROC curves for textural features to predict RECIST response ranged from 0.54 to 0.82. High coarseness was an independent prognostic factor for OS (HR = 4.86); high coarseness, contrast, busyness, and complexity were significantly prognostic for PFS (HR = 2.41; 0.60; 0.97; 0.87)	Textural features were highly predictive for RECIST responders compared to first order features and were prognostic for PFS
(53)	Tumor prognosis with FDG PET	First order and textural features	45	Entropy was determined as a significant independent factor in multivariate analysis (HR = 7.48) for disease-specific survival	Tumor heterogeneity as described by FDG-PET texture is associated with response to radiation therapy in NSCLC
(54)	Tumor prognosis in FDG PET	First order and textural features	201	Internally validated optimism-corrected C-statistic was 0.63 for a model which predicted OS with both high MTV and high SumMean included	A textural feature was identified as prognostic for OS in large tumors only (>93.3 cc)

PET, positron emission tomography; NSCLC, non-small cell lung cancer; SUV, standardized uptake value; PMR, partial metabolic response; PFS, progression free survival; OS, overall survival; MTV, metabolic tumor volume; HR, hazard ratio; OR, odds ratio; CI, concordance index.

PET textural features are calculated on a group of voxels within a region-of-interest (ROI). A common ROI that is used for PET radiomics analysis is MTV. The relationship between MTV and radiomics textural features has been investigated and studies demonstrated that specific PET textural features are closely correlated to MTV (17,19,52,55). Therefore, in these cases prognostic textural features would rather act as a surrogate than as an independent variable. On the other hand, several studies demonstrated that this dependency is decreasing with increasing MTV (17,55), suggesting that for larger tumors these specific PET features may become of relevance. This is probably caused by the partial volume effect for small tumors, making it difficult to obtain reliable PET radiomics features for tumors smaller than 10 cc (56). For instance, to determine the GLCM entropy, tumors were required to be at least more than 10 cc to reduce correlations with MTV (17,55). Unfortunately, these findings were based on only four PET textural features and is not representative for all radiomics features, since this volume dependency differs amongst textural features. A high GLCM entropy (see figure 1 for an example) in combination with a large MTV (>35 cc) led to a worse prognosis in 101 patients with stage I-III NSCLC receiving surgery, chemotherapy, chemoradiotherapy, or a combination (17). Ohri et al. performed PET radiomics analysis in 201 patients with locally advanced NSCLC and concluded that a feature called GLCM SumMean had prognostic value for tumors with an MTV larger than 93 cc (54). On the contrary, Pyka et al. concluded that GLCM entropy was predictive for disease-specific survival in 45 early stage NSCLC patients with mainly small tumors (mean MTV 34 cc, ranging from 1.74 to 178 cc) receiving primary stereotactic radiation therapy (53). However, only 12 patients were used to assess this specific endpoint and divided in two unbalanced groups (2 versus 10) for comparison, which is a serious limitation. These results point in the direction that for larger tumors, PET textural features contain complimentary information above a specific tumor volume. The volume threshold differs per textural feature. We suggest assessing volume dependencies in detail for each textural feature in large patient cohorts, when building prognostic models containing PET textural features.

The literature also reports on correlations between first order SUV metrics and higher order features (texture) (29,52). Additionally, PET textural features also depend on image segmentation and image reconstruction settings (57), SUV binning (58,59), and feature calculation method (17). And although it is clear how these choices influence relationships between PET textural features or their reproducibility, it is not always clear how these factors affect the prognostic value. In the search for independent prognostic PET textural features, some investigators find optimal cutoff values that result in prognostic variables, but do not validate their results (52). Fave et al. (60) argued that testing multiple cutoffs to find the best one without an independent validation dataset for testing, could yield overly optimistic results. In combination with the inclusion of multiple variables that have not been corrected for multiple hypothesis testing, the models will probably not perform

well in other patient cohorts. From the studies shown in Table 1, the issue of multiple-hypothesis testing was solely addressed by Ohri et al. (54), but none of them had validated their results externally. Internal or external validation is lacking in most studies, but is required to strengthen results regarding model performance. In absence of an external validation cohort, it is advised to split the initial cohort into a training and a test set. We would like to emphasize that prognostic models should always be corrected for multiple hypothesis testing and should be validated, preferably with an external patient cohort.

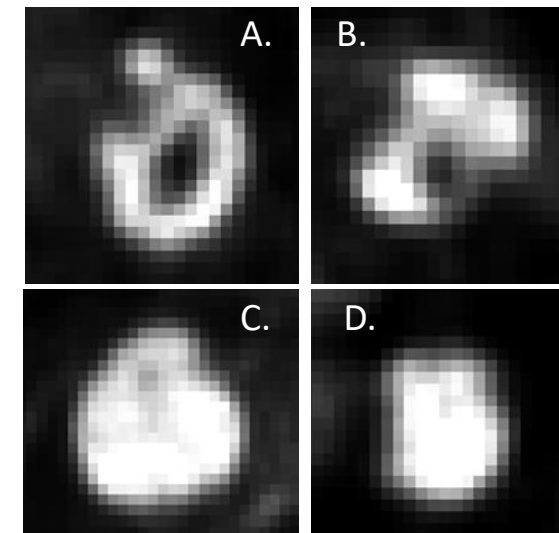


Figure 4. FDG PET images showing primary tumors from different stage III NSCLC patients. (A) and (B) have a high entropy indicating that the variability is high in neighborhood intensity values. (C) and (D) have a more homogeneous distribution of SUV within the tumor, which corresponds to a low entropy value. PET, positron emission tomography; NSCLC, non-small cell lung cancer; SUV, standardized uptake value.

Even though many studies reported on the prognostic value of PET radiomics features, results are difficult to compare amongst studies, and convincing evidence remains poor (61). Many studies about PET radiomics features make the remark that a standardized approach is lacking, which influences generalizability and makes meta-analyses difficult. For radiomics studies, guidelines can be found in the literature that promotes standardization for designated terms, extraction and calculation of textural features (62), and statistical analysis (63). We agree that standardization is necessary and that studies should comply with best-practice procedures to move forward in the field of radiomics. In light of these demands, a Radiomics Quality Score was proposed to evaluate radiomics studies in literature following standardized criteria (64). These criteria can also aid in setting up future radiomics studies and check whether they comply with the written guidelines. The

Radiomics Quality Score will hopefully lead to comparable high quality radiomics studies, and would facilitate future meta-analyses.

Other PET tracers

The majority of evidence on PET/CT for prognostication concerns the use of FDG. Although very sensitive for staging most tumor types, a disadvantage of FDG is that it is not tumor specific. FDG uptake can be seen in inflammatory lesions, increasing the chance of false-positive findings.

The tracer 3'-deoxy-3'-¹⁸F-fluorothymidine for PET imaging (FLT PET), which allows assessment of tumor proliferation, does not seem to accumulate in inflammatory processes. Therefore, FLT is considered a more specific oncological tracer than FDG, but sensitivity is lower (65). Its main role is currently envisioned in evaluating early treatment response (66-68). However, the combination of dual-tracer PET/CT may improve diagnostic accuracy (69) and may also contribute in early response assessment and prediction of clinical outcomes (70). One study on NSCLC compared the prognostic value of FLT PET to FDG PET first order features and showed that both FDG and FLT PET baseline maximum SUV were associated with overall survival. However, clinical studies proving the prognostic value of textural features in FLT PET imaging are scarce. Therefore, the additional value of FLT PET textural features next to well-established prognostic factors for evaluation of treatment response and predicting clinical outcomes in NSCLC has to be further studied (71).

Other tracers which are currently under investigation cover different aspects of lung cancer biology that may enable better phenotypic characterization for improved prognostication. ¹⁴C-methionine or l-3-¹⁸F- α -methyl tyrosine (¹⁸F-FAMT) characterizes amino acid transport and protein metabolism. These processes are upregulated in malignant cells as a consequence of increased cellular proliferation activity. In addition, ¹⁸F-fluoromisonidazole (¹⁸F-FMISO), ¹⁸F-fluoroazomycin arabinoside (¹⁸F-FAZA), and ¹⁸F-flortanidazole (¹⁸F-HX4) are hypoxic tracers. Hypoxia is an important factor in oncology, because this increases radioresistance and chemoresistance, and is related to poor clinical outcome (72). Lastly, there are PET tracers targeting integrins expressed on tumor vasculature, which measure levels of angiogenesis growth factors, to characterize angiogenesis. The clinical usage of these PET tracers in lung cancer has been limited so far, and pre-clinical research is still dominating (67). These upcoming tracers have potential for tumor detection, characterization, prognostication, and response assessment, but efforts need to be made to ensure safety, optimal signal-to-noise, and reproducibility of measurements in order to move forward. Therefore, there is insufficient data on the use of these tracers for prognostication for NSCLC, either as a separate parameter or in combination with FDG PET.

Target volume delineation in locally advanced NSCLC

The definition of target volumes for radiotherapy of NSCLC was traditionally based on anatomical imaging with CT, in combination with findings from e.g. physical examination, endoscopy and biopsies. Multiple advances in CT imaging have further improved its spatial and contrast resolution with benefit for tumor delineation, for example using multi-slice detectors and respiratory motion correction techniques (73-76). In addition, it has been shown that CT images can be enhanced for discrimination of tumor and normal tissues, using e.g. intravenous contrast or dual-energy CT (77-79). However, despite these improvements, CT alone does not provide sufficiently clear information to reliably and consistently discriminate tumor and normal tissues in all situations (80-82).

After the introduction of integrated PET/CT, its additional value for target definition was rapidly acknowledged. The relevance of good and reproducible quality of the images and accurate anatomical registration with planning CT were acknowledged. This resulted in the standardization of PET/CT for radiotherapy TVD (2). Several studies confirmed the value of adding PET imaging to planning CT to reduce inter-observer variation in TVD in RTP in NSCLC patients, and is specifically helpful in TVD when the tumor boundaries are not easily distinguished from surrounding healthy tissue (81-84). Even with the use of PET imaging there is still variability amongst observers (85,86). The remaining question is how to derive the most optimal target volumes using the information gleaned from combinations of PET and CT.

Visual interpretation

Most early clinical studies have used a visual interpretation technique for target definition from combined FDG PET and CT scans (83,84). As in any observer-dependent procedure, it is essential to standardize this interpretation where possible. In one study, the benefit of using strict protocols was shown (85). Another study demonstrated that interdisciplinary cooperation between the radiation oncologist and nuclear medicine specialist is beneficial for consistent contouring (87).

Automatic target volume delineation

Many groups have investigated the use of automated segmentation techniques to either guide or generate the relevant target volume (14,15,88-91). A previous International Atomic Energy Agency (IAEA) publication provided guidance on the use and role of PET/CT imaging for RTP in a range of tumor sites (92) and an update was given in 2015 with additional practical guidelines for the use of FDG PET/CT for the purposes of radiotherapy TVD in NSCLC (3). These guidelines state that target volumes generated following any automatic segmentation algorithm should always be verified, and edited where needed, by a trained

observer. The skepticism towards automatic contouring in general, demonstrates that there are still hurdles to overcome in automatic contouring before it is fully accepted in the clinic.

The most common application of automatic segmentation on PET imaging (PET-AS) is definition of the MTV, however studies tried to go even further and proposed methods for GTV definition. Many different PET-AS algorithms have been proposed to define the edge of the GTV for this purpose (14,93-94). It was thought that these algorithms would provide contours that are more consistent and better represent true tumor borders, when compared to visual interpretation by radiation oncologists. Unfortunately, none of these methods have been tested on large patient datasets and in the presence of a ground truth, probably because pathological correlation has proven difficult (95,96). On top of that, the different PET-AS algorithms all failed under specific circumstances, which prevented recommending a single algorithm (15). A possible solution for this variable performance of PET-AS methods was suggested in the form of a consensus algorithm that combines different PET-AS methods (97). Another difficulty with PET-AS is the variability of SUV values due to factors other than tumor activity alone, such as patient factors and technical factors (98). More recently, the AAPM task group has said to undertake steps to establish a standardized procedure for PET-AS algorithms, which could help with the acceptance and implementation of these delineation methods in the clinic (15). Perhaps the question is whether the focus should be solely on PET imaging for target volume delineation or should be more on the combination of multiple imaging modalities, if there are already many studies that show its shortcomings.

The most important factor is that PET-AS neglects all other available information, like anatomical information from CT and in some cases MRI, the locations of tumor-positive biopsies, and findings at endoscopic examinations. Therefore, the information obtained from PET should be considered complementary. Combining the information contained in the PET and CT scans may lead to more successful auto-contouring (95). Still, resulting contours should be checked visually and edited to other sources of information where needed.

Radiomics for target definition

The application of textural features for use in target volume delineation has yet to garner attention, leading to a minimal number of published studies. Early attempts to automate tumor contouring using textural features were promising. For example Wang et al. (99) designed a radiomics based automatic contouring method by training a ROI-based decision-tree-based K-nearest neighbor (K-nn) classifier using 14 PET and 13 CT textural features. The K-nn classifier is a classical machine learning method, and here its purpose was to detect and classify image voxels in NSCLC. It was demonstrated that a combination of these textural features from PET and CT images could distinguish between abnormal and normal tissue in the head and neck and that their classifier method was able to

generate accurate and consistent delineations compared to other automatic and to manual delineations. Further studies using a larger sample size and with pathologic validation are necessary to determine its clinical value in thoracic diseases.

Interestingly, it is said that the more advanced deep learning method (100), also able to classify objects, outperforms these classical machine learning methods. A recent comparison, however, of lymph node metastasis classifications in NSCLC patients between classical machine learning methods and a deep learning approach showed comparable results (99). Nevertheless, a big advantage of deep learning is its ability to generate data driven features instead of relying on hand crafted features and is thus potentially more powerful (100). This comes with a higher risk of over-fitting if not appropriately trained and validated, and typically needs more data. Its use in medical image analysis is increasing as algorithms become more sophisticated and more data becomes available (101, 102), which might lead to new insights in tissue classification and delineation.

Discussion

The literature presented in this review indicates that FDG PET/CT has a substantial role in prognostication and radiotherapy planning for NSCLC. First order metrics such as SUV_{max} and certain PET textural features were prognostic for survival and treatment response assessment. It is noted, however, that many publications about PET radiomics are based on relatively small datasets without robust internal and/or external validation which challenges clinical translation. Both prognostication and GTV definition should not rely on PET alone, but rather on its addition to and integration with other independently validated sources of information. The optimal implementation of such strategies is expected to continue to evolve over the coming years.

Current challenges

Challenges in prognostication with PET imaging cover a broad range of topics including image acquisition and reconstruction settings, tumor segmentation, image feature calculation, and statistical methods. Standardization of these topics for PET radiomics studies would promote reproducibility of study results, which is typically lacking (103). This lack of standardization challenges prognostication as variations in all these topics may introduce changes that are not due to underlying biology. It is already shown that this may lead to false positive results (61), and therefore standardization of these topics is warranted to move further in the field. A challenge to overcome in image acquisition is in finding optimal image quality, hence preserving heterogeneity information. Respiratory motion during PET imaging causes lesion smearing (104,105), which reduces contrast in the image and therefore affects

the quantification of heterogeneity. Use of partial volume and respiratory motion correction techniques may improve quantification accuracy (56,106). Also differences in matrix grid size can confound results, as larger voxel sizes tend to be more affected by the partial volume effect. This leads to a more uniform intensity distribution, which subsequently has an impact on most radiomics PET features (107), and study designs need to take this into account. Hence, it is known that differences in image quality or matrix grid size affect textural feature outcome, and further studies should focus on its impact on the prognostic value.

The effect of tumor segmentation on the prognostic and predictive value of imaging features is also not yet clear. PET-AS algorithms or manual delineation by radiation oncologists determine which voxels within an image will be analyzed, thus, the variability in segmentation affects reproducibility when extracting imaging parameters from PET and CT scans (108-110). The accuracy and reproducibility of target volume delineation is important as it has an impact on imaging parameters that are used to determine patient prognosis and to predict and monitor response to therapy. Interestingly, in NSCLC it is not yet known whether the GTV used for treatment planning or the MTV results in higher prognostication. The inclusion of low metabolically active regions within a tumor could contain valuable information and also contribute to the heterogeneity in a tumor, whereas on the other side it also holds the risk of including non-pathological tissue. Further studies should assess the impact of these different volumes on prognostication in NSCLC patients.

Interchangeable SUV measurements across centers are very important in PET radiomics, but the methodology used to determine textural features also demonstrates lack of consistency. There are multiple methods to calculate textural features, which may have impact on the prognostic value of these quantitative PET imaging features. In addition, there are textural features that rely on image intensity resampling or SUV discretization. Discretization reduces the large number of intensity values (typically 16-bit in PET imaging (111)) to a smaller number, e.g. 32 or 64 bins. Each SUV discretization method results in a new set of features (58). Also, textural features can be calculated in 2D or 3D, in one direction or in multiple directions and so on. Hence, the number of textural features can easily reach to hundreds. Therefore sufficiently large patient cohorts need to be included combined with sophisticated statistical methods to prevent overfitting. With multiple possibilities to calculate features, it is often not clear how researchers achieved the final results. Hence, there is the urge to create guidelines for standardized PET radiomics analysis, and for reporting study results about prognostication with PET radiomics features in NSCLC.

As more complex PET radiomics features are designed, and hundreds of these complex imaging features are included in PET radiomics studies, it is understandable that studies choose a hypothesis generating strategy. This is a legit strategy, as long as results are validated

and tested for any kind of confounding as described before. It should be avoided that studies suffer from a poor design and result in the publication of overly optimistic results. Although quantitative studies about image heterogeneity in cancer have shown associations with aspects of tumor behavior, it is not fully understood how underlying biology is affecting the PET signal, and it is foreseen that answers to this will facilitate implementation in the clinic (112). Therefore, more studies should investigate the relationship of tumor biology and proven robust independent textural features.

Future perspectives for PET imaging in prognostication and TVD

Innovation in PET technology opens up more possibilities and holds perspective for the future, although not necessarily the solution for all above described issues. Examples include opportunities for PET/MR (113-116) and 4D PET/CT image reconstruction and partial volume correction methods for increased quantitative accuracy (106,117-118). New PET tracers might become available to improve tumor characterization, and potentially lead to new and improved prognostic biomarkers (67,119-122). Newer PET/CT scanners with improved sensitivity and spatial resolution could lead to better tumor detection and target volume definition (123). In addition, new digital PET technology will be used within hybrid PET/MR systems and could facilitate target volume definition by improved motion correction (124). Since almost every patient with lung cancer is scanned with PET/CT, management of this valuable data for quantitative image analysis on a large scale is strongly desired. This allows for studies with larger sample sizes in the future, benefiting radiomics analysis in specific, but also demands proper data management.

Conclusions

PET/CT has an increasing role in prognostication and target volume definition of NSCLC. The implementation of new PET based applications will facilitate the shift from visual interpretation and manual delineation to (semi)automated target volume definition. In order to establish common ground in the clinic, studies about the prognostic value of quantitative PET imaging features will require external validation cohorts and pathological validation. The implementation of quantitative PET imaging features in a clinical setting would require substantial effort to standardize both imaging and methods for radiomics analysis. Current efforts to create larger databases, will hopefully lead to strong evidence in prognostication with PET/CT imaging in NSCLC.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136(5):359-86.
2. Ung JC, Bezjak A, Coakley N, et al. and the Lung Cancer Disease Site Group. Positron Emission Tomography with 18Fluorodeoxyglucose in Radiation Treatment Planning for Non-small Cell Lung Cancer. *J Thorac Oncol* 2011;6:86-97.
3. Konert T, Vogel W, MacManus MP, et al. PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014. *Radiother Oncol* 2015;116(1):27-34.
4. Vogel WV, Lam MG, Pameijer FA, et al. Functional Imaging in Radiotherapy in the Netherlands Availability and Impact on Clinical Practice. *Clin Oncol* 2016;28(12):206-215.
5. Gould MK, Kuschner WG, Rydzak CE, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: A meta-analysis. *Ann Intern Med* 2003;139:879-892.
6. MacManus MP, Hicks RJ, Matthews JP, et al. High rate of detection of unsuspected distant metastases by pet in apparent stage III non-small-cell lung cancer: Implications for radical radiation therapy. *Int J Radiat Oncol Biol Phys* 2001;50:287-293.
7. Cerfolio RJ, Ojha B, Bryant AS, et al. The accuracy of integrated PET-CT compared with dedicated PET alone for the staging of patients with nonsmall cell lung cancer. *Ann Thorac Surg* 2004;78(3):1017-23.
8. Shim SS, Lee KS, Kim BT, et al. Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology* 2005;236:1011-1019.
9. De Wever W, Ceyskens S, Mortelmans L, et al. Additional value of PET-CT in the staging of lung cancer: comparison with CT alone, PET alone and visual correlation of PET and CT. *Eur Radiol* 2007;17:23-32.
10. Antoch G, Stattaus J, Nemat AT, et al. Non-small cell lung cancer: Dual-modality PET-CT in preoperative staging. *Radiology* 2003;229:526-533.
11. Lardinois D, Weder W, Hany TF, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med* 2003;348:2500-7.
12. Shim SS, Lee KS, Kim BT, et al. Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology* 2005;236:1011-1019.
13. Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64(2):435-48.
14. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med* 2010;37(11):2165-87.
15. Hatt M, Lee JA, Schmidlein CR, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. *Med Phys* 2017;44(6):1-42.
16. Paesmans M, Berghmans T, Dusart M, et al. European Lung Cancer Working Party, and on behalf of the IASLC Lung Cancer Staging Project. Primary tumor standardized uptake value measured on fluorodeoxyglucose positron emission tomography is of prognostic value for survival in non-small cell lung cancer: update of a systematic review and meta-analysis by the European Lung Cancer Working Party for the International Association for the Study of Lung Cancer Staging Project. *J Thorac Oncol* 2010;5(5):612-9.
17. Hatt M, Majdoub M, Vallieres M, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* 2015;56:38-44.
18. Shang J, Ling X, Zhang L, et al. Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer. *Eur J Nucl Med* 2016;43(11):1945-53.
19. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
20. Carvalho S, Leijenaar RT, Troost EGC, et al. Early variation of FDG-PET radiomics features in NSCLC is related to overall survival — the “delta radiomics” concept. *Radiother. Oncol* 2016;118:20-21.
21. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48(4):441-6.
22. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology* 2013;269:8-15.
23. Weiss GJ, Ganeshan B, Miles KA, et al. Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. *PLoS One* 2014;9(7):e100244.
24. Yamamoto S, Korn RL, Oklu R, et al. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. *Radiology* 2014;272:568-576.
25. Yip SS, Kim J, Coroller TP, Parmar C, et al. Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer. *J Nucl Med* 2017;58(4):569-576.
26. Del Gobbo A, Pellegrinelli A, Gaudio G, et al. Analysis of NSCLC tumour heterogeneity, proliferative and 18F-FDG PET indices reveals Ki67 prognostic role in adenocarcinomas. *Histopathology* 2016;68(5):746-51.
27. van Baardwijk A, Bosmans G, van Suylen RJ, et al. Correlation of intra-tumour heterogeneity on 18F-FDG PET with pathologic features in non-small cell lung cancer: a feasibility study. *Radiother Oncol* 2008;87:55-58.
28. Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* 2012;264:387-396.
29. Cook GJR, O'Brien ME, Siddique M, et al. Non-small cell lung cancer treated with erlotinib: heterogeneity of 18F-FDG uptake at PET—association with treatment response and prognosis. *Radiology* 2015;276:883-893.
30. Depeursinge A, Yanagawa M, Leung AN, et al. Predicting adenocarcinoma recurrence using computational texture models of nodule components in lung CT. *Med Phys* 2015;42:2054-2063.
31. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 2014;90:834-842.
32. Yu H, Caldwell C, Mah K, et al. Coregistered FDG PET/CT-Based Textural Characterization of Head and Neck Cancer for Radiation Treatment Planning. *IEEE Trans Med Imaging* 2009;28(3):374-383.
33. Aupérin A, Le Péchoux C, Rolland E, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small cell lung cancer. *J Clin Oncol* 2010;28:2181-90.
34. Deterbeck FC, Boffa DJ, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. *Chest* 2017;151(1):193-203.
35. Strom HH, Bremnes RM, Sundstrom SH, et al. Poor prognosis patients with inoperable locally advanced NSCLC and large tumors benefit from palliative chemoradiotherapy: a subset analysis from a randomized clinical phase III trial. *J Thorac Oncol* 2014;9:825-33.
36. Mahar AL, Compton C, McShane LM, et al, on behalf of the Molecular Modellers Working Group of the American Joint Committee on Cancer. Refining prognosis in lung cancer: A report on the quality and relevance of clinical prognostic tools. *J Thorac Oncol* 2015;10(11):1576-1589.
37. Berghmans T, Paesmans M, Sculier JP. Prognostic factors in stage III non-small cell lung cancer: a review of conventional, metabolic and new biological variables. *Ther Adv Med Oncol* 2011;3:127-138.
38. Borst GR, Belderbos JS, Boellaard R, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. *Eur J Cancer* 2005;41(11):1533-1541.
39. Dingemans AM, de Langen AJ, van den Boogaart V, et al. First-line erlotinib and bevacizumab in patients with locally advanced and/or metastatic non-small-cell lung cancer: a phase II study including molecular imaging. *Ann Oncol* 2011;22(3):559-66.
40. Mileskhin L, Hicks RJ, Hughes BG, et al. Changes in 18F-fluorodeoxyglucose and 18F-fluorodeoxythymidine positron emission tomography imaging in patients with non-small cell lung cancer treated with erlotinib. *Clin Cancer Res* 2011;17(10):3304-15.

41. Hyun SH, Ahn HK, Kim H, et al. Volume-based assessment by (18)F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer. *Eur J Nucl Med* 2014;41(1):50-8.
42. Moon SH, Cho SH, Park LC, et al. Metabolic response evaluated by 18F-FDG PET/CT as a potential screening tool in identifying a subgroup of patients with advanced non-small cell lung cancer for immediate maintenance therapy after first-line chemotherapy. *Eur J Nucl Med* 2013;40(7):1005-13.
43. Naqa IE. The role of quantitative PET in predicting cancer treatment outcomes. *Clin Transl Imaging* 2014;2:305-320.
44. Tixier F, Hatt M, Valla C, et al. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *J Nucl Med* 2014;55:1235-1241.
45. Grootjans W, Tixier F, van der Vos CS, et al. The impact of optimal respiratory gating and image noise on evaluation of intra-tumor heterogeneity in 18F-FDG positron emission tomography imaging of lung cancer. *J Nucl Med* 2016;57(11):1692-1698.
46. Salavati A, Duan F, Snyder BS, et al. Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial. *Eur J Nucl Med* 2017;44(12):1969-1983.
47. Downey RJ, Akhurst T, Gonen M, et al. Fluorine-18 fluorodeoxyglucose positron emission tomographic maximal standardized uptake value predicts survival independent of clinical but not pathologic TNM staging of resected non-small cell lung cancer. *J Thorac Cardiovasc Surg* 2007;133:1419-1427.
48. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610-621.
49. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172-179.
50. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans Biomed Eng.* 2014;61:630-637.
51. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989;19:1264-1274.
52. Cook GJR, Yip C, Siddique M, et al. Are pre-treatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J Nucl Med* 2013;54:19-26.
53. Pyka T, Bundschuh RA, Andratschke N, et al. Textural features in pre-treatment (F18)-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol* 2015;10:100.
54. Ohri N, Duan F, Snyder BS, et al. Pre-treatment FDG PET Textural Features in Locally Advanced NSCLC Secondary Analysis of ACRIN 6668/RTOG 0235. *J Nucl Med* 2016;57(6):842-8.
55. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med* 2014;55:37-42.
56. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932-945.
57. van Velden FHP, Kramer GM, Frings V, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer (18F)FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol Imaging Biol* 2016;18(5):788-795.
58. Leijenaar RT, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 2015;5:11075.
59. Desseroit MC, Tixier F, Weber WA, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort. *J Nucl Med* 2017;58(3):406-411.
60. Fave X, Mackin D, Yang J, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med Phys* 2015;42:6784-6797.
61. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS One* 2015;10(5):e0124165.
62. Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative arXiv161207003 2016.
63. Collins G, Reitsma J, Altman D, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Ann Intern Med* 2015;162:55-63.
64. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12):749-762.
65. Yang W, Zhang Y, Fu Z, et al. Imaging of proliferation with 18F FLT PET/CT versus 18F-FDG PET/CT in non-small-cell lung cancer. *Eur J Nucl Med* 2010;37(7):1291-9.
66. Bollineni VR, Kramer GM, Jansma EP, et al. A systematic review on (18F)FLT-PET uptake as a measure of treatment response in cancer patients. *Eur J Cancer* 2016;55:81-97.
67. Szyszko TA, Yip C, Szlosarek , et al. The role of new PET tracers for lung cancer. *Lung Cancer* 2016;94:7-14.
68. Everitt SJ, Ball DL, Hicks RJ, et al. Differential (18)F-FDG and (18)F-FLT Uptake on Serial PET/CT Imaging Before and During Definitive Chemoradiation for Non-Small Cell Lung Cancer. *J Nucl Med* 2014;55(7):1069-74.
69. Tian J, Yang X, Yu L, et al. A multicenter clinical trial on the diagnostic value of dual-tracer PET/CT in pulmonary lesions using 3'-deoxy-3'-18F-fluorothymidine and 18F-FDG. *J Nucl Med* 2008;49(2):186-94.
70. Crandall JP, Tahari AK, Juergens RA, et al. A comparison of FLT to FDG PET/CT in the early assessment of chemotherapy response in stages IB-IIIa resectable NSCLC. *EJNMMI Res* 2017;7:8.
71. Scheffler M, Zander T, Nogova L, et al. Prognostic Impact of (18F)fluorothymidine and (18F)fluoro-D-glucose baseline uptakes in patients with lung cancer treated first-line with erlotinib. *PLoS One* 2013;8(1):e53081.
72. Vaupel P, Mayer A. Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metastasis Rev* 2007;26(2):225-239.
73. Yi CA, Lee KS, Kim EA, et al. Solitary pulmonary nodules: dynamic enhanced multi-detector row CT study and comparison with vascular endothelial growth factor and microvessel density. *Radiology* 2004;233:191-9.
74. Hanna GG, van Sörnsen de Koste JR, Dahele MR, et al. Defining target volumes for stereotactic ablative radiotherapy of early-stage lung tumours: a comparison of three-dimensional 18F-fluorodeoxyglucose positron emission tomography and four-dimensional computed tomography. *Clin Oncol (R Coll Radiol)* 2012;24(6):71-80.
75. Wolthaus JW, Schneider C, Sonke JJ, et al. Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. *Int J Radiat Oncol Biol Phys* 2006;65(5):1560-71.
76. Wolthaus JW, Sonke JJ, van Herk M, et al. Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration. *Med Phys* 2008;35(9):3998-4011.
77. Nanni C, Zompatori M, Ambrosini V, et al. The additional diagnostic value of contemporary evaluation of FDG PET/CT scan and contrast enhanced CT imaging both acquired by a last generation PET/CT system in oncologic patients. *Biomed Pharmacother* 2013;67(2):172-8.
78. Ito R, Iwano S, Shimamoto H, et al. A comparative analysis of dual-phase dual-energy CT and FDG-PET/CT for the prediction of histopathological invasiveness of non-small cell lung cancer. *Eur J Radiol* 2017;95:186-191.
79. Iwano S, Ito R, Umakoshi H, et al. Evaluation of lung cancer by enhanced dual-energy CT: association between three-dimensional iodine concentration and tumour differentiation. *Br J Radiol* 2015;88(1055):20150224.
80. Mansoor A, Bagci U, Foster B, et al. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *Radiographics* 2015;35(4):1056-76.
81. Caldwell CB, Mah K, Ung YC, et al. Observer variation in contouring gross tumor volume in patients with poorly defined nonsmall-cell lung tumors on CT: The impact of 18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys* 2001;51:923-931.
82. Greco C, Rosenzweig K, Cascini GL, et al. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). *Lung Cancer* 2007;57:125-34.

83. Steenbakkens RJ, Duppen JC, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation. *Radiother Oncol* 2005;77:182-90.
84. Fox JL, Rengan R, O'Meara W, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys* 2005;62:70-75.
85. Konert T, Vogel WV, Everitt S, et al. Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol. *Radiother Oncol* 2016;121(1):39-45.
86. Rasch C, Belderbos J, van Giersbergen A, et al. The Influence Of A Multi-disciplinary Meeting For Quality Assurance On Target Delineation In Radiotherapy Treatment Preparation. *Int J Radiat Oncol Biol Phys* 2009;75(3):452-453.
87. Doll C, Duncker-Rohr V, Rucker G, et al. Influence of experience and qualification on PET-based target volume delineation. When there is no expert--ask your colleague. *Strahlenther Onkol* 2014;190(6):555-62.
88. Werner-Wasik M, Nelson AD, Choi W, et al. What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *Int J Radiat Oncol Biol Phys* 2012;82:1164-71.
89. Cui H, Wang X, Feng D. Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis. *Conf Proc IEEE Eng Med Biol Soc* 2012;2012:5384-7.
90. Bayne M, Hicks RJ, Everitt S, et al. Reproducibility of "intelligent" contouring of gross tumor volume in non-small-cell lung cancer on PET/CT images using a standardized visual method. *Int J Radiat Oncol Biol Phys* 2010;77(4):1151-7.
91. Foster B, Bagci U, Mansoor A, et al. A review on segmentation of positron emission tomography images. *Comput Biol Med* 2014;50:76-96.
92. MacManus M, Nestle U, Rosenzweig KE, et al. Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007. *Radiother Oncol* 2009;91(1):85-94.
93. Schinagl DA, Vogel WV, Hoffmann AL, et al. Comparison of five segmentation tools for 18F-fluoro-deoxyglucose-positron emission tomography-based target volume definition in head and neck cancer. *Int J Radiat Oncol Biol Phys* 2007;69:1282-1289.
94. Shepherd T, Teras M, Beichel R, et al. Comparative study with new accuracy metrics for target volume contouring in PET image guided radiation therapy. *IEEE Trans Med Imaging* 2012;31:2006-2024.
95. Wu K, Ung YC, Hornby J, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? *Int J Radiat Oncol Biol Phys* 2010;77(3):699-706.
96. van Loon J, Siedschlag C, Stroom J, et al. Microscopic disease extension in three dimensions for non-small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features. *Int J Radiat Oncol Biol Phys* 2012;82(1):448-56.
97. Schaefer A, Vermandel M, Baillet C, et al. Impact of consensus contours from multiple PET segmentation methods on the accuracy of functional volume delineation. *Eur J Nucl Med* 2016;43(5):911-24.
98. Weiss GJ and Korn RL. Interpretation of PET scans: do not take SUVs at face value. *J Thorac Oncol* 2012;7(12):1744-6.
99. Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res* 2017;7(1):11.
100. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44.
101. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;18:570-584.
102. Kalpathy-Cramer J, Freymann JB, Kirby JS, et al. Quantitative Imaging Network: data sharing and competitive algorithm validation leveraging the Cancer Imaging Archive. *Transl Oncol* 2014;7:147-152.
103. Aide N, Lasnon C, Veit-Haibach P, et al. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med* 2017;44(1):17-31.
104. Kruis MF, van de Kamer JB, Houweling AC, et al. PET motion compensation for radiation therapy using a CT-based mid-position motion model: methodology and clinical evaluation. *Int J Radiat Oncol Biol Phys* 2013;87(2):394-400.
105. Kruis MF, van de Kamer JB, Sonke JJ, et al. Registration accuracy and image quality of time averaged mid-position CT scans for liver SBRT. *Radiother Oncol* 2013;109(3):404-8.
106. Lamare F, Fayad H, Fernandez P, et al. Local respiratory motion correction for PET/CT imaging: Application to lung cancer. *Med Phys* 2015;42(10):5903-12.
107. Yan J, Lim JC-S, Loi HY, et al. Impact of image reconstruction settings on texture features in 18FFDG PET. *J Nucl Med* 2015;56:1667-73.
108. Cheebsumon P, van Velden FH, Yaqub M, et al. Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment. *J Nucl Med* 2011;52(10):1550-8.
109. Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics feature: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol* 2013;52(7):1391-7.
110. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 2014;9(7):e102-107.
111. Larobina M, Murino L. Medical image file formats. *J Digit Imaging* 2014;27:200-6.
112. Wu W, Parmar C, Grossmann P, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front Oncol* 2016;6:71.
113. Antoch G, Bockisch A. Combined PET/MRI: a new dimension in whole-body oncology imaging? *Eur J Nucl Med* 2009;36:s113-s120.
114. Heusch P, Buchbender C, Köhler J, et al. Thoracic staging in lung cancer: prospective comparison of 18F-FDG PET/MR imaging and 18F-FDG PET/CT. *J Nucl Med* 2014;55(3):373-8.
115. Schwenzer NF, Schraml C, Müller M, et al. Pulmonary lesion assessment: comparison of whole-body hybrid MR/PET and PET/CT imaging—pilot study. *Radiology* 2012;264(2):551-558.
116. Ohno Y, Koyama H, Yoshikawa T, et al. Three-way Comparison of Whole-Body MR, Coregistered Whole-Body FDG PET/MR, and Integrated Whole-Body FDG PET/CT Imaging: TNM and Stage Assessment Capability for Non-Small Cell Lung Cancer Patients. *Radiology* 2015;275(3):849-61.
117. Aristophanous M, Berbeco RI, Killoran JH, et al. Clinical utility of 4D FDG-PET/CT scans in radiation treatment planning. *Int J Radiat Oncol Biol Phys*. 2012;82:99-105
118. Bettinardi V, Castiglioni I, De Bernardi E, et al. PET quantification: strategies for partial volume correction. *Clin Transl Imaging* 2014;2:199-218
119. Siva S, Hardcastle N, Kron T, et al. Ventilation/Perfusion Positron Emission Tomography--Based Assessment of Radiation Injury to Lung. *Int J Radiat Oncol Biol Phys* 2015;93(2):408-17.
120. Everitt S, Ball D, Hicks RJ, et al. Prospective Study of Serial Imaging Comparing Fluorodeoxyglucose Positron Emission Tomography (PET) and Fluorothymidine PET During Radical Chemoradiation for Non-Small Cell Lung Cancer: Reduction of Detectable Proliferation Associated With Worse Survival. *Int J Radiat Oncol Biol Phys* 2017;99(4):947-955.
121. Kerner GS, Bollineni VR, Hiltermann TJ, et al. An exploratory study of volumetric analysis for assessing tumor response with (18)F-FAZA PET/CT in patients with advanced non-small-cell lung cancer (NSCLC). *EJNMMI Res* 2016;6(1):33.
122. Bahce I, Yaqub M, Smit EF, et al. Personalizing NSCLC therapy by characterizing tumors using TKI-PET and immuno-PET. *Lung Cancer* 2017;107:1-13
123. Wright CL, Binzel K, Zhang J, et al. Advanced Functional Tumor Imaging and Precision Nuclear Medicine Enabled by Digital PET Technologies. *Contrast Media Mol Imaging* 2017;2017:260305.
124. Kolbitsch C, Neji R, Fenchel M, et al. Respiratory-resolved MR-based attenuation correction for motion-compensated cardiac PET-MR. *Phys Med Biol* 2018;63(13):135008.



Chapter 3

PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014

3

Tom Konert
Wouter V. Vogel
Michael P. MacManus
Ursula Nestle
José Belderbos
Vincent Grégoire
Daniela Thorwarth
Elena Fidarova
Diana Paez
Arturo Chiti
Gerard G. Hanna

Radiotherapy and Oncology. 2015; 116(1): 27-34.

Abstract

This document describes best practice and evidence based recommendations for the use of FDG-PET/CT for the purposes of radiotherapy target volume delineation (TVD) for curative intent treatment of non-small cell lung cancer (NSCLC). These recommendations have been written by an expert advisory group, convened by the International Atomic Energy Agency (IAEA) to facilitate a Coordinated Research Project (CRP) aiming to improve the applications of PET based radiation treatment planning (RTP) in low and middle income countries. These guidelines can be applied in routine clinical practice of radiotherapy TVD, for NSCLC patients treated with concurrent chemoradiation or radiotherapy alone, where FDG is used, and where a calibrated PET camera system equipped for RTP patient positioning is available. Recommendations are provided for PET and CT image visualization and interpretation, and for tumor delineation using planning CT with and without breathing motion compensation.

Introduction

¹⁸F-fluorodeoxyglucose (FDG) positron emission tomography (PET) is recommended as a useful tool in helping staging accuracy and treatment planning [1]. FDG-PET is superior to computed tomography (CT) alone in the staging of lung cancer [2], [3]. It is now considered a routine investigation in the baseline staging evaluation of patients with non-small cell lung carcinoma (NSCLC) who are being considered for radical intent treatment [4]. When PET is acquired in conjunction with a CT (PET/CT), the combined PET/CT information has been shown to have greater staging accuracy than PET imaging alone [5], [6], [7], [8], [9], [10]. A combined PET/CT acquisition is now the standard method of acquiring FDG-PET images for the purposes of baseline staging and for radiotherapy treatment planning (RTP) [1].

The introduction of FDG-PET has been shown to have a significant impact in selecting patients for curative intent or “radical” radiotherapy [11], [12], [13], [14]. PET imaging also has been noted to reduce inter-observer variation when used to guide target volume delineation in RTP in NSCLC patients [15], [16], [17], [18]. Furthermore the acquisition of a dedicated PET/CT scan for the purposes of RTP in patients who have had a previous staging PET/CT has been shown to have further impact in reducing inter-observer variation [19].

A number of techniques have been used to generate RTP target volumes using the information gleaned from PET and CT. Most clinical studies have used a visual interpretation technique, while others have reported the use of automated segmentation techniques to either guide or generate the relevant target volume [1], [20], [21], [22], [23]. A previous International Atomic Energy Agency (IAEA) publication provided guidance on the use and role of PET/CT imaging for RTP in a range of tumor sites [24].

Methods

Following an IAEA Expert Meeting on the use of PET/CT imaging for RTP in Vienna in July 2013 it was decided to update the previous IAEA report to provide clear guidance on target volume delineation (TVD) using PET/CT imaging, specifically for the applications in lung cancer and taking advantage of the considerable research activity that has occurred since the last reports. This publication focuses entirely on the use of FDG-PET/CT in defining the target for RTP in NSCLC and seeks to update the previous guidance in light of emerging evidence and consensus opinion.

To ensure the inclusion of relevant publications the following search was undertaken. The terms “positron emission tomography”, “Non-Small Cell Lung Cancer”, “target volume delineation” and “Radiotherapy”, along with their derivatives were used to search PubMed. All studies relating to PET/CT for target volume delineation in the treatment of NSCLC with radiotherapy and of relevance to this overview were included in the preparation of the review. No limitations were placed on language or year of publication.

Background of PET based radiotherapy target volume delineation in NSCLC

The first human PET scanner was constructed in 1974, but it is only in the last 16 years that clinical studies have examined the impact of using FDG-PET for TVD in NSCLC [25], [26]. Early studies simply described the impact on the treatment volume, often without any quantification [26]. Several staging studies clearly demonstrated the superiority of PET/CT over CT for identification of involved mediastinal lymph nodes [27], [28]. PET based TVD was also shown to improve the inclusion of truly involved mediastinal lymph nodes [29]. In patients with atelectasis, it was apparent from the earliest studies that PET could help discriminate collapsed lung from tumor [30]. This approach is now widely accepted and clinically applied, although few studies have undertaken a direct validation of imaging against pathological specimens due to the difficulties with correlation and processing artefacts [31], [32], [33]. A number of studies have sought to measure the impact of FDG-PET/CT based TVD on inter-observer variation or against a ‘gold standard’. Besides the impact on staging, FDG-PET/CT imaging greatly reduces the undesirable impact of inter-observer variation [19]. In addition, PET/CT based target volume delineations are on average smaller than their non-PET counterparts, thereby reducing the dose to normal structures. This may open the possibility of dose escalation [34], [35].

Acquiring a PET/CT scan for the purposes of target volume delineation

FDG-PET/CT imaging for the purposes of baseline staging is now considered as ‘standard of care’ in patients with NSCLC being considered for treatment with radical intent [4]. Hence any patient considered for radical radiotherapy should have had a staging PET/CT. When used in its most basic form, without specific adaptations for RTP, images from a staging PET/CT scan can be visually correlated with the RTP CT to identify areas of disease for inclusion within the treatment volume. The accuracy of correlating a staging PET/CT scan and a planning CT will be improved by fusing both image sets. However, most staging scans are acquired on a curved top couch, possibly with the patient’s arms positioned down by their side, while an RTP CT is usually acquired on a flat top CT couch with arms immobilized above the patient’s head. The resulting differences in position between the two scans may make anatomical registration of these image sets difficult, leading to interpretation issues and potential inaccuracies in TVD. Therefore, these images should

not be coregistered in a treatment planning system for the purpose of tumor delineation. Deformable registration has been used in an effort to account for differences in patient positioning between imaging studies acquired in different positions. However, elastic registration is difficult due to the differences in acquisition technique between PET and CT, and deformation of PET images would compromise their inherently good representation of average tumor positions in case of breathing motion. The use of deformable registration to combine a diagnostic PET/CT scan with an RTP CT scan has not been consistently validated and at present is not recommended for this purpose [36]. In conclusion, there are no validated strategies available that would allow the use of a diagnostic PET/CT for reliable image fusion with an RTP CT.

Technically the best available option is to acquire a PET/CT scan exclusively for the purpose of RTP. This scan may be performed when a staging PET has already been acquired and the patient is deemed suitable for radical radiotherapy [37]. This approach requires two separate PET scans, which has as advantage that it removes any staging or patient selection issues. However, this approach is expensive and therefore may not be possible in all health care systems because of financial or logistical limitations. Another valid approach is to acquire a single PET/CT scan in radiotherapy treatment position to serve the dual purposes of staging and TVD. This approach avoids the costs, radiation dose and delay due to repeated imaging. It requires a complete RTP procedure on the PET/CT scanner for all potential candidates for radical irradiation despite the fact that a significant number of patients will be classified with stage IV disease after interpreting the PET images. It is imperative that the time interval between any imaging used for the purposes of radiotherapy target volume delineation and the radiotherapy treatment delivery should be as short as possible. Several studies examined the effect of radiotherapy field changes size and the effect of staging accuracy with different time scales from PET/CT scan acquisition [38], [39], [40], [41]. All of them demonstrate that PET/CT scan accuracy reduces with increasing time from the scan acquisition and that some patients may develop more advanced stage disease in the time to treatment, which will affect their chances of long-term survival. Long delays in time to treatment could result in a geographic miss if RT fields based on prior PET/CT scans no longer encompass the entire tumor or all involved lymph node stations. To avoid this issue, it is suggested that radiotherapy treatment should commence no later than 4 weeks after acquisition of the PET/CT scan.

Regardless of the selected approach, the PET/CT scanner has to be equipped with a flat RT table top, RT patient positioning devices and the CT component has to be calibrated to be safely used for RTP and RT dose calculation [42].

It is essential that where PET/CT imaging is used for TVD, that each part of the process has undergone appropriate quality assurance testing and that the entirety of this process has been validated [43]. This includes patient preparation, scan acquisition, review of the images acquired, the alignment of the CT and PET components of the PET/CT scan, transfer of the images to the radiotherapy planning system and the final display of the PET/CT images on the planning computer.

Guidance for PET/CT based visual target volume delineation

The combined procedure consisting of image interpretation, patient staging, treatment selection, and target volume definition requires many different aspects of multidisciplinary clinical expertise. It is recommended that a radiation oncologist (RO) and a nuclear medicine physician (NMP) / PET radiologist are both involved where PET is used for TVD [44], [45]. In any discussion regarding tumor volume delineation based on PET, emphasis should be given to the opinion of the NMP / PET physician in interpretation of the images and to the opinion of the RO in interpretation of all relevant clinical aspects. This also implies that the use of unmodified automated delineation of PET images for TVD is not recommended and that the final contour assessment should be made based on human visual interpretation of the images [24]. To ensure adequate and reproducible visual interpretation and application of PET images for RTP, this procedure should be standardized. The following principles are followed for visual TVD in using PET/CT imaging in NSCLC patients.

General approach of target volume delineation with PET and CT

Target volume definition involves the identification of all recognizable tumor locations, both the primary tumor and involved lymph nodes, to delineate a gross target volume (GTV) as a primary step and secondly the lymph nodes [46]. Depending on the applied strategy, this GTV may also include the full motion path of all tumor locations to create a respiration expanded GTV (reGTV). This volume, analogous to a respiration correlated GTV, contains the tumor at all times of its excursion and with suitable expansions is capable to form the basis for the clinical target volume (CTV) and the planning target volume (PTV) [47], [48]. How the combined information in the PET and CT scans contributes to the generation of a GTV or reGTV depends on the characteristics of both of the available image sets, as described here.

Where a PET/CT scan is not acquired in the RTP position

Where the PET/CT used for interpretation has not been acquired in the treatment planning position and is only visually compared with a 3D radiotherapy planning CT, PET should only be used to identify those tissues which contain tumor. The RTP CT should be used when delineating the edge of the GTV and lymph nodes [37].

The radiation oncologist should work together with the nuclear medicine physician to identify tissues that contain tumor and need to be included in the GTV.

Where a PET/CT scan is acquired in the RTP position, without respiration compensation

Standard CT for RTP is acquired during free breathing without specific measures for compensation of breathing motion, resulting in deformation and misplacement of tumor locations. In some cases a general impression of the breathing motion is identified using fluoroscopy or slow CT, but these approaches are considered insufficient for RTP procedures. Since PET is acquired during free breathing, the images are blurred according to the breathing motion and provide a good impression of the shape and average location of tumor sites [49], [50]. Therefore, in the most common scenario for RTP, where 3DCT and 3D PET scans are acquired, a reGTV approach is suggested. In this approach, the tumor should be delineated using the PET to guide both the location and the boundary of the reGTV. Where suspected disease is located outside the PET based target volume, for example on CT or based on clinical information such as positive biopsy locations, those areas should also be included in the reGTV [51]. When a margin for the CTV is added to this reGTV, an internal target volume (ITV) can be generated. Since PET has a poor resolution of 4–8 mm, it should be noted that 3D PET/CT may not fully define the ITV of highly mobile lung tumors and tumors with low FDG uptake. Hence, in the absence of 4DCT, the approach of defining a reGTV using PET should be used with caution in these circumstances [52]. To compensate for underestimation of motion in these circumstances, larger expansion margins from CTV to PTV in the superior and inferior direction should be considered.

In summary, tumor delineation is a multidisciplinary procedure. The NMP should provide the RO with information about the shape and location of tumor sites from PET imaging during delineation of the GTV or reGTV. The RO should use his/her expertise to detect suspicious tissue outside the PET based target volume and include this in the GTV or reGTV.

Where a PET/CT scan is acquired in RTP position, with breathing compensation

Where adequate respiration compensation is used in the treatment planning CT, such as with 4DCT, the CT images may provide reliable information on the shape and location of tumor sites, except in cases where there is insufficient contrast between tumor and non-tumor tissues, for example where there is atelectasis or postobstructive pneumonia (Fig. 1). With 4DCT acquisition the GTV should be based on the 4DCT, with the PET being used to discriminate tumor and non-tumor sites to adapt the GTV where appropriate. It is important that where a RTP 3D PET/CT scan is registered to a RTP 4DCT scan, the limitations of this registration are appreciated. It is suggested that PET/CT images should

be registered to the average intensity projection (Ave-IP) scan set using a rigid registration focusing on bony anatomy which is not affected by respiratory motion (e.g., spinal column). It is advised that this approach should not be used routinely for gated treatments.

Specific guidance for PET/CT based TVD

The FDG uptake of the primary tumor and any involved lymph node(s) may require evaluation with separate FDG-PET “window/level (W/L) settings.” It is important to standardize these settings, as variations in W/L settings will result in significant differences in the apparent tumor size on PET images and thus in the resulting target volumes. In addition, patients may have significant variations in biological factors, such as renal clearance of FDG, resulting in unpredictable background activity with impact on visual and quantitative strategies to discriminate tumor from physiological FDG uptake. There are no validated quantitative approaches for PET contouring that will result in ideal tumor delineation for all patients and tumor locations. However, the procedure can be standardized to some extent using visual calibration of the W/L settings, for example:

- Standardize signal intensity visually according to the biology of the patient (e.g., always start with a signal brightness of the liver (Fig. 2), vessels or other normal tissue which is familiar to the NMP/PET radiologist as normal background physiological uptake).
- Use a simple linear grayscale (e.g., black to white) for reviewing the PET images alone. For image fusion of PET with CT use a linear scale to one or at most two colors (e.g., black to red to yellow). Avoid polychromatic scales to avoid misleading color scaling contours.

Similarly, the W/L settings of CT images will influence the tumor delineation procedure. Depending on tumor localization, the appropriate CT window should be chosen. For example:

- Where the tumor is surrounded by lung tissue, lung window level settings should be used.
- For delineation of lymph nodes and where tumor invades the chest wall or mediastinum, soft tissue window settings should be used.

The nuclear medicine physician should assist the radiation oncologist in selecting standardized PET W/L settings in case of delineating the reGTV. Variations in W/L settings on PET images will result in significant differences in the apparent tumor size and thus in the resulting target volumes.

Standard delineation procedure for combined PET/CT imaging

When contouring is based on two image sets, discrepancies between the two scans may lead to uncertainty as where to draw the final contour. It is important to acknowledge these issues and to standardize solutions, in order to avoid observer variations and potential geographic miss.

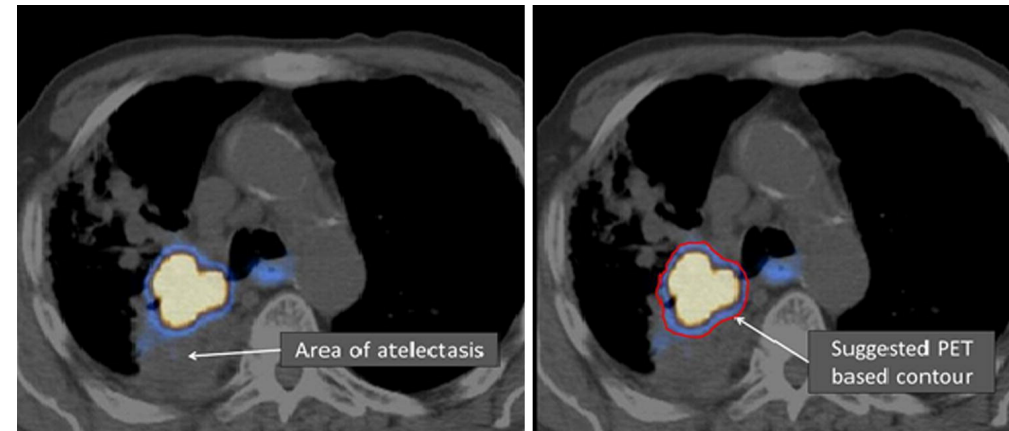


Fig. 1. Area of atelectasis in the right upper lobe. CT images show insufficient contrast between tumor and non-tumor tissue where atelectasis is present, therefore delineation should be defined by PET FDG avid areas.

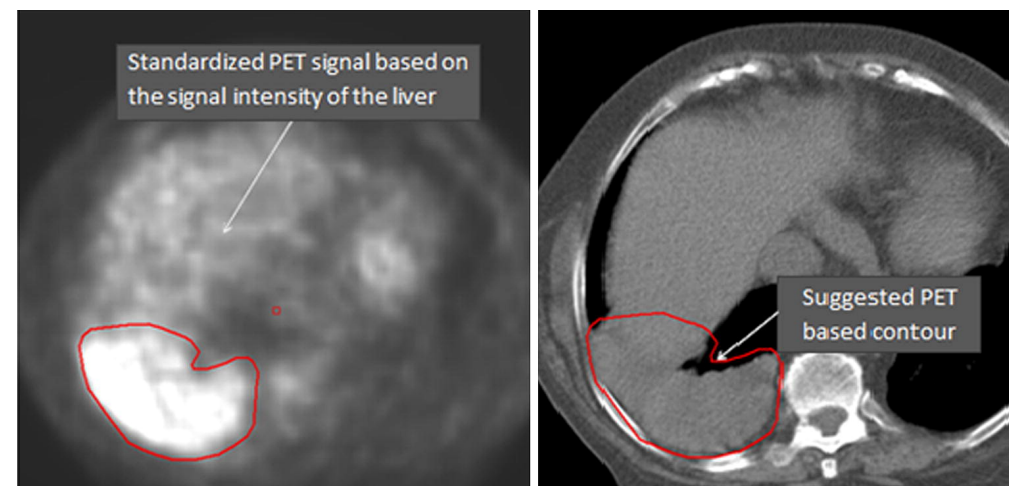


Fig. 2. Example of a standardized signal intensity based on the signal brightness of the liver. The PET signal is very intense in the right lower lobe. When the tumor is contiguous with a non-tumor structure that has a similar density and where no tumor boundary can be distinguished on CT (e.g., when the tumor is adjacent to the liver), the reGTV should be defined by the PET FDG avid areas.

An important question is whether the GTV (or reGTV) may contain areas where PET is positive for tumor but CT shows normal lung tissue. When delineating a reGTV based on PET (e.g., when using 3DCT), all tumor locations should be defined primarily by the FDG avid signal including their full motion paths, and this may include areas at the surface where there is no tumor apparent on the non-respiration correlated CT (Fig. 3). However, when delineating a GTV based on CT (e.g., when using 4DCT), the primary tumor should be defined primarily by the structures as seen on CT and therefore not include air. However, as the PET scan may reveal a so-called “baseline shift” i.e., the change of the basic position of the tumor, in the case of perfect coregistration of bony structures, a clear deviation of the PET signal from the 4DCT created ITV should not be disregarded and might be used to further expand the ITV.

Another common issue is the distinction between tumor and adjacent soft tissues. In areas where the tumor is contiguous with a structure that has a similar density and where no tumor boundary can be distinguished on CT (e.g., in the presence of extrathoracic extensions, see Fig. 4), the reGTV should be defined by the FDG avid areas [30]. Where the PET scan is compromised by FDG uptake that is apparently not explained by tumor (e.g., physiological uptake in the heart, or active infection), the reGTV should be defined by the CT images.

For lymph nodes the same approaches for GTV and reGTV can be applied as for a primary tumor. An additional issue is the identification of the lymph nodes that need to be included in the delineation. A pathological lymph node is defined as a lymph node which is involved on FDG-PET in the opinion of a trained NMP/PET radiologist. Non FDG avid (negative) nodes that appear enlarged on CT and that have a low likelihood of containing macroscopic tumor, do not need to be included in the GTV under certain circumstances [34]. PET negative nodes may be included in the final GTV volume based on information obtained through bronchoscopy, mediastinoscopy, endoscopic ultrasound sampling (EUS), and endobronchial ultrasound sampling (EBUS). Any biopsy proven lymph node should be included in the GTV. Recently, an update of guidelines for preoperative mediastinal lymph node staging is published, recommending EBUS/EUS with fine needle aspiration as the first choice [53]. If EBUS/EUS findings are negative and if uncertainty regarding the involvement of mediastinal lymph nodes remains, video-assisted mediastinoscopy is preferred over mediastinoscopy as the next most appropriate staging procedure. It should be noted that patients with lymph nodes measuring >16 mm on CT and a negative FDG-PET result should undergo mediastinoscopy before possible thoracotomy [54]. Combination of endoscopic staging and surgical staging results in the highest accuracy [55]. In addition, clinical considerations may contribute to identification of suspect lymph nodes, e.g., small FDG-negative lymph nodes that are directly adjacent to the tumor or are located between

other evidently pathological nodes or those which show progression (tumor growth) as determined from multiple (low dose) CT scans over a certain period of time.

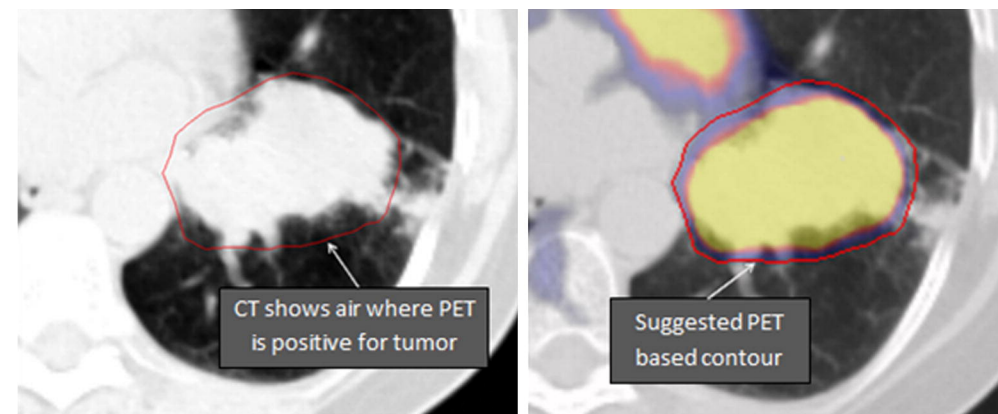


Fig. 3. Example of a delineation of the reGTV based on PET in combination with a 3DCT. A primary tumor can be seen in the left upper lobe with high FDG uptake. The PET signal shows FDG uptake in air when the modalities are combined which is due to the movement of the tumor.

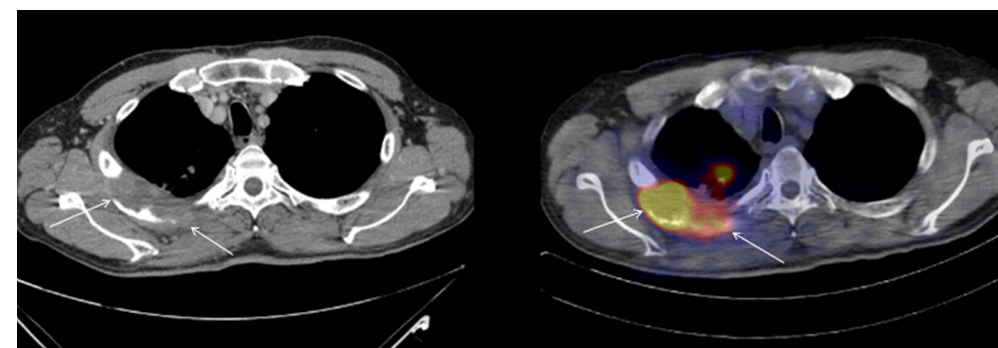


Fig. 4. Example of a histopathologically proven primary tumor with high FDG uptake in the apical segment of the right upper lobe with a diameter of more than 6 cm, and the adjacent pleural mass with extrathoracic growth and destruction of the dorsal third to sixth rib on the right. Difficulties arise when distinguishing between normal tissue and tumor in the thoracic wall using CT alone. In this case PET should be used to define the boundary of the tumor.

Automated delineation methods for PET/CT imaging

As discussed earlier the one source of error and potential miss in TVD is the accuracy of delineation of contours by the oncologist [56]. Given the nature of PET images a number of investigations have examined the use of automated methods to define the edge of the tumor [57], [58]. Auto contours may provide consistent contours, but have difficulty dealing with

normal tissue adjacent to the tumor with high SUV uptake such as the heart. There is also no clear consensus on which method most closely approximates to the tumor position and tumor edge and pathological correlation has proven difficult [59]. Another difficulty with PET based auto-contouring is the variability of SUV values due to factors other than tumor activity such as patient biological factors and technical factors [60]. When delineating a reGTV to include the full motion path of all tumor locations, the value of auto contours is without any supporting evidence. Furthermore the information obtained from the PET component of the scan is complementary to that contained within the CT scan and the use of information from both may lead to more successful auto-contouring [49]. Automated PET based contours can be useful as a starting point for PET based TVD and are worthy of further investigation, particularly in the era of 4D PET/CT imaging. At present the IAEA panel recommendation remains that, outside of a clinical trial context, target volumes generated with the use of PET should be delineated using visual interpretation alone or should be visually edited following any automated target volume delineation [24].

The use of PET to define a respiration expanded GTV for mobile lung tumors

As PET images are acquired over a number of minutes at each table position, it has been suggested that PET could define the entire motion trajectory of a lung tumor. Tumors identified as low risk of macroscopic disease extension (MDE) show lower rates of disease around the GTV than do high-risk tumors. Both PET and CT accurately visualize the CTV(path) in low-risk tumors, but underestimate MDE in high-risk tumors [59]. When a suitable margin is added for microscopic extensions of a moving lung tumor, this volume is also known as the internal target volume (ITV). According to ICRU 62, the ITV is defined as the “internal margin” plus a CTV expansion [47]. Phantom studies have demonstrated that the PET target volume may contain all of the respiratory motion (respiration expanded GTV) of a moving lung tumor [50], [61]. Hence, in the absence of other respiration motion compensation techniques (e.g., respiratory gating or real-time tracking), PET based target volumes may be used to approach a PET based ITV approach, namely the respiration expanded GTV. Unfortunately clinical investigations using 3D PET/CT imaging have not shown consistently that a PET based ITV is identical to a 4DCT based ITV for small tumors [52]. A recent study using 4D PET/CT imaging has demonstrated that a 4D PET based ITV closely approximates to a 4DCT [62]. Using 4D PET/CT imaging may lead to better quantification of tumor motion during prolonged radiotherapy treatment times but further investigation and clinical validation are required.

Four dimensional PET/CT imaging

With a 4D PET/CT the 4DCT and 4D PET scan are retrospectively binned into a number of respiratory phases correlated with the breathing cycle using a respiratory tracking system. Each phase of the 4D PET is corrected for attenuation with the respective phase of the 4DCT.

4D PET/CT imaging may overcome some of the inaccuracies associated with a free breathing PET/CT scan. One such factor is tumor motion. In 3D PET/CT imaging, the CT component is acquired as fast CT which may catch a mobile lung tumor at an extreme of the ITV or cause artefacts, while PET scans are acquired over a number of minutes. Hence the SUV for a given pixel is an average of the SUV over that time period. Furthermore, in integrated PET/CT acquisition the attenuation correction is based on the CT data and as mentioned above this may misrepresent the average density for a given pixel position. Hence, in essence, for mobile lung tumors the PET component is more akin to 4D imaging while the CT component is more akin to a 3D imaging technique. A number of studies have shown sizeable differences in SUV calculation between 3D PET/CT and 4D PET/CT imaging [63], [64]. Using 4D PET/CT imaging may provide more accurate SUV quantification for moving lung cancer and has implications for auto-contouring which may lead to new methods of PET based TVD.

CTV and PTV expansions to a PET derived GTV

These guidelines have focused on the delineation of standardized GTV or reGTV contours. Subsequently, these volumes need to be expanded to a CTV and to a PTV. The CTV expansion is based on pathological tumor characteristics, and therefore not dependent on the imaging technique or GTV delineation strategy. Clinically applied expansions from the GTV to CTV are generally in the range of 5–8 mm [65]. The PTV expansion can be calculated in a probabilistic approach by considering all systematic uncertainties, all random uncertainties and also the width of treatment beam penumbra [66]. This may also include patient characteristics such as breathing motion, if this has not already been incorporated in the GTV definition. A more basic approach can be used when a PET based respiration expanded GTV has been created, and the PTV expansion is applied to compensate for setup variations alone (e.g., 1 cm in all directions). When a GTV based on 4DCT has been created, PTV expansions are primarily based on the characteristics derived from CT and other systematic errors (e.g., errors from image fusion and target volume delineation) of which all can be taken into account with the van Herk formula [66].

PET combined with MR imaging

When envisioning the future, it is interesting to follow the recent advances in hybrid imaging systems which made it possible to combine PET and (functional) MR information. With the combination of functional MR and PET information new possibilities in functional cancer imaging are emerging [67]. However, one of the first few studies reported no significant difference in diagnostic performance when PET/MR was compared to PET/CT [68], [69]. Interestingly, a recent study showed that PET/MR imaging could have an advantage in lymph node detection [70]. No study yet exists about the use of PET/MR in tumor volume delineation but further research is awaited.

Conclusions

It remains the recommendation of an IAEA expert panel that an appropriately timed and technically adequate PET/CT imaging is an essential component in the radiotherapy treatment planning process for lung cancer. Specific guidance regarding the interpretation of PET/CT imaging for TVD is provided. It is also recognized that further research is needed in the fields of 4D PET/CT imaging and automated TVD techniques.

References

1. Ung JC, Bezjak A, Coakley N, Evans WK and the Lung Cancer Disease Site Group. *Positron Emission Tomography with ¹⁸Fluorodeoxyglucose in Radiation Treatment Planning for Non-small Cell Lung Cancer*. *J Thorac Oncol*. 2011; 6: 86–97.
2. Gould MK, Kuschner WG, Rydzak CE, et al. *Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: A meta-analysis*. *Ann Intern Med*. 2003; 139: 879–892.
3. MacManus MP, Hicks RJ, Matthews JP, et al. *High rate of detection of unsuspected distant metastases by pet in apparent stage III non-small-cell lung cancer: Implications for radical radiation therapy*. *Int J Radiat Oncol Biol Phys*. 2001; 50: 287–293.
4. *NICE Clinical Guideline: Lung cancer: The diagnosis and treatment of lung cancer*. National Institute for Health and Care Excellence, April 2011. Available at <http://www.nice.org.uk/guidance/CG121/chapter/introduction> accessed 17th July 2014.
5. Cerfolio RJ, Ojha B, Bryant AS, Raghuvver V, Mountz JM, Bartolucci AA. *The accuracy of integrated PET-CT compared with dedicated PET alone for the staging of patients with nonsmall cell lung cancer*. *Ann Thorac Surg*. 2004; 78(3): 1017–23.
6. De Wever W, Ceyskens S, Mortelmans L, et al. *Additional value of PET-CT in the staging of lung cancer: comparison with CT alone, PET alone and visual correlation of PET and CT*. *Eur Radiol*. 2007; 17: 23–32.
7. Sung SS, Kyung SL, Byung-Tae K, et al. *Non-small cell lung cancer: prospective comparison of integrated FDG PET-CT and CT alone for preoperative staging*. *Radiology*. 2005; 236: 1011–19.
8. Antoch G, Stattaus J, Nemat AT, et al. *Non-small cell lung cancer: Dual-modality PET-CT in preoperative staging*. *Radiology*. 2003; 229: 526–533.
9. Lardinois D, Weder W, Hany TF, et al. *Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography*. *N Engl J Med*. 2003; 348: 2500–7.
10. Shim SS, Lee KS, Kim BT, Chung MJ, Lee EJ, Han J, Choi JY, Kwon OJ, Shim YM, Kim S. *Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging*. *Radiology*. 2005; 236: 1011–1019.
11. Mac Manus MP, Wong K, Hicks RJ, Matthews JP, Wirth A and Ball DL. *Early mortality after radical radiotherapy for non-small-cell lung cancer: comparison of PET-staged and conventionally staged cohorts treated at a large tertiary referral center*. *Int J Radiat Oncol Biol Phys*. 2002; 52(2): 351–361.
12. Fischer B, Lassen U, Mortensen J, Larsen S, Loft A, Bertelsen A, et al. *Preoperative staging of lung cancer with combined PET-CT*. *N Engl J Med*. 2009; 361(1): 32–9.
13. Hillner BE, Siegel BA, Lui D, et al. *Impact of positron emission tomography (PET) alone on expected management of patients with cancer: Initial results from the National Oncologic PET Registry*. *J Clin Oncol*. 2008; 26: 2155–2161.
14. Gregory DL, Hicks RJ, Hogg A, Binns DS, Shum PL, Milner A, Mac Manus MP. *Effect of PET/CT on Management of Patients with Non-Small Cell Lung Cancer: Results of a Prospective Study with 5-Year Survival Data*. *Journal of Nuclear Medicine*. 2012; 53(7): 1007–1015.
15. Caldwell CB, Mah K, Ung YC, et al. *Observer variation in contouring gross tumor volume in patients with poorly defined nonsmall-cell lung tumors on CT: The impact of 18FDG-hybrid PET fusion*. *Int J Radiat Oncol Biol Phys*. 2001; 51: 923–931.
16. Greco C, Rosenzweig K, Cascini GL, Tamburrini O. *Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC)*. *Lung Cancer*. 2007; 57: 125–34.
17. Steenbakkers RJ, Duppen JC, Fitton I, et al. *Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation*. *Radiother Oncol*. 2005; 77: 182–90.
18. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, Leibel SA, Rosenzweig KE. *Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer?* *Int J Radiat Oncol Biol Phys*. 2005; 62: 70–75.

19. Hanna GG, McAleese J, Carson KJ, et al. (18)F-FDG PET-CT simulation for non-small-cell lung cancer: effect in patients already staged by PET-CT. *Int J Radiat Oncol Biol Phys.* 2010; 77(1): 24-30.
20. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging.* 2010; 37(11): 2165-87.
21. Werner-Wasik M, Nelson AD, Choi W, et al. What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *Int J Radiat Oncol Biol Phys.* 2012; 82: 1164-71.
22. Cui H, Wang X, Feng D. Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis. Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, Calif, USA. 2012; 5384-5387.
23. Bayne M, Hicks RJ, Everitt S, Fimmell N, Ball D, Reynolds J, Lau E, Pitman A, Ware R, MacManus M. Reproducibility of "intelligent" contouring of gross tumor volume in non-small-cell lung cancer on PET/CT images using a standardized visual method. *Int J Radiat Oncol Biol Phys.* 2010; 77(4): 1151-7.
24. MacManus M, Nestle U, Rosenzweig KE, et al. Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007. *Radiother Oncol.* 2009; 91(1): 85-94.
25. Ter-Pogossian MM, Phelps ME, Hoffman EJ, Mullani NA. A positron-emission transaxial tomograph for nuclear imaging (PETT). *Radiology.* 1975; 114(1): 89-98.
26. Kiffer JD, Berlangieri SU, Scott AM, et al. The contribution of 18F-fluoro-2-deoxy-glucose positron emission tomographic imaging to radiotherapy planning in lung cancer. *Lung Cancer.* 1998; 19(3): 167-77.
27. De Ruyscher D, Wanders S, van Haren E, Hochstenbag M, Geeraedts W, Utama I, Simons J, Dohmen J, Rhami A, Buell U, Thimister P, Snoep G, Boersma L, Verschueren T, van Baardwijk A, Minken A, Bentzen SM, Lambin P. Selective mediastinal node irradiation based on FDG-PET scan data in patients with non-small-cell lung cancer: a prospective clinical study. *Int J Radiat Oncol Biol Phys.* 2005; 62(4): 988-94.
28. Belderbos JSA, Heemsbergen WD, De Jaeger K, Baas P, Lebesque JV. Final results of a phase I/II dose escalation trial in non-small cell lung cancer using three-dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys.* 2006; 66(1): 126-34.
29. Bradley J, Thorstad WL, Mutic S, et al. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys.* 2004; 59(1): 78-86.
30. Nestle U, Walter K, Schmidt S, et al. 18F-deoxyglucose positron emission tomography (FDG-PET) for the planning of radiotherapy in lung cancer: high impact in patients with atelectasis. *Int J Radiat Oncol Biol Phys.* 1999; 44(3): 593-7.
31. Faria SL, Menard S, Devic S, et al. Impact of FDG-PET/CT on radiotherapy volume delineation in non-small-cell lung cancer and correlation of imaging stage with pathologic findings. *Int J Radiat Oncol Biol Phys.* 2008; 70(4): 1035-8.
32. Wanet M, Lee JA, Weynand B, De Bast M, Poncelet A, Lacroix V, Coche E, Grégoire V, Geets X. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. *Radiother Oncol.* 2011; 98(1): 117-25.
33. Hellwig D, Baum RP, Kirsch C. FDG-PET, PET/CT and conventional nuclear medicine procedures in the evaluation of lung cancer: a systematic review. *Nuklearmedizin.* 2009; 48(2): 59-69.
34. De Ruyscher D, Wanders S, Minken A, et al. Effects of radiotherapy planning with a dedicated combined PET-CT-simulator of patients with non-small cell lung cancer on dose limiting normal tissues and radiation dose-escalation: a planning study. *Radiother Oncol.* 2005; 77: 5-10.
35. van Elmpt W, De Ruyscher D, van der Salm A, Lakeman A, van der Stoep J, Emans D, Damen E, Öllers M, Sonke JJ, Belderbos J. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. *Radiother Oncol.* 2012; 104(1): 67-71.
36. Hanna GG, Van Sörnsen De Koste JR, Carson KJ, O'Sullivan JM, Hounsell AR and Senan S. Conventional 3D staging PET/CT in CT simulation for lung cancer: impact of rigid and deformable target volume alignments for radiotherapy treatment planning. *Br J Radiol.* 2011; 84(1006): 919-29.
37. Jarritt PH, Hounsell AR, Carson KJ et al. Use of combined PET/CT images for radiotherapy planning: initial experiences in lung cancer. *Br J Radiol suppl* 2005; 28: 33-40.
38. Everitt S, Plumridge N, Herschtal A, Bressel M, Ball D, Callahan J, Kron T, Schneider-Kolsky M, Binns D, Hicks RJ, Mac Manus M. The impact of time between staging PET/CT and definitive chemo-radiation on target volumes and survival in patients with non-small cell lung cancer. *Radiother Oncol.* 2013; 106(3): 288-91.
39. Geiger GA, Kim MB, Xanthopoulos EP, Pryma DA, Grover S, Plastaras JP, Langer CJ, Simone CB 2nd, Rengan R. Stage migration in planning PET/CT scans in patients due to receive radiotherapy for non-small-cell lung cancer. *Clin Lung Cancer.* 2014; 15(1): 79-85.
40. Booth K, Hanna GG, McGonigle N, McManus KG, McGuigan J, O'Sullivan J, Lynch T, McAleese J. The mediastinal staging accuracy of 18F-Fluorodeoxyglucose positron emission tomography/computed tomography in non-small cell lung cancer with variable time intervals to surgery. *Ulster Med J.* 2013; 82(2): 75-81.
41. Belderbos J, Hope A, Kestin L, Guckenberger M, Werner-Wasik M, Sonke J, Bissonnette J, Xiao Y, Yan D, Grills I. Time Interval between Staging FDG Positron Emission Tomography (PET) and Initiation of Stereotactic Lung Radiotherapy (SBRT) Impacts the Risk of Recurrence and Metastasis in Non-Small Cell Lung Cancer (NSCLC).
42. Thorwarth D, Beyer T, Boellaard R, de Ruyscher D, Grgic A, Lee JA, Pietrzyk U, Sattler B, Schaefer A, van Elmpt W, Vogel W, Oyen WJ, Nestle U. Integration of FDG-PET/CT into external beam radiation therapy planning: technical aspects and recommendations on methodological approaches. *Nuklearmedizin.* 2012; 51(4): 140-53.
43. Boellaard R, O'Doherty MJ, Krause BJ, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging.* 2010; 37(1): 181-200.
44. Hanna GG, Carson KJ, Lynch T, McAleese J, Cosgrove VP, Eakin RL, Stewart DP, Zafari A, O'Sullivan JM, Hounsell AR. 18F-fluorodeoxyglucose positron emission tomography/computed tomography-based radiotherapy target volume definition in non-small-cell lung cancer: delineation by radiation oncologists vs. joint outlining with a PET radiologist? *Int J Radiat Oncol Biol Phys.* 2010; 78(4): 1040-51.
45. Doll C, Duncker-Rohr V, Rücker G, Mix M, MacManus M, De Ruyscher D, Vogel W, Eriksen JG, Oyen W, Grosu AL, Weber W, Nestle U. Influence of experience and qualification on PET-based target volume delineation. When there is no expert--ask your colleague. *Strahlenther Onkol.* 2014; 190(6): 555-62.
46. ICRU (1993) ICRU report vol. 50. Bethesda: International Commission on Radiation Units and Measurements. *Prescribing, Recording and Reporting Photon Beam Therapy.*
47. ICRU (1999) ICRU report vol. 62. Bethesda: International Commission on Radiation Units and Measurements. *Prescribing, Recording and Reporting Photon Beam Therapy.*
48. Grégoire V, Mackie R. ICRU (2010) ICRU report Vol. 83. Bethesda: International Commission on Radiation Units and Measurements. *Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT).*
49. Wu K, Ung YC, Hornby J, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? *Int J Radiat Oncol Biol Phys.* 2010; 77(3): 699-706.
50. Okubo M, Nishimura Y, Nakamatsu K, Okumura M, Shibata T, Kanamori S, Hanaoka K, Hosono M. Static and moving phantom studies for radiation treatment planning in a positron emission tomography and computed tomography (PET/CT) system. *Ann Nucl Med.* 2008; 22(7): 579-86.
51. MacManus MP and Hicks RJ. Where do we draw the line? Contouring tumors on positron emission tomography/computed tomography. *Int J Radiat Oncol Biol Phys.* 2008; 71(1): 2-4.
52. Hanna GG, van Sörnsen de Koste JR, Dahele MR, et al. Defining target volumes for stereotactic ablative radiotherapy of early-stage lung tumours: a comparison of three-dimensional 18F-fluorodeoxyglucose positron emission tomography and four-dimensional computed tomography. *Clin Oncol (R Coll Radiol).* 2012; 24(6): 71-80.
53. De Leyn P, Dooms C, Kuzdzal J, Lardinois D, Passlick B, Rami-Porta R, Turna A, Van Schil P, Venuta F, Waller D, Weder W, Zielinski M. Preoperative mediastinal lymph node staging for non-small cell lung cancer: 2014 update of the 2007 ESTS guidelines. *Transl Lung Cancer Res.* 2014; 3(4): 225-233.
54. de Langen AJ, Raijmakers P, Riphagen I, Paul MA, Hoekstra OS. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. *Eur J Cardiothorac Surg.* 2006; 29(1): 26-9.
55. Annema JT, van Meerbeeck JP, Rintoul RC, Dooms C, Deschepper E, Dekkers OM, De Leyn P, Braun J, Carroll NR, Praet M, de Ryck F, Vansteenkiste J, Vermassen F, Versteegh MI, Veselić M, Nicholson AG, Rabe KF, Tournoy KG. Mediastinoscopy vs endosonography for mediastinal nodal staging of lung cancer: a randomized trial. *JAMA.* 2010; 304(20): 2245-52.

56. Rasch C, Belderbos J, van Giersbergen A, De Kok I, Laura T, Boer M, de Boer P, Gilles R, Teertstra J, Verheij M. *The Influence Of A Multi-disciplinary Meeting For Quality Assurance On Target Delineation In Radiotherapy Treatment Preparation International Journal of Radiation Oncology.* 2009; 75(3): 452–453.
57. Schinagl D. A., Vogel W. V., Hoffmann A. L., van Dalen J. A., Oyen W. J., and Kaanders J. H. *Comparison of five segmentation tools for 18F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer.* Int J Radiat Oncol Biol. Phys. 2007; 69: 1282–1289.
58. Shepherd T, Teras M, Beichel R *et al.* *Comparative study with new accuracy metrics for target volume contouring in PET image guided radiation therapy.* IEEE Trans Med Imaging. 2012; 31: 2006–2024.
59. van Loon J, Siedschlag C, Stroom J, Blauwgeers H, van Suylen RJ, Kneijens J, Rossi M, van Baardwijk A, Boersma L, Klomp H, Vogel W, Burgers S, Gilhuijs K. *Microscopic disease extension in three dimensions for non-small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features.* Int J Radiat Oncol Biol Phys. 2012; 82(1): 448–56.
60. Weiss GJ and Korn RL. *Interpretation of PET scans: do not take SUVs at face value.* J Thorac Oncol. 2012; 7(12): 1744–6.
61. Caldwell CB, Mah K, Skinner M, and Danjoux CE. *Can PET provide the 3D extent of tumor motion for individualized internal target volumes? A phantom study of the limitations of CT and the promise of PET.* Int J Radiat Oncol Biol Phys. 2003; 55: 1381–1393.
62. Callahan J, Kron T, Schneider-Kolsky M, *et al.* *Validation of a 4D-PET maximum intensity projection for delineation of an internal target volume.* Int J Radiat Oncol Biol Phys. 2013; 86(4): 749–54.
63. Aristophanous M, Berbeco RI, Killoran JH, *et al.* *Clinical utility of 4D FDG-PET/CT scans in radiation treatment planning.* Int J Radiat Oncol Biol Phys. 2012; 82(1): 99–105.
64. Callahan J, Kron T, Schneider ME, Hicks RJ. *A prospective investigation into the clinical impact of 4D-PET/CT in the characterisation of solitary pulmonary nodules.* Cancer Imaging. 2014 Jun 5;14(1):24.
65. Giraud P, Antoine M, Larrouy A, Milleron B, Callard P, De Rycke Y, Carette MF, Rosenwald JC, Cosset JM, Housset M, Touboul E. *Evaluation of microscopic tumor extension in non-small-cell lung cancer for three-dimensional conformal radiotherapy planning.* Int J Radiat Oncol Biol Phys. 2000; 48(4): 1015–24.
66. van Herk M. *Errors and margins in radiotherapy.* Semin Radiat Oncol. 2004 Jan;14(1):52–64.
67. Antoch G, Bockisch A. *Combined PET/MRI: a new dimension in whole-body oncology imaging?* Eur J Nucl Med Mol Imaging. 2009;36:S113–S120.
68. Heusch P, Buchbender C, Köhler J, Nensa F, Gauler T, Gomez B, Reis H, Stamatis G, Köhl H, Hartung V, Heusner TA. *Thoracic staging in lung cancer: prospective comparison of 18F-FDG PET/MR imaging and 18F-FDG PET/CT.* J Nucl Med. 2014;55(3):373–8.
69. Schwenzer NF, Schraml C, Müller M, *et al.* *Pulmonary lesion assessment: comparison of whole-body hybrid MR/PET and PET/CT imaging—pilot study.* Radiology 2012;264(2):551–558.
70. Ohno Y, Koyama H, Yoshikawa T, Takenaka D, Seki S, Yui M, Yamagata H, Aoyagi K, Matsumoto S, Sugimura K. *Three-way Comparison of Whole-Body MR, Coregistered Whole-Body FDG PET/MR, and Integrated Whole-Body FDG PET/CT Imaging: TNM and Stage Assessment Capability for Non-Small Cell Lung Cancer Patients.* Radiology 2015;275(3):849–61.

Summary of recommendations for PET based Radiation Treatment Planning (RTP)

• Applicability of the guidelines

- These guidelines apply to patients with NSCLC. The process of TVD is different for other tumor entities; these recommendations should not be applied to e.g. the head/neck or abdominal areas.
- The presented strategies apply to patients planned for curative intent treatment with concurrent chemoradiation or with radiotherapy alone. If neo-adjuvant (sequential) chemotherapy has been applied before radiotherapy, it is not possible to acquire a timely and reliable PET scan for RTP purposes.
- These guidelines apply to FDG avid tumors and lymph nodes. For tumors with no FDG uptake current locally accepted standard CT scan based procedures should be applied.
- This document discusses the use of FDG-PET. There currently is no definitive evidence for the use of other PET tracers in routine TVD.
- The guidelines assume a properly calibrated PET/CT system (according to the EANM Research 4 Life (EARL) accreditation specifications) with capabilities for RTP positioning and adequate delineation software capable of image fusion of PET and planning CT.

• Image acquisition recommendations

- These guidelines apply to PET/CT scans which are acquired in radiotherapy treatment position. An RTP PET/CT is usually acquired on flat top CT couch with arms immobilized above the patient's head. Any differences in patient positioning between two scans make anatomical registration of these image sets difficult, leading to interpretation issues and potential inaccuracies in TVD.
- Image fusion of a staging PET/CT scan and a planning CT is only recommended when both scans are acquired in radiotherapy treatment position.

• General visualization recommendations

- CT window / level: Standardize according to the anatomical location. Use a soft tissue setting for tumor parts adjacent to mediastinum and thoracic wall, and lung window setting for tumor parts adjacent to lung tissue.
- PET window / level: Standardize visually according to the biology of the patient (e.g. always start with a reproducible signal brightness for normal structures such as the liver or mediastinal structures e.g. great vessels).

- PET visualization: Use a simple linear grayscale (e.g. black to white) for reviewing. For image fusion of PET with CT use a linear scale to one or at most two colors (e.g. black to red to yellow). Avoid polychromatic scales to avoid misleading color scaling contours.
- **Tumor delineation approach for 3D free breathing CT**
 - In this case PET provides the best estimation of the location, shape and motion path of tumor lesions. Contour a GTV around all suspect areas of the PET signal, to create a respiration expanded GTV (reGTV) that includes the tumor locations in all parts of the respiratory cycle.
 - Adapt the PET based contour according to CT and clinical data as appropriate (e.g. include FDG-negative tumor parts in case of suspicion on CT and positive biopsy areas).
- **Tumor delineation for motion compensated or 4DCT**
 - 4DCT can provide an adequate estimation of the location and motion path of tumor. Contour a GTV using all CT information.
 - Adapt the CT based GTV contour according to PET and clinical data as appropriate (e.g. exclude FDG-negative areas of atelectasis or include positive biopsy areas).
- **Tumor identification and inclusion in GTV**
 - Include all tissues considered having a high probability for tumor involvement, either on CT or PET, to avoid geographic miss.
 - Include all FDG avid lymph nodes that are considered positive in the opinion of a nuclear medicine physician / PET radiologist. This may include small lymph nodes with relatively low FDG avidity and may exclude enlarged lymph nodes that are not FDG avid.
 - Include non FDG avid lymph nodes after cytological/histological confirmation of lymphadenopathy.
- **Derivation of the CTV and PTV**
 - CTV expansions do not depend on PET images, and can remain unchanged.
 - PTV expansion depends on the GTV delineation strategy. When a reGTV was created using PET and 3DCT, breathing motion has been largely included within the reGTV and the PTV expansion for respiratory motion may be limited to reflect this (e.g. 1 cm in all directions). When a GTV has been created using a motion compensated CT

(4DCT) and for an un-gated treatment delivery, expansions from CTV and PTV do not need to account for respiratory motion.

• **Other recommendations**

- Complex cases of GTV delineation in lung cancer patients should always be discussed in a multi-disciplinary quality control meeting.
- The time interval between staging procedure and start of treatment delivery should be as short as possible, preferably not more than 4 weeks.
- EUS/EBUS with needle aspiration should be used as a first choice in primary mediastinal staging if local expertise in these minimal invasive techniques is available. Video assisted mediastinoscopy is used as a second choice, in case EUS/EBUS findings are negative.



Chapter 4

Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol

4

Tom Konert
Wouter V. Vogel
Sarah Everitt
Michael P. MacManus
Daniela Thorwarth
Elena Fidarova
Diana Paez
Jan-Jakob Sonke
Gerard G. Hanna

Radiotherapy and Oncology. 2016; 121(1): 39-45

Abstract

Background and purpose

To assess the impact of a standardized delineation protocol and training interventions on PET/CT-based target volume delineation (TVD) in NSCLC in a multicenter setting.

Material and methods

Over a one-year period, 11 pairs, comprised each of a radiation oncologist and nuclear medicine physician with limited experience in PET/CT-based TVD for NSCLC from nine different countries took part in a training program through an International Atomic Energy Agency (IAEA) study (NCT02247713). Teams delineated gross tumor volume of the primary tumor, during and after training interventions, according to a provided delineation protocol. In-house developed software recorded the performed delineations, to allow visual inspection of strategies and to assess delineation accuracy.

Results

Following the first training, overall concordance indices for 3 repetitive cases increased from 0.57 ± 0.07 to 0.66 ± 0.07 . The overall mean surface distance between observer and expert contours decreased from -0.40 ± 0.03 cm to -0.01 ± 0.33 cm. After further training overall concordance indices for another 3 repetitive cases further increased from 0.64 ± 0.06 to 0.80 ± 0.05 ($p=0.01$). Mean surface distances decreased from -0.34 ± 0.16 cm to -0.05 ± 0.20 cm ($p=0.01$).

Conclusion

Multiple training interventions improve PET/CT-based TVD delineation accuracy in NSCLC and reduces interobserver variation.

Introduction

Lung cancer is the most common cause of death from cancer worldwide, estimated to be responsible for nearly 17% of the total [1] and it is estimated that more than 80% of patients in low and middle income countries are diagnosed with lung cancer in an advanced stage (III and IV) [2, 3]. The use of fused ^{18}F -fluorodeoxyglucose Positron Emission Tomography/Computed Tomography (FDG-PET/CT) imaging is now the standard method of acquiring FDG-PET images for the purpose of baseline staging and RT treatment preparation [4], since it has been shown to be superior to either PET or CT alone [5, 6]. The number of PET/CT scanners has increased in low and middle income countries in the last decade [7] and additional training in the use of PET/CT in radiotherapy planning (RTP) is vital to ensure appropriate interpretation of PET/CT with the hope, that the use of PET/CT will improve outcomes for patients treated with radiotherapy.

Due to advancements in radiotherapy techniques, accuracy in treatment delivery is improving and precise target volume definition has become more important, particularly in the era of dose escalation [8, 9]. However, gross tumor volume (GTV) delineation is very sensitive to observer variation [10] and hence there is a potential risk of geographic miss of tumor [11]. PET has been shown to have a significant impact when used in the radiation treatment planning process and in particular when used for target volume delineation (TVD), where a significant reduction in interobserver variability (IOV) has been noted [11-15]. It is recommended that a radiation oncologist (RO) and a nuclear medicine physician (NMP) / PET radiologist should be both involved where PET is used for TVD [16,17]. Complex cases of GTV delineation in lung cancer patients should always be discussed in a multi-disciplinary quality control meeting. Most clinical studies have used a visual interpretation technique, while others have reported the use of a range of automated segmentation techniques to either guide or generate the relevant target volume [18-21]. There is no clear consensus on which method most closely approximates to the tumor position and tumor edge, and pathological correlation has proven difficult [22]. Preoperative PET imaging shows a remarkably good correlation with resected pathological specimens [23], although it is acknowledged that those specimens are affected by processing artefacts. Most recent guidance advises the use of visual interpretation of the PET signal when drawing the final contour, even in cases where auto-contouring is used to generate an initial draft for editing, if PET is to be used to inform the target volume [17].

Factors causing IOV in TVD are variable interpretation of guidelines, lack of differentiation between normal structures and tumor, incorrect interpretation of radiological images, lack of knowledge in cross sectional radiological anatomy, and suboptimal imaging techniques e.g. lack of IV contrast [24-26]. The use of a rigorous contouring protocol in

which clinicians follow a detailed set of instructions and the use of a teaching intervention may help in minimizing IOV [21, 31, 32, 34, 35]. To ensure adequate and reproducible visual interpretation and application of PET images for RTP, this procedure should be standardized. A recent publication provided guidance on the use and role of PET/CT imaging for RTP in NSCLC patient [17]. This study evaluates the impact on the use of these practical guidelines through active teaching using multiple training interventions involving multiple centers with minimal experience in PET/CT-based TVD.

Methods

Target Volume Delineation Assignments

PET/CT-based TVD was assessed through the use of repeated delineation assignments. In all contouring assignments a team consisting of a RO and a NMP were asked to delineate tumor volumes of primary tumor (GTV). Before the training, participants were asked to delineate as per their local delineation protocol and then again after the first training intervention according to a standardized delineation protocol [17]. Fully anonymized patient cases were used, including three dimensional FDG-PET and CT image data sets acquired for the purpose of radiotherapy planning. No intravenous contrast agent was used. Comprehensive case specific medical reports were included in all assignments to avoid bias due to incorrect diagnosis. An overview of the patient cases used during this training program is given in table 1. In each case two senior ROs and a senior NMP delineated one reference ‘expert’ contour (GTV_{exp}) in agreement in the absence of a histopathologically proven gold standard.

Participants

The participants in this study were from eleven medical centers from nine different countries (Brazil, Estonia, India, Jordan, Pakistan, Poland, Turkey, Uruguay, and Vietnam). Each center was represented by a RO and a NMP. Before the training program centers were asked if they already performed PET/CT-based RTP. Five out of eleven centers already had limited experience in TVD with PET/CT. Other centers used PET/CT imaging for diagnostic and staging purposes only. Participants were not able to see delineations of other centers.

Big Brother target volume delineation software

Software developed in the Netherlands Cancer Institute, called Big Brother, was used throughout this multicenter study as platform for image viewing and analysis, and TVD in FDG PET/CT imaging [10]. As soon as the observer starts the Big Brother software and initiates TVD, any interaction with the software is recorded such as mouse motion, window/level and use of delineation tools. This feature allows visual inspection of strategies and comparison with expert contours to assess delineation accuracy.

Table 1. Sequence of events and characteristics of the included patients. Abbreviations: T = Primary tumor, N = Regional Lymph Nodes according to the 7th edition TNM classification.

	Case Number	T	N	M	Stage	Lymph nodes
Contouring Assignment 1	1	2	1	0	IIB	11L
	2	2	2	0	IIIA	10R, 7, 8, 4R
Contouring Assignment 2	3	3	0	0	IIB	
	4	2	0	0	IIA	
	5	4	2	0	IIIB	7, 4R, 2R
Training 1	6	3	2	0	IIIA	4R, 2R
	7	3	0	0	IIB	
	8	1	2	0	IIIA	10R, 4R
Contouring Assignment 3	Consisting of case 3, 4 and 5 (repeat assignment)					
Practice	9	2	2	0	IIIA	10R, 7, 4R
	10	2	2	0	IIIA	10R, 7
	11	2	2	0	IIIA	7, 4R
Training 2	Webinar / feedback reports					
Contouring Assignment 4	Consisting of case 1, 6 and 7 (repeat assignment)					

Target Volume Delineation Training program

The training program consisted of four contouring assignments, two training events and three additional clinical cases for practice (see figure 1). Contouring assignment 1 and 2 were performed before the first training event without the use of a standardized delineation protocol and were used as a baseline measurement. The first training event was face-to-face over a three-day period and included various lectures about relevant topics in nuclear medicine and radiation oncology and a delineation workshop on the use of PET/CT for RTP in NSCLC. The delineation workshop contained three more clinical cases which were again performed without the IAEA delineation protocol. The delineation protocol as described in the IAEA consensus document was introduced during the workshop [17]. The differences between the results and the IAEA protocol constituted the basis for a teaching discussion, consequentially clarifying protocol ambiguities. More contouring assignments followed after this training to evaluate its impact on delineation accuracy and IOV. Contouring assignment 3 was performed three months after contouring assignment 2 and contained the same clinical cases. To allow the participants to practice more with the delineation protocol three additional clinical cases were added.

After results were obtained from the above described assignments, an interim analysis was performed with the aim of identifying difficult areas in TVD and to ensure delineation occurred following the standardized approach. Detailed personal feedback reports were written with the aim of correcting misinterpretations of the delineation protocol and to advise on specific areas prone to deviation from the IAEA expert contour. This served as a preliminary to the webinar which was held as a second training event. An update on PET/CT-based TVD in RTP and general feedback was given in the webinar. Afterwards the content was discussed with the group. As a final step participants performed contouring assignment 4 with three clinical cases, which they had performed earlier during the training program, eight months after the first training event.

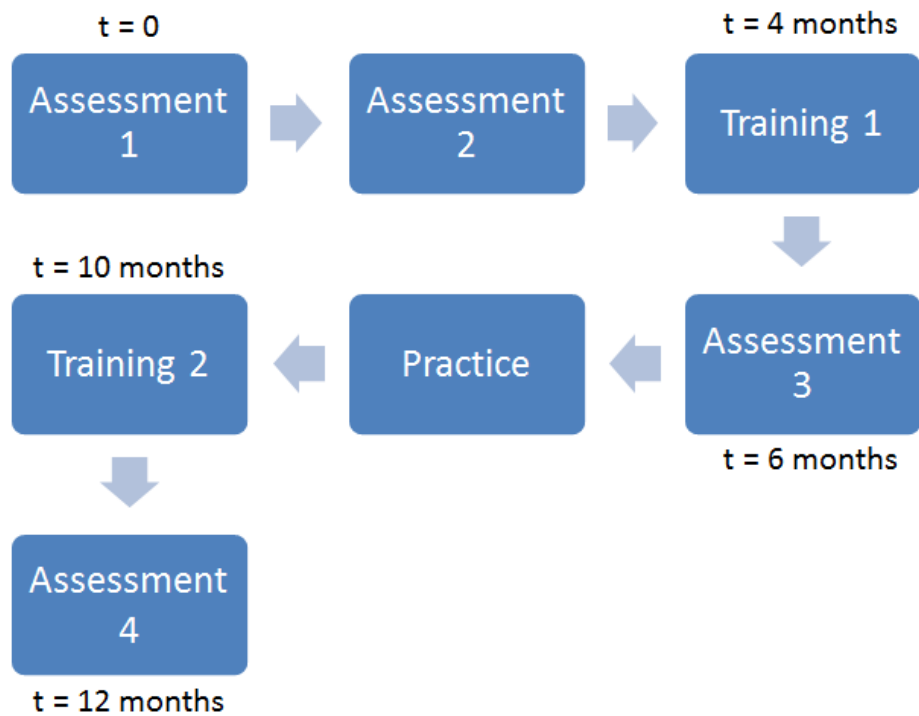


Figure 5. Schematic view of the training program. Over a one year period, 11 pairs of a radiation oncologist and a nuclear medicine physician performed four contouring assignments and three more cases for practice, and also attended two training events. Assessment 1, 2 and cases in Training 1 functioned as baseline measurements. Assessment 2 and 3 contained the same cases and results were compared to assess the impact of the first training event. Assessment 4 contained one case from Assessment 1 and two from Training 1 and the results were compared to assess the impact of the complete training.

Data analysis

To examine the impact of the training interventions the contours from the participants were analyzed. Various parameters such as the GTV size, miss of GTV_{exp} , Concordance Index (CI) and surface mean distance were calculated, and also volumetric and 3-dimensional analysis was performed as described by Deurloo et al. [27]. The CI is defined as the intersection of two delineated volumes divided by their union:

$$CI = \frac{(A \cap B)}{A \cup B}$$

The CI can vary between 0 and 1 whereas 0 means there is a complete disagreement between the observers and a CI of 1 indicates a perfect agreement [28]. It was calculated for measuring the delineation accuracy relative to the expert contour (CI_{expert}). Intragroup agreement (CI_{group}) was also calculated using the surface median as a reference. The surface median was obtained as described by Rasch et al. [29]. The mean (absolute) surface distance between the observers' GTV and expert GTV and the distance between each observers' GTV and surface median were both calculated. For all parameters, the mean \pm SD are reported unless stated otherwise. Wilcoxon signed rank tests were used to estimate the significance of any differences after the training events and p-values of 0.05 or less were considered significant.

Results

In all contouring assignments, teams were asked to delineate GTV of the primary tumor as per study protocol. One of the pairs with a RO, who was not board certified at the time of the training, was excluded from the analysis. After the first training event the overall CI_{expert} slightly increased from 0.57 ± 0.07 to 0.66 ± 0.07 . The mean CI_{expert} and CI_{group} per case are given in figure 2. Observer volumes were larger after the training and miss of GTV_{exp} was significantly reduced from 127.32 ± 42.43 cc to 59.94 ± 48.94 cc. A detailed summary with p-values is given in table 2. The overall mean surface distance and mean absolute surface distance compared to the reference contour decreased from -0.40 ± 0.03 cm to -0.01 ± 0.33 cm and from 0.47 ± 0.08 cm to 0.45 ± 0.17 cm respectively. The overall CI_{group} decreased from 0.81 ± 0.07 to 0.75 ± 0.10 .

Table 2. Comparison of results from contouring the GTV before and after the first training event, and before and after a complete training in the use of a standardized delineation protocol. CI_{expert} = median concordance index between the observers' GTV and expert GTV. Mean distance = mean surface distance between the observers' GTV and expert GTV. Mean |distance| = mean absolute surface distance.

	Case No.	Expert Volume (cc)	Observer Volume (cc ± SD)	Miss (cc ± SD)	CI_{expert} (± SD)	Mean distance (cm ± SD)	Mean distance (cm ± SD)
Results per case before and after Training 1 (contouring assignment 2 versus 3)							
Before	3	388.38	282.75 ± 46.28	127.32 ± 42.43	0.67 ± 0.10	-0.41 ± 0.14	0.44 ± 0.13
After			409.48 ± 115.12	59.94 ± 48.94	0.70 ± 0.07	0.05 ± 0.40	0.37 ± 0.23
			(p=0.03)	(p=0.03)	(p=0.34)	(p=0.03)	(p=0.92)
Before	4	50.86	30.58 ± 6.99	20.43 ± 6.09	0.59 ± 0.11	-0.27 ± 0.18	0.33 ± 0.10
After			51.35 ± 20.94	8.88 ± 9.34	0.65 ± 0.10	-0.11 ± 0.29	0.30 ± 0.10
			(p=0.03)	(p=0.03)	(p=0.06)	(p=0.17)	(p=0.14)
Before	5	164.46	84.93 ± 11.67	84.26 ± 11.66	0.49 ± 0.07	-0.64 ± 0.11	0.66 ± 0.12
After			108.49 ± 41.23	62.18 ± 23.14	0.58 ± 0.12	-0.43 ± 0.64	0.50 ± 0.45
			(p=0.05)	(p=0.05)	(p=0.08)	(p=0.05)	(p=0.46)
Overall results for 3 repeated cases (contouring assignment 2 versus 3)							
Before			123.96 ± 18.35	79.01 ± 17.04	0.57 ± 0.07	-0.40 ± 0.03	0.47 ± 0.08
After			191.38 ± 57.29	42.86 ± 25.02	0.66 ± 0.07	-0.01 ± 0.33	0.45 ± 0.17
			(p=0.03)	(p=0.05)	(p=0.12)	(p=0.03)	(p=0.75)
Results per case before and after the complete training program (contouring assignment 1 versus 4)							
Before	6	377.99	280.39 ± 92.41	115.45 ± 52.16	0.68 ± 0.11	-0.41 ± 0.35	0.43 ± 0.19
After			370.78 ± 37.06	26.73 ± 18.34	0.85 ± 0.03	-0.01 ± 0.16	0.20 ± 0.09
			(p=0.07)	(p=0.01)	(p=0.01)	(p=0.07)	(p=0.02)
Before	7	254.68	157.99 ± 80.15	103.70 ± 43.86	0.59 ± 0.10	-0.55 ± 0.40	0.60 ± 0.16
After			220.24 ± 81.53	39.72 ± 25.87	0.82 ± 0.10	-0.18 ± 0.34	0.24 ± 0.18
			(p=0.04)	(p=0.04)	(p=0.05)	(p=0.07)	(p=0.07)
Before*	1	134.07	78.08 ± 17.60	55.66 ± 9.58	0.56 ± 0.05	-0.44 ± 0.17	0.50 ± 0.06
After*			110.20 ± 14.51	26.23 ± 9.68	0.80 ± 0.04	-0.19 ± 0.11	0.21 ± 0.04
			(p=0.08)	(p=0.03)	(p=0.05)	(p=0.08)	(p=0.05)
Overall results for 3 repeated cases (contouring assignment 1 versus 4)							
Before			190.82 ± 38.07	78.89 ± 22.95	0.64 ± 0.06	-0.34 ± 0.16	0.49 ± 0.10
After			246.32 ± 43.11	42.86 ± 25.01	0.80 ± 0.05	-0.05 ± 0.20	0.27 ± 0.09
			(p=0.01)	(p=0.01)	(p=0.01)	(p=0.01)	(p=0.01)

*For this case (before and after) two observers were excluded from analysis, because PET positive lymph nodes adjacent to the tumor were included in the GTV of the primary tumor.

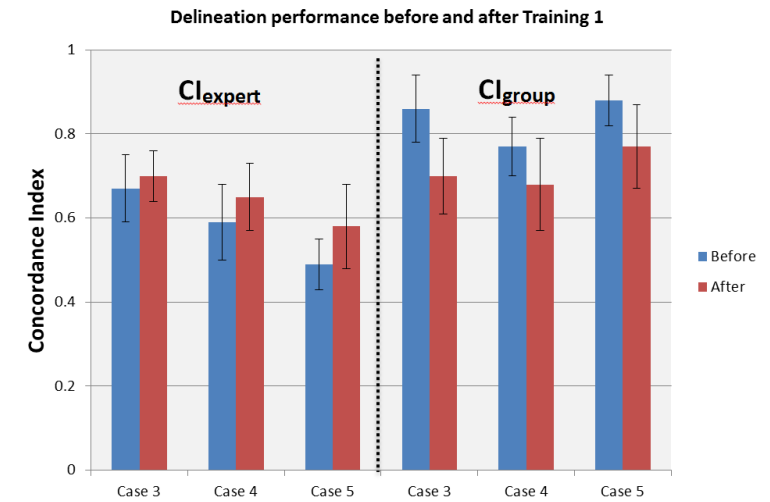


Figure 6. Delineation accuracy relative to expert versus intragroup with 95%-CI before and after Training 1 using a standardized delineation protocol. Cl_{group} = median concordance index between the observers' GTV and the median surface contour. Cl_{expert} = median concordance index between the observers' GTV and expert GTV.

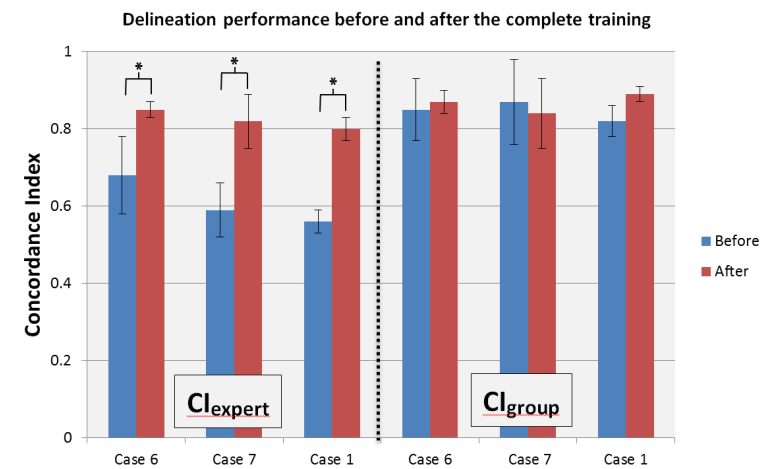


Figure 7. Delineation accuracy relative to expert versus intragroup with 95%-CI before (blue bar) and after (red bar) a complete training in the use of a standardized delineation protocol. Cl_{group} = median concordance index between the observers' GTV and the median surface contour. Cl_{expert} = median concordance index between the observers' GTV and expert GTV.

After the second teaching event overall CI_{expert} for another 3 repetitive cases increased from 0.64 ± 0.06 to 0.80 ± 0.05 (see figure 3 for more details). A reduction of GTV_{exp} miss from 78.89 ± 22.95 cc to 42.86 ± 25.01 cc was observed, next to an increase in observer volume. Overall mean surface distances between observers and the expert contour decreased from -0.34 ± 0.16 cm to -0.05 ± 0.20 cm. A decrease from 0.49 ± 0.10 cm to 0.27 ± 0.09 cm in overall mean absolute surface distance was observed. The overall CI_{group} increased from 0.80 ± 0.08 to 0.85 ± 0.08 . Examples of improvement in IOV and delineation accuracy before and after the training program are given in figure 4 and in supplementary figure 1.

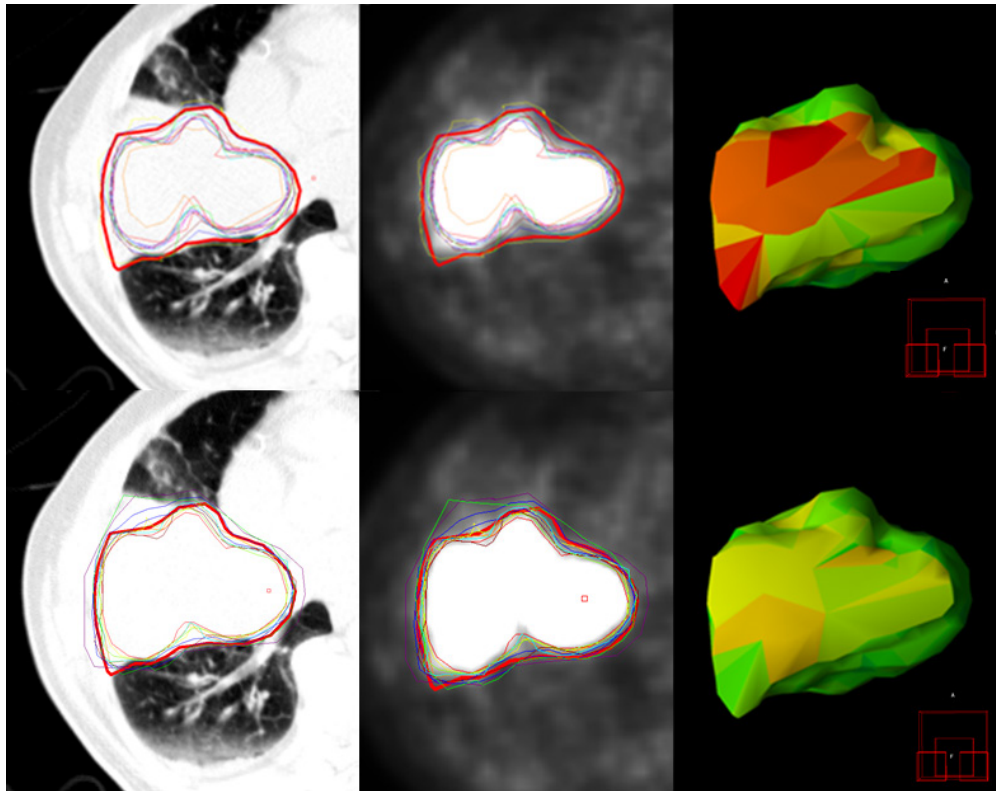


Figure 4. Results from contouring the GTV in case 6 before (top images) and after (bottom images) the complete training. From left to right an image slice in the axial plane from CT, PET, and a 3D model of the expert contour with mean surface distance errors projected (from blue to red, corresponding with a mean absolute surface distance of 0 to >1 cm). The bold red line represents the GTV_{exp} .

Discussion

Many studies have reported on the effect of guidelines or protocols and testing of the effect of specific teaching and measured contouring variability before and after an intervention [30]. This study is the first to report on the impact of a training program about the use of the recently published IAEA guidelines for PET/CT-based radiotherapy planning in lung cancer patients in an international multicenter trial. To the best of our knowledge this is the first study about the use of multiple teaching interventions using not only face-to-face training, but also providing an online learning platform in the form of a webinar which showed to play an effective role in harmonizing the delineation process globally. In terms of practicalities of delivering this type of training, the use of innovative technology such as the delivery of live webinars may significantly reduce cost without significantly reducing educational quality.

Before any training was given, centers delineated their GTV based on local delineation protocols and thus a variety of approaches were observed. Results after the first training suggest that the use of a delineation protocol increased delineation accuracy, however a significant reduction in IOV and better adherence of the outlining protocol was only achieved through additional extended training.

A major contribution in reducing IOV may have been feedback participants received from the interim analysis that was performed before the second training intervention. Whilst it is acknowledged that the cases selected for the practice case (see table 1) may have been more challenging to delineate than the test cases, the use of these helped identify several areas that caused difficulty in TVD for the participants. These areas of variation included atelectasis, PET window level settings, nearby normal structures (e.g. pulmonary veins or arteries) that were seen as tumor, and/or suspicious areas on CT showing low FDG uptake. These areas of variation were documented in the individual feedback reports, and were also included in the general feedback and discussed during the webinar. This may have contributed to the significant impact on delineation accuracy seen in the last assignment. This emphasizes the importance of additional training events to correct errors in delineation still occurring after the first training. Schimek-Jasch et al. found out that the use of a dummy-run and study group meetings as part of Quality Assurance (QA) in multicenter clinical trials helps to identify misinterpretations of a standardized delineation method which helped in reducing IOV [32]. However, the outcome of that study did not show a significant effect which may underline the take-home message of our study that changing behavior requires multiple and multi-faceted interventions. Lack of other studies investigating the impact of multiple training interventions in TVD makes comparison of outcomes difficult. Further studies with more observers would be needed to

validate the results in this study. Currently the IAEA conducts a multicenter international study investigating the impact of blended distance learning (with additional training interventions) in the field of RT contouring on quality of delineation (CRP E33040).

Several studies concluded that the use of a standardized delineation protocol helps in decreasing IOV [31, 32, 34] and our results concur with that. There was a non-significant increase of IOV seen after the first training event, however a decrease was noted after the complete training program. In spite of the first training, visual inspection with Big Brother showed that some observers misinterpreted or did not comply with the guidelines thus their contours did not more closely resemble the CI_{expert} contour. Due to the fact that some observers drew contours more similar to that of the expert in contrary to others in the group, the CI_{expert} still increased slightly. This increase in variation between observers reduced the CI_{group} (see figure 2). The second training event was necessary to correct any misinterpretations of practical guidelines and this approach of having training interventions may seem more effective than a single event training program. This again highlights the difficulty of changing behavior in order to obtain reproducibility and underlines the importance of teaching through multiple interventions to improve adherence to contouring guidelines. However, participants still had difficulties in determining the tumor boundary in cases with suspicious areas showing FDG uptake comparable to background activity. Decisions, as to whether or not to include these suspect areas within the GTV, are namely based on experience and expertise and this contributes to a degree of IOV. This emphasizes the importance of multidisciplinary meetings in case the RO experiences difficulties in contouring.

In contrary to the delineation accuracy, GTV size did increase significantly among the observers after the first training. This was mainly due to specific training in standardization of PET window level settings [17], which is possible in most commercially available radiotherapy planning and contouring systems. The difficulty with manual delineation in PET/CT imaging is that the apparent boundaries of the FDG avid tumor are highly dependent on the chosen PET window level settings. Tumors will appear larger when delineation occurs using a high window setting and smaller with a low window setting on the PET display. Observers were trained in using standardized PET level window settings consequently showing an apparently larger tumor than before and this contributed to an increase in GTV size. Before the 1st training intervention most delineations were drawn too tightly around the tumor, which may possibly lead to geographical tumor miss. Caution has to be taken in circumstances where, in the absence of respiration correlated CT, PET/CT may be substantially less accurate in defining the motion pathway of highly mobile lung tumors and in tumors with low FDG uptake [33].

When observers delineate the same tumor repeatedly this creates systematic errors and contributes to intra-observer variability, since it is unlikely that any manual contour would be reproduced identically at different time points. In this study we did not examine what contribution this effect had on intra-observer variability when repeating the same case in a short timeframe. However it is hypothesized this effect is negligible compared to learning effects over a longer timeframe.

There have been a number of studies which have examined IOV in TVD using PET based delineation with a number of these studies focusing on the comparison of automatic delineation methods with manual delineation [17-21]. Doll et al. used one patient case and found overall concordance indices between experts, interdisciplinary pairs and single field specialists ranging from 0.49 to 0.67 which is similar to our results after the first training [16]. The experts showed the highest intragroup concordance of 0.67 and if that is representative then the outcome of our training program could be seen as successful. However, since the expert group in Doll et al. only performed one case and the study did not use a standardized delineation protocol this comparison is not valid. This study used only one expert contour per case as reference, which limited the conclusion whether an observer met a certain minimum level of quality in TVD. An intragroup expert concordance value could help in determining such a minimum required level of quality. Further research has to determine which deviation from the intragroup concordance value would be acceptable.

Another limitation was the amount of cases available for the repeated assignments. Not all cases in Training 1 were suitable for inclusion in Assessment 4 due to the small tumor size in case 8 and therefore one case of the first assignment (case 1) had to be included. It is acknowledged that observers were less familiar with the software in the beginning than later in the training program, however the delineation software is similar to any other delineation tool used in clinic and it is hypothesized that if there is any learning effect present, this only has an impact on the delineation speed. Above that, a similar increase is seen in all cases after the complete training suggesting that the learning effect of the delineation software was negligible (see figure 3).

Spoelstra et al. had seen that a significant IOV in contouring confounded interpretation of post-operative radiotherapy and concluded that quality assurance (QA) procedures would need to be incorporated to tackle this problem [34]. A German multicenter PET study also covered a similar interesting topic i.e. harmonization of diagnostic viewing and reporting and also outlined the importance of QA [35]. They concluded that a structured interventional harmonization process significantly improved the IOV in their expert panel. However, no focus had been given on target volume delineation. In our study, additional training led to an increased delineation accuracy and decreased the IOV. In clinic, the IOV

should also be assessed in order to see if it is necessary to provide more training in order to achieve reproducible results among ROs. Therefore it is recommended that assessment of IOV should be performed frequently next to having multi-disciplinary quality control meetings as part of the QA on TVD. The impact of repeated IOV assessment and its optimal frequency in a clinical setting should be further investigated.

A study examining the influence of experience and qualification in PET based TVD concluded that IOV may be dependent on qualification, but not on years of experience [16]. It is known that some centers already had minimal experience in PET/CT-based TVD and that there were centers with no experience in this field. However, no significant difference in performance was seen after comparison of inexperienced versus minimal experienced participants.

Conventional 3-dimensional PET/CT imaging was chosen as the modality for TVD, since not all participants in the study had experience in PET/CT-based radiotherapy planning in NSCLC patients. The impact of 4-dimensional PET/CT has not been investigated and may be of interest to further increase accuracy. Furthermore, if a PET/CT acquired for diagnostic or staging purposes is used to inform the TVD process, care must be taken when registering a diagnostic PET/CT with a planning CT. Guidance regarding this has been described in the IAEA study protocol [17].

Conclusion

ROs and NMPs with limited experience in PET/CT-based TVD for lung cancer benefit significantly from receiving multiple training interventions with a standardized delineation protocol. Future research within a larger population should validate the results in this study to provide more evidence on the impact of multiple training interventions about PET/CT-based TVD for NSCLC.

Acknowledgements

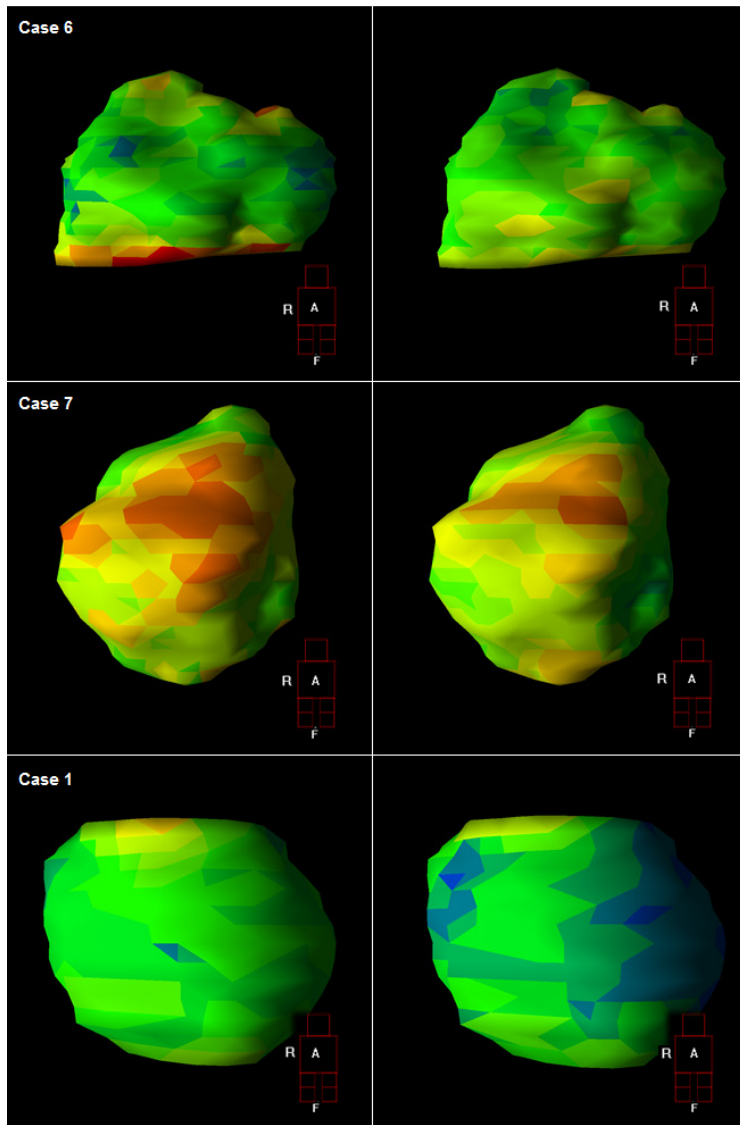
The authors would like to thank Heloisa Carvalho and Paolo Duarte (Instituto do Câncer do Estado de São Paulo), Camilla Mosci and Carlos Zuliani (Hospital das Clínicas UNICAMP), Darja Altuhhova and Liina Karusoo (North Estonia Medical Center), Rakesh Kapoor and Ashwani Sood (Postgraduate Institute of Medical Education and Research), Jamal Khader and Akram Al-Ibraheem (King Hussein Cancer Center), Shahid Abubaker and Muhammad Numair (Institute of Nuclear Medicine and Oncology), Bozena Birkenfeld and

Bartłomiej Masojc (Zachodniopomorskie Centrum Onkologii), Cigdem Soydal and Tuğçe Kütük (Ankara University School of Medicine), Aldo Quarnetti and Omar Alonso (Centro Uruguayo de Imagenología Molecular), Bui Quang Bieu and Le Ngoc Ha (Tran Hung Dao General Hospital), Nguyen Xuan Canh and Tuan Anh Le (Cho Ray Hospital) for their active cooperation in the study. The authors also wish to thank Marcel van Herk for the support with the Big Brother software in this study.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Int J Cancer. 2015; 136(5): 359-86.
2. US National Institutes of Health. National Cancer Institute. SEER Cancer Statistics Review, 1975-2011. 2014 – Accessed on 02-03-2016.
3. National Cancer Intelligence Network. Stage Breakdown by CCG 2013. London: NCIN; 2015. Accessed on 02-03-2016
4. Ung JC, Bezjak A, Coakley N, Evans WK and the Lung Cancer Disease Site Group. Positron Emission Tomography with 18Fluorodeoxyglucose in Radiation Treatment Planning for Non-small Cell Lung Cancer. J Thorac Oncol. 2011; 6: 86–97.
5. Cerfolio RJ, Ojha B, Bryant AS, Raghuvver V, Mountz JM, Bartolucci AA. *The accuracy of integrated PET-CT compared with dedicated PET alone for the staging of patients with non small cell lung cancer*. Ann Thorac Surg. 2004; 78(3): 1017-23.
6. De Wever W, Ceyskens S, Mortelmans L, et al. *Additional value of PET-CT in the staging of lung cancer: comparison with CT alone, PET alone and visual correlation of PET and CT*. Eur Radiol. 2007; 17: 23–32.
7. Dondi M, Kashyap R, Paez D, Pascual T, Zaknun J, Bastos FM, Pynda Y. *Trends in nuclear medicine in developing countries*. J Nucl Med. 2011; 52(2): 16-23.
8. De Ruyscher D, Wanders S, Minken A, et al. *Effects of radiotherapy planning with a dedicated combined PET-CT-simulator of patients with non-small cell lung cancer on dose limiting normal tissues and radiation dose-escalation: a planning study*. Radiother Oncol. 2005; 77: 5-10.
9. van Elmpt W, De Ruyscher D, van der Salm A, Lakeman A, van der Stoep J, Emans D, Damen E, Öllers M, Sonke JJ, Belderbos J. *The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer*. Radiother Oncol. 2012; 104(1): 67-71.
10. Steenbakkers RJ, Duppen JC, Fitton I, Deurloo KE, Zijp L, Uitterhoeve AL, Rodrigus PT, Kramer GW, Bussink J, De Jaeger K, Belderbos JS, Hart AA, Nowak PJ, van Herk M, Rasch CR. *Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation*. Radiother Oncol. 2005; 77(2): 182-90.
11. Rasch C, Belderbos J, van Giersbergen A, De Kok I, Laura T, Boer M, de Boer P, Gilles R, Teertstra J, Verheij M. *The Influence Of A Multi-disciplinary Meeting For Quality Assurance On Target Delineation In Radiotherapy Treatment Preparation International Journal of Radiation Oncology*. 2009; 75(3): 452–453.
12. Hanna GG, McAleese J, Carson KJ, Stewart DP, Cosgrove VP, Eakin RL, et al. *(18)F-FDG PET-CT simulation for non-small-cell lung cancer: effect in patients already staged by PET-CT*. Int J Radiat Oncol Biol Phys. 2010; 77(1): 24–30.
13. Caldwell CB, Mah K, Ung YC, et al. *Observer variation in contouring gross tumor volume in patients with poorly defined non-small cell lung tumors on CT: The impact of 18FDG-hybrid PET fusion*. Int J Radiat Oncol Biol Phys. 2001; 51: 923–931.
14. Greco C, Rosenzweig K, Cascini GL, Tamburrini O. *Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC)*. Lung Cancer. 2007; 57: 125–34.
15. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, Leibel SA, Rosenzweig KE. *Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer?* Int J Radiat Oncol Biol Phys. 2005; 62: 70–75.
16. Doll C, Duncker-Rohr V, Rücker G, Mix M, MacManus M, De Ruyscher D, Vogel W, Eriksen JG, Oyen W, Grosu AL, Weber W, Nestle U. *Influence of experience and qualification on PET-based target volume delineation. When there is no expert—ask your colleague*. Strahlenther Onkol. 2014; 190(6): 555-62.
17. Konert T, Vogel W, MacManus MP, Nestle U, Belderbos J, Grégoire V, Thorwarth D, Fidarova E, Paez D, Chiti A, Hanna GG. *PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014*. Radiother Oncol. 2015; 116(1): 27-34.
18. Zaidi H, El Naqa L. *PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques*. Eur J Nucl Med Mol Imaging. 2010; 37(11): 2165-87.
19. Werner-Wasik M, Nelson AD, Choi W, et al. *What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom*. Int J Radiat Oncol Biol Phys. 2012; 82: 1164-71.
20. Cui H, Wang X, Feng D. *Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis*. Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, Calif, USA. 2012; 5384–5387.
21. Bayne M, Hicks RJ, Everitt S, Fimmell N, Ball D, Reynolds J, Lau E, Pitman A, Ware R, MacManus M. *Reproducibility of "intelligent" contouring of gross tumor volume in non-small-cell lung cancer on PET/CT images using a standardized visual method*. Int J Radiat Oncol Biol Phys. 2010; 77(4): 1151-7.
22. van Loon J, Siedschlag C, Stroom J, Blauwgeers H, van Suylen RJ, Kneigiens J, Rossi M, van Baardwijk A, Boersma L, Klomp H, Vogel W, Burgers S, Gilhuijs K. *Microscopic disease extension in three dimensions for non-small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features*. Int J Radiat Oncol Biol Phys. 2012; 82(1): 448-56.
23. van Baardwijk A, Bosmans G, Boersma L, et al. *PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes*. Int J Radiat Oncol Biol Phys. 2007; 68: 771–8.
24. Van de Steene J, Linthout N, de Mey J, et al. *Definition of gross tumor volume in lung cancer: interobserver variability*. Radiother Oncol 2002; 62:37-49.
25. Giraud P, Elles S, Helfre S, De Rycke Y, Servois V, Carette MF, Alzieu C, Bondiaud PY, Dubray B, Touboul E, Housset M, Rosenwald JC, Cosset JM. *Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists*. Radiother Oncol. 2002; 62(1): 27-36.
26. Kepka L, Buijko K, Garmol D, Palucki J, Zolciak-Siwinska A, Guzel-Szczepiorkowska Z, Pietrzak L, Komosinska K, Sprawka A, Garbaczevska A. *Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy*. Radiother Oncol. 2007; 85(3): 450-5.
27. Deurloo KE, Steenbakkers RJ, Zijp LJ. *Quantification of shape variation of prostate and seminal vesicles during external beam radiotherapy*. Int J Radiat Oncol Biol Phys. 2005; 61: 228–238.
28. Hanna GG, Hounsell AR, O'Sullivan JM. *Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods*. Clin Oncol (R Coll Radiol). 2010 Sep; 22(7):515-25.
29. Rasch CR, Steenbakkers RJ, Fitton I, Duppen JC, Nowak PJ, Pameijer FA, Eisbruch A, Kaanders JH, Paulsen F, van Herk M. *Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer*. Radiat Oncol. 2010; 5: 21.
30. Vinod SK, Min M, Jameson MG, Holloway LC. *A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology*. J Med Imaging Radiat Oncol. 2016; 60(3): 392-406.
31. Bowden P, Fisher R, Mac Manus M, Wirth A, Duchesne G, Millward M, McKenzie A, Andrews J, Ball D. *Measurement of lung tumor volumes using three-dimensional computer planning software*. Int J Radiat Oncol Biol Phys. 2002; 53(3): 566-73.
32. Schimek-Jasch T, Troost EG, Rücker G, Prokic V, Avlar M, Duncker-Rohr V, Mix M, Doll C, Grosu AL, Nestle U. *A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: Reducing the interobserver variability in multicentre clinical studies*. Strahlenther Onkol. 2015; 191(6): 525-33.
33. Hanna GG, van Sörnsen de Koste JR, Dahele MR, et al. *Defining target volumes for stereotactic ablative radiotherapy of early-stage lung tumours: a comparison of three-dimensional 18F-fluorodeoxyglucose positron emission tomography and four-dimensional computed tomography*. Clin Oncol (R Coll Radiol). 2012; 24(6): 71-80.
34. Spoelstra FO, Senan S, Le Péchoux C, Ishikura S, Casas F, Ball D, Price A, De Ruyscher D, van Sörnsen de Koste JR. *Variations in target volume definition for postoperative radiotherapy in stage III non-small cell lung cancer: analysis of an international contouring study*. Lung Adjuvant Radiotherapy Trial Investigators Group. Int J Radiat Oncol Biol Phys. 2010; 76(4): 1106-13.
35. Nestle U, Rischke HC, Eschmann SM, Holl G, Tosch M, Miederer M, Plotkin M, Essler M, Puskas C, Schimek-Jasch T, Duncker-Rohr V, Rühl F, Leifert A, Mix M, Grosu AL, König J, Vach W. *Improved inter-observer agreement of an expert review panel in an oncology treatment trial - Insights from a structured interventional process*. Eur J Cancer. 2015; 51(17): 2525-33.

Supplementary data



Supplementary figure 1. Three dimensional model of the expert contour (frontal view) for cases 1, 6 and 7 with mean surface distance errors projected (from blue to red, corresponding with a mean absolute surface distance of 0 to >1 cm).



Chapter 5

Introducing FDG PET/CT guided chemoradiotherapy for stage III NSCLC in low and middle-income countries: preliminary results from the IAEA PERTAIN trial

5

Tom Konert
Wouter V Vogel
Diana Paez
Jose A. Polo-Rubio
Elena Fidarova
Heloisa Carvalho
Paulo S. Duarte
Carlos A. Zuliani
Alan O. Santos
Darja Altuhhova,
Liina Karusoo
Rakesh Kapoor
Ashwani Sood
Jamal Khader

Akram Al-Ibraheem
Muhammad Numair
Shahid Abubaker
Cigdem Soydal
Tuğçe Kütük
Nguyen X. Canh
Le T. Anh
Bui Q. Bieu
Le N. Ha
José S.A. Belderbos
Michael P. MacManus
Daniela Thorwarth
Gerard G. Hanna

European Journal of Nuclear Medicine and Molecular Imaging. 2019; 46(11): 2235-2243.

Abstract

Purpose

Patients with stage III non-small-cell lung cancer (NSCLC) treated with chemoradiotherapy (CRT) in low- and middle-income countries (LMIC) continue to have a poor prognosis. It is known that FDG PET/CT improves staging, treatment selection and target volume delineation (TVD), and although its use has grown rapidly, it is still not widely available in LMIC. CRT is often used as sequential treatment, but is known to be more effective when given concurrently. The aim of the PERTAIN study was to assess the impact of introducing FDG PET/CT-guided concurrent CRT, supported by training and quality control (QC), on the overall survival (OS) and progression-free survival (PFS) of patients with stage III NSCLC.

Methods

The study included patients with stage III NSCLC from nine medical centres in seven countries. A retrospective cohort was managed according to local practices between January 2010 and July 2014, which involved only optional diagnostic FDG PET/CT for staging (not for TVD), followed by sequential or concurrent CRT. A prospective cohort between August 2015 and October 2018 was treated according to the study protocol including FDG PET/CT in treatment position for staging and multimodal TVD followed by concurrent CRT by specialists trained in protocol-specific TVD and with TVD QC. Kaplan–Meier analysis was used to assess OS and PFS in the retrospective and prospective cohorts.

Results

Guidelines for FDG PET/CT image acquisition and TVD were developed and published. All specialists involved in the PERTAIN study received training between June 2014 and May 2016. The PET/CT scanners used received EARL accreditation. In November 2018 a planned interim analysis was performed including 230 patients in the retrospective cohort with a median follow-up of 14 months and 128 patients in the prospective cohort, of whom 69 had a follow-up of at least 1 year. Using the Kaplan–Meier method, OS was significantly longer in the prospective cohort than in the retrospective cohort (23 vs. 14 months, $p=0.012$). In addition, median PFS was significantly longer in the prospective cohort than in the retrospective cohort (17 vs. 11 months, $p=0.012$).

Conclusion

In the PERTAIN study, the preliminary results indicate that introducing FDG PET/CT-guided concurrent CRT for patients with stage III NSCLC in LMIC resulted in a significant improvement in OS and PFS. The final study results based on complete data are expected in 2020.

Introduction

Worldwide, lung cancer is the most commonly diagnosed cancer (2.1 million new cases in 2018) and the leading cause of cancer death (1.8 million deaths estimated in 2018) [1]. Five-year survival of lung cancer was 20–33% in countries such as Japan, Canada, USA, China, Korea, Israel, Sweden, Switzerland and Austria. However, most other countries had a 5-year survival ranging between 10% and 20%. Survival was less than 10% in countries such as Brazil, India and Thailand. Globally, lung cancer survival rates between 1995 and 1999 and between 2000 and 2014 indicate no improvement with time, but in high-income countries 5-year overall survival (OS) has increased by 5–10% in absolute terms over the same time period [2].

Currently, 18F-fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) is widely used for staging patients with non-small-cell lung cancer (NSCLC) and to a lesser extent for radiotherapy (RT) target volume delineation (TVD) [3, 4]. PET/CT scanners have also become available in several low- and middle-income countries (LMIC), although FDG PET/CT is mainly used for staging purposes rather than as a part of treatment planning in NSCLC [5]. The International Atomic Energy Agency (IAEA) convened an expert panel to appraise the clinical utility of FDG PET/CT for staging and RT planning (RTP) in patients with lung cancer. This coordinated research programme resulted in the design of the international PET/CT in RTP (PERTAIN) study (NCT02247713) to assess the feasibility of including FDG PET/CT in the RTP process in patients with stage III NSCLC in LMIC.

The current standard treatment for stage III NSCLC is concurrent chemoradiotherapy (CRT) [6]. In order to take advantage of the recent developments in RT techniques which have improved the accuracy of treatment delivery, it is essential to ensure TVD is as accurate as possible to avoid geographic miss of disease. Advanced RT techniques have improved local tumour control and have reduced treatment toxicity by enabling the delivery of higher radiation doses to the tumour while sparing adjacent normal tissue [7]. Examples include intensity-modulated RT (IMRT) [8], and image-guided RT, which improves the precision of treatment delivery and allows the use of smaller expansion margins [9].

TVD involves contouring the gross tumour volume (GTV), as specified in ICRU report 50 [10]. GTV delineation is sensitive to interobserver variability (IOV) [11, 12]. A significant reduction in IOV can be achieved with information from both PET and CT [12, 13, 14, 15]. Automatic PET segmentation methods have also been proposed to reduce IOV [16], but always need verification by a radiation oncologist (RO) [17]. PET is specifically helpful in TVD when the tumour is not easily distinguished from surrounding healthy tissue on CT images, due to its higher soft tissue contrast [18]. Even with the use of PET imaging

there is still IOV due to differences in the TVD method and to FDG uptake in normal structures adjacent to the tumour [19, 20]. The use of a rigorous contouring protocol in which a multidisciplinary team including a RO and a nuclear medicine physician (NMP) follow a detailed set of instructions has been shown to help minimize IOV [21]. A recent IAEA publication has provided guidance on the use and role of FDG PET/CT imaging for RTP in NSCLC patients [17]. The impact of the use of the IAEA study protocol on TVD accuracy and reproducibility has been evaluated in multiple centres in LMIC. Multiple training interventions on PET/CT-based TVD in NSCLC improves delineation accuracy and reduces IOV [16]. Hence, we hypothesized that TVD following the IAEA study protocol would increase accuracy and reproducibility of TVD in the clinic, leading to improvement in local control.

There are many reports describing IOV within and outside the context of clinical trials, but few studies have investigated the impact of IOV on clinical outcome [20, 22, 23] or methods that could minimize IOV by means of training [16, 24, 25, 26]. The clinical impact of such training remains unknown. Hence, we hypothesized that TVD following the IAEA study protocol would increase the accuracy and reproducibility of TVD and lead to improvement in local control and thus OS. We present the preliminary results of the PERTAIN study. The aim of this study was to assess the impact of introducing FDG PET/CT-guided concurrent CRT, supported by training and quality control (QC), on the OS and PFS in patients with stage III NSCLC.

Materials and methods

Ethical aspects

The PERTAIN study was approved by the medical ethics committee of Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital (ref. M14PRI). In addition, each centre received ethical clearance from their local medical research ethics committee. Written informed consent was obtained from all patients included in the prospective phase.

Study framework

Nine medical centres in seven countries met the technical requirements to participate in the PERTAIN study, including six middle-income countries: Brazil, India, Jordan, Pakistan, Turkey and Vietnam. The seventh country, Estonia, is classified as a high-income country by the World Bank, but did not routinely use PET/CT for RTP. The first component of the study was data collection from a retrospective cohort, which included consecutive patients with stage III NSCLC who had been treated in the participating centres between January 2010 and July 2014 according to existing local protocols. Over a 1-year period, nine

pairs of trainees each including a RO and a NMP with limited experience in PET/CT-based TVD in NSCLC from seven different countries took part in multiple training interventions. Teams were given hands-on training in delineating the primary tumour according to IAEA protocol guidelines. An online webinar training session was held on TVD in NSCLC, and lectures for ROs and NMPs on current best practice in NSCLC were given [19]. All PET/CT scanners received annual European Association of Nuclear Medicine (EANM) Research Ltd. (EARL) FDG PET/CT accreditation. After the training intervention and scanner calibration, patients with stage III NSCLC were included in the prospective cohort between August 2015 and October 2018. The study entry criteria are summarized in the Supplementary material S1 (Form 1). Patients who did not meet the study entry criteria were excluded from the study.

Case report forms

Patient data were collected using electronic case report forms (eCRFs). Five different eCRFs were designed to collect information on patient eligibility, before and after treatment, and follow-up. More details on eCRFs and their format can be found in the Supplementary material S1.

Clinical endpoints

The primary endpoint was OS, defined as the time between the start of treatment and date of death or loss to follow-up. The secondary endpoint was progression-free survival (PFS). PFS was defined as the time from the start of treatment to local failure, time to regional failure, and/or time to distant failure. Local failure was defined as progression in the primary tumour, and regional failure as progression in involved lymph nodes as assessed on follow-up scans. Distant failure was defined according to the 8th edition of the TNM classification for NSCLC [27]. The intervals for the follow-up assessments and imaging were as per local follow-up guidelines.

Chemotherapy and radiotherapy details

Patients in the retrospective cohort were treated according to respective institutional practice with sequential concurrent CRT, neoadjuvant chemotherapy or RT alone, but with curative intent. In the prospective cohort, patients were treated with concurrent CRT to a total dose of at least 60 Gy in fractions of 2 Gy over 6 weeks. Centres were free to select chemotherapy regimens according to local practice.

PET/CT image acquisition

Patients underwent whole-body FDG PET/CT using one of the following scanners: Discovery ST, Discovery 710, Discovery STE (GE Medical Systems, Chicago, IL, USA), Biograph 40 mCT, and Biograph 64 mCT (Siemens Medical Solutions, Erlangen, Germany).

The reconstruction voxel size of the PET data varied from 2.0 × 2.0 × 3.3 mm to 5.5 × 5.5 × 3.3 mm. Patients fasted for at least 8 h to ensure low levels of serum glucose. The total injected dose ranged between 226 MBq and 441 MBq (data not available for all patients). Patients were scanned approximately 60 min after injection of 18F-FDG according to EANM guidelines [28]. The acquisition times of the PET/CT scanners were in the range 2–5 min per bed position.

Assessment of the retrospective and prospective cohort

To assess the overall impact of the multiple training interventions and the routine use of FDG PET/CT-based concurrent CRT, survival outcomes in the retrospective cohort were compared with those in the prospective cohort. Although the training programme focused mainly on standardized PET/CT-based TVD, in general, the whole RTP procedure was also standardized to ensure the use of current treatment standards. Differences in the RTP procedures between the retrospective and prospective cohorts are summarized in Table 1.

Quality control of target volume delineation

To ensure that participating centres in the prospective study complied with the IAEA study protocol, central QC review of TVD was performed for the first three patients included per centre, and thereafter as needed. In the QC process anonymized PET/CT data and RT structure sets were made available through a secure online storage service and were reviewed by at least two members of the study trial management group.

Table 1. Differences in staging, radiotherapy planning, treatment and target volume delineation procedures between the retrospective and prospective cohorts.

Comparison	Retrospective cohort	Prospective cohort
Staging	With or without PET/CT	With PET/CT
RTP	With or without PET/CT	PET/CT in RTP-position
Time interval	Per local protocol, delays of >1 month possible	Within 4 weeks of last PET/CT
Delivered dose	Per local protocol	≥60 Gy
Treatment	RT, sequential CRT, CCRT	CCRT only
TVD	Per local protocol	Per IAEA study protocol (PET/CT-based)
PET/CT quality assurance	EARL accreditation not compulsory	EARL accreditation compulsory
Nodal irradiation	Both elective and involved nodal RT	Involved nodal RT

All procedures in the prospective cohort were standardized in all centers in accordance with the IAEA study guidelines [17]. (C)CRT = (concurrent) chemoradiotherapy, RT radiotherapy, RTP radiotherapy planning, TVD tumour volume delineation.

Statistical analysis

Any differences in continuous variables between the retrospective and prospective cohorts were assessed using the independent t test. Any differences in categorical variables were assessed using the chi-squared test. Strong prognostic factors were identified using univariate Cox regression analysis. Kaplan–Meier analysis was performed to assess OS and PFS in the retrospective and prospective groups. The log-rank statistic was used to assess the significance of any differences. Statistical analysis was performed using IBM SPSS statistics for Windows, version 22.0 (IBM Corp., Armonk, NY). Values of p less than 0.05 were considered significant.

Results

Patient inclusion

The retrospective cohort included 230 patients with stage III NSCLC treated with sequential or concurrent CRT or RT alone. The prospective cohort included 69 patients with stage III NSCLC. In all centres, a high percentage of patients (up to 51%) were upstaged to stage IV after staging with PET/CT became the standard. Overall, five patients did not meet the study inclusion criteria, and were therefore excluded. Reasons for exclusion were inability to provide informed consent (one patient), unable to start treatment within 4 weeks of PET/CT (two patients), and an ECOG performance status (PS) of 2 (two patients). An overview of the patient and tumour characteristics is given in Table 2.

Quality control of target volume delineation

All participating centres completed the first step of the QC procedure, in which the first three patients were accepted in the PERTAIN study. In total, 35 patients were submitted for TVD QC. Nine patients (26%) were excluded after evaluation. The reasons for not accepting the TVD as acceptable were: incorrect staging (two patients), involved lymph nodes not included (two patients), tumours too large to treat radically (≥60 Gy) without exceeding dose constraints (three patients), and noncompliance with IAEA study guidelines (two patients). All other patients in the prospective cohort were accepted for inclusion and treatment with concurrent CRT.

Table 2. Patient and tumour characteristics.

	Retrospective cohort	Prospective cohort	p value ^a
No. of patients	230	69	–
Mean age (range)	61 (31–86)	64 (43–86)	0.136
Gender			
Male	191 (83%)	57 (83%)	0.831
Female	39 (17%)	12 (17%)	
Smoker	182 (79%)	67 (97%)	<0.001
COPD	75 (33%)	47 (68%)	<0.001
ECOG performance status			
0	72 (31%)	21 (30%)	0.841
1	158 (69%)	48 (70%)	
Disease stage			
IIIA	145 (63%)	29 (42%)	0.021
IIIB	53 (23%)	27 (39%)	
IIIC	32 (14%)	13 (19%)	
T stage			
1	4 (2%)	3 (4%)	0.369
2	43 (19%)	11 (16%)	
3	87 (38%)	21 (30%)	
4	96 (42%)	34 (49%)	
N stage			
0	16 (7%)	4 (6%)	0.082
1	39 (17%)	4 (6%)	
2	135 (59%)	42 (61%)	
3	40 (17%)	19 (27%)	
Histology			
Squamous cell carcinoma	90 (39%)	32 (46%)	0.013
Adenocarcinoma	97 (42%)	35 (51%)	
Large cell carcinoma	15 (7%)	0 (0%)	
Not otherwise specified	28 (12%)	2 (3%)	

COPD chronic obstructive pulmonary disease, ECOG Eastern Cooperative Oncology Group

^aCalculated using the independent t test for continuous variables or the chi-squared test for categorical variables

Treatment parameters

In the retrospective cohort, 18 patients were treated with RT only (8%), 65 patients received sequential CRT (28%), and 147 patients (64%) received concurrent CRT. In contrast, all patients in the prospective cohort received concurrent CRT with curative intent. In both the retrospective and prospective cohorts various chemotherapy regimens were intravenously administered weekly: either carboplatin-based or cisplatin-based in combination with paclitaxel, etoposide, docetaxel, pemetrexed or gemcitabine. Of the 230 patients in the retrospective cohort, 32 (14%) were treated using an IMRT technique and 198 (86%)

using three-dimensional conformal RT (3DCT). By comparison, of the 69 patients in the prospective cohort, 29 (42%) were treated with IMRT, 2 (3%) with volumetric modulated arc therapy (VMAT), and 38 (55%) with 3DCT. The prescribed dose fractionation scheme varied between 50 Gy in 30 fractions and 70 Gy in 35 fractions in the retrospective cohort, with a mean prescribed dose of 61.4 ± 2.8 Gy. In the prospective cohort the dose fractionation scheme varied between 60 Gy in 30 fractions and 66 Gy in 33 fractions, with a mean prescribed dose of 60.7 ± 1.7 Gy.

Impact on survival

Prognostic factors were evaluated in the retrospective and prospective cohorts separately. In the retrospective cohort, age and ECOG PS were significant prognostic factors ($p = 0.039$ and 0.024 , respectively), and T stage demonstrated borderline significance ($p = 0.053$). In the prospective cohort, univariate Cox regression analysis showed no significant prognostic factors. No significant differences between the retrospective and the prospective cohorts in any of these prognostic variables were found, and therefore these variables were considered balanced. However, TNM staging was significantly higher in the prospective cohort than in the retrospective cohort ($p = 0.021$), and histological subtype was significantly different between the cohorts ($p = 0.013$). The difference in histological subtype was due to the absence of large-cell and lack of not otherwise specified types in the prospective cohort (see Table 2).

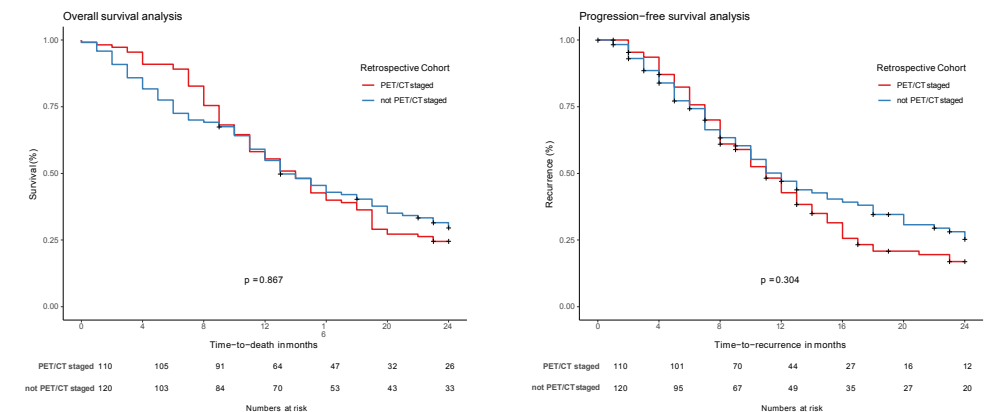


Fig. 1. Kaplan–Meier analysis of the difference in overall survival (left) and progression-free survival (right) in the retrospective cohort between patients who were and were not PET/CT-staged. No significant differences were observed in overall survival ($p = 0.867$) or progression-free survival ($p = 0.304$)

In the retrospective cohort, Kaplan–Meier analysis showed no significant differences in OS or PFS between patients who were and were not PET/CT-staged ($p=0.867$ and 0.304 , respectively; Fig. 1). Only 18.1% of the retrospective data were censored; in contrast, 52.2% of the prospective data were censored. Median survival was 14 months (95% CI 12–15 months) in the retrospective cohort and 23 months (95% CI 15–30 months) in the prospective cohort ($p=0.012$, log-rank test). Two-year OS was 27% in the retrospective cohort and 47% in the prospective cohort. The corresponding Kaplan–Meier analysis is shown in Fig. 2. Two-year PFS was 22% in the retrospective cohort and 45% in the prospective cohort.

Kaplan–Meier analysis of PFS in the retrospective and the prospective cohorts is shown in Fig. 3. Median time to progression was 11 months (95% CI 9–12 months) in the retrospective cohort and 17 months (95% CI 10–23 months) in the prospective cohort ($p=0.012$, log-rank test). Two-year PFS was 22% in the retrospective cohort and 45% in the prospective cohort.

Discussion

This study investigated the impact of introducing FDG PET/CT-guided concurrent CRT, supported by training and QC, on the OS in patients with stage III NSCLC. Preliminary results demonstrated a positive trend in a cohort comparison towards improved OS and PFS in the prospective cohort, suggesting a benefit from implementing FDG PET/CT-guided concurrent CRT in patients with stage III NSCLC in centres in LMIC with limited experience with PET/CT. TVD QC showed that IAEA study guidelines were implemented successfully in the clinic in 74% of patients. This demonstrates compliance with the study guidelines in the clinic, but also emphasizes the importance of QC in multi-centre trials to ensure compliance with the study protocol. Using TVD QC we were therefore able to confirm compliance with IAEA study guidelines in the clinic. This procedure led to the removal of five patients ineligible for curative CCRT who could otherwise have influenced the outcome of this study. Four issues were observed in TVD QC: incorrect staging, involved lymph nodes not included, tumours too large to treat radically without exceeding dose constraints, and tumours not delineated following the reGTV approach. The QC reviews continued to inform study participants during patient accrual and served as educational material when these issues occurred, which emphasizes once more the importance of QC during clinical studies.

Significant differences in TNM stage were observed between the retrospective and prospective cohorts. The retrospective cohort included predominantly stage IIIA patients, whereas the prospective cohort had a more balanced distribution of patients with stages IIIA, B and C. Despite a survival benefit of stage IIIA over IIIB and IIIC, results

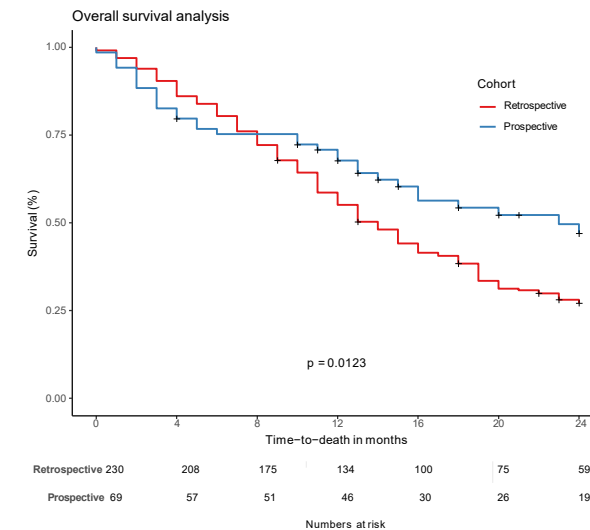


Fig. 2. Kaplan–Meier analysis of the difference in overall survival between the retrospective and the prospective cohorts. A survival benefit was observed in the prospective patient cohort ($p=0.012$)

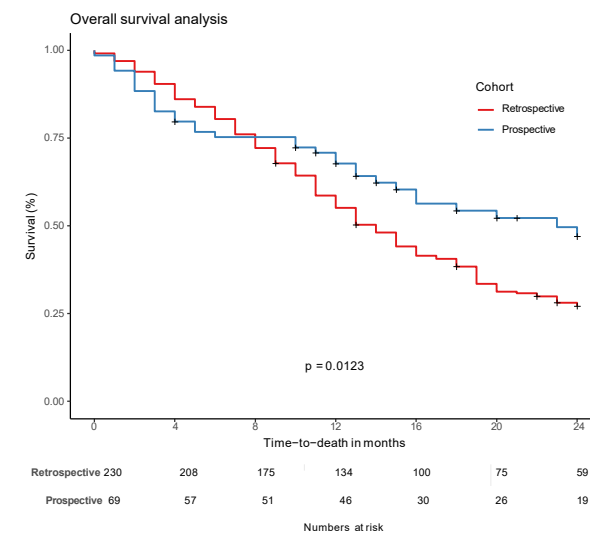


Fig. 3. Kaplan–Meier analysis of the difference in progression-free survival between the retrospective and the prospective cohort. A survival benefit was observed in the prospective patient cohort ($p=0.012$)

demonstrated better survival in the prospective cohort. Nevertheless, 48% of the patients in the retrospective cohort were not PET/CT-staged, and could have been incorrectly staged. Indeed, in all centres, a high percentage (up to 51%) of patients were upstaged to stage IV after staging with PET/CT became the standard.

The improved OS and PFS were possibly due to several factors. Besides the introduction of PET/CT for TVD and improved patient selection with FDG PET/CT, patients in the prospective cohort all received CCRT (64% in the retrospective cohort; 100% in the prospective cohort) and were also treated with more advanced RT techniques such as IMRT and VMAT (14% in the retrospective cohort; 45% in the prospective cohort), which could also have led to survival benefits [6]. Even so, there was no significant difference in the prescribed doses between the cohorts, and evidence is lacking on the survival benefit of IMRT/VMAT versus 3DCT in lung cancer patients [29]. On the other hand, more confidence was gained in PET/CT-based contouring, increasing delineation accuracy which hopefully resulted in reduced geographic miss of tumour. This may explain the improved local control seen in Fig. 2. In addition, the PERTAIN trial improved or reaffirmed collaborative working relationships between nuclear medicine and radiation oncology departments in the participating centres. This collaboration may not only have led to improved delineation, but also may have improved patient management by streamlining the patient pathway from diagnosis to treatment. Evaluation of the impact of this collaboration on outcome was beyond the scope of this study, and further research is required to obtain definitive evidence [30, 31].

In the PERTAIN trial there was a heterogeneous group of participating centres with different levels of experience in PET/CT scan acquisition. Furthermore, the training interventions were limited to RO and NMP chief scientific investigators. A train-the-trainers approach was used to disseminate the knowledge further in the departments involved in the PERTAIN study. The improvement in survival outcomes shown in this analysis suggests that this training approach had a clinically meaningful impact in the participating centres. We suggest that it is feasible to disseminate education regarding new radiation oncology techniques using the multiple intervention method we used [17].

One potential confounding impact in the comparison of outcomes between the cohorts may have been the impact of PET/CT staging alone. It is interesting to note that in the retrospective cohort, no significant difference in OS was observed between patients who were and were not PET/CT-staged (Fig. 1), and hence this confounding effect may have been negligible in this cohort, but it is acknowledged that because of the size of these groups, the study may not have been powered to detect a true difference. Another potential confounding factor was the selection bias that may have been present between the retrospective cohort and prospective cohort with regard to patients who died during treatment. This may

explain the worse survival seen in the first months in the prospective cohort. Five patients died during RT in the prospective cohort after being included in the analysis, and patients who died during treatment were not selected for the retrospective cohort. However, this would have had the effect of reducing the apparent survival difference between the two arms of the study. Another influence that may have contributed to worse survival in the first months may have been the higher incidence of smokers and patients with COPD in the prospective cohort.

Conclusion

The initial analysis of the PERTAIN study showed that a combined package of FDG PET/CT-planned RT, the routine use of concurrent CRT with training support, and a robust QC process led to improved OS and PFS in patients with stage III NSCLC patients in low- and middle-income countries.

Acknowledgements

The authors acknowledge the following persons in providing technical assistance for this study: Elba Etchebehere, Ana Brito, Nauman Amjad, Camila Mosci, Valdelania Lamounier, Flavia Gabriella, Diyaa Juaidi, Olga Morozova, Yaroslav Pynda, Nguyen Tan Chau, and Terez Sera. We would like to express our gratitude towards Vincent van der Noort for providing statistical support. The authors also would like to acknowledge Arturo Chiti, Ursula Nestle, and Vincent Grégoire for contributing to the design of the PERTAIN study and providing expert opinion on PERTAIN study guidelines.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet.* 2018;391(10125):1023–1075.
3. Ung YC, Bezjak A, Coakley N, Evans WK; the Lung Cancer Disease Site Group of Cancer Care Ontario. Positron emission tomography with 18fluorodeoxyglucose in radiation treatment planning for non-small cell lung cancer: a systematic review. *J Thorac Oncol.* 2011;6:86–97.
4. Gregory DL, Hicks RJ, Hogg A, Binns DS, Shum PL, Milner A, et al. Effect of PET/CT on management of patients with non-small cell lung cancer: results of a prospective study with 5-year survival data. *J Nucl Med.* 2012;53:1007–1015.
5. Badar F, Mahmood S, Yusuf MA, Sultan F. Epidemiology of cancers in Lahore, Pakistan, 2010–2012: a cross-sectional study. *BMJ Open.* 2016;6:e011828.
6. Aupérin A, Le Péchoux C, Rolland E, Curran WJ, Furuse K, Fournel P, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small cell lung cancer. *J Clin Oncol.* 2010;28:2181–2190.
7. Liao ZX, Komaki RR, Thames HD, Jr, Liu HH, Tucker SL, Mohan R, et al. Influence of technologic advances on outcomes in patients with unresectable, locally advanced non-small-cell lung cancer receiving concomitant chemoradiotherapy. *Int J Radiat Oncol Biol Phys.* 2010;76(3):775–781.
8. Chun SG, Hu C, Choy H, Komaki RU, Timmerman RD, Schild SE, et al. Comparison of 3-D conformal and intensity modulated radiation therapy outcomes for locally advanced non-small cell lung cancer in NRG oncology/RTOG 0617. *Int J Radiat Oncol Biol Phys.* 2015;93(3 Suppl):S1–2.
9. Ahmad SS, Duke S, Jena R, Williams MV, Burnet NG. Advances in radiotherapy. *BMJ.* 2012;345:e7765.
10. International Commission on Radiation Units and Measurements . ICRU Report No. 50. Prescribing, Recording and Reporting Photon Beam Therapy. Washington, DC: International Commission on Radiation Units and Measurements; 1993.
11. Steenbakkers RJ, Duppen JC, Fitton I, Deurloo KE, Zijp L, Uitterhoeve AL, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'big brother' evaluation. *Radiother Oncol.* 2005;77:182–190.
12. Rasch C, Belderbos J, van Giersbergen A, De Kok I, Laura T, Boer M, et al. The influence of a multi-disciplinary meeting for quality assurance on target delineation in radiotherapy treatment preparation. *Int J Radiat Oncol.* 2009;75(3 Suppl):S452–S453.
13. Hanna G, McAleese J, Carson KJ, Stewart DP, Cosgrove VP, Eakin RL, et al. (18)F-FDG PET-CT simulation for non-small-cell lung cancer: effect in patients already staged by PET-CT. *Int J Radiat Oncol Biol Phys.* 2010;77:24–30.
14. Greco C, Rosenzweig K, Cascini GL, Tamburrini O. Current status of PET/CT for tumor volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC) *Lung Cancer.* 2007;57:125–134.
15. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys.* 2005;62:70–75.
16. Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, De Bernardi E, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: report of AAPM task group no. 211. *Med Phys.* 2017;44(6):e1–e42.
17. Konert T, Vogel W, MacManus MP, Nestle U, Belderbos J, Grégoire V, et al. PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014. *Radiother Oncol.* 2015;116:27–34.
18. Caldwell CB, Mah K, Ung YC, Danjoux CE, Balogh JM, Ganguli SN, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys.* 2001;51:923–931.
19. Konert T, Vogel WV, Everitt S, MacManus MP, Thorwarth D, Fidarova E, et al. Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol. *Radiother Oncol.* 2016;121:39–45.
20. Van de Steene J, Linthout N, de Mey J, Vinh-Hung V, Claassens C, Noppen M, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol.* 2002;62(1):37–49.
21. Bayne M, Hicks RJ, Everitt S, Fimmell N, Ball D, Reynolds J, et al. Reproducibility of “intelligent” contouring of gross tumor volume in non-small-cell lung cancer on PET/CT images using a standardized visual method. *Int J Radiat Oncol Biol Phys.* 2010;77:1151–1157.
22. Jameson MG, Kumar S, Vinod SK, Metcalfe PE3, Holloway LC. Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. *Radiother Oncol.* 2014;112:332–6.
23. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol.* 2010;28(18):2996–3001.
24. Schimek-Jasch T, Troost EG, Rücker G, Prokic V, Avlar M, Duncker-Rohr V, et al. A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: reducing the interobserver variability in multicentre clinical studies. *Strahlenther Onkol.* 2015;191:525–533.
25. Spoelstra FO, Senan S, Le Péchoux C, Ishikura S, Casas F, Ball D, et al. Variations in target volume definition for postoperative radiotherapy in stage III non-small cell lung cancer: analysis of an international contouring study. *Int J Radiat Oncol Biol Phys.* 2010;76:1106–1113.
26. Nestle U, Rischke HC, Eschmann SM, Holl G, Tosch M, Miederer M, et al. Improved inter-observer agreement of an expert review panel in an oncology treatment trial – insights from a structured interventional process. *Eur J Cancer.* 2015;51:2525–2533.
27. Dettmerbeck FC, Boffa JB, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. *Chest.* 2017;151(1):193–203.
28. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumor imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42(2):328–354.
29. Ball D, Mac Manus M, Siva S, Plumridge N, Bressel M, Kron T. Routine use of intensity-modulated radiotherapy for locally advanced non-small-cell lung cancer is neither choosing wisely nor personalized medicine. *J Clin Oncol.* 2017;35(13):1492–1493.
30. Prabhu Das I, Baker M, Altice C, Castro KM, Brandys B, Mitchell SA. Outcomes of multidisciplinary treatment planning in US cancer care settings. *Cancer.* 2018;124(18):3656–3667.
31. Pillay B, Wootten AC, Crowe H, Corcoran N, Tran B, Bowden P, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: a systematic review of the literature. *Cancer Treat Rev.* 2016;42:56–72.

Supplementary Material S1

In the retrospective study, a total of four different eCRFs were used to collect patient data (Table 1). In Form 1, eligibility criteria of each patient were checked whereas Form 2 consisted of detailed information on delivered treatment. Form 3 was used in case the patient was alive and did not finish two-year follow up, and provided data on survival status. Otherwise, Form 4 was filled in and consisted of information on survival and cause of death. As for the prospective part, Form 2 was filled in before - and Form 3 after treatment, to check for possible discrepancies between prescribed and delivered treatment. Form 4 contained survival data on follow-up moments (3, 6, 9, 12, 24 months). Form 5 was filled in if the patient went off-study for any reason or if the patient completed 24 months of follow-up.

Table 1. Overview of the electronic case report forms used to collect patient data.

	Retrospective	Prospective
Form 1	Checking for eligibility: inclusion/exclusion criteria	Checking for eligibility: inclusion/exclusion criteria
Form 2	Treatment data	Pre-treatment data
Form 3	Follow up data	Treatment data
Form 4	Off-study data	Follow-up data
Form 5	n/a	Off-study data



RETROSPECTIVE	FORM 1 INITIAL DATA	
Identification		
01.01. Patient ID		
01.01.01 Country Code	Please select your centre...	
01.01.02 Site Sequence Number		
01.01.03 Patient initials	<small>(first letters of first and last name)</small>	
01.01.04 Date of birth	<small>(dd/mm/yyyy)</small>	
01.01.05 Gender of patient	<input type="radio"/> Male <input type="radio"/> Female	
01.01.06 Date of registration	<small>(dd/mm/yyyy)</small>	
01.01.07 Begin of the treatment	<small>(dd/mm/yyyy)</small>	
Inclusion Criteria		
01.02.01 Pathologically confirmed NSCLC	<input type="radio"/> Yes <input type="radio"/> No	
01.02.02 Stage III <small>(as per local diagnostic protocol or agreement of local multidisciplinary team)</small>	<input type="radio"/> Yes <input type="radio"/> No	
01.02.03 Suitable for treatment with a Radical Target Volume	<input type="radio"/> Yes <input type="radio"/> No <small>(in the opinion of a RC)</small>	
01.02.04 ECOG performance status 0 or 1	<input type="radio"/> Yes <input type="radio"/> No	
01.02.05 Available for clinical follow-up for at least 2 years	<input type="radio"/> Yes <input type="radio"/> No	
Exclusion Criteria		
01.03.01 Age under 18 years	<input type="radio"/> Yes <input type="radio"/> No	
01.03.02 Other neoplasms in the last 5 years	<input type="radio"/> Yes <input type="radio"/> No <small>(except non-melanoma skin cancer)</small>	
01.03.03 Uncontrolled diabetes mellitus (consistent blood sugar level > 100 mg/dL) or morning fasting blood glucose level (> 200 ng/dl)	<input type="radio"/> Yes <input type="radio"/> No	
01.03.04 Neo-adjuvant chemotherapy	<input type="radio"/> Yes <input type="radio"/> No	
01.03.05 Pregnant or breast feeding mother	<input type="radio"/> Yes <input type="radio"/> No	
Validate Data		Submit Data



RETROSPECTIVE		FORM 2 TREATMENT DATA	
Patient Demographics			
02.01. Patient ID			
02.01.01 Country Code	Please select your centre...		
02.01.02 Site Sequence Number			
02.01.03 Age at diagnosis		years	
02.01.04 Smoker	<input type="radio"/> Yes <input type="radio"/> No		
02.01.05 COPD	<input type="radio"/> Yes <input type="radio"/> No		
General Status			
02.02.01 ECOG Performance score	<input type="radio"/> PS0 <input type="radio"/> PS1 <input type="radio"/> PS2 <input type="radio"/> PS3		
02.02.02 Height		cm	
02.02.03 Actual weight		kg	
Diagnostic Surgical Procedures			
02.03.01 Performed	<input type="radio"/> Yes <input type="radio"/> No		
02.03.02 Please specify diagnostic surgical procedures			
Histology			
02.04.01 Proven NSCLC	<input type="radio"/> Yes <input type="radio"/> No		
02.04.02 Date of histological assessment		(dd/mm/yyyy)	
02.04.03 Histopathological type	<input type="radio"/> SQ <input type="radio"/> AC <input type="radio"/> Large Cell <input type="radio"/> Other		
02.04.04 Please specify histopathological type			
Tumour			
02.05.01 T-stage	T-St <input type="text"/>		
02.05.02 N-stage	N-St <input type="text"/>		
02.05.03 M-stage	M-St <input type="text"/>		
02.05.04 Stage	Stag <input type="text"/>		
Treatment			
02.06.01 Treatment completed as planned	<input type="radio"/> Yes <input type="radio"/> No		
02.06.02 If no chemotherapy given, specify reason	<input type="radio"/> Acute Toxicity <input type="radio"/> Compliance <input type="radio"/> Comorbidity <input type="radio"/> Other		
02.06.03 Please describe further the reason(s)			
Chemotherapy			
02.07.01 Chemotherapy given	<input type="radio"/> Yes <input type="radio"/> No		
02.07.02 Please specify reason	<input type="radio"/> Age <input type="radio"/> Co.Morb. <input type="radio"/> Other		
02.07.03 Schedule	<input type="radio"/> Neoadjuvant <input type="radio"/> Concurrent <input type="radio"/> Adjuvant		
02.07.04 Please specify regimen			
EBRT			
02.08.01 Date EBRT started		(dd/mm/yyyy)	
02.08.02 Date EBRT ended		(dd/mm/yyyy)	
02.08.03 Technique	<input type="radio"/> 3-DCT <input type="radio"/> IMRT <input type="radio"/> VMAT		
02.08.04 Prescribed total dose		Gy	
02.08.05 Delivered total dose		Gy	
02.08.06 Number of fractions			
02.08.07 Dose per fraction		Gy	
02.08.08 Dmax		Maximum dose received by 2cc of tissue in Gy	
02.08.09 D95		Percentage of PTV receiving 95% of the prescribed dose	
02.08.10 Elective nodal fields inclusion	<input type="radio"/> Yes <input type="radio"/> No		
02.08.11 Elective nodal field area			

IAEA-CRP E13042 "PERTAIN" - Case Report Forms



Vital Status	
02.09.01 Date of last information on survival status	(dd/mm/yyyy)
02.09.02 Survival status	<input type="radio"/> Alive without progression or relapse <input type="radio"/> Alive with progression or relapse <input type="radio"/> Dead
02.09.03 Any evidence of relapse?	<input type="radio"/> Yes <input type="radio"/> No
02.09.04 Date of recurrence	(dd/mm/yyyy)
02.09.05 Type of recurrence	<input type="checkbox"/> Local <input type="checkbox"/> Regional <input type="checkbox"/> Distant <input type="checkbox"/> Unknown
02.09.06 Cause of death	<input type="radio"/> Progression <input type="radio"/> Treatment related toxicity/morbidity <input type="radio"/> Both <input type="radio"/> Other
02.09.07 Please specify	
<input type="button" value="Validate Data"/> <input type="button" value="Submit Data"/>	

IAEA-CRP E13042 "PERTAIN" - Case Report Forms



RETROSPECTIVE		FORM 3 FOLLOW-UP DATA	
Follow-Up			
03.01. Patient ID			
03.01.01 Country Code	Please select your centre...		
03.01.02 Site Sequence Number			
03.01.03 Date of Completion			(dd/mm/yyyy)
03.01.04 Evaluating investigator initials			
03.01.05 Date of follow-up			(dd/mm/yyyy)
03.01.06 Evaluation moment	Evaluation mom		months
Recurrence			
03.02.01 Recurrent disease detected	<input type="radio"/> Yes <input type="radio"/> No		
03.02.02 Date of recurrence			(dd/mm/yyyy)
03.02.03 Type of recurrence	<input type="checkbox"/> Local <input type="checkbox"/> Regional <input type="checkbox"/> Distant		
03.02.04 Survival status	<input type="radio"/> Alive <input type="radio"/> Dead		
03.02.05 Date of last information on survival status			(dd/mm/yyyy)
03.02.06 Cause of death	<input type="radio"/> Progression <input type="radio"/> Treatment related toxicity/morbidity <input type="radio"/> Both <input type="radio"/> Other		
03.02.07 Please specify			
Validate Data		Submit Data	



RETROSPECTIVE		FORM 4 OFF STUDY DATA	
Off Study			
04.01. Patient ID			
04.01.01 Country Code	Please select your centre...		
04.01.02 Site Sequence Number			
04.01.03 Date off study			(dd/mm/yyyy)
04.01.04 Reason	<input type="radio"/> Lost to follow-up <input type="radio"/> Recurrence <input type="radio"/> Death <input type="radio"/> Other		
04.01.05 Please specify			
Survival Status			
04.02.01 Survival status	<input type="radio"/> Alive <input type="radio"/> Dead		
04.02.02 Date of last information on survival status			(dd/mm/yyyy)
04.02.03 Cause of death	<input type="radio"/> Progression <input type="radio"/> Treatment related toxicity/morbidity <input type="radio"/> Both <input type="radio"/> Other		
04.02.04 Please specify			
Validate Data		Submit Data	



FORM 1 INITIAL DATA	
Identification	
01.01. Patient ID	
01.01.01 Country Code	Please select your centre...
01.01.02 Site Sequence Number	
01.01.03 Patient initials	<small>(first letters of first and last name)</small>
01.01.04 Date of birth	<small>(dd/mm/yyyy)</small>
01.01.05 Gender of patient	<input type="radio"/> Male <input type="radio"/> Female
01.01.06 Date of registration	<small>(dd/mm/yyyy)</small>
01.01.07 Begin of the treatment	<small>(dd/mm/yyyy)</small>
Inclusion Criteria	
01.02.01 Pathologically confirmed NSCLC	<input type="radio"/> Yes <input type="radio"/> No
01.02.02 Stage III <small>(as per local diagnostic protocol or agreement of local multidisciplinary team)</small>	<input type="radio"/> Yes <input type="radio"/> No
01.02.03 Suitable for treatment with a Radical Target Volume	<input type="radio"/> Yes <input type="radio"/> No <small>(in the opinion of a RO)</small>
01.02.04 ECOG performance status 0 or 1	<input type="radio"/> Yes <input type="radio"/> No
01.02.05 Ability to start radiotherapy within 4 weeks	<input type="radio"/> Yes <input type="radio"/> No <small>(of acquisition of the RTP PET/CT)</small>
01.02.06 Stage III confirmed by FDG PET/CT	<input type="radio"/> Yes <input type="radio"/> No
01.02.07 Suitable for concurrent chemoradiotherapy	<input type="radio"/> Yes <input type="radio"/> No
Exclusion Criteria	
01.03.01 Age under 18 years	<input type="radio"/> Yes <input type="radio"/> No
01.03.02 Inability to provide informed consent	<input type="radio"/> Yes <input type="radio"/> No
01.03.03 Other neoplasms in the last 5 years	<input type="radio"/> Yes <input type="radio"/> No <small>(except non-melanoma skin cancer)</small>
01.03.04 Uncontrolled diabetes mellitus (consistent blood sugar level > 100 mg/dL) or morning fasting blood glucose level (> 200 ng/dl)	<input type="radio"/> Yes <input type="radio"/> No
01.03.05 Stage IV disease diagnosed before acquisition of staging PET/CT	<input type="radio"/> Yes <input type="radio"/> No
01.03.06 Tuberculosis	<input type="radio"/> Yes <input type="radio"/> No
01.03.07 Pregnant or breast feeding mother	<input type="radio"/> Yes <input type="radio"/> No
<input type="button" value="Validate Data"/> <input type="button" value="Submit Data"/>	



FORM 2 PRE-TREATMENT DATA	
Patient Demographics	
02.01 Patient ID	
02.01.01 Country Code	Please select your centre...
02.01.02 Site Sequence Number	
02.01.03 Age at diagnosis	<small>years</small>
02.01.04 Smoker	<input type="radio"/> Yes <input type="radio"/> No
02.01.05 COPD	<input type="radio"/> Yes <input type="radio"/> No
General Status	
02.02.01 ECOG Performance score	<input type="radio"/> PS0 <input type="radio"/> PS1 <input type="radio"/> PS2 <input type="radio"/> PS3
02.02.02 Height	<small>cm</small>
02.02.03 Actual weight	<small>kg</small>
Diagnostic Surgical Procedures	
02.03.01 Performed	<input type="radio"/> Yes <input type="radio"/> No
02.03.02 Please specify diagnostic surgical procedures	
Histology	
02.04.01 Proven NSCLC	<input type="radio"/> Yes <input type="radio"/> No
02.04.02 Date of histological assessment	<small>(dd/mm/yyyy)</small>
02.04.03 Histopathological type	<input type="radio"/> SQ <input type="radio"/> AC <input type="radio"/> Large Cell <input type="radio"/> Other
02.04.04 Please specify histopathological type	
Tumour	
02.05.01 T-stage	<input type="radio"/> T0 <input type="radio"/> T1 <input type="radio"/> T2 <input type="radio"/> T3 <input type="radio"/> T4
02.05.02 N-stage	<input type="radio"/> N0 <input type="radio"/> N1 <input type="radio"/> N2 <input type="radio"/> N3
02.05.03 M-stage	<input type="radio"/> M0 <input type="radio"/> M1
02.05.04 Stage	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3A <input type="radio"/> 3B <input type="radio"/> 4
CT	
02.06.01 Performed	<input type="radio"/> Yes <input type="radio"/> No
02.06.02 Date of diagnostic CT	<small>(dd/mm/yyyy)</small>
Staging PET/CT	
02.07.01 Performed	<input type="radio"/> Yes <input type="radio"/> No
02.07.02 Date of diagnostic PET/CT	<small>(dd/mm/yyyy)</small>
02.07.03 Performed in treatment position	<input type="radio"/> Yes <input type="radio"/> No
<input type="button" value="Validate Data"/> <input type="button" value="Submit Data"/>	



FORM 3 TREATMENT DATA	
Overall Assessment	
03.01 Patient ID	
03.01.01 Country Code	Please select your centre...
03.01.02 Site Sequence Number	
03.01.03 Treatment completed as planned	<input type="radio"/> Yes <input type="radio"/> No
03.01.04 Specify Reason	<input type="radio"/> Acute Toxicity <input type="radio"/> Compliance <input type="radio"/> Comorbidity <input type="radio"/> Other
03.01.05 Please describe further the reason(s)	
EBRT	
03.02.01 Date EBRT started	(dd/mm/yyyy)
03.02.02 Date EBRT ended	(dd/mm/yyyy)
03.02.03 Technique	<input type="radio"/> 3-DCT <input type="radio"/> IMRT <input type="radio"/> VMAT
03.02.04 Prescribed total dose	Gy
03.02.05 Delivered total dose	Gy
03.02.06 Number of fractions	
03.02.07 Dose per fraction	Gy
03.02.08 Dmax	Maximum dose received by 2CC of tissue in Gy
03.02.09 D95	Percentage of PTV receiving 95% of the prescribed dose
03.02.10 Elective nodal fields inclusion	<input type="radio"/> Yes <input type="radio"/> No
03.02.11 Elective nodal field area	
Chemotherapy	
03.03.01 Chemotherapy given <input type="radio"/> Yes <input type="radio"/> No	
03.03.02 Please specify reason <input type="radio"/> Age <input type="radio"/> Co.Morb. <input type="radio"/> Other	
03.03.03 Schedule <input type="radio"/> Neoadjuvant <input type="radio"/> Concurrent <input type="radio"/> Adjuvant	
03.03.04 Please specify regimen	
Radiation Treatment Planning	
03.04.01 PET/CT in treatment position for RTP <input type="radio"/> Yes <input type="radio"/> No	
Validate Data	Submit Data



FORM 4 FOLLOW-UP DATA	
Follow-Up	
04.01 Patient ID	
04.01.01 Country Code	Please select your centre...
04.01.02 Site Sequence Number	
04.01.03 Date of Form Completion	Today (dd/mm/yyyy)
04.01.04 Evaluating investigator initials	
04.01.05 Date of follow-up	(dd/mm/yyyy)
04.01.06 Evaluation moment	Evaluation momen
Recurrence	
04.02.01 Recurrent disease detected <input type="radio"/> Yes <input type="radio"/> No	
04.02.02 Date of recurrence	(dd/mm/yyyy)
04.02.03 Type of recurrence <input type="checkbox"/> Local <input type="checkbox"/> Regional <input type="checkbox"/> Distant	
04.02.04 Survival status <input type="radio"/> Alive <input type="radio"/> Dead	
04.02.05 Date of last information on survival status (dd/mm/yyyy)	
04.02.06 Cause of death <input type="radio"/> Progression <input type="radio"/> Treatment related toxicity/morbidity <input type="radio"/> Both <input type="radio"/> Other	
04.02.07 Please specify	
Validate Data	Submit Data



FORM 5 OFF STUDY DATA	
Off Study	
05.01 Patient ID	
05.01.01 Country Code	Please select your centre...
05.01.02 Site Sequence Number	
05.01.03 Date off study	{dd/mm/yyyy}
05.01.04 Reason	<input type="radio"/> Lost to follow-up <input type="radio"/> Recurrence <input type="radio"/> Death <input type="radio"/> Other
05.01.05 Please specify	
Survival Status	
05.02.01 Survival status	<input type="radio"/> Alive <input type="radio"/> Dead
05.02.02 Date of last information on survival status	{dd/mm/yyyy}
05.02.03 Cause of death	<input type="radio"/> Progression <input type="radio"/> Treatment related toxicity/morbidity <input type="radio"/> Both <input type="radio"/> Other
05.02.04 Please specify	
Validate Data	Submit Data



Chapter 6

Robust, independent, and relevant prognostic

^{18}F -fluorodeoxyglucose positron emission tomography radiomics features in non-small cell lung cancer: are there any?

Tom Konert
Sarah Everitt
Matthew D. La Fontaine
Jeroen B. van de Kamer
Michael P. MacManus
Wouer V. Vogel
Jason Callahan
Jan-Jakob Sonke

6

Abstract

In locally advanced lung cancer, established baseline clinical variables show limited prognostic accuracy and ^{18}F -fluorodeoxyglucose positron emission tomography (FDG PET) radiomics features may increase accuracy for optimal treatment selection. Their robustness and added value relative to current clinical factors are unknown. Hence, we identify robust and independent PET radiomics features that may have complementary value in predicting survival endpoints. A 4D PET dataset ($n=70$) was used for assessing the repeatability (Bland-Altman analysis) and independence of PET radiomics features (Spearman rank: $|\rho| < 0.5$). Two 3D PET datasets combined ($n=252$) were used for training and validation of an elastic net regularized generalized logistic regression model (GLM) based on a selection of clinical and robust independent PET radiomics features (GLM_{all}). The fitted model performance was externally validated ($n=40$). The performance of GLM_{all} (measured with area under the receiver operating characteristic curve, AUC) was highest in predicting 2-year overall survival (0.66 ± 0.07). No significant improvement was observed for GLM_{all} compared to a model containing only PET radiomics features or only clinical variables for any clinical endpoint. External validation of GLM_{all} led to AUC values no higher than 0.55 for any clinical endpoint. Robust independent FDG PET radiomics features did not have complementary value in predicting survival endpoints in lung cancer patients. Improving risk stratification and clinical decision making based on clinical variables and PET radiomics features has still been proven difficult in locally advanced lung cancer patients.

Introduction

Despite the emergence of new technologies and treatment options such as tyrosine kinase inhibitors targeted towards mutations, and immune checkpoint inhibitors, the global survival of lung cancer patients has improved only gradually in the last decades [1-4]. Locally advanced non-small cell lung cancer (NSCLC) is a highly heterogeneous disease where only modest improvements in survival have been observed, with the exception of chemoradiotherapy (CRT) patients treated with the anti-PD-L1 antibody Durvalumab whose overall and progression-free survival significantly improved compared to those receiving CRT alone [5]. New approaches are urgently needed for the selection of treatment strategies for NSCLC patients, which are currently determined mainly by TNM staging [6, 7]. In addition to TNM staging, other well-established, reproducible, independent prognostic factors are used to guide clinicians in making treatment decisions, such as Eastern Cooperative Oncology Group (ECOG) performance status [8, 9], weight loss [10], and gender [11]. Numerous other biomarkers have been investigated, although less reproducible, such as histology [12], age [13], serum blood levels [14, 15], mutation status [16], and protein expression levels [17, 18]. In locally advanced NSCLC, treatment selection based on TNM staging and other clinical variables may not be accurate enough for survival probability prediction [19, 20]. Therefore, the search for more accurate reproducible independent prognostic factors is warranted in the context of personalized medicine.

A current field of interest is the assessment of quantitative image features and its complementary value to well-established clinical prognostic models. Radiomics has been introduced as a sophisticated way to extract and mine a large number of quantitative image features, primarily using anatomical CT information [21]. The basic assumption of radiomics is that underlying tumour biology could be captured [22]. This information may actually be better characterized with functional imaging such as ^{18}F -fluorodeoxyglucose Positron Emission Tomography (FDG PET), the gold standard in NSCLC diagnosis and staging, which is able to characterize molecular heterogeneity in lung cancer [23,24]. It is therefore worthwhile to investigate the prognostic performance of radiomics features from functional imaging such as PET.

Basic PET radiomics features have provided clinically relevant prognostic information for NSCLC patients. Examples include standardized uptake value (SUV) based metrics like maximum, peak, and mean SUV (SUV_{max} , SUV_{peak} , and SUV_{mean} , respectively), metabolic tumour volume (MTV), and total lesion glycolysis (TLG) [25-32]. The more advanced PET texture features employed for quantification of tumour heterogeneity, have also been reported to be of prognostic value [33-41]. However, the variable nature of PET imaging makes it difficult to reproduce these results [42,43].

Furthermore, PET texture features can also be subject to differences in reconstruction settings and delineation methods [44], SUV binning methods [45,46], and feature calculation methods [47]. It is not yet clear which PET radiomics features are insensitive to all of these factors, and also to what degree.

Regardless of the issues with variability, complementary PET radiomics features should be independent from well-known prognostic SUV metrics, such as MTV and SUV_{max} . Some investigators reported specific PET texture features that were associated with MTV [37,39,47-49]. In these cases, prognostic texture features would rather act as a surrogate than as an independent variable. Such an association is also not warranted for clinical variables. Hence, the relationship of PET texture features with well-known prognostic factors has to be thoroughly studied too.

With all the confounding factors described above, in combination with the high number of possible radiomics features, it is not surprising that false discovery rates are high amongst FDG PET and CT studies on texture features [18]. Without proven, robust, and independent prognostic PET texture features, it will be challenging to move further in the field. Therefore, this study aims to investigate the repeatability of PET radiomics features, and also assesses the relationship with well-known prognostic factors in PET, such as MTV and SUV_{max} . The rationale is to identify a group of radiomics features derived from pre-treatment PET imaging that are robust, independent, and prognostic, with possible additional value to current clinical prognostic variables.

Materials and methods

Patient data

Three NSCLC patient cohorts from the Netherlands Cancer Institute (NKI) and one from the Peter MacCallum Cancer Centre (PMCC) were included in this study to develop and validate a radiomics signature. Peter MacCallum Cancer Centre Ethics and Clinical Research Committees approval was granted and all research was performed in accordance with relevant guidelines/regulations. Patient's written, informed consent was obtained. An overview of the datasets is given in Table 1. Patients were excluded if the primary tumour was smaller than 10 cc or if the patient had stage IV NSCLC at baseline. To detect brain metastases at baseline, the NKI patients were scanned with MR imaging and the PMCC performed FLT baseline scans before treatment.

Table 1. Overview of the four patient cohorts used in the study. Unless otherwise stated, values represent the median with the range in parentheses.

	4D PET lung	NKI lung1	NKI lung2	PMCC lung1
No. of patients	70	228	24	40
Age (year)	n/a	64 (36-87)	63 (39-82)	68 (53-86)
Gender				
Male	n/a	142	13	29
Female		86	11	11
Disease stage		IA-IIIIC	IIB-IIIIC	IB-IIIIC
IA		1	0	0
IB		4	0	1
IIA	n/a	0	0	3
IIB		16	2	5
IIIA		102	7	15
IIIB		82	13	12
IIIIC		23	2	4
T stage				
1		6	1	5
2	n/a	78	4	24
3		63	7	9
4		81	12	2
N stage				
0		42	4	5
1	n/a	20	2	5
2		130	13	17
3		36	5	13
Histology				
Adeno		80	12	16
Squamous cell	n/a	83	7	13
Large cell		8	1	5
Nos or other		57	4	6
GTV (cc)	n/a	118 (15-906)	84 (10-351)	49 (12-544)
MTV _{2.5} (cc)	62 cc (10-545)	72 (10-693)	91 (11-337)	51 (8-478)
MTV ₄₀ (cc)	27 cc (4-169)	31 (3-394)	34 (5-289)	31 (4-378)
SUV _{max}	11.5 (4.3-55.1)	14.6 (5.9-44.8)	15.7 (6.9-28.3)	16.5 (6.3-33.2)
Median follow-up time (months)	n/a	17	22	24
2-year OS	n/a	40%	46%	54%
2-year PFS	n/a	29%	21%	20%
1-year PFS	n/a	50%	54%	35%
1-year LRS	n/a	58%	71%	47%
1-year DMS	n/a	54%	54%	45%

MTV_{2.5}, metabolic tumour volume obtained using a SUV threshold of 2.5; MTV₄₀, metabolic tumour volume obtained using a threshold of 40% of the maximum intensity; SUV_{max}, maximum SUV uptake; OS, overall survival; PFS, progression-free survival; LRS, local recurrence-free survival; DMS, distant metastases-free survival; Nos, not otherwise specified.

The repeatability and independence of PET radiomics features was assessed using a 4D PET/CT dataset (4D PET lung) consisting of 70 stage III NSCLC patients. No clinical data was collected for these patients. The second cohort (NKI lung 1) contained 228 patients treated with concurrent chemoradiotherapy (CCRT) for stage IA-IIIC NSCLC in the NKI between 2007 and 2011 as described earlier [51]. The third cohort, also from the NKI (NKI lung 2), consisted of 24 patients with stage IIB-IIIC NSCLC treated between 2013 and 2016, similar as NKI lung 1. The fourth cohort was from the PMCC (PMCC lung 1) and involved 40 stage IB-IIIC NSCLC patients treated with CCRT as previously reported [32].

Clinical endpoints for prognostic model

The primary endpoint used for the prognostic model was two-year overall survival (2-year OS). Overall survival was defined as the time between the start of treatment and date of death. In addition, two-year progression-free survival (2-year PFS), one-year PFS (1-year PFS), one-year local recurrence-free survival (1-year LRS), and one-year distant metastases-free survival (1-year DMS) were also studied. Progression was defined as growth of tumour cells in the primary tumour or involved lymph nodes, or metastases to other organs, or death. LRS was defined as progression in the primary tumour and/or involved lymph nodes as assessed on follow-up scans. DMS was described according to the 8th edition of the TNM classification for NSCLC [52] as evaluated on follow-up scans.

Data acquisition and image reconstruction

Patients from the NKI lung 1 and 2 dataset both underwent a whole-body FDG PET/CT using a Gemini TF or Gemini TF Big Bore scanner (Philips Medical Systems, Cleveland, OH). The reconstruction voxel size of the PET data was $4 \times 4 \times 4 \text{ mm}^3$. Patients fasted for at least 8 h to ensure low levels of serum glucose. Patients with a Body Mass Index (BMI) ≤ 28 were intravenously injected with 190 MBq ^{18}F -FDG, or 240 MBq in case of a BMI > 28 . Patients were scanned 60 minutes after injection of ^{18}F -FDG. The acquisition time of the PET/CT scanner was 2 minutes per bed position.

In the PMCC lung 1 cohort, whole-body FDG PET/CT scans were acquired on a GE STE (GE Medical Systems, Milwaukee, WI) or Biograph (Siemens Medical Solutions, Erlangen, Germany) scanner. The reconstructed voxel size of the PET data was $4.3 \times 4.3 \times 3.3 \text{ mm}^3$ for the GE STE scanner, and $4.1 \times 4.1 \times 3.0 \text{ mm}^3$ for the Siemens Biograph scanner. Patients fasted for more than 6 hours before ^{18}F -FDG scans. Patients were intravenously injected with 4.2 MBq/kg ^{18}F -FDG. Baseline emission scans were initiated 60 minutes after injection. The acquisition time of the PET/CT scanner was 3 minutes per bed position.

For the 4D PET lung dataset, scans were acquired on a Gemini TF scanner (Philips Medical Systems, Cleveland, OH). The reconstruction voxel size of the PET data was $4 \times 4 \times 4 \text{ mm}^3$. The 4D PET/CT data were reconstructed in 10 phases, and the attenuation in each frame of the 4D PET data was corrected with the corresponding 4D CT frame. The acquisition time of the 4D PET was kept the same as that used for 3D PET [53].

Mid-position scans from 4D PET lung dataset for repeatability testing

The 4D PET/CT data were reconstructed in 10 phases, and from these phases two new mid-position scans were derived [53]. The first mid-position scan was created from the even phases (0, 2, 4, 6, and 8) and is named 'Mid-P even', and the odd phases (1, 3, 5, 7, and 9) were used to create the second mid-position scan 'Mid-P odd'. The even and odd number of frames were selected to keep the amount of tumour motion balanced in both scans. Figure 1 gives an overview of the workflow.

The source of variability was different in these two mid-position scans compared to a test-retest setting, since the biological tumour variability has been eliminated. In this case, the variability was mostly caused by minor differences in noise-levels and tumour motion, hence robust quantitative features should not differ substantially in outcome.

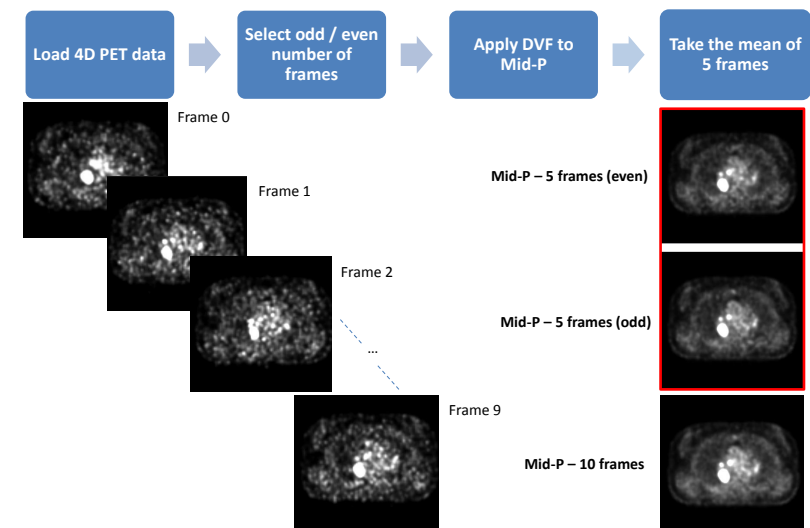


Figure 1. Workflow of the PET mid-position scans. A 4D PET scan was loaded for each patient consisting of 10 frames, where the odd or even number of frames were selected. A 4D deformation vector field (DVF) was applied to these frames to deform them to the mid-position. Lastly, the mean of the 5 deformed frames was calculated to obtain the PET mid-position scan. For comparison, the PET mid-position scan obtained from 10 frames has been included in the image too. Mid-P = PET mid-position scan.

Tumour segmentation

For each patient in the NKI lung 1, NKI lung 2, and PMCC lung 1 cohort, a volume-of-interest (VOI) enveloping the primary tumour was manually drawn by radiation oncologists using information from both PET and CT imaging. From this VOI, the MTV was auto-segmented on the FDG PET scan. Two auto-segmentation methods were applied: a metabolic tumour region delineation that included all voxel intensities above 2.5 ($SUV_{2.5}$), and a high intensity delineation that included all voxel intensities that were at least 40% of the SUV_{max} (SUV_{40}). Auto-segmentation was performed with in-house developed software named Match42 (version 1.0.0) using a Python plug-in. The metabolic tumour volume obtained from $SUV_{2.5}$ and SUV_{40} were named $MTV_{2.5}$ and MTV_{40} , respectively. In the 4D PET lung dataset, a VOI was manually drawn around the primary tumour in one PET mid-position scan, and copied to the second PET mid-position scan. The auto-segmentation was performed on both PET mid-position scans independently.

PET radiomics features

The Pyradiomics toolkit was used for radiomics feature extraction [54]. With this toolkit a total of 105 features were available for feature calculations. These were divided into 18 first-order features, 13 shape features, and 74 texture features describing the spatial distribution of voxel intensities. The texture features were derived from the gray level co-occurrence matrix (GLCM; 23 features) [55], gray level run-length matrix (GLRLM; 16 features) [56], gray level size-zone matrix (GLSZM; 16 features) [57], gray level dependence matrix (GLDM; 14 features) [58], and neighbourhood gray tone difference matrix (NGTDM; 5 features) [59]. The mathematical definitions of these features were in compliance with feature definitions as described by the Imaging Biomarker Standardization Initiative (IBSI) [60].

SUV discretization and matrix calculation

Before texture features were extracted, pre-processing steps were required in the form of SUV binning and matrix definition. SUV discretization is an intensity-resampling step, before building the texture matrices on which texture features rely. SUV discretization or binning was applied with the fixed bin count method (e.g. 64 bins) and an alternative method using a fixed bin width (e.g. 0.25 SUV). All texture features were calculated from a single matrix taking into account all 13 directions simultaneously. A more detailed description on SUV binning and matrix calculation can be found in Supplementary material S1 and S2, respectively.

Repeatability

The repeatability assessment was performed within the same patient comparing two different PET mid-position scans. For each patient, the PET mid-position scan obtained from the even numbered frames (Mid-P even) was compared with the PET mid-position scan from the odd numbered frames (Mid-P odd). This resulted in four comparisons: 2 SUV binning methods and 2 thresholding methods were applied.

The repeatability of each PET radiomics feature was assessed with the Coefficient of Repeatability (CR) [61]. See Supplementary material S3 for more details. The CR was reported as a percentage: $100\% \times \frac{CR}{mean}$, where *mean* is the average of the PET radiomics feature value within the patient cohort. The threshold for poor repeatability was set to a value of 30%, corresponding to PET Response Criteria in Solid Tumours (PERCIST) [62].

Independence testing

To determine whether the features were correlated with the two commonly reported prognostic PET features MTV and SUV_{max} , the Spearman's rank correlation coefficient (ρ) was calculated on one of the Mid-P scans, using the same set-up as for the repeatability testing. PET radiomics features that had a $|\rho| \geq 0.5$ were considered to have a correlation with MTV or SUV_{max} , and were discarded from further analysis. The choice of $|\rho| < 0.5$ as limit for independent features was validated with the 'elbow method' using hierarchical clustering [63]. An overview of the radiomics workflow and feature selection procedure is given in Figure 2.

Model training

An elastic net regularized generalized logistic regression model (GLM) was built with PET radiomics features derived from pre-treatment PET imaging (GLM_{rad}). To increase the sample size in the training and test sets, for the purpose of building a GLM, NKI lung 1 and lung 2 were combined. In this study, 80% of the NKI data was used for training the model, and 20% for validation. Different ratios of training/validation were also tested, but were not reported as there was no major differences seen in the results. Elastic net regression analysis using the R package 'glmnet' was performed on the training set [64]. With 20-fold cross validation (CV), the most optimal fitted GLM_{rad} with minimal CV error was determined and selected for model validation.

Model validation

To validate the fitted model of the training set, the area under the receiver operating characteristic curve (AUC) was calculated between the predicted outcome and the observed outcome in the validation set. To reduce randomness introduced by selecting a random subset of the complete data for training and validation, the procedure for model training and validation was repeated 100 times. This yields a better estimate of the true validation set

performance by randomly simulating many scenarios with varying training and validation set compositions [65]. From the 100-times-repeated training/validation procedure, results were averaged, and the best performing GLM_{rad} was externally validated for each clinical endpoint on PMCC lung 1.

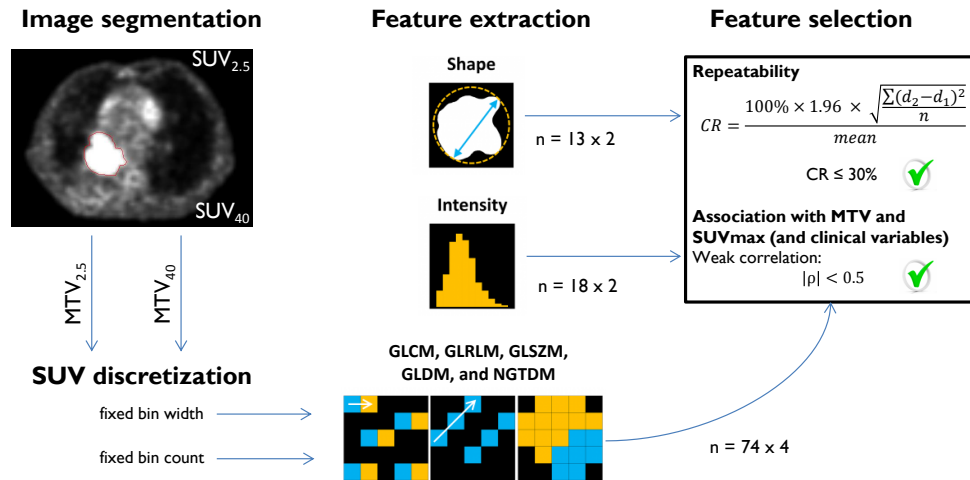


Figure 8. Radiomics feature selection workflow: from PET image segmentation to selected features. Features from $MTV_{2.5}$ and MTV_{40} were seen as a separate set of features, doubling the amount of features in the analysis. This also counts for features calculated with fixed bin width and fixed bin count, except for most intensity and shape features that were not affected by SUV discretization. An exception was observed for first-order features Uniformity and Entropy. A total of 360 PET radiomics features were entered into the analysis, including SUV_{max} , $MTV_{2.5}$, and MTV_{40} . PET radiomics features were selected for further analysis when two criteria were met: high repeatability and low association with MTV and SUV_{max} . $SUV_{2.5}$ = SUV threshold of 2.5; SUV_{40} = SUV threshold of 40% of maximum SUV; $MTV_{2.5}$ = metabolic tumour volume obtained from use of $SUV_{2.5}$; MTV_{40} = metabolic tumour volume obtained from use of SUV_{40} . GLCM = gray level co-occurrence matrix; GLRLM = gray level run-length matrix; GLSZM = gray level size-zone matrix; GLDM = gray level dependence matrix; NGTDM = neighbourhood gray tone difference matrix; CR = coefficient of repeatability.

During 100-times-repeated training/validation procedure, per iteration, the fitted model was stored to keep track of the PET radiomics features that were selected by elastic net in the fitted model [66]. PET radiomics features and clinical variables were ranked based on the frequency of inclusion in the fitted model.

Model comparison

Clinical variables such as PET/CT-based GTV, TNM staging, histology, gender, and age were also introduced into the radiomics signature to create a prognostic model containing PET radiomics features and clinical variables (GLM_{all}). In addition, a model based on only the clinical variables was calculated using elastic net regression (GLM_{clin}). To assess the

complementary value of PET radiomics features with clinical variables, the mean AUC was calculated from 100 iterations for each model and compared. The Mann Whitney U Test was used to assess any significant differences between the predictive performance of GLM_{all} , GLM_{clin} , and GLM_{rad} , and p-values below 0.05 were seen as significant.

Results

Repeatability

Results of the repeatability test were based on the 4D PET lung dataset and an overview of notable PET radiomics features and their corresponding CR is given in Table 2. All first-order features were repeatable when extracted from $MTV_{2.5}$ irrespective of SUV binning method. In contrast, 13 out of 18 first-order features were repeatable when extracted from MTV_{40} . Furthermore, around 50 texture features were repeatable when extracted from $MTV_{2.5}$ regardless of SUV discretization method, versus 28 repeatable texture features extracted from MTV_{40} . With regards to shape features, only MTV_{40} was not repeatable.

Amongst the four comparisons, 211 out of 360 PET radiomics features were repeatable. An overview of all PET radiomics features and their corresponding CR are given in Supplementary Table 1 and Supplementary Table 2. The impact of large delineation inaccuracies on repeatability was studied between contours generated by the two different SUV thresholds, though only reported as supplementary data (Supplementary material S4).

Relationship of PET radiomics features with MTV and SUV_{max}

The Spearman's Rank correlation coefficient was calculated to assess the relationship of 211 repeatable PET radiomics features with MTV and SUV_{max} . Four assessments were performed in total on one of the mid-position scans, with groups consisting of a combination of either one of the SUV binning methods and one of the tumour volumes ($MTV_{2.5}$ or MTV_{40}). Not all repeatable PET radiomics features were found to be independent from MTV and SUV_{max} . From the first-order features, only Kurtosis and Skewness extracted from $MTV_{2.5}$ were independent from MTV and SUV_{max} . There were no independent repeatable first-order features for MTV_{40} . Regarding the fixed bin count method, 17 out of 50 texture features extracted from $MTV_{2.5}$ were not strongly associated with MTV and SUV_{max} . This also counted for 5 texture features extracted from MTV_{40} . With regards to the fixed bin width method, there were no texture features independent from either SUV_{max} or MTV. Elongation, Flatness, and Sphericity were the only independent shape features when extracted from $MTV_{2.5}$, though only Elongation and Flatness remained independent for MTV_{40} . A complete overview of independence testing for all PET radiomics features are given in Supplementary Table 4-6.

Table 2. An overview of categorized notable PET radiomics features that are commonly reported in literature with their coefficient of repeatability (CR, %).

CR (%)	Notable features	Fixed bin width		Fixed bin count	
		MTV _{2.5}	MTV ₄₀	MTV _{2.5}	MTV ₄₀
First-order features	18/18	13/18	18/18	13/18	
	Entropy*	3.4	5.5	3.8	6.0
	Kurtosis	26.8	34.7	26.8	34.8
	Skewness	23.1	50.4	23.1	51.3
	SUVmax*	13.2	13.2	13.2	13.2
	SUVmean*	6.0	12.9	6.0	12.7
	Uniformity	17.9	41.9	21.1	37.0
Texture features	49/74	28/74	50/74	28/74	
	GLCM Contrast*	23.2	28.1	28.8	29.9
	GLCM Correlation*	2.6	11.9	2.7	11.2
	GLCM DifferenceAverage*	9.9	13.1	14.1	17.2
	GLCM JointEntropy*	2.8	4.5	3.3	5.7
	GLCM SumEntropy*	2.7	4.5	2.8	4.2
	GLRLM GrayLevelNonUniformity	13.7	59.1	18.7	55.4
	NGTDM Busyness	75.5	91.3	33.0	81.9
	NGTDM Coarseness	12.0	41.5	16.8	35.2
	NGTDM Contrast	23.3	68.5	31.9	64.9
Shape features	13/13	12/13	13/13	12/13	
	Elongation*	4.7	10.8	4.7	10.7
	Flatness*	7.1	15.0	7.1	13.7
	MetabolicTumourVolume	5.9	45.5	5.9	45.3
	Sphericity*	3.2	8.9	3.2	8.7

The asterisk (*) represents features that were repeatable in all four different settings. Per category, the total number of PET radiomics features that met the study repeatability criterion is added.

An overview of correlations amongst the selected robust independent PET radiomics features and clinical variables is given in Figure 3. More details on robust and independent PET radiomics features can be viewed in Supplementary Table 7. The robust independent PET radiomics features did not show any strong correlation with the other clinical variables, such as age, ECOG PS, gender, histology, and TNM stage. However, there were associations present amongst the PET texture features.

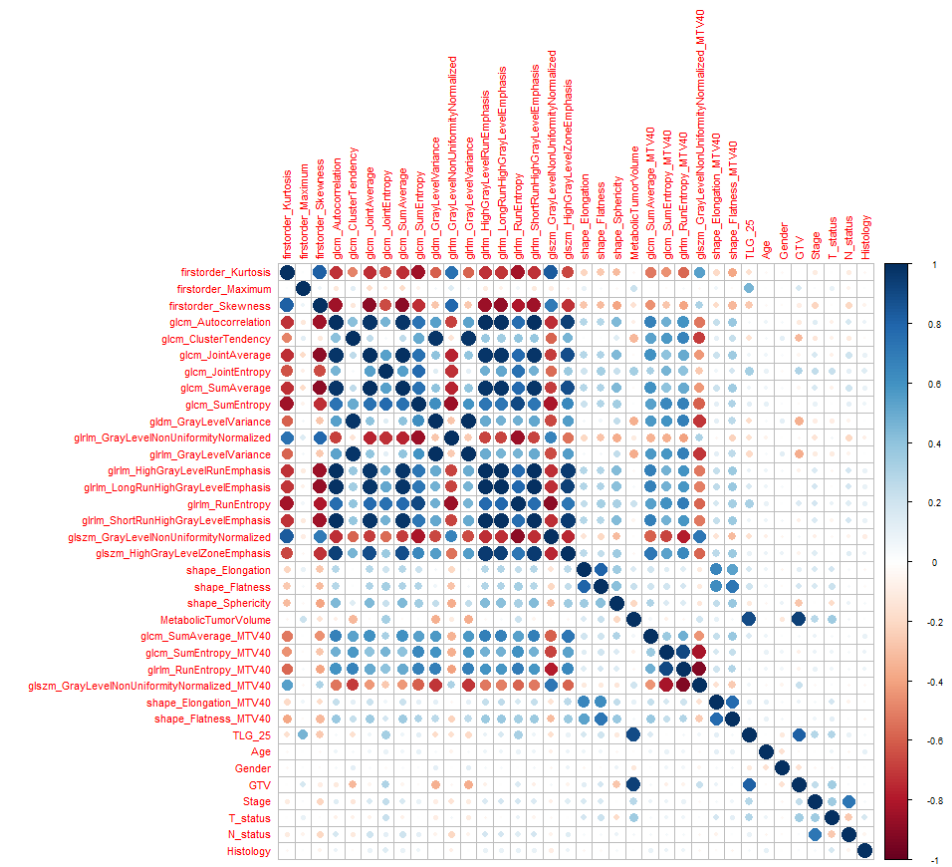


Figure 9. Correlation coefficients of the robust independent PET radiomics features and clinical variables. Positive correlation coefficients are displayed in blue and negative correlation coefficients in red color. Color intensity and the size of the circle are proportional to the correlation coefficients. A distinction was made between features calculated from MTV_{2.5} and MTV₄₀.

Building the radiomics signature

Based on the feature selection criteria, 31 PET radiomics features were selected for the next steps (see Figure 3). Three elastic net regularized GLMs were built per endpoint: GLM_{rad}, GLM_{clin}, and GLM_{all}. Results of the model performances are shown in Figure 4, showing that GLM_{rad} does not significantly outperform GLM_{clin} for any clinical endpoints. The GLM_{clin} has a significantly better predictive performance compared to GLM_{rad} in 2-year OS (p<0.0001), and in 1-year LRS (p<0.001). GLM_{all} did not show a significantly better performance to both

GLM_{rad} and GLM_{clin} simultaneously in any endpoint. External validation of GLM_{all} led to AUC values ranging from 0.51 to 0.59 for any clinical endpoint. When GLM_{clin} was externally validated, the highest predictive performance was 0.60 for 2 year OS. For GLM_{rad}, the highest predictive performance was 0.71 for 2-year PFS.

Promising features

Table 3 shows selected features for each fitted GLM, and how frequent these features were chosen in the fitted model over 100 iterations. The feature shape Sphericity was present in 100% of the iterations for 2-year OS. From the 100 repetitions, GLCM ClusterTendency was selected in more than 95% for predicting 1-year PFS and 1-year DMS. Clinical variables such as age and GTV were prominent in predicting 2-year OS and 1-year LRS, next to shape Sphericity. As can be seen in Table 3, age, shape Sphericity, and GLCM ClusterTendency are present amongst the most selected features for all clinical endpoints.

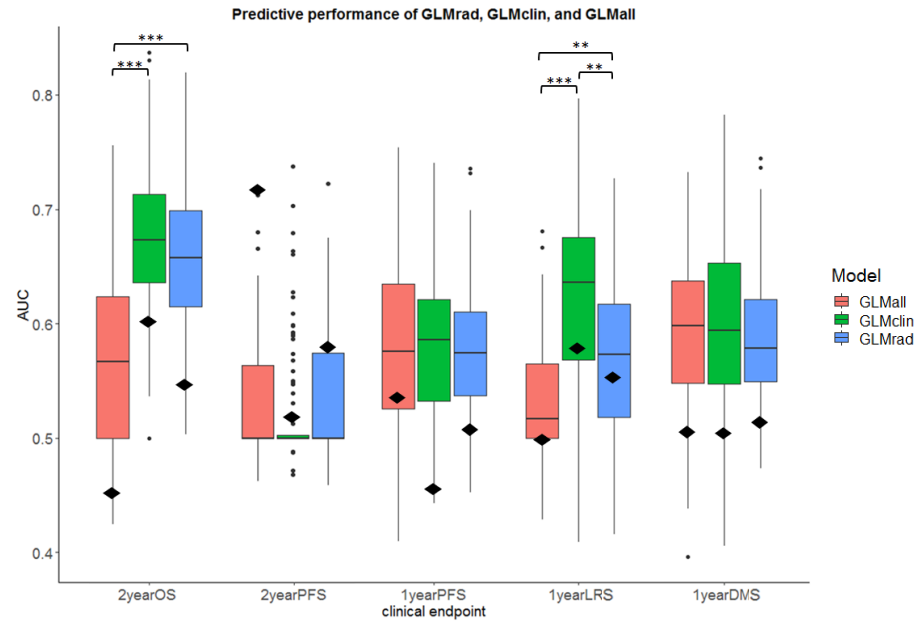


Figure 10. Model performance for the PET radiomics model (GLM_{rad}), the model containing clinical variables (GLM_{clin}), and a combination of radiomics and clinical variables (GLM_{all}). The median AUC values from 100-times-repeated training/validation are depicted per model, per clinical endpoint. The lower and upper hinges correspond to the 25th and 75th percentiles. The whiskers depict the 1.5*IQR from the lower and upper hinge. Data beyond the end of the whiskers are shown as outlier points. AUC values corresponding to the external validation set are shown as a black diamond. Significance levels, **p<0.001, ***p<0.0001.

Table 3. The most selected features in the model by elastic net, ranked by the number of times selected in the generalized linear model.

Endpoint	GLM _{all} selected features by elastic net	Frequency
2-year OS	Age	100
	GTV	100
	Shape_Sphericity	100
	MTV _{2.5}	78
	glcm_ClusterTendency	56
	SUV _{max}	39
	Gender	34
	glcm_JointEntropy	34
	glrlm_GrayLevelNonUniformityNormalized	33
	glrlm_GrayLevelVariance	29
2-year PFS	Age	50
	SUV _{max}	50
	glrlm_GrayLevelNonUniformityNormalized	49
	shape_Sphericity	47
	Histology	42
	MTV _{2.5}	38
	glcm_ClusterTendency	30
	N_status	28
	T_status	25
	shape_Elongation_MTV40	20
1-year PFS	GTV	99
	glcm_ClusterTendency	95
	shape_Sphericity	76
	Age	63
	T_status	49
	MTV _{2.5}	40
	glcm_SumEntropy_MTV40	39
	shape_Elongation	37
	SUV _{max}	31
	Histology	29
1-year LRS	GTV	83
	Age	82
	glcm_ClusterTendency	65
	glcm_SumEntropy_MTV40	63
	shape_Sphericity	57
	Gender	48
	glcm_GrayLevelVariance	47
	N_status	27
	Stage	25
	glrlm_GrayLevelVariance	24

1-year DMS	GTV	99
	<i>glcm_ClusterTendency</i>	96
	<i>shape_Sphericity</i>	57
	MTV _{2.5}	52
	Histology	34
	T_status	33
	Age	32
	glcm_SumEntropy_MTV40	32
	shape_Elongation	29
	SUV _{max}	23

Only the top 10 most selected PET radiomics are shown. The features written in italic bold are present in all endpoints.

Discussion

The rationale of this study was to identify a group of FDG PET radiomics features for NSCLC patients that are robust, independent, prognostic, and complementary to well-established clinical variables. We found PET radiomics features that met the study criteria of robustness and independence, and that also exhibited prognostic value. However, results demonstrated that PET radiomics features are not complementary to clinical variables for predicting clinical endpoints in NSCLC patients that were treated with CCRT. This indicates that clinical variables provide more prognostic information than robust independent PET radiomics features, and that the prognostic value in PET radiomics features is minimal. This study did take into account shortcomings of other studies on PET radiomics features [50] with the use of a feature selection method that reduces overfitting and external validation of results.

Feature selection based on the repeatability of PET radiomics features was feasible with the use of different phases from 4D PET imaging, in the absence of test-retest data. Larue et al. showed that in 4D CT, the majority of the features have a high agreement between radiomics feature stability based on 4D CT and test–retest data in lung cancer [67]. It was therefore hypothesized that 4D PET scans could also be used for repeatability testing. To determine robust PET radiomics features, a CR of 30% was chosen as limit for repeatability, based on PERCIST. However, a limitation of using 4D PET for repeatability testing is the absence of biological tumour variability, and PERCIST takes this variability into account. Hence, the use of a 30%-limit could be seen as too tolerant, and 15%, as commonly used in phantom studies, could be more appropriate. Even under these stricter circumstances, 12 first-order features, 24 out of 74 texture features, and all shape features would still meet that criterion as can be seen in Supplementary Table 2. Besides that, the most prominent PET radiomics features in the fitted GLMs were SUV_{max} (CR=13.2%), shape Sphericity (CR=3.2%), GLCM

ClusterTendency (CR=21.9%), GLRLM GrayLevelNonUniformityNormalized (CR=18.4%), and MTV_{2.5} (CR=5.9%) as seen in Table 3. This shows that repeatable PET radiomics features with a CR>15% are also frequently present in the fitted models. Even though there is literature reporting on stability of PET radiomics features in a test-retest setting [45,46], there is no objective limit for the level of repeatability for each PET radiomics feature. Determining such an objective limit is only relevant if the studied PET radiomics feature contains clinically useful information. Hence, in the absence of an objective limit for each PET radiomics feature, the 30%-limit of PERCIST was applied to all.

Another step of the feature selection procedure was to assess the independence of PET radiomics features, to identify possible prognostic features that could complement basic SUV metrics and volumetric features. In this context, changes in PET radiomics features would be independent from changes in basic SUV metrics and volume, increasing their utility in longitudinal studies. Therefore, the use of a fixed bin width for SUV binning should be avoided as this method resulted in PET radiomics features that were all strongly correlated to either maximum SUV or MTV. While the choice of $|\rho| < 0.5$ for independence testing may seem arbitrary, a $|\rho| < 0.7$ was also studied and did not improve results (see Supplementary material S7 for more details). Independence testing had the most impact in the feature pre-selection procedure as it resulted in a substantial decrease of PET radiomics features. Unfortunately, results demonstrated that independence testing could not guarantee that remaining robust independent PET radiomics features exhibited complementary value next to clinical variables. Even so, we strongly advise assessing the relationship of radiomics features with current established prognostic factors in any study considering PET radiomics features for prognostication as this is the first important step in showing their potential added value in the clinic.

A final selection of features in the GLM was performed by elastic net regression, robust to collinearity amongst features [66]. More feature selection/classification methods exist [68], though comparing multiple methods was beyond the scope of this study. However, in literature, elastic net regression yielded one of the highest discriminative performances in chemoradiotherapy outcome prediction in 12 patient datasets containing in total 1053 lung cancer patients [65]. Interestingly, elastic net regression could also be used as a standalone feature selection method. A comparison of the feature selection method based on repeatability, independence, and elastic net regression (GLM_{all}), and a method using only elastic net regression (GLM_{elnet}) was performed, see Supplementary material S8. Pre-selection of PET radiomics features is worthwhile, because the number of PET radiomics features in GLM_{elnet} was often high (>20 features) and many were highly correlated to volume or SUV_{max}. In contrast, the average number of features in GLM_{all} was 9. Even so, it was observed that elastic net tends to keep all of the correlated and presumably prognostic

features in the fitted model or shrinks all to zero, whereby increasing the number of (correlated) features resulted in a decrease of the predictive performance. This decrease of predictive performance seen in the validation set suggests that overfitting, although reduced, may still be present. This shows the value of dimensionality reduction in order to optimize predictive performance in rather small sample sizes.

The predictive performance of PET texture features in NSCLC has been studied widely, but clear evidence that PET texture features are complementary to clinical variables is lacking [69]. This study has extensively studied PET texture features and did not find any evidence for added value next to current clinical variables. Supplementary material S9 provides a complete overview of all assessed model performances, including additional investigations with TLG. In literature, typically, only one or two PET texture features have been significantly associated with predicting various survival endpoints [39-41, 47, 70-72]. However, of all the prognostic PET texture features from those studies, such as GLCM Joint Entropy, Correlation, Contrast, Dissimilarity (or Difference Average), NGTDM Coarseness, Busyness, and Contrast, only GLCM Joint Entropy was both repeatable and independent from SUV_{max} or volume in our dataset. In this study, GLCM Joint Entropy was selected 34 times out of 100 by elastic net regression for predicting 2-year OS, and its value in overall survival was also previously shown [47]. Nonetheless, in our study the average predictive performance for GLM_{all} in all clinical endpoints ranged from 0.50 to 0.66. For comparison, other studies predicting outcome with both PET radiomics and clinical variables in NSCLC found predictive performances of 0.63 for predicting OS [41], 0.72 for local recurrence [71], and 0.71 for distant metastases [72]. Even with those results, neglecting any limitations of those studies, there is still no strong evidence that PET texture features exhibit complementary information.

Results from the external validation demonstrated even lower AUC values in most cases than the internal validation set. Besides the limitation of the use of a small external dataset, differences were observed between institutes regarding patients, treatment, and image acquisition and reconstruction settings, that also can influence outcome [44,73], and could have resulted in poor generalizability. To overcome the issue of poor generalizability, a prognostic model should be trained on a combination of well-balanced patient cohorts from multiple institutes, and PET acquisition and reconstruction protocols should be harmonized across centers in multi-centre studies. Alternatively, a post-reconstruction harmonization method proposed by Orhac et al. may also aid in removing the multicenter effect for textural features and SUV [74].

Furthermore, limitations of this paper include the relatively small sample size for machine learning methods that could have affected the predictive performance [75], and the impact of tumour motion on PET radiomics features, especially in lower lobe tumours [76]. Although

Grootjans et al. showed that there are specific PET radiomics features whose prognostic accuracy was not affected by respiratory motion and varying noise-levels [29].

To overcome the limitations of this study, and to be certain that there is no complimentary information in PET radiomics features, future studies need to set up large scale multi-centre cohorts to allow for multiple independent validation datasets. To further improve predictive performance, studies could investigate elastic net-Cox proportional hazard models [77], non-linear relationships by applying data transformation on PET radiomics features [21,78,79], or assess computer engineered features with neural networks or deep learning networks [80, 81]. Currently, deep learning is under investigation for use in lung nodule detection, tumour segmentation, and tumour classification with histopathology images [82]. Its use in medical image analysis is increasing as algorithms become more sophisticated and more data becomes available, which might lead to new insights in survival prediction. A step further would be to combine radiomics features from multimodal imaging such as PET, CT and MRI [83,84], where the combination of anatomical and biological features may of added value for providing a personalized treatment strategy.

Conclusion

In conclusion, robust independent PET radiomics features, identified with 4D PET imaging, did not have complementary value in predicting overall survival and progression-free survival in NSCLC patients treated with concurrent chemoradiotherapy. Improving risk stratification and clinical decision making based on clinical variables and PET radiomics features has still been proven difficult in locally advanced lung cancer patients. New approaches should be investigated in large scale multi-centre studies to deal with current challenges in the field of radiomics before translation to the clinic becomes realistic.

Acknowledgments

The authors would like to thank Simon van Kranen and Jonas Teuwen for helping out with programming, Michel van den Heuvel for providing NKI lung 1 dataset, Natascha Bruin for providing delineations and clinical data for the NKI lung 2 dataset, and Erik Vegt, Rod Hicks, Nick Hardcastle, David Ball, and Tomas Kron for scientific editing of the manuscript.

References

- Aupérin A, Le Péchoux C, Rolland E, et al. *Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small cell lung cancer.* J Clin Oncol 2010; 28: 2181–90.
- Ferlay J, Soerjomataram I, Dikshit R, et al. *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer 2015; 136(5): 359–86.
- Chan BA, Hughes BGM. *Targeted therapy for non-small cell lung cancer: current standards and the promise of the future.* Transl Lung Cancer Res 2015; 4(1): 36–54.
- Remon J, Vilariño N, Reguart N. *Immune checkpoint inhibitors in non-small cell lung cancer (NSCLC): Approaches on special subgroups and unresolved burning questions.* Cancer Treat Rev 2018; 64:21–29.
- Antonia SJ, Villegas A, Daniel D, et al.; PACIFIC Investigators. *Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC.* N Engl J Med 2018. [Epub ahead of print]
- Detterbeck FC, Boffa DJ, Kim AW, et al. *The Eighth Edition Lung Cancer Stage Classification.* Chest 2017; 151(1): 193–203.
- Rami-Porta R, Bolejack V, Giroux DJ, et al, and International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Advisory Board Members and Participating Institutions. *The IASLC Lung Cancer Staging Project: the new database to inform the eighth edition of the TNM classification of lung cancer.* J Thorac Oncol 2014; 9: 1618–1624.
- Paesmans M. *Prognostic and predictive factors for lung cancer.* Breathe. 2012; 9: 112–121.
- Berghmans T, Paesmans M, Sculier JP. *Prognostic factors in stage III non-small cell lung cancer: a review of conventional, metabolic and new biological variables.* Ther Adv Med Oncol 2011; 3: 127–138.
- G Buccheri, D Ferrigno. *Importance of weight loss definition in the prognostic evaluation of non-small-cell lung cancer.* Lung Cancer 2001; 34: 433–440.
- Nakamura H, Ando K, Shinmyo T, et al. *Female gender is an independent prognostic factor in non-small-cell lung cancer: a meta-analysis.* Ann Thorac Cardiovasc Surg 2011; 17: 469–480.
- Yu KH, Zhang C, Berry GJ, et al. *Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features.* Nat Commun 2016; 7(7): 12474.
- Pallis AG, Gridelli C. *Is age a negative prognostic factor for the treatment of advanced/metastatic non-small-cell lung cancer?* Cancer Treat Rev 2010; 36(5): 436–41.
- Yu Z, Zhang G, Yang M, et al. *Systematic review of CYFRA 21-1 as a prognostic indicator and its predictive correlation with clinicopathological features in Non-small Cell Lung Cancer: A meta-analysis.* Oncotarget 2017; 8(3): 4043–4050.
- Jiang AG, Chen HL, Lu HY. *The relationship between Glasgow Prognostic Score and serum tumour markers in patients with advanced non-small cell lung cancer.* BMC Cancer 2015; 15: 386.
- Steels E, Paesmans M, Berghmans T, et al. *Role of p53 as prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis.* Eur Respir J 2001; 18: 705–719.
- Tong J, Sun X, Cheng H, et al. *Expression of p16 in non-small cell lung cancer and its prognostic significance: a meta-analysis of published literatures.* Lung Cancer. 2011; 74: 155–163.
- Martin B, Paesmans M, Mascaux C, et al. *KI-67 expression and patients survival in lung cancer: systematic review of the literature with meta-analysis.* Br J Cancer. 2004; 91: 2018–2025.
- Strom HH, Bremnes RM, Sundstrom SH, et al. *Poor prognosis patients with inoperable locally advanced NSCLC and large tumours benefit from palliative chemoradiotherapy: a subset analysis from a randomized clinical phase III trial.* J Thorac Oncol 2014; 9: 825–33.
- Mahar AL, Compton C, McShane LM, et al, on behalf of the Molecular Modellers Working Group of the American Joint Committee on Cancer. *Refining prognosis in lung cancer: A report on the quality and relevance of clinical prognostic tools.* J Thorac Oncol 2015; 10(11): 1576–1589.
- Aerts HJ, Velazquez ER, Leijenaar RT, et al. *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.* Nat Commun 2014; 5: 4006.
- Lambin P, Rios-Velazquez E, Leijenaar R, et al. *Radiomics: extracting more information from medical images using advanced feature analysis.* Eur J Cancer. 2012; 48(4): 441–6.
- Szyszkowski TA, Yip C, Szlosarek P, et al. *The role of new PET tracers for lung cancer.* Lung Cancer 2016; 94: 7–14.
- Cremonesi M, Gilardi L, Ferrari ME, et al. *Role of interim 18F-FDG-PET/CT for the early prediction of clinical outcomes of Non-Small Cell Lung Cancer (NSCLC) during radiotherapy or chemoradiotherapy. A systematic review.* Eur J Nucl Med Mol Imaging 2017; 44(11): 1915–1927.
- Dingemans AM, de Langen AJ, van den Boogaart V, et al. *First-line erlotinib and bevacizumab in patients with locally advanced and/or metastatic non-small-cell lung cancer: a phase II study including molecular imaging.* Ann Oncol 2011; 22(3): 559–66.
- Mileshkin L, Hicks RJ, Hughes BG, et al. *Changes in 18F-fluorodeoxyglucose and 18F-fluorodeoxythymidine positron emission tomography imaging in patients with non-small cell lung cancer treated with erlotinib.* Clin Cancer Res 2011; 17(10): 3304–15.
- Hyun SH, Ahn HK, Kim H, et al. *Volume-based assessment by (18)F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer.* Eur J Nucl Med Mol Imaging 2014; 41(1): 50–8.
- Moon SH, Cho SH, Park LC, et al. *Metabolic response evaluated by 18F-FDG PET/CT as a potential screening tool in identifying a subgroup of patients with advanced non-small cell lung cancer for immediate maintenance therapy after first-line chemotherapy.* Eur J Nucl Med Mol Imaging 2013; 40(7): 1005–13.
- Grootjans W, Tixier F, van der Vos CS, et al. *The impact of optimal respiratory gating and image noise on evaluation of intra-tumour heterogeneity in 18F-FDG positron emission tomography imaging of lung cancer.* J Nucl Med 2016; 57(11): 1692–1698.
- Salavati A, Duan F, Snyder BS, et al. *Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial.* Eur J Nucl Med Mol Imaging 2017; 44(12): 1969–1983.
- Paesmans M, Berghmans T, Dusart M, et al; European Lung Cancer Working Party, and on behalf of the IASLC Lung Cancer Staging Project. *Primary tumour standardized uptake value measured on fluorodeoxyglucose positron emission tomography is of prognostic value for survival in non-small cell lung cancer: update of a systematic review and meta-analysis by the European Lung Cancer Working Party for the International Association for the Study of Lung Cancer Staging Project.* J Thorac Oncol. 2010; 5(5): 612–9.
- Everitt S, Ball D, Hicks RJ, Callahan J, Plumridge N, Trinh J, et al. *Prospective study of serial imaging comparing fluorodeoxyglucose positron emission tomography (PET) and fluorothymidine PET during radical chemoradiation for non-small cell lung cancer: reduction of detectable proliferation associated with worse survival.* Int J Radiat Oncol 2017; 99:947–55.
- Weiss GJ, Ganeshan B, Miles KA, et al. *Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic.* PLoS ONE. 2014; 9(7): e100244.
- Yip SS, Kim J, Coroller TP, et al. *Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer.* J Nucl Med. 2017; 58(4): 569–576.
- Del Gobbo A, Pellegrinelli A, Gaudioso G, et al. *Analysis of NSCLC tumour heterogeneity, proliferative and 18F-FDG PET indices reveals Ki67 prognostic role in adenocarcinomas.* Histopathology 2016; 68(5): 746–51.
- van Baardwijk A, Bosmans G, van Suylen RJ, et al. *Correlation of intra-tumour heterogeneity on 18F-FDG PET with pathologic features in non-small cell lung cancer: a feasibility study.* Radiother Oncol 2008; 87: 55–58.
- Cook GJR, O'Brien ME, Siddique M, et al. *Non-small cell lung cancer treated with erlotinib: heterogeneity of 18F-FDG uptake at PET—association with treatment response and prognosis.* Radiology 2015; 276: 883–893.
- Fried DV, Tucker SL, Zhou S, et al. *Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer.* Int J Radiat Oncol Biol Phys 2014; 90: 834–842.
- Cook GJR, Yip C, Siddique M, et al. *Are pre-treatment 18F-FDG PET tumour textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy?* J Nucl Med 2013; 54: 19–26.

40. Pyka T, Bundschuh RA, Andratschke N, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol* 2015;10:100.
41. Ohri N, Duan F, Snyder BS, et al. Pre-treatment FDG PET Textural Features in Locally Advanced NSCLC Secondary Analysis of ACRIN 6668/RTOG 0235. *J Nucl Med*. 2016; 57(6): 842-8.
42. Keyes JW Jr. SUV: standard uptake or silly useless value? *J Nucl Med* 1995; 36(10): 1836-9.
43. de Jong EEC, van Elmpt W, Hoekstra OS, et al. Quality assessment of positron emission tomography scans: recommendations for future multicenter trials. *Acta Oncol* 2017; 56(11): 1459-1464.
44. van Velden FHP, Kramer GM, Frings V, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol Imaging Biol*. 2016; 18(5): 788–795.
45. Leijenaar RTH, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumour texture analysis. *Sci Rep* 2015; 5: 11075.
46. Desseroit MC, Tixier F, Weber WA, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumours: A Repeatability Analysis in a Prospective Multicenter Cohort. *J Nucl Med* 2017; 58(3): 406-411.
47. Hatt M, Majdoub M, Vallières M, et al. ¹⁸F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumour volume in a multi-cancer site patient cohort. *J Nucl Med* 2015; 56: 38–44.
48. Brooks FJ, Grigsby PW. The effect of small tumour volumes on studies of intratumoural heterogeneity of tracer uptake. *J Nucl Med*. 2014; 55: 37–42.
49. Orlhac F, Soussan M, Maisonobe J, et al. Tumour texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* 2014; 55:414–422.
50. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS One*. 2015; 10(5): e0124165.
51. Walraven I, van den Heuvel M, van Diessen J, et al. Long-term follow-up of patients with locally advanced non-small cell lung cancer receiving concurrent hypofractionated chemoradiotherapy with or without cetuximab. *Radiother Oncol* 2016;118:442-446.
52. Detterbeck FC, Boffa JB, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. *CHEST* 2017; 151(1):193-203.
53. Kruis MF, van de Kamer JB, Houweling AC, et al. PET motion compensation for radiation therapy using a CT-based mid-position motion model: methodology and clinical evaluation. *Int J Radiat Oncol Biol Phys* 2013;87:394-400.
54. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* 2017;77(21):e104–e107.
55. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610–621.
56. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172–179.
57. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans Biomed Eng* 2014;61:630–637.
58. Sun CJ, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vision Graph Image Process* 1983;23:341–52.
59. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989;19:1264–1274.
60. Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative - feature definitions. 2016. In eprint arXiv:1612.07003.
61. Bland JM, Altman DG. *Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement*. *Lancet* 1986;1(8476):307-10.
62. Shang J, Ling X, Zhang L, et al. Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging* 2016;43:1945-53.
63. Zambelli AE. A data-driven approach to estimating the number of clusters in hierarchical clustering. *Fl000Res* 2016;
64. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Soft* 2010;33(1):1-22.
65. Deist TM, Dankers FJWM, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. *Med Phys* 2018;45(7):3449-3459.
66. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc B* 2005;67(2):301–320.
67. Larue RTHM, Van De Voorde L, van Timmeren JE, et al. 4DCT imaging to assess radiomics feature stability: an investigation for thoracic cancers. *Radiother Oncol* 2017;125(1):147–153.
68. Parmar C, Grossmann P, Bussink J, et al. Machine learning methods for quantitative radiomics biomarkers. *Sci Rep*. 2015;5:13087.
69. Konert T, van de Kamer JB, Sonke JJ, et al. The developing role of FDG PET imaging for prognostication and radiotherapy target volume delineation in non-small cell lung cancer. *J Thorac Dis* 2018;10(21):2508–2521.
70. Lovinfosse P, Janvary ZL, Coucke P, et al. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *Eur J Nucl Med Mol Imaging* 2016;43:1453–60.
71. Takeda K, Takanami K, Shirata Y, et al. Clinical utility of texture analysis of 18F-FDG PET/CT in patients with Stage I lung cancer treated with stereotactic body radiotherapy. *J Radiat Res* 2017;58(6):862-869.
72. Wu J, Aguilera T, Shultz D, et al. Early-Stage Non-Small Cell Lung Cancer: Quantitative Imaging Characteristics of (18)F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. *Radiology* 2016;281(1):270-8.
73. Yan J, Chu-Shern JL, Loi HY, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. *J Nucl Med* 2015;56(11):1667-73.
74. Orlhac F, Boughdad S, Philippe C, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. *J Nucl Med*. 2018 Aug;59(8):1321-1328.
75. Pavlou M, Ambler G, Seaman S, et al. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med* 2015;35:1159–1177.
76. Yip S, McCall K, Aristophanous M, et al. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. *PLoS One* 2014;9:e115510.
77. Yu K-H, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* 2016;7:12474.
78. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563–577.
79. Parmar C, Leijenaar RT, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Sci Rep* 2015;5:11044.
80. Bertolaccini L, Solli P, Pardolesi A, et al. An overview of the use of artificial neural networks in lung cancer research. *J Thorac Dis* 2017;9:924–931.
81. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018 Aug;18(8):500-510.
82. Murphy A, Skalski M, Gaillard F. The utilisation of convolutional neural networks in detecting pulmonary nodules: a review. *Br J Radiol* 2018;91(1090):20180028.
83. Vaidya M, Creach KM, Frye J, et al. Combined PET/CT image characteristics for radiotherapy tumour response in lung cancer. *Radiother Oncol* 2012;102(2):239-45.
84. Vallières M, Freeman CR, Skamene SR, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015 Jul 21; 60(14):5471-96.

Supplementary Material

S1. SUV discretization

SUV discretization was applied to the image as an intensity-resampling step, before building the texture matrices on which texture features rely. These matrix dimensions were determined by the number of discrete intensity values (bins) obtained after this resampling. SUV discretization or binning was achieved with the fixed bin count method using the following equation, as suggested in literature [1]:

$$I_B = B \times \frac{I - I_{\min}}{I_{\max} - I_{\min}},$$

where I_{\max} and I_{\min} denote the maximum and minimum SUV intensities and B, the number of bins.

It has been suggested that an alternative SUV binning method using a fixed bin width (e.g. 0.5 SUV) would allow for a better inter- and intra-patient comparison. The fixed bin width method was calculated following the equation below as previously described by Desseroit *et al.* [2]:

$$I_w = \left\lceil \frac{I}{W} \right\rceil - \left\lceil \frac{I_{\min}}{W} \right\rceil + 1,$$

where W is the bin width.

While we showed only results using a fixed bin count of 64, we do not believe changing the number of bins to 32 would impact our results as we found a strong correlation (mean $\rho > 0.95$) between features calculated with a fixed bin count of 32 and 64.

S2. Matrix calculation

All texture features were calculated in a single matrix taking into account all 13 directions simultaneously. Features were then calculated on the resultant matrix.

The gray level co-occurrence matrices were weighted by weighting factor W and then summed and normalized. Weighting factor W is calculated for the distance between neighboring voxels by:

$$W = e^{-\|d\|^2},$$

where d is the Euclidean distance for the associated direction.

In addition, GLCM and GLRLM features were calculated according to another method: using 13 matrices, one for each spatial direction, followed by averaging the values calculated separately in each matrix [3]. No weighting was applied for this method.

In this paper, we only presented results from feature calculations based on 1 matrix, as each GLCM and GLRLM feature calculated with the average of 13 matrices did strongly correlate (mean $\rho = 0.98$) with the corresponding GLCM and GLRLM feature calculated with all spatial directions in 1 matrix simultaneously.

S3. Repeatability testing

The repeatability assessment was performed within the same patient, using two different mid-position scans, using slightly different delineations due to random variations. For each patient, the PET mid-position scan obtained from the even number of frames (Mid-P even) was compared with the PET mid-position scan including the odd number of frames (Mid-P odd). This resulted in four comparisons, since also either one of the SUV binning methods and one of the delineation methods were applied (see Supplementary Figure 1). The repeatability of each PET radiomics feature was assessed with the Coefficient of Repeatability (CR):

$$CR = 1.96 \times \sqrt{\frac{\sum(d_2 - d_1)^2}{n}}$$

where the CR was calculated as 1.96 times the standard deviation of the differences between the two measurements (d_2 and d_1) [4]. The CR was directly related to the 95% limits of agreement proposed by Bland and Altman that contain 95% of differences between repeated measurements on same subjects. The CR was preferred over intraclass correlation coefficients as this measures reliability, not repeatability [5]. The CR was reported as a percentage: $100\% \times \frac{CR}{mean}$, where mean is the average of the PET radiomics feature value within the patient cohort. The threshold for poor repeatability was set to a value of 30%, corresponding to PERCIST [6].



Supplementary Figure 1. Four assessments that assess repeatability of PET radiomics features. The blue arrows link the groups that are used for repeatability testing.

Supplementary Table 1. Results from the repeatability test comparing two mid-position (Mid-P even versus Mid-P odd) scans from the same patient. The values given are the Coefficient of Repeatability in percentage (CR%). Values in green are defined as highly repeatable.

Name	Fixed bin width		Fixed bin count	
	MTV2.5	MTV40	MTV2.5	MTV40
First-order features				
10Percentile	4.5	19.4	4.5	19.0
90Percentile	9.7	10.0	9.7	10.6
Energy	14.2	27.5	14.2	27.0
Entropy	3.4	5.5	3.8	6.0
InterquartileRange	9.7	19.5	9.7	19.8
Kurtosis	26.8	34.7	26.8	34.8
Maximum	13.2	13.3	13.2	13.2
Mean	6.0	12.9	6.0	12.7
MeanAbsoluteDeviation	10.5	15.4	10.5	15.7
Median	5.3	13.9	5.3	14.0
Minimum	19.2	45.5	19.2	43.4
Range	15.8	21.1	15.8	20.6
RobustMeanAbsoluteDeviation	11.0	17.8	11.0	18.1
RootMeanSquared	6.9	12.1	6.9	12.0
Skewness	23.1	50.4	23.1	51.3
TotalEnergy	14.2	27.5	14.2	27.0
Uniformity	17.9	41.9	21.1	37.0
Variance	22.1	30.2	22.1	30.4
Texture features				
Gray Level Co-occurrence Matrix (GLCM)				
Autocorrelation	29.0	91.5	27.4	37.8
ClusterProminence	50.1	64.6	36.3	52.4
ClusterShade	50.2	137.6	32.3	64.2
ClusterTendency	23.5	34.6	21.9	32.6
Contrast	23.2	28.1	28.8	29.9
Correlation	2.6	11.9	2.7	11.2
DifferenceAverage	9.9	13.1	14.1	17.2
DifferenceEntropy	3.5	4.4	4.3	5.4
DifferenceVariance	25.4	33.5	27.2	30.3
Id	8.7	13.5	11.9	18.1
Idm	14.2	21.8	19.3	30.8
Idmn	0.5	1.0	0.5	0.9
Idn	1.2	2.0	1.2	2.0

Imc1	8.7	19.8	10.2	17.0
Imc2	2.5	4.6	1.1	2.0
InverseVariance	12.9	20.9	19.4	34.2
JointAverage	14.5	41.4	15.6	21.6
JointEnergy	39.7	102.4	61.1	61.8
JointEntropy	2.8	4.5	3.3	5.7
MaximumProbability	53.1	126.7	108.1	86.4
SumAverage	14.5	41.4	15.6	21.6
SumEntropy	2.7	4.5	2.8	4.2
SumSquares	23.3	31.4	22.2	30.2
Gray Level Dependence Matrix (GLDM)				
DependenceEntropy	2.7	5.1	2.2	5.2
DependenceNonUniformity	8.7	39.7	11.7	42.1
DependenceNonUniformityNormalized	11.6	25.0	17.1	26.1
DependenceVariance	51.7	84.3	89.5	55.8
GrayLevelNonUniformity	15.6	62.1	22.0	56.7
GrayLevelVariance	22.1	30.0	22.9	31.2
HighGrayLevelEmphasis	28.4	93.8	27.3	38.3
LargeDependenceEmphasis	29.3	41.2	43.9	33.7
LargeDependenceHighGrayLevelEmphasis	41.8	117.4	47.9	76.9
LargeDependenceLowGrayLevelEmphasis	161.2	174.6	285.3	172.4
LowGrayLevelEmphasis	61.8	67.5	69.9	83.1
SmallDependenceEmphasis	10.7	18.8	12.4	15.5
SmallDependenceHighGrayLevelEmphasis	36.8	86.2	35.3	36.7
SmallDependenceLowGrayLevelEmphasis	59.5	76.3	52.9	92.0
Gray Level Run-Length Matrix (GLRLM)				
GrayLevelNonUniformity	13.7	59.1	18.7	55.4
GrayLevelNonUniformityNormalized	18.0	36.9	18.4	36.1
GrayLevelVariance	22.2	29.8	22.6	31.0
HighGrayLevelRunEmphasis	28.5	93.8	27.1	38.2
LongRunEmphasis	2.9	3.7	3.7	2.7
LongRunHighGrayLevelEmphasis	27.5	95.4	26.3	40.1
LongRunLowGrayLevelEmphasis	69.4	72.3	88.5	86.6
LowGrayLevelRunEmphasis	59.9	66.9	65.2	82.6
RunEntropy	3.0	4.8	3.1	5.4
RunLengthNonUniformity	6.8	42.0	6.0	43.8
RunLengthNonUniformityNormalized	1.5	2.1	1.7	1.7
RunPercentage	0.8	1.1	1.0	0.9

RunVariance	23.5	34.9	40.4	41.2
ShortRunEmphasis	0.6	0.8	0.7	0.7
ShortRunHighGrayLevelEmphasis	28.9	93.4	27.4	37.8
ShortRunLowGrayLevelEmphasis	58.1	65.7	61.5	82.0
Gray Level Size-Zone Matrix (GLSZM)				
GrayLevelNonUniformity	12.6	27.8	11.1	41.6
GrayLevelNonUniformityNormalized	32.4	31.1	15.9	26.5
GrayLevelVariance	24.5	28.6	21.2	29.7
HighGrayLevelZoneEmphasis	31.7	93.2	25.9	37.0
LargeAreaEmphasis	234.2	150.7	144.2	59.7
LargeAreaHighGrayLevelEmphasis	53.5	123.5	40.9	89.4
LargeAreaLowGrayLevelEmphasis	444.7	231.4	454.2	214.8
LowGrayLevelZoneEmphasis	72.4	73.2	54.8	84.7
SizeZoneNonUniformity	14.9	37.8	19.6	45.0
SizeZoneNonUniformityNormalized	15.3	21.2	12.6	16.6
SmallAreaEmphasis	9.9	12.1	6.2	7.7
SmallAreaHighGrayLevelEmphasis	37.0	91.2	30.3	35.9
SmallAreaLowGrayLevelEmphasis	100.9	94.4	65.4	100.4
ZoneEntropy	3.8	7.1	2.4	5.1
ZonePercentage	9.3	14.6	11.2	12.2
ZoneVariance	295.3	195.9	198.0	138.6
Neighborhood Gray Tone Difference Matrix (NGTDM)				
Busyness	75.5	91.3	33.0	81.9
Coarseness	12.0	41.5	16.8	35.2
Complexity	46.0	64.7	27.2	34.3
Contrast	23.3	68.5	31.9	64.9
Strength	33.3	45.7	19.3	28.6
Shape features				
Elongation	4.7	10.8	4.7	10.7
Flatness	7.1	15.0	7.1	13.7
LeastAxis	3.4	14.2	3.4	14.2
MajorAxis	3.1	11.8	3.1	11.7
Maximum2DDiameterColumn	6.1	16.1	6.1	16.2
Maximum2DDiameterRow	7.0	13.0	7.0	13.0
Maximum2DDiameterSlice	6.9	14.2	6.9	14.1
Maximum3DDiameter	6.7	14.8	6.7	14.8
MinorAxis	2.8	10.4	2.8	10.4
Sphericity	3.2	8.9	3.2	8.7

Chapter 6 - Robust independent prognostic FDG PET radiomics features in NSCLC—Are there any?

SurfaceArea	4.6	28.5	4.6	28.4
SurfaceVolumeRatio	4.2	17.8	4.2	18.3
MetabolicTumorVolume	5.9	45.5	5.9	45.3

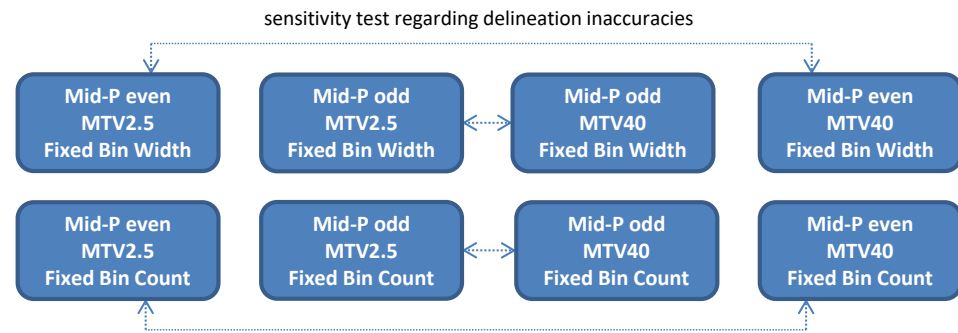
Supplementary Table 2. Overview of PET radiomics features that met the study repeatability criterion. PET radiomics were categorized and ordered by increasing Coefficient of Repeatability in percentage (CR%). The values shown were based on calculations with a fixed bin count and the use of MTV_{2.5}. PET radiomics features with a CR% > 30 were discarded, and were not shown in the table.

First-order features (18/18)	CR%	Texture features (continued)	CR%
Entropy	3.8	GLSZM SizeZoneNonUniformityNormalized	12.6
10Percentile	4.5	GLCM DifferenceAverage	14.1
Median	5.3	GLCM JointAverage	15.6
Mean	6.0	GLCM SumAverage	15.6
RootMeanSquared	6.9	GLSZM GrayLevelNonUniformityNormalized	15.9
90Percentile	9.7	NGTDM Coarseness	16.8
InterquartileRange	9.7	GLDM DependenceNonUniformityNormalized	17.1
MeanAbsoluteDeviation	10.5	GLRLM GrayLevelNonUniformityNormalized	18.4
RobustMeanAbsoluteDeviation	11.0	GLRLM GrayLevelNonUniformity	18.7
Maximum (SUV _{max})	13.2	GLCM Idm	19.3
Energy	14.2	NGTDM Strength	19.3
TotalEnergy	14.2	GLCM InverseVariance	19.4
Range	15.8	GLSZM SizeZoneNonUniformity	19.6
Minimum	19.2	GLSZM GrayLevelVariance	21.2
Uniformity	21.1	GLCM ClusterTendency	21.9
Variance	22.1	GLDM GrayLevelNonUniformity	22.0
Skewness	23.1	GLCM SumSquares	22.2
Kurtosis	26.8	GLRLM GrayLevelVariance	22.6
Texture features (50/74)	CR%	GLDM GrayLevelVariance	22.9
GLCM Idmn	0.5	GLSZM HighGrayLevelZoneEmphasis	25.9
GLRLM ShortRunEmphasis	0.7	GLRLM LongRunHighGrayLevelEmphasis	26.3
GLRLM RunPercentage	1.0	GLRLM HighGrayLevelRunEmphasis	27.1
GLCM Imc2	1.1	GLCM DifferenceVariance	27.2
GLCM Idn	1.2	NGTDM Complexity	27.2
GLRLM RunLengthNonUniformityNormalized	1.7	GLDM HighGrayLevelEmphasis	27.3
GLDM DependenceEntropy	2.2	GLCM Autocorrelation	27.4
GLSZM ZoneEntropy	2.4	GLRLM ShortRunHighGrayLevelEmphasis	27.4
GLCM Correlation	2.7	NGTDM Contrast	28.8
GLCM SumEntropy	2.8	Shape features (13/13)	CR%

GLRLM RunEntropy	3.1	MinorAxis	2.8
GLCM JointEntropy	3.3	MajorAxis	3.1
GLRLM LongRunEmphasis	3.7	Sphericity	3.2
GLCM DifferenceEntropy	4.3	LeastAxis	3.4
GLRLM RunLengthNonUniformity	6.0	SurfaceVolumeRatio	4.2
GLSZM SmallAreaEmphasis	6.2	SurfaceArea	4.6
GLCM Imc1	10.2	Elongation	4.7
GLSZM GrayLevelNonUniformity	11.1	MetabolicTumorVolume (MTV _{2.5})	5.9
GLSZM ZonePercentage	11.2	Maximum2DDiameterColumn	6.1
GLDM DependenceNonUniformity	11.7	Maximum3DDiameter	6.7
GLCM Id	11.9	Maximum2DDiameterSlice	6.9
GLDM SmallDependenceEmphasis	12.4	Maximum2DDiameterRow	7.0
		Flatness	7.1

S4. Sensitivity of PET radiomics features regarding different delineation methods

The impact of large delineation inaccuracies on repeatability was studied between contours generated by the two different SUV thresholds (see Supplementary Figure 2). While most PET radiomics features were not influenced by small delineation inaccuracies, in our study, large delineation inaccuracies (between $MTV_{2.5}$ and MTV_{40}) did have a strong influence on the repeatability of PET radiomics features (see Supplementary Table 3). Comparing PET radiomics features calculated on $MTV_{2.5}$ and MTV_{40} could be hypothetically seen as comparing two sets of independent features in specific cases, and therefore, no features were discarded based on the sensitivity regarding delineation methods. Further studies are warranted to test this hypothesis.



Supplementary Figure 2. Four assessments that assess the sensitivity of PET radiomics features towards large delineation inaccuracies. The blue dotted arrows link the groups that are used for testing the sensitivity towards variations caused by different delineation methods. For each group, it was investigated if PET radiomics features were independent from metabolic tumor volume and maximum SUV.

Supplementary Table 3. Results from the sensitivity test regarding delineation methods. Two different delineation methods were compared using either one of the SUV binning methods and either the Mid-P even or Mid-P odd. The values given are the Coefficient of Repeatability in percentage. Values in green are defined as highly repeatable.

Name	Fixed bin width		Fixed bin count	
	Mid-P Even	Mid-P Odd	Mid-P Even	Mid-P Odd
First-order features				
10Percentile	256.2	253.6	256.2	253.6
90Percentile	55.6	53.8	55.6	53.8
Energy	97.2	90.9	97.2	90.9
Entropy	13.3	12.7	8.5	7.5
InterquartileRange	81.7	82.1	81.7	82.1
Kurtosis	59.9	49.8	59.9	49.8
Maximum	0.0	0.0	0.0	0.0
Mean	135.8	134.3	135.8	134.3
MeanAbsoluteDeviation	81.3	82.4	81.3	82.4
Median	172.6	173.9	172.6	173.9
Minimum	284.7	284.4	284.7	284.4
Range	77.7	78.0	77.7	78.0
RobustMeanAbsoluteDeviation	83.6	84.7	83.6	84.7
RootMeanSquared	104.4	102.9	104.4	102.9
Skewness	88.9	88.6	88.9	88.6
TotalEnergy	97.2	90.9	97.2	90.9
Uniformity	59.1	39.9	69.5	63.1
Variance	257.1	260.6	257.1	260.6
Texture features				
Autocorrelation	112.1	112.8	79.0	74.1
ClusterProminence	1405.2	1355.7	82.0	79.4
ClusterShade	1107.1	1111.9	137.4	145.2
ClusterTendency	319.0	321.5	56.8	53.1
Contrast	108.3	98.0	153.4	146.2
Correlation	47.0	47.0	46.8	46.8
DifferenceAverage	45.9	41.4	78.6	74.9
DifferenceEntropy	9.0	8.6	21.8	21.8
DifferenceVariance	59.4	55.2	138.0	137.6
Id	24.1	21.7	59.4	57.5
Idm	34.6	31.4	90.0	87.6
Idmn	3.5	3.4	3.5	3.4
Idn	7.6	7.2	7.6	7.3
Imct	45.9	43.3	52.6	52.5

Imc2	8.5	8.1	5.8	5.5				
InverseVariance	32.4	30.5	78.6	79.3				
JointAverage	51.4	50.1	55.6	53.1				
JointEnergy	123.4	64.1	155.4	130.9				
JointEntropy	14.7	14.1	9.1	8.8				
MaximumProbability	121.0	79.2	271.1	219.8				
SumAverage	51.4	50.1	55.6	53.1				
SumEntropy	13.4	12.8	6.3	5.9				
SumSquares	273.4	277.3	54.6	49.1				
DependenceEntropy	19.5	19.4	15.4	14.7				
DependenceNonUniformity	249.1	242.3	173.0	170.1				
DependenceNonUniformityNormalized	57.2	51.4	74.3	63.2				
DependenceVariance	97.5	109.3	280.9	248.8				
GrayLevelNonUniformity	171.7	186.5	291.6	300.8				
GrayLevelVariance	256.9	260.6	53.9	47.7				
HighGrayLevelEmphasis	115.6	115.0	78.9	74.2				
LargeDependenceEmphasis	57.9	65.0	168.3	158.0				
LargeDependenceHighGrayLevelEmphasis	117.6	117.4	86.6	87.2				
LargeDependenceLowGrayLevelEmphasis	180.5	234.9	734.1	550.7				
LowGrayLevelEmphasis	129.4	119.0	185.8	177.4				
SmallDependenceEmphasis	38.1	35.0	54.8	50.4				
SmallDependenceHighGrayLevelEmphasis	119.9	124.9	90.6	85.1				
SmallDependenceLowGrayLevelEmphasis	121.9	135.6	115.6	132.4				
GrayLevelNonUniformity	169.3	181.9	280.5	289.8				
GrayLevelNonUniformityNormalized	55.5	39.7	61.3	55.9				
GrayLevelVariance	258.1	261.5	52.9	47.0				
HighGrayLevelRunEmphasis	117.3	116.5	77.6	72.8				
LongRunEmphasis	7.5	7.7	14.3	13.5				
LongRunHighGrayLevelEmphasis	116.5	115.5	76.4	72.0				
LongRunLowGrayLevelEmphasis	135.0	125.5	231.1	208.5				
LowGrayLevelRunEmphasis	128.5	116.9	176.9	171.1				
RunEntropy	13.7	13.5	6.3	5.6				
RunLengthNonUniformity	265.9	264.2	247.1	246.9				
RunLengthNonUniformityNormalized	4.2	4.2	7.3	7.0				
RunPercentage	2.4	2.4	4.3	4.1				
RunVariance	63.2	66.1	182.8	169.4				
ShortRunEmphasis	1.7	1.7	2.9	2.8				
ShortRunHighGrayLevelEmphasis	117.5	116.8	78.0	73.1				
ShortRunLowGrayLevelEmphasis	127.2	115.5	168.5	164.8				
GrayLevelNonUniformity	149.4	148.6	167.4	166.3				
GrayLevelNonUniformityNormalized	57.5	47.9	28.4	25.6				
GrayLevelVariance	263.2	266.0	42.3	39.5				
HighGrayLevelZoneEmphasis	142.2	141.8	61.5	57.8				
LargeAreaEmphasis	253.7	276.2	573.1	489.5				
LargeAreaHighGrayLevelEmphasis	161.2	152.4	118.0	119.5				
LargeAreaLowGrayLevelEmphasis	455.9	444.0	1554.0	1227.4				
LowGrayLevelZoneEmphasis	124.8	107.9	116.3	125.3				
SizeZoneNonUniformity	260.0	252.8	123.2	125.7				
SizeZoneNonUniformityNormalized	30.8	28.6	45.8	41.2				
SmallAreaEmphasis	13.8	13.2	22.3	20.7				
SmallAreaHighGrayLevelEmphasis	151.4	153.8	63.7	61.6				
SmallAreaLowGrayLevelEmphasis	120.6	124.3	115.2	132.7				
ZoneEntropy	21.2	21.0	16.4	15.9				
ZonePercentage	33.7	31.7	48.8	45.8				
ZoneVariance	295.9	330.1	858.8	722.7				
Busyness	175.7	141.5	199.0	194.4				
Coarseness	147.4	141.1	139.1	135.3				
Complexity	283.6	279.9	68.9	63.1				
Contrast	257.7	272.8	241.9	224.4				
Strength	223.3	233.1	46.4	44.6				
Shape features								
Elongation	23.3	22.4	23.3	22.4				
Flatness	24.9	26.1	24.9	26.1				
LeastAxis	61.8	62.9	61.8	62.9				
MajorAxis	46.3	45.0	46.3	45.0				
Maximum2DDiameterColumn	55.6	57.7	55.6	57.7				
Maximum2DDiameterRow	57.8	57.6	57.8	57.6				
Maximum2DDiameterSlice	54.2	56.9	54.2	56.9				
Maximum3DDiameter	51.5	52.8	51.5	52.8				
MinorAxis	55.3	55.6	55.3	55.6				
Sphericity	25.4	24.4	25.4	24.4				
SurfaceArea	123.8	127.1	123.8	127.1				
SurfaceVolumeRatio	77.2	77.0	77.2	77.0				
MetabolicTumorVolume	266.235	266.791	266.2	266.8				

S5. Independence testing

Supplementary Table 4. Overview of Spearman Rank Correlation Coefficients of first-order and shape features. First-order features are not influenced by SUV binning methods. Values in green are defined as features independent from MTV or SUV_{max}. Bold values are correspond to a feature who is independent from both MTV and SUV_{max}. Group 1: mid-p even MTV_{2.5}; group 2: mid-p even MTV₄₀; group 3: mid-p odd MTV_{2.5}; group 4: mid-p odd MTV₄₀. MTV = metabolic tumor volume; SUVmax = maximum SUV.

	even_mtv25		even_mtv40		odd_mtv25		odd_mtv40	
	group 1		group 2		group 3		group 4	
	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax
MetabolicTumor Volume	1	0.43	1	-0.015	1	0.464	1	0.053
Maximum SUV	0.43	1	-0.015	1	0.464	1	0.053	1
10Percentile	0.457	0.598	-0.007	0.988	0.465	0.576	0.04	0.985
90Percentile	0.361	0.934	-0.046	0.982	0.381	0.938	0.01	0.985
Energy	0.845	0.792	0.585	0.765	0.859	0.801	0.612	0.783
Entropy	0.492	0.93	0.11	0.922	0.508	0.931	0.196	0.914
InterquartileRange	0.372	0.896	-0.097	0.893	0.389	0.892	-0.01	0.903
Kurtosis	-0.084	0.058	0.021	-0.058	-0.087	0.044	0.039	-0.051
Mean	0.422	0.915	-0.024	0.979	0.435	0.913	0.029	0.98
MeanAbsoluteDeviation	0.355	0.939	-0.11	0.942	0.379	0.938	-0.039	0.954
Median	0.502	0.826	-0.002	0.972	0.515	0.822	0.052	0.973
Minimum	-0.629	-0.118	-0.29	0.838	-0.576	-0.283	-0.243	0.807
Range	0.449	0.999	0.111	0.972	0.48	0.999	0.164	0.977
RobustMeanAbsoluteDeviation	0.354	0.907	-0.098	0.912	0.382	0.903	-0.029	0.92
RootMeanSquared	0.407	0.931	-0.029	0.979	0.424	0.932	0.028	0.981
Skewness	-0.16	0.146	-0.228	-0.118	-0.153	0.149	-0.202	-0.131
TotalEnergy	0.845	0.792	0.585	0.765	0.859	0.801	0.612	0.783
Uniformity	-0.526	-0.869	-0.12	-0.888	-0.541	-0.867	-0.183	-0.879
Variance	0.345	0.952	-0.088	0.962	0.378	0.954	-0.021	0.972
Elongation	-0.045	0.162	-0.024	0.19	-0.037	0.185	-0.102	0.15
Flatness	-0.138	0.195	-0.104	0.12	-0.159	0.164	-0.182	0.076
LeastAxis	0.954	0.448	0.95	0.02	0.957	0.472	0.947	0.069
MajorAxis	0.865	0.279	0.789	0.011	0.869	0.314	0.791	0.077
Maximum2DDiameterColumn	0.911	0.358	0.852	0.061	0.91	0.391	0.855	0.109
Maximum2DDiameterRow	0.941	0.31	0.923	-0.004	0.938	0.331	0.916	0.046
Maximum2DDiameterSlice	0.931	0.312	0.887	0.013	0.932	0.345	0.889	0.062
Maximum3DDiameter	0.919	0.298	0.851	0.021	0.91	0.327	0.844	0.072
MinorAxis	0.977	0.417	0.951	0.086	0.98	0.458	0.942	0.145
Sphericity	-0.447	0.094	-0.63	0.006	-0.457	0.042	-0.631	-0.035
SurfaceArea	0.98	0.373	0.977	-0.019	0.981	0.409	0.974	0.063
SurfaceVolumeRatio	-0.919	-0.536	-0.768	0.035	-0.927	-0.539	-0.754	-0.001

Supplementary Table 5. Overview of Spearman Rank Correlation Coefficients of PET texture features calculated with a fixed bin width. Values in green are defined as features independent from MTV or SUV_{max}. Bold values are correspond to a feature who is independent from both MTV and SUV_{max}. Only one of the PET texture features calculated with a fixed bin width was independent from MTV_{2.5} and SUV_{max}. Group 1: mid-p even MTV_{2.5}; group 2: mid-p even MTV₄₀; group 3: mid-p odd MTV_{2.5}; group 4: mid-p odd MTV₄₀. MTV = metabolic tumor volume; SUVmax = maximum SUV.

Name	even_mtv25		even_mtv40		odd_mtv25		odd_mtv40	
	group 1		group 2		group 3		group 4	
	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax
MetabolicTumorVolume	1	0.43	1	-0.015	1	0.464	1	0.053
Maximum SUV	0.43	1	-0.015	1	0.464	1	0.053	1
Autocorrelation	0.442	0.944	0.126	0.888	0.46	0.95	0.172	0.908
ClusterProminence	0.363	0.983	-0.036	0.982	0.397	0.986	0.04	0.986
ClusterShade	0.302	0.94	-0.135	0.776	0.35	0.946	-0.112	0.72
ClusterTendency	0.353	0.956	-0.051	0.968	0.386	0.962	0.02	0.978
Contrast	0.112	0.897	-0.32	0.914	0.14	0.893	-0.25	0.913
Correlation	0.835	0.586	0.783	-0.21	0.85	0.606	0.803	-0.112
DifferenceAverage	0.085	0.87	-0.336	0.898	0.103	0.861	-0.271	0.897
DifferenceEntropy	0.122	0.898	-0.272	0.928	0.149	0.894	-0.207	0.929
DifferenceVariance	0.143	0.924	-0.281	0.937	0.185	0.923	-0.215	0.934
Id	-0.02	-0.804	0.38	-0.851	-0.035	-0.796	0.287	-0.864
Idm	-0.002	-0.785	0.384	-0.843	-0.021	-0.772	0.289	-0.851
Idmn	0.85	0.474	0.855	-0.141	0.855	0.483	0.874	-0.034
Idn	0.82	0.45	0.832	-0.128	0.829	0.48	0.858	-0.015
Imc1	-0.043	-0.643	0.59	-0.457	-0.05	-0.645	0.544	-0.515
Imc2	0.157	0.859	-0.442	0.726	0.184	0.858	-0.384	0.759
InverseVariance	0.008	-0.79	0.376	-0.837	-0.016	-0.777	0.35	-0.833
JointAverage	0.445	0.924	0.139	0.87	0.47	0.928	0.18	0.888
JointEnergy	-0.473	-0.856	-0.316	-0.832	-0.493	-0.849	-0.382	-0.828
JointEntropy	0.545	0.917	0.381	0.81	0.56	0.913	0.446	0.812
MaximumProbability	-0.396	-0.654	-0.267	-0.815	-0.452	-0.655	-0.366	-0.781
SumAverage	0.445	0.924	0.139	0.87	0.47	0.928	0.18	0.888
SumEntropy	0.446	0.945	0.112	0.938	0.452	0.946	0.151	0.95
SumSquares	0.33	0.956	-0.101	0.967	0.367	0.958	-0.033	0.976
DependenceEntropy	0.889	0.736	0.839	0.401	0.903	0.75	0.845	0.447
DependenceNonUniformity	0.946	0.636	0.895	0.369	0.951	0.658	0.908	0.399
DependenceNonUniformityNormalized	-0.128	0.698	-0.475	0.796	-0.139	0.672	-0.397	0.801
DependenceVariance	0.34	-0.437	0.528	-0.71	0.312	-0.419	0.453	-0.731
GrayLevelNonUniformity	0.825	-0.041	0.857	-0.424	0.815	-0.024	0.846	-0.398
GrayLevelVariance	0.345	0.952	-0.088	0.963	0.378	0.954	-0.026	0.972
HighGrayLevelEmphasis	0.452	0.945	0.139	0.882	0.47	0.949	0.182	0.901
LargeDependenceEmphasis	0.171	-0.653	0.477	-0.786	0.162	-0.627	0.391	-0.808
LargeDependenceHighGrayLevelEmphasis	0.709	0.778	0.516	0.631	0.687	0.82	0.476	0.711
LargeDependenceLowGrayLevelEmphasis	-0.382	-0.514	-0.292	-0.732	-0.389	-0.561	-0.321	-0.756



Chapter 6 - Robust independent prognostic FDG PET radiomics features in NSCLC—Are there any?

LowGrayLevelEmphasis	-0.74	-0.578	-0.451	-0.619	-0.7	-0.643	-0.502	-0.658
SmallDependenceEmphasis	-0.021	0.794	-0.42	0.838	-0.004	0.78	-0.324	0.842
SmallDependenceHighGrayLevelEmphasis	0.337	0.955	-0.006	0.927	0.377	0.957	0.074	0.935
SmallDependenceLowGrayLevelEmphasis	-0.854	-0.373	-0.71	-0.255	-0.84	-0.427	-0.679	-0.379
GrayLevelNonUniformity	0.844	-0.016	0.872	-0.405	0.832	-0.004	0.855	-0.382
GrayLevelNonUniformityNormalized	-0.524	-0.876	-0.119	-0.888	-0.548	-0.873	-0.179	-0.884
GrayLevelVariance	0.344	0.951	-0.091	0.963	0.379	0.954	-0.019	0.973
HighGrayLevelRunEmphasis	0.451	0.945	0.138	0.883	0.468	0.953	0.181	0.903
LongRunEmphasis	0.147	-0.687	0.462	-0.804	0.112	-0.688	0.396	-0.806
LongRunHighGrayLevelEmphasis	0.469	0.939	0.167	0.871	0.487	0.947	0.216	0.894
LongRunLowGrayLevelEmphasis	-0.699	-0.586	-0.425	-0.654	-0.667	-0.655	-0.471	-0.681
LowGrayLevelRunEmphasis	-0.748	-0.591	-0.451	-0.622	-0.711	-0.655	-0.5	-0.662
RunEntropy	0.556	0.934	0.199	0.903	0.578	0.934	0.26	0.904
RunLengthNonUniformity	0.992	0.504	0.996	0.048	0.994	0.522	0.995	0.114
RunLengthNonUniformityNormalized	-0.087	0.732	-0.436	0.816	-0.06	0.723	-0.368	0.818
RunPercentage	-0.119	0.707	-0.451	0.807	-0.101	0.694	-0.385	0.81
RunVariance	0.191	-0.656	0.476	-0.795	0.168	-0.651	0.408	-0.805
ShortRunEmphasis	-0.091	0.729	-0.438	0.815	-0.071	0.714	-0.371	0.817
ShortRunHighGrayLevelEmphasis	0.449	0.947	0.128	0.886	0.466	0.953	0.17	0.906
ShortRunLowGrayLevelEmphasis	-0.756	-0.591	-0.458	-0.617	-0.719	-0.655	-0.505	-0.654
GrayLevelNonUniformity	0.951	0.27	0.956	-0.12	0.959	0.334	0.947	-0.081
GrayLevelNonUniformityNormalized	-0.519	-0.937	-0.101	-0.917	-0.549	-0.934	-0.171	-0.917
GrayLevelVariance	0.343	0.97	-0.099	0.968	0.382	0.973	-0.019	0.978
HighGrayLevelZoneEmphasis	0.438	0.961	0.118	0.896	0.479	0.965	0.174	0.922
LargeAreaEmphasis	0.204	-0.636	0.465	-0.797	0.143	-0.641	0.4	-0.803
LargeAreaHighGrayLevelEmphasis	0.744	0.374	0.703	0.229	0.749	0.444	0.685	0.316
LargeAreaLowGrayLevelEmphasis	-0.182	-0.603	-0.242	-0.764	-0.25	-0.631	-0.257	-0.794
LowGrayLevelZoneEmphasis	-0.771	-0.7	-0.416	-0.68	-0.747	-0.745	-0.487	-0.71
SizeZoneNonUniformity	0.833	0.787	0.676	0.643	0.848	0.787	0.706	0.639
SizeZoneNonUniformityNormalized	0.035	0.836	-0.379	0.863	0.09	0.837	-0.289	0.847
SmallAreaEmphasis	0.045	0.839	-0.374	0.865	0.1	0.84	-0.272	0.851
SmallAreaHighGrayLevelEmphasis	0.397	0.968	0.056	0.921	0.433	0.969	0.123	0.937
SmallAreaLowGrayLevelEmphasis	-0.805	-0.617	-0.515	-0.575	-0.781	-0.671	-0.547	-0.602
ZoneEntropy	0.839	0.792	0.757	0.515	0.868	0.796	0.788	0.536
ZonePercentage	-0.043	0.779	-0.437	0.827	-0.02	0.764	-0.346	0.832
ZoneVariance	0.246	-0.587	0.496	-0.76	0.179	-0.591	0.431	-0.777
Busyness	-0.024	-0.835	0.297	-0.803	-0.039	-0.826	0.23	-0.809
Coarseness	-0.962	-0.555	-0.956	-0.095	-0.966	-0.58	-0.966	-0.171
Complexity	0.318	0.977	-0.031	0.975	0.36	0.976	0.058	0.978
Contrast	-0.221	0.598	-0.601	0.683	-0.181	0.621	-0.572	0.68
Strength	-0.108	0.8	-0.571	0.78	-0.06	0.817	-0.483	0.807

Supplementary Table 6. Overview of Spearman Rank Correlation Coefficients of PET texture features calculated with a fixed bin count. Values in green are defined as features independent from MTV or SUV_{max}. Bold values are correspond to a feature who is independent from both MTV and SUV_{max}. More PET texture features calculated with a fixed bin count did show independent behavior, namely 15 out of 76 texture features. Group 1: mid-p even MTV_{2.5}; group 2: mid-p even MTV₄₀; group 3: mid-p odd MTV_{2.5}; group 4: mid-p odd MTV₄₀. MTV = metabolic tumor volume; SUV_{max} = maximum SUV.

Name	even_mtv25		even_mtv40		odd_mtv25		odd_mtv40	
	group 1	group 2	group 3	group 4	group 1	group 2	group 3	group 4
	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax	MTV	SUVmax
MetabolicTumorVolume	1	0.43	1	-0.015	1	0.464	1	0.053
Maximum SUV	0.43	1	-0.015	1	0.464	1	0.053	1
Autocorrelation	-0.118	-0.267	0.123	0.109	-0.113	-0.262	0.108	0.116
ClusterProminence	-0.42	0.091	-0.608	-0.135	-0.506	0.035	-0.651	-0.23
ClusterShade	-0.399	0.136	-0.432	-0.181	-0.447	0.106	-0.393	-0.23
ClusterTendency	-0.265	0.061	-0.549	-0.098	-0.29	0.013	-0.592	-0.165
Contrast	-0.853	-0.479	-0.86	0.136	-0.857	-0.497	-0.879	0.021
Correlation	0.84	0.577	0.781	-0.218	0.848	0.598	0.791	-0.135
DifferenceAverage	-0.832	-0.486	-0.842	0.11	-0.834	-0.51	-0.861	-0.008
DifferenceEntropy	-0.831	-0.465	-0.82	0.146	-0.841	-0.5	-0.848	0.026
DifferenceVariance	-0.888	-0.46	-0.868	0.18	-0.882	-0.469	-0.868	0.095
Id	0.784	0.543	0.808	-0.09	0.792	0.572	0.814	0.07
Idm	0.761	0.56	0.794	-0.073	0.782	0.571	0.774	0.094
Idmn	0.852	0.478	0.86	-0.125	0.855	0.498	0.88	-0.011
Idn	0.827	0.482	0.836	-0.107	0.83	0.51	0.855	0.019
Imc1	0.613	-0.018	0.935	0.006	0.612	0.012	0.949	0.066
Imc2	-0.6	-0.053	-0.891	-0.011	-0.629	-0.116	-0.92	-0.079
InverseVariance	0.769	0.482	0.759	-0.094	0.797	0.552	0.784	0.114
JointAverage	-0.107	-0.356	0.146	0.149	-0.112	-0.353	0.136	0.152
JointEnergy	-0.061	0.396	-0.675	0.068	-0.061	0.356	-0.652	0.055
JointEntropy	0.318	-0.158	0.778	-0.077	0.284	-0.138	0.772	-0.016
MaximumProbability	0.074	0.641	-0.603	0.135	0.089	0.568	-0.577	0.115
SumAverage	-0.107	-0.356	0.146	0.149	-0.112	-0.353	0.136	0.152
SumEntropy	-0.003	-0.164	0.148	-0.168	0.028	-0.162	0.148	-0.159
SumSquares	-0.412	-0.037	-0.667	-0.049	-0.446	-0.101	-0.723	-0.11
DependenceEntropy	0.962	0.409	0.94	-0.046	0.962	0.412	0.94	0.032
DependenceNonUniformity	0.974	0.34	0.987	0.004	0.971	0.373	0.986	0.037
DependenceNonUniformityNormalized	-0.791	-0.636	-0.829	0.054	-0.807	-0.642	-0.825	-0.062
DependenceVariance	0.779	0.657	0.811	-0.01	0.786	0.634	0.81	0.094
GrayLevelNonUniformity	0.95	0.556	0.956	-0.029	0.954	0.576	0.964	0.054
GrayLevelVariance	-0.342	-0.042	-0.602	-0.045	-0.345	-0.08	-0.637	-0.116
HighGrayLevelEmphasis	-0.073	-0.255	0.138	0.118	-0.063	-0.242	0.146	0.115
LargeDependenceEmphasis	0.791	0.664	0.83	-0.039	0.806	0.64	0.82	0.097
LargeDependenceHighGrayLevelEmphasis	0.667	0.145	0.63	0.098	0.615	0.163	0.564	0.144
LargeDependenceLowGrayLevelEmphasis	0.185	0.66	-0.401	-0.327	0.279	0.666	-0.456	-0.2



LowGrayLevelEmphasis	-0.372	0.33	-0.603	-0.219	-0.299	0.321	-0.616	-0.237
SmallDependenceEmphasis	-0.794	-0.625	-0.823	0.055	-0.83	-0.611	-0.82	-0.038
SmallDependenceHighGrayLevelEmphasis	-0.527	-0.411	-0.294	0.122	-0.521	-0.434	-0.239	0.034
SmallDependenceLowGrayLevelEmphasis	-0.837	-0.357	-0.738	-0.078	-0.781	-0.365	-0.668	-0.187
GrayLevelNonUniformity	0.956	0.544	0.958	-0.027	0.959	0.568	0.966	0.056
GrayLevelNonUniformityNormalized	-0.165	0.326	-0.233	0.026	-0.125	0.311	-0.207	0.069
GrayLevelVariance	-0.345	-0.037	-0.606	-0.045	-0.356	-0.084	-0.642	-0.115
HighGrayLevelRunEmphasis	-0.063	-0.239	0.135	0.117	-0.062	-0.234	0.146	0.118
LongRunEmphasis	0.797	0.657	0.83	-0.048	0.805	0.644	0.835	0.081
LongRunHighGrayLevelEmphasis	0.043	-0.179	0.193	0.131	0.037	-0.184	0.216	0.13
LongRunLowGrayLevelEmphasis	-0.265	0.411	-0.591	-0.234	-0.197	0.391	-0.602	-0.239
LowGrayLevelRunEmphasis	-0.384	0.321	-0.603	-0.214	-0.317	0.309	-0.617	-0.235
RunEntropy	0.396	-0.064	0.42	-0.056	0.352	-0.076	0.369	-0.086
RunLengthNonUniformity	0.998	0.421	1	-0.006	0.999	0.452	1	0.055
RunLengthNonUniformityNormalized	-0.793	-0.643	-0.828	0.055	-0.809	-0.632	-0.827	-0.073
RunPercentage	-0.797	-0.65	-0.831	0.052	-0.806	-0.638	-0.832	-0.072
RunVariance	0.804	0.66	0.83	-0.041	0.806	0.661	0.832	0.077
ShortRunEmphasis	-0.792	-0.644	-0.83	0.054	-0.809	-0.632	-0.828	-0.075
ShortRunHighGrayLevelEmphasis	-0.082	-0.251	0.127	0.116	-0.078	-0.236	0.128	0.118
ShortRunLowGrayLevelEmphasis	-0.403	0.304	-0.61	-0.209	-0.34	0.288	-0.616	-0.239
GrayLevelNonUniformity	0.986	0.375	0.977	-0.025	0.986	0.409	0.978	0.05
GrayLevelNonUniformityNormalized	-0.316	-0.012	-0.2	0.068	-0.281	0.003	-0.154	0.111
GrayLevelVariance	-0.478	-0.049	-0.655	-0.046	-0.51	-0.139	-0.679	-0.141
HighGrayLevelZoneEmphasis	-0.03	-0.066	0.114	0.111	-0.016	-0.097	0.105	0.09
LargeAreaEmphasis	0.798	0.677	0.828	-0.027	0.787	0.692	0.831	0.102
LargeAreaHighGrayLevelEmphasis	0.848	0.358	0.686	0.1	0.845	0.405	0.606	0.186
LargeAreaLowGrayLevelEmphasis	0.354	0.726	-0.389	-0.305	0.41	0.726	-0.448	-0.229
LowGrayLevelZoneEmphasis	-0.641	-0.136	-0.629	-0.171	-0.56	-0.17	-0.599	-0.234
SizeZoneNonUniformity	0.941	0.287	0.976	0.019	0.938	0.335	0.979	0.037
SizeZoneNonUniformityNormalized	-0.816	-0.569	-0.818	0.062	-0.849	-0.536	-0.812	-0.024
SmallAreaEmphasis	-0.816	-0.57	-0.818	0.063	-0.847	-0.535	-0.809	-0.023
SmallAreaHighGrayLevelEmphasis	-0.36	-0.184	-0.089	0.104	-0.341	-0.242	-0.066	0.057
SmallAreaLowGrayLevelEmphasis	-0.73	-0.414	-0.686	-0.091	-0.672	-0.42	-0.607	-0.199
ZoneEntropy	0.965	0.499	0.921	-0.046	0.97	0.488	0.919	0.003
ZonePercentage	-0.802	-0.636	-0.822	0.047	-0.819	-0.639	-0.826	-0.068
ZoneVariance	0.794	0.68	0.825	-0.016	0.781	0.684	0.824	0.097
Busyness	0.789	0.656	0.556	-0.037	0.807	0.655	0.617	-0.008
Coarseness	-0.966	-0.527	-0.958	-0.039	-0.969	-0.547	-0.967	-0.093
Complexity	-0.737	-0.396	-0.471	0.05	-0.719	-0.395	-0.469	-0.082
Contrast	-0.817	-0.422	-0.893	0.072	-0.826	-0.444	-0.909	-0.03
Strength	-0.899	-0.191	-0.947	-0.081	-0.914	-0.243	-0.951	-0.149

S6. An overview of all robust independent PET radiomics features

Supplementary Table 7. An overview of robust independent PET radiomics features per cohort. A distinction was made between features that were calculated from $MTV_{2.5}$ and MTV_{40} . Note that the texture features were all calculated with the fixed bin count method, unless otherwise stated. FBW = fixed bin width.

4D PET lung (odd and even mid-position scans)		
Feature types	$MTV_{2.5}$	MTV_{40}
First-order	Kurtosis	
	Skewness	
	SUV_{max}	
Shape	Elongation	Elongation
	Flatness	
	Sphericity	
Texture	GLCM Autocorrelation	
	GLCM ClusterTendency	
	GLCM JointAverage	GLCM JointAverage
	GLCM JointEntropy	
	GLCM SumAverage	GLCM SumAverage
	GLCM SumEntropy	
	GLCM SumSquares	
	GLDM GrayLevelVariance	
	GLDM HighGrayLevelEmphasis	
	GLRLM GrayLevelNonUniformityNormalized	
	GLRLM GrayLevelVariance	
	GLRLM HighGrayLevelRunEmphasis	
GLRLM LongRunHighGrayLevelEmphasis		
GLRLM RunEntropy	GLRLM RunEntropy	
GLRLM ShortRunHighGrayLevelEmphasis		
GLSZM GrayLevelNonUniformityNormalized	GLSZM GrayLevelNonUniformityNormalized	
GLSZM HighGrayLevelZoneEmphasis		
NKI lung 1 and 2 (3D whole-body FDG PET scan)		
Feature types	$MTV_{2.5}$	MTV_{40}
First-order	10Percentile	
	Entropy	Entropy
Kurtosis		
Skewness		
SUV_{max}		
Uniformity		
Shape	Elongation	Elongation

	Flatness	Flatness
	Sphericity	Sphericity
Texture	GLCM Autocorrelation	
	GLCM ClusterTendency	
	GLCM Correlation	
	GLCM Correlation (FBW)	
	GLCM Imc1 (FBW)	
	GLCM JointAverage	GLCM JointAverage
	GLCM JointEntropy	
	GLCM SumAverage	GLCM SumAverage
	GLCM SumEntropy	GLCM SumEntropy
	GLCM SumSquares	
	GLDM GrayLevelVariance	
	GLDM HighGrayLevelEmphasis	
	GLDM LargeDependenceEmphasis (FBW)	
	GLRLM GrayLevelNonUniformityNormalized	
	GLRLM GrayLevelVariance	
	GLRLM HighGrayLevelRunEmphasis	
	GLRLM LongRunHighGrayLevelEmphasis	
	GLRLM RunEntropy	GLRLM RunEntropy
	GLRLM RunPercentage	GLRLM RunPercentage
	GLRLM RunPercentage (FBW)	GLRLM RunPercentage (FBW)
	GLRLM ShortRunHighGrayLevel-Emphasis	
	GLSZM GrayLevelNonUniformityNormalized	GLSZM GrayLevelNonUniformityNormalized
	GLSZM HighGrayLevelZoneEmphasis	
PMCC lung1 (3D whole-body FDG PET scan)		
Feature types	MTV_{2.5}	MTV₄₀
First-order	Entropy	Entropy
	Kurtosis	
	Minimum	
	Skewness	
	SUV _{max}	
	Uniformity	
Shape	Elongation	Elongation
	Flatness	Flatness
	Sphericity	
Texture	GLCM Autocorrelation	
	GLCM ClusterTendency	

	GLCM Correlation	GLCM Correlation
	GLCM DifferenceVariance	
	GLCM JointAverage	GLCM JointAverage
	GLCM JointEntropy	
	GLCM SumAverage	GLCM SumAverage
	GLCM SumSquares	
	GLDM GrayLevelVariance	
	GLDM HighGrayLevelEmphasis	GLDM HighGrayLevelEmphasis
	GLRLM GrayLevelNonUniformityNormalized	
	GLRLM GrayLevelVariance	
	GLRLM HighGrayLevelRunEmphasis	
	GLRLM LongRunHighGrayLevelEmphasis	
	GLRLM RunEntropy	GLRLM RunEntropy
	GLRLM RunPercentage (FBW)	GLRLM RunPercentage (FBW)
	GLRLM ShortRunHighGrayLevel-Emphasis	
	GLSZM GrayLevelNonUniformityNormalized	GLSZM GrayLevelNonUniformityNormalized
	GLSZM HighGrayLevelZoneEmphasis	
	GLSZM GrayLevelVariance	
	GLSZM HighGrayLevelZoneEmphasis	
	NGTDM Complexity	

S7. Testing a different level of independence

An argument for choosing $|\rho| < 0.7$ for the level of independence could be that the variance explained is approximately 50%, instead of 25% for $|\rho| < 0.5$. With $|\rho| < 0.7$ more features met the criterion for independence. With this new selection of repeatable independent PET radiomics features, elastic net regression was applied and results are summarized in Supplementary Table 8. Marginal differences are observed between the two assessments, with a slight benefit seen for $|\rho| < 0.5$.

Supplementary Table 8. Predictive performance of GLM_{all} and GLM_{rad} with varying level of independence. The top results are median AUC values (+SD) resulting from an independence level of $|\rho| < 0.7$. The results at the bottom are from an independence level of $|\rho| < 0.5$, and marginal differences are observed. Validation 1 represents the average of the 100-times repeated training/validation procedure. Below Validation 2 is the AUC calculated on the external validation set using the best performing fitted model from Validation 1.

Endpoint	GLM_{all}		GLM_{rad}	
	Validation 1	Validation 2	Validation 1	Validation 2
$\rho < 0.7$				
2 year OS	0.64±0.08	0.52	0.50±0.05	0.50
2 year PFS	0.50±0.04	0.50	0.50±0.04	0.50
1 year PFS	0.56±0.06	0.61	0.57±0.06	0.59
1 year LR	0.53±0.06	0.57	0.51±0.05	0.54
1 year DM	0.58±0.07	0.54	0.60±0.07	0.56
$\rho < 0.5$				
2 year OS	0.66±0.07	0.55	0.57±0.07	0.55
2 year PFS	0.50±0.06	0.52	0.50±0.05	0.50
1 year PFS	0.57±0.06	0.51	0.58±0.07	0.51
1 year LR	0.57±0.07	0.55	0.52±0.05	0.50
1 year DM	0.58±0.06	0.52	0.60±0.07	0.52

S8. Elastic Net Regression for feature selection

As stated before, elastic net regression could also be used as a standalone feature selection method [7]. A comparison of the feature selection method based on repeatability, independence, and elastic net regression (GLM_{all}), and a method using only elastic net regression (GLM_{elnet}) was performed, see Supplementary Figure 3, and Supplementary Table 9 and 10 for more information. As an example, 2-year overall survival was used as clinical outcome.

As a first step, any difference in predictive performance between both feature selection methods was tested. As a second step, a sensitivity study was performed to assess the impact of sample size on predictive performance for both methods.

In Supplementary Table 9, the predictive performance is shown for GLM_{all} and GLM_{elnet} , and a marginal benefit was observed in this study for pre-selection of PET radiomics features before applying elastic net regression.

Supplementary Table 9. Model performance for a combination of PET radiomics features and clinical variables chosen by the pre-selection procedure (GLM_{all}) compared to the model performance of a set-up using elastic net regression only (GLM_{elnet}). The AUC values are depicted per clinical endpoint, per model. Values below Validation 1 represent the average of 100 AUC values with standard deviation. Below Validation 2 is the AUC calculated on the external validation set using the best performing fitted model from Validation 1.

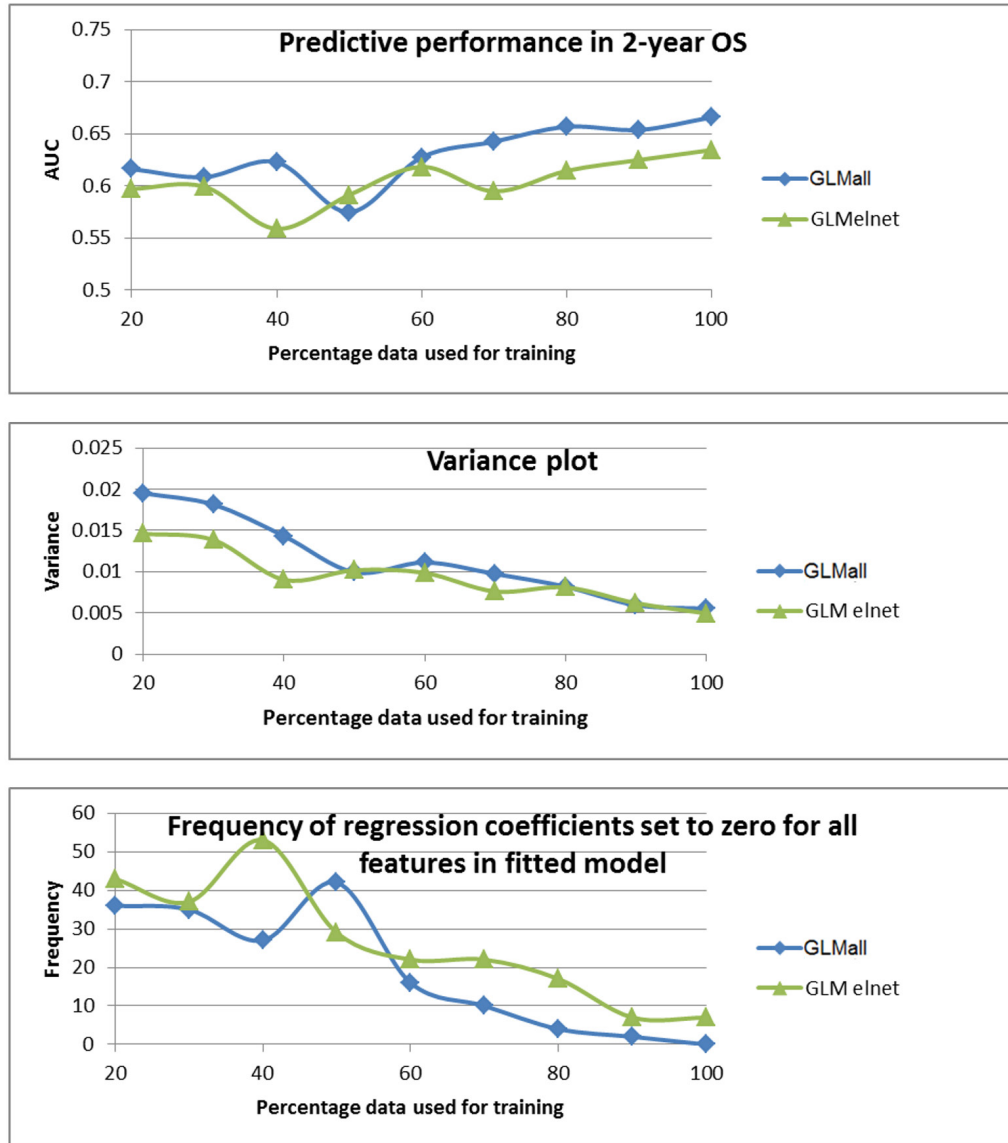
Endpoint	Pre-selection (GLM_{all})		Without pre-selection (GLM_{elnet})	
	Validation 1	Validation 2	Validation 1	Validation 2
2 year OS	0.66±0.07	0.55	0.64±0.07	0.48
2 year PFS	0.50±0.06	0.52	0.53±0.06	0.63
1 year PFS	0.57±0.06	0.51	0.56±0.07	0.59
1 year LR	0.57±0.07	0.55	0.53±0.05	0.54
1 year DM	0.58±0.06	0.52	0.58±0.08	0.54

Supplementary Table 10. The most selected features in the model using only elastic net regression for feature selection (GLM_{elnet}). Features are ranked by the number of times selected in the fitted model. Only the top 10 most selected PET radiomics are shown. FBW = fixed bin width, FBC = fixed bin count. A distinction was made between features calculated from $MTV_{2.5}$ and MTV_{40} .

Endpoint	Selected features by elastic net	Frequency
2-year OS	Age	96
	glrlm_GrayLevelNonUniformity_MTV25_FBW	90
	glszm_GrayLevelNonUniformity_MTV25_FBW	80
	shape_Sphericity_MTV25_FBC	64
	shape_Sphericity_MTV25_FBW	64
	shape_Maximum3DDiameter_MTV40_FBC	61
	shape_Maximum3DDiameter_MTV40_FBW	61
	glcm_MaximumProbability_MTV25_FBC	42
	glcm_Imc2_MTV40_FBW	37
	gldm_GrayLevelNonUniformity_MTV25_FBW	30
2-year PFS	glcm_MaximumProbability_MTV25_FBC	24
	glcm_Imc2_MTV40_FBW	21
	glrlm_RunPercentage_MTV40_FBW	20
	Age	19
	glcm_DifferenceEntropy_MTV40_FBW	18
	shape_Maximum3DDiameter_MTV40_FBC	14
	shape_Maximum3DDiameter_MTV40_FBW	14
	gldm_DependenceVariance_MTV40_FBW	11
	firstorder_10Percentile_MTV25_FBC	9
firstorder_10Percentile_MTV25_FBW	9	
1-year PFS	glrlm_GrayLevelNonUniformity_MTV25_FBW	87
	ngtdm_Busyness_MTV40_FBW	78
	glcm_Imc2_MTV40_FBW	59
	gldm_GrayLevelNonUniformity_MTV25_FBW	46
	glcm_ClusterTendency_MTV25_FBC	45
	glcm_MaximumProbability_MTV40_FBC	41
	shape_SurfaceArea_MTV25_FBC	39
	shape_SurfaceArea_MTV25_FBW	39
	Age	37
	firstorder_Minimum_MTV25_FBW	35
1-year LR	Age	52
	glcm_Imc2_MTV40_FBW	34
	glszm_GrayLevelNonUniformity_MTV25_FBW	31
	firstorder_Minimum_MTV25_FBC	20
	firstorder_Minimum_MTV25_FBW	20
	glcm_MaximumProbability_MTV40_FBC	18
	Gender	16
	GTV	16
	glrlm_RunPercentage_MTV25_FBW	15
glszm_GrayLevelNonUniformity_MTV40_FBW	15	

1-year DM	glrlm_GrayLevelNonUniformity_MTV25_FBW	94
	glcm_Imc2_MTV40_FBW	86
	ngtdm_Busyness_MTV40_FBW	62
	glcm_MaximumProbability_MTV40_FBC	59
	glcm_ClusterTendency_MTV25_FBC	44
	shape_SurfaceArea_MTV25_FBW	41
	shape_SurfaceArea_MTV25_FBC	40
	glszm_GrayLevelNonUniformity_MTV25_FBW	39
	gldm_GrayLevelNonUniformity_MTV25_FBW	33
	shape_Maximum2DDiameterColumn_MTV40_FBW	29

Supplementary Figure 3 demonstrates that GLM_{all} performs slightly better than GLM_{elnet} and reaches a 'plateau' when 80% of the data is used for training. The percentage data used for training started at 20% in order to minimize the occurrence that one of the folds for cross validation contained 0 events. GLM_{elnet} may require a larger dataset to obtain a similar predictive performance as GLM_{all} . The middle graph shows a variance plot, where the variance is plotted against the percentage data used for training. Both feature selection methods show the same trend for variance with an increasing number of patients used for training. The graph at the bottom displays the number of times that zero prognostic features were found with elastic net regression. Elastic net regression is able to find more prognostic features when the number of patients used for training is increasing, and this is also correlated to predictive performance. GLM_{all} is able to find prognostic features in 100% of the iterations when using 100% of the dataset, in contrast to GLM_{elnet} .



Supplementary Figure 3. A comparison of the feature selection method based on repeatability, independence, and elastic net regression (GLM_{all}), and a method using only elastic net regression (GLM_{eNet}) was performed in predicting 2-year overall survival. The upper graph shows a sensitivity study of the predictive performance against the percentage data used for training of the model. The middle graph shows a variance plot, where the variance is plotted against the percentage data used for training. The graph at the bottom displays the number of times that zero prognostic features were found with elastic net regression.

S9. Overview of the predictive performance of different assessed models

Besides GLM_{all}, GLM_{rad}, and GLM_{clin}, also other more simplistic models were assessed, such as a model that could only select SUV_{max} and GTV. In addition, also the effect of adding solely SUV_{max} to the clinical variables was investigated. Similar assessments were performed with total lesion glycolysis (TLG), a prognostic feature in NSCLC used to describe tumor burden [8]. The effect of TLG was investigated by adding it to GLM_{all}, GLM_{clin}, and GLM_{rad}, but results did not improve significantly. Results are summarized in Supplementary Table 11.

Supplementary Table 11. An overview of 9 different assessments on predictive performance using elastic net regression. The AUC values are depicted per clinical endpoint, per model. Values below Validation 1 represent the average of 100 AUC values with standard deviation. Below Validation 2 is the AUC calculated on the external validation set using the best performing fitted model from Validation 1.

Endpoint	GLM _{all}		GLM _{clin}		GLM _{rad}	
	Validation 1	Validation 2	Validation 1	Validation 2	Validation 1	Validation 2
2 year OS	0.66±0.07	0.55	0.67±0.06	0.59	0.57±0.07	0.45
2 year PFS	0.50±0.06	0.52	0.50±0.05	0.50	0.50±0.05	0.71
1 year PFS	0.57±0.06	0.51	0.59±0.07	0.54	0.58±0.07	0.54
1 year LR	0.57±0.07	0.55	0.64±0.08	0.58	0.52±0.05	0.50
1 year DM	0.58±0.06	0.52	0.59±0.08	0.52	0.60±0.07	0.50

Endpoint	GLM _{eNet}		SUV _{max} + GTV		GLM _{clin} + SUV _{max}	
	Validation 1	Validation 2	Validation 1	Validation 2	Validation 1	Validation 2
2 year OS	0.64±0.07	0.48	0.55±0.07	0.55	0.67±0.08	0.60
2 year PFS	0.53±0.06	0.63	0.50±0.01	0.50	0.50±0.07	0.50
1 year PFS	0.56±0.07	0.59	0.59±0.07	0.62	0.59±0.06	0.55
1 year LR	0.53±0.05	0.54	0.52±0.06	0.51	0.62±0.06	0.58
1 year DM	0.58±0.08	0.54	0.60±0.07	0.57	0.60±0.07	0.52

Endpoint	GLM _{all} + TLG		GLM _{clin} + TLG		GLM _{rad} + TLG	
	Validation 1	Validation 2	Validation 1	Validation 2	Validation 1	Validation 2
2 year OS	0.65±0.08	0.59	0.67±0.08	0.62	0.55±0.07	0.57
2 year PFS	0.50±0.06	0.52	0.50±0.04	0.53	0.50±0.06	0.69
1 year PFS	0.57±0.06	0.58	0.59±0.07	0.52	0.59±0.06	0.51
1 year LR	0.57±0.07	0.58	0.61±0.07	0.60	0.52±0.05	0.50
1 year DM	0.60±0.07	0.58	0.59±0.07	0.51	0.61±0.07	0.52

Supplementary References

1. Leijenaar RTH, Nalbantov G, Carvalho S, et al. *The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis.* Sci Rep 2015; 5: 11075.
2. Desserot MC, Tixier F, Weber WA, et al. *Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort.* J Nucl Med 2017; 58(3): 406-411.
3. Hatt M, Majdoub M, Vallieres M, et al. *18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort.* J Nucl Med 2015; 56: 38–44.
4. Bland JM, Altman DG. *Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement.* Lancet 1986;1(8476):307-10.
5. Zaki R, Bulgiba A, Ismail R, et al. *Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review.* PLoS One. 2012;7:e37908.
6. Shang J, Ling X, Zhang L, et al. *Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer.* Eur J Nucl Med Mol Imaging 2016;43:1945-53.
7. Zou H, Hastie T. *Regularization and variable selection via the elastic net.* J Royal Stat Soc B 2005;67(2):301–320.
8. Chen HH, Chiu NT, Su WC, et al. *Prognostic value of whole-body total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer.* Radiology 2012;264:559–566.

Chapter 7

General discussion and future perspectives



General discussion and future perspectives

Hospitals in low- and middle-income countries (LMIC) slowly start to gain more access to ^{18}F -fluorodeoxyglucose positron emission tomography / computed tomography (FDG PET/CT) technology. FDG PET/CT plays an essential role in the management of patients with non-small cell lung cancer (NSCLC) and became the standard in high-income countries for patient selection for radical therapy and radiotherapy tumor volume delineation (TVD). We investigated methods to reduce inter-observer variability (IOV) in FDG PET/CT based TVD of NSCLC. As a first step we developed pragmatic guidelines to standardize contouring of primary lung carcinoma and involved lymph nodes in patients with NSCLC. Secondly, these guidelines were introduced to radiation oncologists (ROs) and nuclear medicine physicians (NMPs) from LMIC with minimal experience in utilizing FDG PET/CT for radiotherapy planning purposes. Our guidelines contributed to increased delineation accuracy and reduced IOV in TVD. Together with other factors such as patient selection with FDG PET/CT, treating patients with concurrent chemoradiotherapy (CRT), and improved collaboration between departments, our standardized TVD approach showed in a cohort comparison a positive trend towards improved overall survival (OS) and progression-free survival (PFS) in patients with locally advanced NSCLC. Next to improving TVD, there is wide interest in improving prognostic models in locally advanced NSCLC that can aid in treatment decision making. Locally advanced NSCLC is a highly heterogeneous disease and requires a more individualized treatment strategy to avoid ineffective therapy. Such a strategy demands the use of more accurate prognostic factors than that are currently used in clinic. Despite some studies claiming to have found these accurate prognostic factors through PET radiomics analysis, our findings indicate that these quantitative features from FDG PET imaging contain no complementary prognostic information next to well-known clinical variables. This suggests that the addition of PET radiomics features in prognostic models do not improve treatment decision making in patients with locally advanced NSCLC, although this claim is limited by sample size. The work in this thesis underlines the difficulty of reproducing results in PET radiomics studies, and emphasizes the need for extensive evaluation of patient cohorts and validation datasets.

This thesis demonstrated the implications of introducing FDG PET/CT guided TVD in clinical practice in LMIC and provided more insight in the prognostic value of FDG PET radiomics features. The next paragraphs will elaborate on several other approaches reported in literature that focused on reducing IOV in TVD and improving prognostication in lung cancer. Our methods will be weighed against those approaches with regards to applicability in LMIC and in addition, future perspectives are discussed.

Interobserver variability in tumor volume delineation

Delineation of the tumor volume is of prime importance for lung cancer treatment. Inadequate TVD introduces a systematic error that could potentially result in a decrease of the dose delivered to the tumor, reduced local control and/or increased patient morbidity. TVD represents the largest uncertainty in the radiotherapy process compared to patient set-up error, inter- and intra-fraction organ movement, and patient movement [1]. IOV is a result of uncertainties in TVD and can be reduced by optimization of current imaging techniques [2], use of multi-modality imaging [3], implementation of standardized protocols and delineation guidelines including training [4-6] and semi-automated delineation methods [7]. We have chosen to investigate methods to decrease IOV using a pragmatic approach taking into account the limited resources in LMIC.

Many guidelines on PET/CT based TVD were published prior to our work, but were not addressing practicalities such as contouring of tumors in specific situations and to what extend to use information from PET and CT [8,9]. A comprehensive set of guidelines covering standardized PET/CT guided TVD practicalities in lung cancer was developed in this thesis with the aim to reduce IOV in delineation of lung cancer. *Chapter 4* demonstrates that our pragmatic guidelines contributed to reduced IOV. In the absence of 4D CT techniques in most LMIC, a different delineation approach to take into account motion was proposed in *Chapter 2*: the respiration expanded Gross Tumour Volume (reGTV). This served as alternative for the International Committee on Radiation Units and Measurements' (ICRU) defined internal target volume (ITV). The suggested 3D PET/CT based reGTV in our study guidelines is quite similar to the ITV, however, ITV relies on information from 4D CT to account for different tumor positions during the respiratory cycle [10]. LMIC have limited access to 4D CT. Alternatively, LMIC could use slow or fast 3D CT scanning techniques, although these techniques are subject to motion-induced artefacts [11]. Therefore, the PET/CT based reGTV approach seems the best pragmatic alternative to take into account motion and avoid geographic tumor miss [12].

While the use of our IAEA study guidelines reduced IOV, it is not known if the IOV could be further reduced with even more additional training. It is common knowledge that the learning of new methods is subject to a learning curve. The shape and impact of this learning curve on IOV was not studied in this thesis. It is worthwhile to investigate when a 'plateau' is reached during training. The plateau resembles the remaining uncertainty that cannot be solved with additional training. Subsequently, our guidelines need to be improved or alternative methods as described in *Chapter 2* and below should be studied to deal with remaining uncertainty.

As observed in *Chapter 4* and *Chapter 5*, the IOV in manual PET/CT based TVD stems from differences in window/level (W/L) settings, non-tumor specific FDG uptake adjacent to tumor, and tissue with low to zero FDG uptake though highly suspicious for being malignant (in the opinion of the NMP or RO). One could argue whether the latter deviates from the definition of GTV: the demonstrable extent seen or felt with standard examination techniques. In reality, not only the demonstrable extent, but also the likeliness of the extent of the disease has impact on TVD. As a result, the experience from the observer is contributing to IOV in clinic as observers also include tissue with a high likelihood of containing tumor in their opinion [13]. Measures to counteract on this subjectivity are to 1) validate PET/CT based contours using histopathology results of excised lesions, or 2) to use intelligent computer techniques that can automatically generate contours, also known as Artificial Intelligence (AI). The first approach gives us insight in the ‘ground truth’ that could not only eliminate uncertainties in TVD, but also opens up opportunities to gain more insight in subpopulations within the tumor useful for ‘dose painting’ [14]. Many research groups have shown that specimen handling techniques used to obtain a histopathology-derived ground truth are subject to geometrical uncertainties due to specimen deformation or shrinkage [15]. In addition, locally advanced NSCLC is seldom treated with surgery, hence, in our patient cohorts the number of PET image datasets accompanied with histopathology information is limited. Validating PET images and their corresponding contours against tumor pathology remains challenging due to the significant difficulties in performing these studies and robust methods to provide the ground truth are still awaited. A second approach that is commonly studied is the use of AI applications for TVD. AI applications, such as deep learning networks, are in theory better than humans in solving the statistically raised questions stated by the ROs and NMPs e.g. on the likelihood of tumor presence in suspicious areas on PET/CT scans during TVD. Deep learning networks can automatically learn feature representations from a given dataset and have been shown to match and even surpass human performance in task-specific applications [16]. To match or even surpass human performance, AI methods are dependent on sufficient and high-quality data and their models need to be further optimized to achieve clinically relevant results in TVD and treatment decision making. Despite suggesting manual delineation in *Chapter 2* and *Chapter 3* instead of using automatic delineation due to its current inaccuracies, it is expected that using AI for TVD will be at the same level as a human expert in the near future [17]. AI for TVD has the potential to reduce efforts significantly as manual TVD is a time-consuming task and could even aid in malignancy detection [18]. I believe it is rather the question of when instead of which of the uncertainties mentioned above in TVD will be solved using an (affordable) technical solution. Hence, I foresee that multi-disciplinary meetings on TVD will be on reaching a consensus on a highly standardized automatically generated contour instead of discussing a highly subjective manually drawn contour.

PET guided prognostication in lung cancer

Accurate definition of target volumes forms the basis of an effective treatment. However, even if we can define the exact location of the tumor in all lung cancer patients, this does not mean each patient will cure from receiving a given treatment. For example, locally advanced NSCLC is highly heterogeneous meaning each tumor has its own characteristics and responds differently to concurrent CRT. To treat locally advanced NSCLC more effectively, prognostic factors should be discovered that could describe these tumor characteristics and predict response to treatments. Subsequently, these features could increase the accuracy of current prognostic models that would improve treatment decision making and avoid giving patients ineffective treatments.

We investigated the prognostic value of PET radiomics features to improve the current prognostic model in lung cancer management. The rationale of the work described in *Chapter 6* was to identify FDG PET radiomics features for NSCLC patients that were robust, independent, prognostic, and complementary to well-established clinical variables. We found PET radiomics features that were robust and prognostic, and did not show any association with well-known prognostic clinical variables. However, results were not convincing that PET radiomics features provided complementary prognostic information next to clinical variables in patients with NSCLC treated with concurrent CRT. Consequently, there is no clear clinical benefit in the use of PET radiomics features as described in this thesis for prognostication.

There are differences between tumors of the same type in different patients, and between cancer cells within a tumor of the same patient, called tumor heterogeneity. Both can lead to different responses to therapy. In PET radiomics we want to measure FDG uptake patterns describing tumor heterogeneity. However, in *Chapter 6* we observed that the majority of PET radiomics features are highly correlated to volume and basic SUV metrics and may therefore not be suitable to provide us the information that we wish to capture. While volumetric properties and SUV metrics are indeed clinically useful, they do not represent the FDG uptake patterns capturing underlying tumor biology as supposedly sought with PET radiomics analysis. Brooks et al. also observed the association with volume on a limited number of PET radiomics features and placed a critical note that assessment of underlying tumor biology with PET radiomics features may be impossible [19]. Namely, the magnitudes for PET radiomics feature differences were so small in comparison to those resulting from differences in e.g. volume, that it would be possible that tumor would exhibit distinctly different tumor biology without capturing this with PET radiomics analysis [20]. This assumes that all PET radiomics features are associated in a way to volume and SUV metrics, although this is not in line with our findings in *Chapter 6* where independent PET

features were observed. However, further investigations should assess any relationship of these independent PET features with underlying tumor biology in detail..

Although we have shown that improving risk stratification and clinical decision making based on clinical variables and PET radiomics features has still been proven difficult, we believe there is more to gain from (pre-treatment) PET imaging than its current use that could improve prognostication in patients with NSCLC. Several other approaches could be studied to assess the prognostic value of PET imaging.

The use of “human engineered” PET radiomics features is highly debatable due to many confounding factors, such as tumor volume and simple SUV metrics that can influence their outcome as described previously. These confounding factors make results hard to reproduce and moreover difficult to compare amongst institutions. Another way to assess the added value of PET imaging in prognostication is to study the prognostic value of features found by deep learning networks. As noted before, deep learning networks require a large sample size of high quality, but do have the potential to quickly process large amounts of data and identify independent prognostic features. Recently, methods have been proposed to use a small set of medical images to fine-tune the deep network learned from a large natural image set, which could solve the problem of insufficient medical training data [21]. It has not been studied if such a method would work with PET images. Even so, this study demonstrated that their deep learning features were more discriminative than the radiomics texture features, mainly because this method seems to be more robust to small volumes. Further studies should investigate the prognostic value of deep learning features from PET imaging.

Other studies have investigated multi-modality radiomics features instead of processing the radiomics features per imaging modality separately to improve prognostic models. Zhou et al. observed that their multi-modality deep learning approach fusing PET/CT and clinical parameters could predict high risk NSCLC patients for early-stage distant failure with an AUC value of 0.83 [22]. A few other research groups have shown that multimodality-based prognostic biomarkers can be used to predict disease recurrence and survival in the context of NSCLC, but most of them do not provide sufficient evidence, because of limited sample size and absence of validation cohorts [23].

As can be learned from this thesis, improving healthcare requires a multidisciplinary approach, hence for prognostication it is interesting to also look further than radiomics. Researchers outside the field of radiomics have investigated the prognostic value of circulating cell-free DNA (cfDNA) and/or circulating tumor DNA (ctDNA) in the blood of patients with NSCLC. Non-invasive assessment of cfDNA and/or ctDNA in plasma and

serum of lung cancer patients is possible by sensitive molecular biology techniques [24]. The use of so called liquid biopsies to assess ctDNA could enhance the understanding of disease progression, and has already been shown to carry prognostic information [25-27]. Limited data has been published regarding cfDNA and ctDNA in locally advanced NSCLC and its ability to correlate with treatment outcomes. Hence, outcomes of future studies on prognostic value of ctDNA in patients with locally advanced NSCLC must be awaited to validate promising early results. With regards to LMIC, the highly sensitive techniques required to analyze cfDNA and ctDNA in combination with the use of targeted drug therapies may be too costly at the moment for clinical use.

Other studies indicate that, when radiomics features are combined with clinical and genomic features, radiomics could help optimize treatment decision making [23]. Ultimately, the search for more of these prognostic factors in conjunction with radiomics features could contribute to improved patient selection to avoid ineffective therapy and improve quality of life.

Recommendations for PET/CT in radiotherapy TVD and prognostication

The work performed in this thesis helped to establish a key role for FDG PET/CT in the management of patients with NSCLC in LMIC. Also in these countries PET/CT is now becoming the standard for patient selection for radical therapy and radiotherapy TVD. The various approaches described in this chapter are promising tools for reducing uncertainties in TVD and could contribute to improved patient selection for radical therapy. Further studies are necessary to deal with uncertainties in TVD and improve accuracy of prognostic models to aid in treatment decision making. In LMIC, the other approaches presented in this chapter are not ready to be implemented yet in clinic as they are still in the validation phase. Therefore we suggest to first continue disseminating expertise on recent advancements in radiation treatment planning in LMIC using hands-on training and webinars as provided by IAEA [28]. The training program focused mainly on delineation of primary lung cancers, but as a next step, could also be extended to lymph node delineation and towards other disease sites such as head and neck cancer. More recent guidelines from ESTRO and ACROP covered the topic on lymph node management in more detail [29]. Including these detailed and clear guidelines in future IAEA guidelines may result in improved lymph node management and delineation, which may also benefit survival of patients with NSCLC.

For most potential solutions for improved target volume definition and prognostication, researchers would benefit if there was a vast amount of data publicly available on request to increase the number of patients in their study. Brooks et al. suggested to overcome challenges in PET radiomics analysis by creating homogeneous patient cohorts with equivalent tumor volumes [20]. This requires great efforts in data collection and data management, which is currently insufficient in most centers. We believe it is the most pragmatic and economic solution on the long run, and centers should be encouraged to start standardizing ways of gathering imaging data and sharing data with other investigators. The creation of a large imaging database including other clinical information such as treatment outcome and radiotherapy plans would enable the next step in medical image analysis for TVD and prognostication, and could thus lead to improved local control and treatment decision making. In the case of target volume delineation, building an ‘atlas’ or database with a wide variety of delineation examples could also provide tools for education. A database with clinical cases would move us in the direction of creating a database for use of training deep learning networks. Currently, a limited number of medical imaging datasets is already available online (e.g. Cancer Imaging Archive and Quantitative Imaging Network), although limited to certain imaging modalities, not always ready-for-use and often of suboptimal quality. On the other hand, these first available datasets are highly appreciated and will encourage other investigators to share data as well.

Institutions may be limited in their data, but are not limited in asking for collaborations. Hospitals in LMIC should be encouraged to look outside their institute when it comes to collecting data. Recommendations are to set up a large network of researchers within their country or between countries, where IAEA can serve as coordinator, that share the same vision, and are open to share data on similar topics of interest. Hospitals in LMIC would be wise to start preparing their IT infrastructure as a first step. It seems that collection of a vast amount of high quality data is necessary to overcome current hurdles in target volume delineation and prognostication. Therefore, the most pragmatic way to be part of this evolution is to contribute with supportive high-quality data.

Conclusions

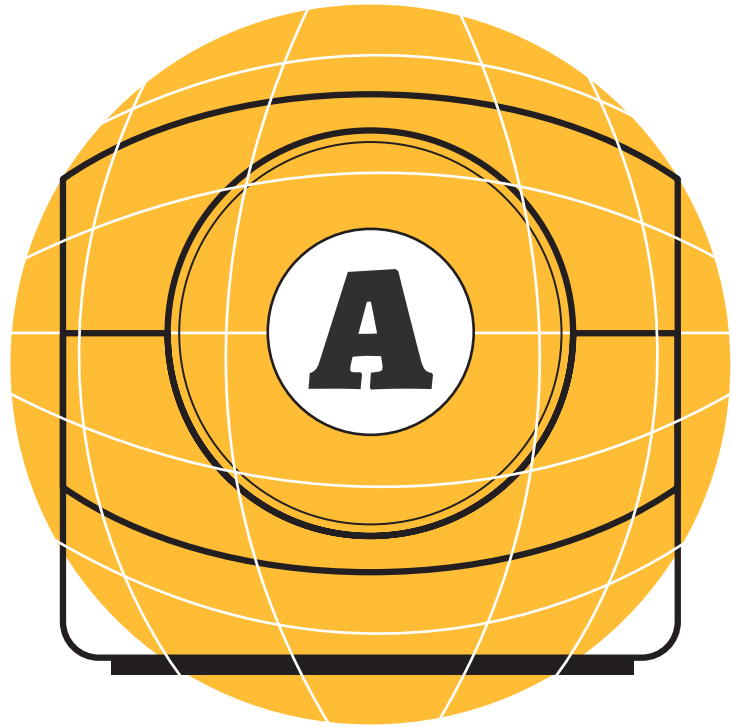
The work in this thesis focused on improved patient selection for concurrent CRT, and increased and standardized treatment accuracy using PET/CT applications, with the goal to improve survival in patients with NSCLC. Our work contributed to a broader use of PET/CT in LMIC by providing 1) pragmatic standardized guidelines for TVD, 2) training on PET/CT based concurrent CRT and 3) evidence on the effectiveness of PET/CT guided concurrent CRT in their centers. This thesis also contributed to the debate of improving

prognostic accuracy in locally advanced NSCLC patients with the use of PET radiomics features, and further research is warranted in the field of prognostication. In order to establish common ground in the clinic, future studies should investigate if the prognostic accuracy can be further improved by combining or fusing imaging features from multiple imaging modalities with clinical and genomic features. A prognostic model that integrates features using a multidisciplinary approach is needed to fully address the challenges of radiomics toward practical applications. Current efforts to create larger databases will pave the way for AI applications in TVD and prognostication for improved lung cancer management improving patients’ quality of life worldwide.

References

1. Weiss, E. and Hess, C.F. *The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences.* *Strahlen Onkol.* 2003;179: 21–30.
2. Vaquero JJ, Kinahan P. *Positron Emission Tomography: Current Challenges and Opportunities for Technological Advances in Clinical and Preclinical Imaging Systems.* *Annu Rev Biomed Eng.* 2015;17:385–414.
3. MacManus M, Everitt S, Schimek-Jasch T, et al. *Anatomic, functional and molecular imaging in lung cancer precision radiation therapy: treatment response assessment and radiation therapy personalization.* *Transl Lung Cancer Res.* 2017; 6(6): 670–688.
4. Nestle U, Rischke HC, Eschmann SM, et al. *Improved inter-observer agreement of an expert review panel in an oncology treatment trial – Insights from a structured interventional process.* *Eur J Cancer* 2015;51:2525–2533.
5. Schimek-Jasch T, Troost EG, Rücker G, et al. *A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: Reducing the interobserver variability in multicentre clinical studies.* *Strahlenther Onkol* 2015;191:525–533.
6. Doll C, Duncker-Rohr V, Rücker G, et al. *Influence of experience and qualification on PET-based target volume delineation. When there is no expert—ask your colleague.* *Strahlenther Onkol* 2014;190:555–562.
7. Zaidi H, El Naqa I. *PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques.* *Eur J Nucl Med Mol Imaging* 2010;37:2165–2187.
8. Spoelstra FO, Senan S, Le Pechoux C, et al., Lung Adjuvant Radiotherapy Trial Investigators Group. *Variations in target volume definition for postoperative radiotherapy in stage III non-small-cell lung cancer: analysis of an international contouring study.* *Int J Radiat Oncol Biol Phys* 2010;76:1106–1113.
9. MacManus M, Nestle U, Rosenzweig KE, et al. *Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006–2007.* *Radiother Oncol* 2009 Apr;91(1):85–94.
10. Ehrbar S, Jöhl A, Tartas A, et al. *ITV, mid-ventilation, gating or couch tracking - A comparison of respiratory motion-management techniques based on 4D dose calculations.* *Radiother Oncol* 2017;124(1):80–88.
11. Chen GT, Kung JH, Beaudette KP. *Artifacts in computed tomography scanning of moving objects.* *Semin Radiat Oncol.* 2004;14(1):19–26.
12. Caldwell CB, Mah K, Skinner M, et al. *Can PET provide the 3D extent of tumor motion for individualized internal target volumes? A phantom study of the limitations of CT and the promise of PET.* *Int J Radiat Oncol Biol Phys* 2003;55(5):1381–93.
13. Doll C, Duncker-Rohr V, Rücker G, et al. *Influence of experience and qualification on PET-based target volume delineation. When there is no expert—ask your colleague.* *Strahlenther Onkol.* 2014;190(6):555–62.
14. Even AJ, van der Stoep J, Zegers CM, Reymen B, Troost EG, Lambin P, et al. *PET-based dose painting in non-small cell lung cancer: comparing uniform dose escalation with boosting hypoxic and metabolically active sub-volumes.* *Radiother Oncol* 2015; 116: 281–6.
15. Kirov AS, Fanchon L. *Pathology-validated PET image data sets and their role for PET segmentation.* *Clin Trans Imaging.* 2014;2:253–267.
16. Hosny, A., Parmar, C., Quackenbush, J. et al. *Artificial intelligence in radiology.* *Nat Rev Cancer* 2018;18:500–510.
17. Thompson RF, Valdes G, Fuller CD, et al. *The Future of Artificial Intelligence in Radiation Oncology.* *International Journal of Radiation Oncology Biology Physics.* 2018;102(2):247–248.
18. Weikert T, Akinci D'Antonoli T, Bremerich J, et al. *Evaluation of an AI-Powered Lung Nodule Algorithm for Detection and 3D Segmentation of Primary Lung Tumors.* *Contrast Media Mol Imaging.* 2019:1545747.
19. Brooks FJ, Grigsby PW. *The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake.* *J Nucl Med* 2014;55:37–42.
20. Brooks FJ, Grigsby PW. *Low-Order Non-Spatial Effects Dominate Second-Order Spatial Effects in the Texture Quantifier Analysis of 18F-FDG-PET Images.* *PLoS One* 2015;10(2):e0116574.
21. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. *Convolutional neural networks for medical image analysis: full training or fine tuning?* *IEEE Trans Med Imag.* 2016;35:1299–312
22. Zhou Z, Zhou ZJ, Hao H et al. *Constructing multi-modality and multi-classifier radiomics predictive models through reliable classifier fusion.* 2017. arXiv:1710.01614
23. Liu Z, Wang S, Dong D, et al. *The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges.* *Theranostics* 2019;9:1303–22
24. Gedvilaitė V, Schweigert D, Cicėnas S. *Cell-free DNA in non-small cell lung cancer.* *Acta Med Litu.* 2017;24(2):138–144.
25. Winther-Larsen A, Demuth C, Fledelius J, et al. *Correlation between circulating mutant DNA and metabolic tumour burden in advanced non-small cell lung cancer patients.* *Br J Cancer* 2017;117:704–9.
26. Hyun MH, Sung JS, Kang EJ, et al. *Quantification of circulating cell-free DNA to predict patient survival in non-small-cell lung cancer.* *Oncotarget* 2017;8:94417–30.
27. Lee Y, Park S, Kim WS, et al. *Correlation between progression-free survival, tumor burden, and circulating tumor DNA in the initial diagnosis of advanced-stage EGFR-mutated non-small cell lung cancer.* *Thorac Cancer* 2018;9:1104–10.
28. Konert T, Vogel WV, Everitt S, et al. *Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol.* *Radiother Oncol* 2016;121(1):39–45.
29. Nestle U, De Ruysscher D, Ricardi U, et al. *ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer.* *Radiother Oncol* 2018;127:1–5.

Appendices



Summary

Towards worldwide use of FDG PET/CT applications for optimal treatment of lung cancer

In low- and middle-income countries (LMIC), an expected increase in lung cancer cases in the coming years is a major concern. Currently, there are already insufficient treatment facilities to provide the necessary healthcare services. Sites that are suitably equipped to provide nuclear medicine (NM) and radiation oncology (RO) services have minimal experience in the multidisciplinary use of PET/CT imaging for concurrent chemoradiotherapy (CRT), resulting in many patients receiving suboptimal staging, therapy selection, and treatment. The work in this thesis was performed in collaboration with the International Atomic Energy Agency (IAEA). Our aim was to achieve accurate treatment delivery and improved prognostic tumor characterization for optimized lung cancer management with PET/CT in patients with locally advanced NSCLC treated with CRT. Ultimately, the goal was to improve the survival of patients with locally advanced NSCLC. The following objectives were defined:

- To develop pragmatic guidelines for TVD;
- To reduce the interobserver variability in TVD;
- To improve the current prognostic models for patients with lung cancer.

Our work focused on introducing FDG PET/CT applications for patients with locally advanced non-small cell lung cancer (NSCLC). Patients with locally advanced NSCLC would benefit from the multidisciplinary use of FDG PET/CT imaging for concurrent CRT, the current standard treatment in high-income countries (HIC). In LMIC, there is an increased demand for more expertise in RO and NM in these centers to be able to provide the current standard treatment. Concurrent CRT requires accurate treatment delivery as it is more intensive than sequential CRT and comes with increased risk of adverse events and longer hospitalization. FDG PET/CT serves as a valuable tool in radiotherapy planning to accurately and precisely delineate tumor. Hence, nuclear medicine physicians and radiation oncologists from Brazil, India, Jordan, Pakistan, Turkey, Vietnam, and Estonia, were trained in delineating target volumes using combined PET and CT images. **Pragmatic IAEA guidelines for PET/CT-based tumor volume delineation (TVD) were developed**, dealing with the uncertainties of PET/CT for management of respiration motion, and the PET/CT-based respiration expanded GTV approach was introduced. This delineation protocol served as comprehensive guidance on using PET/CT for radiotherapy planning in NSCLC, which had never been provided to this extent. These guidelines were incorporated in training events on PET/CT-based TVD in lung cancer, which were given over a period of

one year. Our participants were given contouring assignments which were performed before and after training events to investigate the impact of the IAEA guidelines on delineation accuracy and interobserver variability. In the absence of a histopathological proven gold standard, two senior ROs and a senior NM physician delineated one reference 'expert' contour per assignment and this was compared with the contours from the participants. From our results we observed that **PET/CT based contouring following IAEA protocol established a 25% increase in delineation accuracy and a 45% reduction in interobserver variability** after multiple training interventions. As a next step, we investigated whether the training interventions on PET/CT based TVD in lung cancer had any impact on local control and patient survival. The assessment of the clinical impact of the multiple training interventions included the comparison of the overall survival (OS) and progression-free survival (PFS) between a retrospective cohort of patients who were treated prior to the trainings, and a prospective cohort of patients who were treated after the trainings following IAEA protocol. A planned interim analysis showed a significant improvement in the prospective cohort: **OS increased from 14 to 23 months and PFS increased from 11 to 17 months.**

To further improve survival in patients with locally advanced NSCLC, we investigated whether FDG PET could provide prognostic information to aid in therapy selection. In this thesis we looked at whether quantitative imaging features from PET, called PET radiomics features, would exhibit prognostic information. PET radiomics features could potentially complement biopsy based histopathologic examination by characterization of interpatient tumor heterogeneity, which could aid in treatment decision making. Currently, the choice of treatment is largely dependent on the disease stage and patient condition, and this has not always been accurate enough when selecting the optimal treatment. Patients with the same disease stage and condition can still respond differently to the same treatment due to interpatient tumor heterogeneity. Therefore, the search for additional prognostic factors is warranted in order to strive for maximum therapeutic efficacy and to reduce the number of futile treatments, thus optimizing use of resources. Unfortunately, we observed that most PET radiomics features in our study were strongly correlated to well-established prognostic factors. On the other hand, we did find PET radiomics features that were robust, independent, and that also exhibited prognostic value. However, there was no significant difference observed between the prognostic value of combined PET radiomics features and clinical variables and clinical variables alone when predicting clinical endpoints in NSCLC patients treated with concurrent CRT. Hence, **we were unable to increase the accuracy of current prognostic models in locally advanced NSCLC by adding PET radiomics features.** Improving risk stratification and clinical decision making based on clinical variables and PET radiomics features has still been proven difficult in locally advanced lung cancer patients and further investigation in the area of prognostication is warranted.

The work in this thesis focused on improved patient selection for concurrent CRT, and increased and standardized treatment accuracy using PET/CT applications, with the goal to improve survival in patients with NSCLC. Our work contributed to a broader use of PET/CT in LMIC by providing 1) pragmatic standardized guidelines for TVD, 2) training on PET/CT based concurrent CRT and 3) evidence on the effectiveness of PET/CT guided concurrent CRT in their centers. This thesis also contributed to the debate of improving prognostic accuracy in locally advanced NSCLC patients with the use of PET radiomics features, and further research is warranted in the field of prognostication. Future studies should investigate if the prognostic accuracy can be further improved by using imaging features from multiple imaging modalities with clinical and genomic data.

Samenvatting

Wereldwijde inzet van FDG PET/CT applicaties voor optimale behandeling van longkanker

In de meeste landen met een laag of midden inkomen (LMIC) is de zorg voor longkanker niet op het niveau van landen met een hoog inkomen (HIC), en is de overleving bij lokaal gevorderde niet-kleincellig longcarcinoom (NSCLC) met kliermetastasen erg slecht. Een gebrek aan kennis, kunde, en resources spelen een rol bij vrijwel alle aspecten van diagnostiek tot therapie. Er zijn in het bijzonder problemen bij het innovatieve en multidisciplinaire gebruik van PET-scans voor radiotherapie, een medische afbeeldingstechniek waarmee het mogelijk is om moleculaire processen (bijvoorbeeld metabolisme) in het lichaam kwantitatief af te beelden. Dit heeft als gevolg dat patiënten in LMIC vaak niet het juiste stadium en behandelplan krijgen of een suboptimaal uitgevoerde behandeling ondergaan. Het werk in dit proefschrift is uitgevoerd in samenwerking met het internationaal atoomenergie agentschap (IAEA). Ons gezamenlijke doel is de nauwkeurigheid te verbeteren van radiotherapie met behulp van PET/CT bij patiënten met lokaal gevorderde NSCLC in LMIC. Daarnaast streven we ernaar om karakteristieke tumor eigenschappen met een prognostische waarde te kunnen meten met behulp van PET scans. Deze meetbare prognostische eigenschappen kunnen worden ingezet ter verbetering van het huidige prognostische model, waar momenteel de behandelkeuze op gebaseerd is. Het huidige prognostische model blijkt nochtans onnauwkeurig voor patiënten met lokaal gevorderde NSCLC. De prognostische eigenschappen afkomstig van PET scans kunnen mogelijk dienen als extra ondersteuning voor de klinici bij het maken van een behandeling op maat. Uiteindelijk is het hogere doel om de overleving van patiënten met lokaal gevanceerde NSCLC te verbeteren. De volgende doelstellingen zijn gedefinieerd:

- Het ontwikkelen van pragmatische richtlijnen voor intekenen van tumoren tijdens het maken van een bestralingsplan;
- Het verminderen van de variatie van ingetekende doelgebieden tussen de klinici;
- Het verbeteren van het huidige prognostische modellen voor patiënten met longkanker.

Ons werk richtte zich op de introductie van PET/CT toepassingen voor de behandeling van patiënten met lokaal gevorderde NSCLC. Patiënten met lokaal gevorderde NSCLC hebben baat bij het multidisciplinaire gebruik van PET/CT voor gecombineerde chemoradiotherapie (CRT), de standaard aanbevolen behandeling in HIC. Bij medische centra in LMIC is er een toenemende vraag naar meer expertise op de afdeling radiotherapie en nucleaire geneeskunde om gecombineerde CRT te kunnen bieden. Gecombineerde CRT vereist een nauwkeurige behandeling, omdat het intensiever is dan sequentiële CRT waardoor het risico hierdoor op bijwerkingen en een langere ziekenhuisopname groter is. PET/CT

dient als een waardevol instrument bij de voorbereiding van radiotherapie, aangezien men hiermee de doelgebieden nauwkeurig kan bepalen en intekenen. Daarom werden nucleair geneeskundigen en radiotherapeuten uit Brazilië, India, Jordanië, Pakistan, Turkije, Vietnam en Estland getraind in het intekenen van doelgebieden met behulp van gefuseerde PET/CT scans. **Er werden pragmatische IAEA richtlijnen geschreven voor het intekenen van tumoren gebaseerd op PET/CT.** Het IAEA protocol diende als uitgebreide leidraad bij het gebruik van PET/CT voor radiotherapie planning bij NSCLC, dat nog nooit in deze mate was verstrekt. Deze richtlijnen zijn opgenomen in een trainingsprogramma over een periode van een jaar. De deelnemers kregen intekenopdrachten voor en na trainingssessies om de impact te onderzoeken van de IAEA richtlijnen op de nauwkeurigheid van het intekenen en de variatie tussen de klinici. Bij gebrek aan een histopathologisch bewezen gouden standaard, hebben twee ervaren radiotherapeuten en een ervaren nucleair geneeskundige per intekenopdracht één referentie 'expert'-contour getekend en dit werd vergeleken met de contouren van de deelnemers. **De overeenkomst van de intekeningen van doelgebieden met de referentie bleek na afronding van het trainingsprogramma significant te zijn toegenomen met 25% en de variatie tussen de ingetekende doelgebieden van de deelnemers was verkleind met 45%.** Als volgende stap onderzochten we of de trainingssessies over PET/CT applicaties bij de behandeling van longkanker enige invloed hadden op de lokale tumorcontrole en de overleving van de patiënt. Hiervoor werd een vergelijking uitgevoerd tussen een retrospectieve groep patiënten die voorafgaand aan de trainingen is behandeld in de betreffende centra, en een prospectieve groep die na de training met verbeterde technieken en volgens IAEA protocol werd behandeld. Resultaten uit dit onderzoek tonen een significante verbetering van de overleving aan: **de algehele overleving steeg van 14 naar 23 maanden, en de progressie-vrije overleving steeg van 11 naar 17 maanden.**

Om de overleving bij patiënten met lokaal gevorderde NSCLC verder te verbeteren, hebben we onderzocht of PET/CT scans over prognostische informatie beschikken die een behandeling op maat mogelijk maakt. In dit promotieonderzoek hebben we gekeken of zogenaamde *radiomics features* uit PET scans over deze prognostische informatie zouden beschikken. PET radiomics features vertegenwoordigen distributiepatronen te zien op PET beelden, die weer gelinkt kunnen worden aan de tumorheterogeniteit. Daarom onderzochten we of er prognostische PET radiomics features bestonden ter aanvulling van de klinische besluitvorming. Helaas bleken de meeste PET radiomics features sterk gecorreleerd te zijn met bekende prognostische factoren en hierdoor waren ze niet van toegevoegde waarde. Anderzijds vonden we een aantal PET radiomics features die robuust en onafhankelijk waren en ook prognostische waarde vertoonden. Er werd echter geen significant verschil waargenomen in prognostische waarde tussen PET radiomics features en klinische variabelen bij het voorspellen van de overlevingskans van patiënten met

lokaal gevorderde NSCLC behandeld met gecombineerde CRT. **We hebben het huidige prognostische model dus niet kunnen verbeteren door toevoeging van PET radiomics features.**

Het werk in dit proefschrift heeft bijgedragen aan een intensiever gebruik van PET/CT in LMIC door 1) het bieden van pragmatische gestandaardiseerde richtlijnen voor het intekenen van longtumoren, 2) het geven van training in PET/CT applicaties voor gecombineerde CRT en 3) het leveren van bewijs over de effectiviteit van multidisciplinaire gebruik van PET/CT voor gecombineerde CRT. Dit proefschrift heeft ook bijgedragen aan onderzoek naar verbeteringen van de prognostische nauwkeurigheid bij patiënten met lokaal gevorderde NSCLC en de waarde hierin van PET radiomics features. Toekomstige studies zullen gaan uitwijzen of prognostische modellen verder kunnen worden verbeterd door radiomics features van meerdere afbeeldingstechnieken te combineren met genomics en andere klinische variabelen.

List of publications and author contributions

Konert T, Vogel W, MacManus MP, Nestle U, Belderbos J, Grégoire V, Thorwarth D, Fidarova E, Paez D, Chiti A, Hanna GG.

PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014.

Radiotherapy and Oncology. 2015; 116(1): 27-34.

Author contributions: guarantor of integrity of entire study, GH; conceptualization, TK, WV, MM, DT, EF, DP, AC, GH; study design, all authors; literature research, TK; data acquisition or data analysis/interpretation, TK, WV, GH; manuscript writing, TK, WV, GH; and manuscript editing, all authors.

Konert T, Vogel WV, Everitt S, MacManus MP, Thorwarth D, Fidarova E, Paez D, Sonke JJ, Hanna GG.

Multiple training interventions significantly improve reproducibility of PET/CT-based lung cancer radiotherapy target volume delineation using an IAEA study protocol.

Radiotherapy and Oncology. 2016; 121(1): 39-45.

Author contributions: guarantor of integrity of entire study, GH; conceptualization; TK, WV, EF, DP, GH; study design, TK, WV, EF, DP, GH; literature research, TK; data acquisition or data analysis/interpretation, TK; manuscript writing, TK; and manuscript editing, all authors.

Konert T, van de Kamer JB, Sonke JJ, Wouter Vogel.

The developing role of FDG PET imaging for prognostication and radiotherapy target volume delineation in non-small cell lung cancer.

Journal of Thoracic Disease. 2018; 10(21): 2508–2521.

Author contributions: guarantor of integrity of entire study, WV; conceptualization, all authors; study design, all authors; literature research, TK, WV; manuscript writing, TK; and manuscript editing, all authors.

Konert T, Vogel WV, Paez D, Polo A, Fidarova E, Carvalho H, Duarte PS, Zuliani AC, Santos AO, Altuhhova D, Karusoo L, Kapoor R, Sood A, Khader J, Al-Ibraheem A, Numair Y, Abubaker S, Soydal C, Kütük T, Le TA, Canh NX, Bieu BQ, Ha LN, Belderbos JSA, MacManus MP, Thorwarth D, Hanna GG.

Introducing FDG PET/CT-guided chemoradiotherapy for stage III NSCLC in low- and middle-income countries: preliminary results from the IAEA PERTAIN trial.

European Journal of Nuclear Medicine and Molecular Imaging. 2019; 46(11): 2235-2243.

Author contributions: guarantor of integrity of entire study, GH; conceptualization, TK, WV, MM, DP, AP, EF, DT, GH; study design, TK, WV, MM, GH; literature research, TK; data acquisition or data analysis/interpretation, TK; manuscript writing, TK; and manuscript editing, all authors.

Konert T, Everitt S, La Fontaine MD, van de Kamer JB, MacManus MP, Vogel WV, Callahan J, Sonke JJ.

Robust, independent and relevant prognostic 18F-fluorodeoxyglucose positron emission tomography radiomics features in non-small cell lung cancer: Are there any?

PLoS One. 2020; 15(2): e0228793.

Author contributions: guarantor of integrity of entire study, JJS; conceptualization, TK, SE, MM, JJS; study design, TK, JJS, WV, JvdK; literature research, TK; data acquisition or data analysis/interpretation, TK, SE, MLF, JC; manuscript writing, TK; and manuscript editing, all authors.



Dankwoord

Voor menig lezer zal dit het leukste deel zijn van mijn proefschrift. Zelf ben ik zeer verheugd op het moment van schrijven, aangezien het schrijven van het dankwoord betekent dat er een eind zit te komen aan dit wetenschappelijke avontuur. Mijn promotietraject was een reis om de wereld met vele uitdagende bergen die ik moest overwinnen, maar ook dalen waar ik weer moest uitklimmen. Ik maak graag van deze gelegenheid gebruik om mijn dankbaarheid te uiten aan alle personen die mij op deze reis gegidst hebben en/of gezelschap hebben gehouden. Een reis maken als deze, doe je nooit alleen.

Prof. dr. Jan-Jakob Sonke. Beste **JJ**, jouw supervisie en sturing hebben het proefschrift gemaakt tot wat het nu is. Na een tweetal voortvarende jaren dreigde mijn onderzoek te stagneren, echter was het voor mij een geruststelling dat ik kon terugvallen op iemand met veel 'brainpower' die mij altijd weer in beweging kon brengen met briljante suggesties en oplossingen. Ik heb daarnaast veel gehad aan je kritische houding die vaak tot nieuwe inzichten leidden en het werk naar een hoger niveau heeft getild. Naast het serieus bedrijven van wetenschap, was er ook ruimte voor ontspanning en humor. De borrels en congressen die we hebben beleefd zullen mij dan ook altijd bijblijven, waarbij jij altijd mijn voorbeeld was van "s avonds een man, 's ochtends een man". Hartelijk dank voor de plezierige samenwerking.

Mijn co-promotores Dr. Wouter Vogel en Dr. Marcel Stokkel. Beste **Wouter**, na mijn afstuderen heb je mij het vertrouwen gegeven om het internationale onderzoek met de IAEA voort te zetten. Vanaf het begin hebben wij altijd prettig en efficiënt kunnen samenwerken en zat je vol goede ideeën voor vervolgonderzoeken. Ik was ook erg onder de indruk van je taalkundige oplossingen als ik zoekende was naar de juiste woorden in mijn manuscripten. Ook ben ik je dankbaar voor je tijd als klinische begeleider op de afdeling nucleaire geneeskunde waar ik veel van je heb mogen leren tijdens het meedraaien in de kliniek. Gedurende mijn promotie hebben we vele memorabele uitstapjes gemaakt naar Wenen, waarbij ik nooit meer zal vergeten hoe we al rondstruinend per ongeluk ongezien binnen wandelden bij een feestelijke gelegenheid van een Oostenrijkse regeringspartij en we er het beste van hebben gemaakt. Het laatste jaar van mijn promotie was je helaas minder betrokken doordat het onderwerp wat verder van je af stond, maar nog steeds was je bereid om mijn manuscripten te reviseren. Veel dank dat je er altijd voor me was. Beste **Marcel**, de afdeling nucleaire geneeskunde heeft een zeer positieve ontwikkeling doorgemaakt door jouw inspanningen. Het was fijn daar onderdeel van te zijn, ondanks dat mijn werkplek zich bevond op de afdeling radiotherapie. Dank voor de vrijheid en vertrouwen die je me hebt gegeven om mijn eigen weg te bepalen. Ook ben ik je zeer dankbaar dat je een onvergetelijke reis naar Down Under mogelijk hebt gemaakt, zodat ik onderzoek kon doen in het Peter MacCallum Cancer Center in Melbourne.

Geachte leden van de leescommissie, uiteraard wil ik jullie graag bedanken dat jullie de tijd hebben genomen om mijn proefschrift kritisch onder de loep te nemen. Ik kijk ernaar uit om met jullie van gedachten te wisselen over dit proefschrift.

Dr. Jeroen van de Kamer. Beste **Jeroen**, dit avontuur heb ik ook grotendeels onder jouw vleugels gemaakt. Net als Wouter gaan wij al een lange tijd terug. Dank voor je betrokkenheid, je humor, en co-promotorwaardige begeleiding.

PERTAIN management committee. Dear **Diana, Elena, and Alfredo**, a standing ovation would be appropriate for all the work you put in organizing and hosting meetings in Vienna and for bringing together a wonderful group of people that share the same vision towards a better future for cancer patients worldwide. Because of you, coming to Vienna felt like coming home. Dear **Gerry, Daniela, and Michael**, we shared memorable moments together over plenty of coffee, beers, wines and even some shots of strong liquor. I am truly grateful to have been amongst such bright and inspiring people like you. Your presence was sufficient to make people smile. Many thanks for sharing your valuable knowledge and experience. It was an honor to work with you. Dear **Arturo, Vincent, Ursula, and José**, unfortunately I have not been able to meet two of you in person. However, I would still like to thank you all for your willingness to provide data and feedback whenever requested.

Participants of the PERTAIN trial, friends from abroad. Dearest **Heloisa, Paulo, Antonio, Allan, Darja, Liina, Rakesh, Ashwani, Jamal, Akram, Younis, Shahid, Çiğdem, Tuğçe, Tuan Anh, Nguyen Xuan, Bui Quang, Le Ngoc**, and all the other colleagues from your departments that supported the work in this thesis. You deserve a round of applause for being the backbone of the PERTAIN trial. I have enjoyed working together with you despite the distance between us most of the time. It is a pity we did not share more time with each other face-to-face, because I believe our fruitful meetings in Vienna boosted our spirit. It was incredible to see you progress year after year, to see you gain more confidence in putting newly gained knowledge and skills into practice, and how that also had a positive impact on the collaboration between your departments. You can be proud of yourselves!

Dr. Sarah Everitt. Dear **Sarah**, thank-you for making me feel at home and for guiding me around during my time at the PeterMac. You made sure I was integrating well with the Australian way of life and made this research trip truly worthwhile. It was a pleasure to work with such a cheerful and optimistic person and I hope our paths will cross again in the future. Of course I would also like to thank all the other colleagues at the PeterMac for their hospitality during my stay and their contributions to this thesis. I enjoyed experiencing your laid-back, welcoming, curious attitudes and your serious love for nature, sports, food and coffee (what else)!

Mijn paranimfen en vriendinnen voor het leven. Lieve **Merel** en **Susanne**, als *ex-roomies* van Huize Geduld hebben we elkaar door de jaren heen zowel leren kennen in meest sobere toestand, alsook elk denkbare andere toestand. Alsof dat niet genoeg was, bleven we elkaar door de jaren heen opzoeken, ondanks dat we allang niet meer elkaars gekke, vrolijke, actieve, ondernemende huisgenoten waren. Het was wellicht al duidelijk, maar toch, jullie zijn die zusjes die ik nooit heb gehad, en daarom ben ik erg aan jullie gehecht. Ondanks dat mijn sociale leven op en neer golft en zeker in de laatste fase van mijn PhD op een lager pitje stond, zijn jullie een constante factor in mijn leven. Ik ben jullie onwijs dankbaar voor jullie vriendschap en blijf uitkijken naar de komende reünietjes, en wie weet wanneer impulsieve daadkracht ons weer een mooie trip zal opleveren.

Alle oud collega's van de afdeling radiotherapie. Het zijn er teveel om op allemaal op te noemen, maar toch bedankt dat ik zoveel mooie en speciale momenten met jullie heb mogen delen. Een aantal wil ik graag in het speciaal bedanken. My office buddies **Artem** and **Matthew**, many thanks for the heated discussions, whether it was about sexism in the words 'handsome' and 'pretty' or about American politics, we often ended saying "how did that escalate so quickly?". I also enjoyed how we changed our office into a game room each Friday afternoon before the borrel started. Thank-you for reminding me frequently that we should appreciate a proper work-life balance. **Natasja** en **Sander**, veel dank voor de gezellige etentjes samen en de koffiepauzes op werk, de hardloop sessies en de drankjes erna (Natasja), en het bekijken van spannende Europese wedstrijden van Ajax in de kroeg (Sander). **Zeno**, **Tessa**, **Iban**, **Bruno**, **Kleopatra**, **Catarina**, **Roel**, **Cris**, **Celia**, **Uros**, **Lukas**, **Vineet**, **Takahiro**, **Anja**, **Jonas**, **Barbara** and **Simon**, many thanks for being a bunch of great personalities all together and making my PhD more fun than my *accelerated imagination* could think of. To all of the above, I will always cherish the moments we shared, whether hilarious or dead-serious, at conferences, dinners, borrels and afterparties, doing karaoke on Japanese nights, playing poker, sailing on various lakes, and so on. I am grateful to have known you not only as colleagues, but also as friends.

Alle oud collega's van de afdeling nucleaire geneeskunde. Graag wil ik **Daan**, **Linda**, **Else**, **Daphne**, **Judith**, **Bas**, en **Suzana** in het zonnetje zetten, en verder is mijn dank groot aan alle nucleair geneeskundigen en laboranten, in het bijzonder **Natascha**, voor hun ondersteuning en feedback tijdens de klinische activiteiten, onderzoeken en de journal clubs. Maar ook bedankt voor de gezelligheid tijdens de vrimibo's, de Sinterklaas-vieringen, en de pubquiz-avonden.

Diedie en **Patricia**. Lieve dames, dank voor jullie onvoorwaardelijke steun en gezellige afleidingen in de vorm van een praatje, de meest actuele roddels, of het plagen van Artem. Volgens mij hebben jullie al zoveel meegemaakt op de afdeling door de jaren heen dat niets meer te gek voor jullie is. Zo ook niet toen ik mijn paspoort was kwijtgeraakt een paar dagen

voor de ESTRO begon in Turijn (het was een erg gezellige Koningsdag zullen we zeggen!) en jullie begripvol hebben meegedacht aan oplossingen en de juiste mensen op de hoogte stelden dat ik iets later zou verschijnen. Jullie zorgen ervoor dat iedereen vlekkeloos zijn weg vindt op de afdeling en ook makkelijk over allerlei logistische en administratieve hobbels heen komt. Mijn dank is groot, ga astublieft zo door!

Ik dank ook mijn vrienden voor het leven. Mijn **middelbare schoolvrienden**, van losbandige pubers naar 'volwassen' burgers, wij hebben elkaar zien opgroeien en samen de wereld ontdekt, waarbij ieder op zijn eigen unieke wijze iets toevoegt aan de groep en dat voor een onlosmakelijke band heeft gezorgd. Dat de toekomst nog veel moois voor ons in petto heeft. **AKFST-ers**, *shit is about to get real!* Jullie zijn de meest ongerijmde en tegelijk meest wijze mensen bij elkaar en daarom hou ik ook zo van jullie. Zoals één van jullie zo prachtig verwoorde tijdens één van de vele memorabele weekenden weg: *"it is one of the blessings of old friends that you can afford to be stupid with them"*. De cultuursnuivende **Paratrippers** en tevens oude studievrienden, mijn **oud-huisgenoten van Huize Rynstroom** (ik ga zeker onze restaurant roulette en avondjes Ozark missen), de altijd relaxte doch ambitieuze **Pythianen**, en de levensgenieters van **Heeren Zeeven**, bedankt dat jullie voor gezelligheid en ontspanning zorgden tijdens deze wetenschappelijke reis.

Mijn broers, aanhang, en mijn grootse warme familie. **Maarten**, **Jan** en **Anne**. We zien elkaar niet altijd even vaak door de afstand, maar de momenten dat we compleet zijn geniet ik volop. Ik kijk uit naar nog meer mooie vakanties samen met het hele gezin om toch weer die oude tijden te herleven. **Kiko**, ook jij verdient lof voor de gezelligheid die je ons gezin brengt met je vrolijke gefluit. Ik hoop dat je nog heel lang in ons midden kan zijn. Lieve ooms en tantes, neven en nichten (in het bijzonder **Judith**), neefjes en nichtjes, wat een geweldig lieve mensen zijn jullie bij elkaar en het is bijzonder om te zien dat we ondanks de toenemende grote, toch nog steeds goed op de hoogte zijn van elkaars leven en in het geval van de Veldmannen, elkaar nog jaarlijks zien door het familieweekend. Het was dan ook een groot gemis dat het familieweekend niet doorging dit jaar door COVID-19 toestanden. Dank voor jullie support, warmte en gastvrijheid.

Mijn lieve ouders. *"Wie je bent en wat je doet daar ben je zelf verantwoordelijk voor, maar je ouders zorgen voor de basis waarmee je het volwassen leven instapt."* Lieve **pap** en **mam**, met die basis zat het wel goed en daar ben ik jullie zeer dankbaar voor. Ik wil jullie enorm bedanken dat ik in zo'n vrij, warm en reislustig gezin ben opgegroeid, en dat jullie altijd alles hebben gegeven voor mij en mijn broers. Ook al was het misschien niet altijd even duidelijk waar ik mee bezig was in Enschede, Nijmegen, Utrecht, Amsterdam, Pisa, en Melbourne, jullie onvoorwaardelijke steun heeft mij altijd kracht en heel veel rust gegeven. Ik zeg het veel te weinig voor mijn gevoel, maar ik hou onwijs van jullie.

Lieve **Silvie**, wat voel ik me toch een ontzettende bofkont met jou aan mij zij. Het toeval wil zijn dat je er al vanaf het begin was, ware het op de achtergrond werkzaam in de kliniek, en af toe zagen we elkaar op de borrels. Nadat ik het AvL verliet voor een vroegtijdig avontuur in het bedrijfsleven, kwamen we elkaar tegen op een festival, en een korte tijd later tijdens een koffiedate sloeg de vonk pardoes over. Dit boekje moest toen nog wel worden afgerond en dat ging niet als vanzelf, maar gelukkig zorgde jij dat ik optimaal kon presteren. Ik ben je zeer dankbaar voor je hartverwarmende, attente zorgzaamheid wat mij altijd motiveert om verder te gaan. Daarnaast is jouw liefdevolle kalmte een goede remedie tegen frustratie en stress, en laat je mij altijd weer stralen met je lieve lach. Nu we samen een heerlijke thuisbasis hebben gecreëerd, kan ik niet wachten om samen met jou nieuwe avonturen te beginnen. Mijn allerliefste, *ik ook van jou*.

Curriculum Vitae

Tom Konert werd op 2 juni 1988 geboren te Seria, Brunei. Op zijn zesde verhuisde hij naar Nederland na een buitenlands avontuur in Brunei en Oman. Na het behalen van zijn VWO diploma aan het Oranje Nassau College in Zoetermeer in 2006, begon hij aan de opleiding Technische Geneeskunde aan de Universiteit Twente. Gedurende zijn driejarige Master programma specialiseerde hij in 'Robotics and Imaging' en verrichtte hij (klinisch) onderzoek naar medisch technische toepassingen in het Radboud UMC, het Universitair Medisch Centrum Utrecht, en The BioRobotics Institute in Pontedera, Italië.



Tussen de Master stages door deed hij drie maanden vrijwilligerswerk in Nepal waar hij een arts assisteerde op een medische post. Ook gaf hij voorlichting op verschillende scholen over overdraagbare ziektes en (orale) hygiëne.

Afstuderen deed hij in het Nederlands Kanker Instituut op de afdeling Nucleaire Geneeskunde. Het wetenschappelijk onderzoek dat hij daar verrichtte werd afgerond op november 2014 en vormde een fundering voor dit promotieonderzoek. In 2017 deed hij twee maanden onderzoek naar de kwantificatie van FDG en FLT PET tracers in het Peter MacCallum Cancer Center in Melbourne onder begeleiding van Sarah Everitt en Michael MacManus.

Sinds 1 april 2019 werkt hij bij Microsure, een Nederlands medtech bedrijf die robots ontwikkelt voor microchirurgische toepassingen. Hier is hij onderdeel van het klinisch team die de brug vormt tussen de kliniek en product ontwikkeling.