Journal of Nuclear Medicine, published on May 10, 2019 as doi:10.2967/jnumed.119.227744

Prospective evaluation of a tumor control probability model based on dynamic 18F-FMISO PET for head-and-neck cancer radiotherapy

Daniela Thorwarth[1,2]*, Stefan Welz[3], David Mönnich[1], Christina Pfannenberg[4], Konstantin Nikolaou[4], Matthias Reimold[5], Christian La Fougère[5], Gerald Reischl[6], Paul-Stefan Mauz[7], Frank Paulsen[3], Markus Alber[1,8], Claus Belka[2,9], Daniel Zips[2,3]

[1]Section for Biomedical Physics, Department of Radiation Oncology, University of Tübingen, Germany
[2]German Cancer Consortium (DKTK), partner site Tübingen, Tübingen, Germany; and German Cancer Research Center (DKFZ), Heidelberg, Germany
[3]Department of Radiation Oncology, University of Tübingen, Germany
[4]Department of Radiology, Diagnostic and Interventional Radiology, University of Tübingen, Germany
[5]Department of Nuclear Medicine, University of Tübingen, Germany
[6]Department of Preclinical Imaging and Radiopharmacy, University of Tübingen, Germany
[7]Department of Otorhinolaryngology, University of Tübingen, Germany
[8]Department of Radiation Oncology, University of Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany
[9]Department of Radiation Oncology, LMU Munich, Marchioninistr. 15, 81377 München, Germany

*Corresponding author:
Daniela Thorwarth, Section of Biomedical Physics, University Hospital for Radiation Oncology, University of Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany.
E-mail: daniela.thorwarth@med.uni-tuebingen.de

Manuscript Details:

# of words Abstract:	272
# of words Manuscript:	3001 (excl. Literature, Figures, Tables)
# of pages:	26
# of tables:	2
# of figures:	3

Acknowledgements:

This project has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC starting grant agreement no. 335367.

Running title: A TCP model based on dynamic 18F-FMISO PET

Keywords: Hypoxia, Radiotherapy, Head and neck cancer, TCP model, outcome prediction, 18F-FMISO PET/CT

ABSTRACT

Purpose: To evaluate an imaging parameter response relationship between the extent of tumor hypoxia quantified by dynamic 18F-Fluoromisonidazole (18F-FMISO) PET/CT and the risk of relapse after radiotherapy (RT) in patients with head-and-neck cancer (HNC).

Methods: A prospective cohort of n=25 HNC patients was examined before starting radiotherapy with dynamic 18F-FMISO PET/CT 0-240min post tracer injection (pi). 18F-FMISO image parameters including a hypoxia metric M_{FMISO} derived from pharmacokinetic modelling of dynamic 18F-FMISO as well as maximum tumor-to-muscle ratio at 4h pi (TMR_{max}), gross tumor volume (V_{GTV}), relative hypoxic volume (rHV) based on M_{FMISO} and a logistic regression model combining V_{GTV} and TMR_{max} were assessed and compared to a previous training cohort (n=15). Dynamic 18F-FMISO was used to validate a tumor-control-probability (TCP) model based on M_{FMISO}.

The prognostic potential with respect to local control of all potential parameters was validated using the concordance index (ci) for uni-variate cox regression models determined from the training cohort, in addition to Kaplan-Meier analysis including log-rank test.

Results: The TCP model was confirmed indicating that dynamic 18F-FMISO allows to stratify patients into different risk groups according to radiotherapy outcome.

In this study, the hypoxia metric M_{FMISO} was the only parameter which was confirmed as prognostic in the independent validation cohort (ci=0.71, p=0.004). All other investigated parameters, such as TMR_{max}, V_{GTV}, rHV, and the combination of V_{GTV} and TMR_{max} were not able to stratify patient groups according to outcome in this validation cohort (p=n.s.).

Conclusion: In this study, the relationship between the hypoxia parameter M_{FMISO} and the risk of relapse was prospectively validated. The data supports further evaluation and external validation of dynamic 18F-FMISO PET/CT as a promising method for patient stratification and hypoxia-based radiotherapy personalization including dose painting.

INTRODUCTION

Tumor hypoxia is a major cause of resistance to radiotherapy as well as to other treatments such as chemotherapy and consequently leads to poor outcome (*1-4*). Several studies have shown that in locally advanced primary head and neck cancer (HNC), tumor hypoxia is associated with poor response to radiotherapy (*5-9*). Consequently, strategies to overcome hypoxia-induced treatment resistance, such as increasing the radiation dose to the whole tumor or to tumor subvolumes, i.e. dose painting, have been proposed earlier (*10-12*).

However, a robust and accurate selection of patients for hypoxia-based radiotherapy interventions is crucial. Different methods of hypoxia detection have been proposed recently, such as hypoxia gene classifiers (*9,13*), biopsy or blood based biomarkers (*14,15*) or non-invasive positron emission tomography (PET) using dedicated hypoxia tracers such as 18F-Fluoromisonidazole (18F-FMISO) (*5,7,8,16-20*), 18F-Fluoroazomycin-arabinoside (18F-FAZA) (*2,6*), or 18F-Fluortanidazole (18F-HX4) (*21,22*). As the non-invasive, three-dimensional measurement of tumor hypoxia is very complex and depends sensitively on the image analysis approach, dynamic hypoxia PET imaging has been proposed by several groups in order to obtain a measure of tumor hypoxia quantitatively from the kinetics of tracer uptake (*23-29*).

Based on earlier findings about the prognostic value of 18F-FMISO PET (*7,18,29,30*), a randomized study to investigate the effectiveness of hypoxia dose painting in HNC has been initiated in our institution (NCT 02352792). Results of a planned interim analysis were published, indicating the clinical feasibility of hypoxia dose painting (*8*). Importantly, dynamic 18F-FMISO PET was shown to stratify patients into two groups with different risks of loco-regional failure. However, predictive biomarkers including imaging parameters to modify radiotherapy require independent prospective validation before implementation in the clinic (*18*). The aim of the present study is to compare and validate hypoxia imaging parameters derived from a previous training cohort (*29,30*) with data from an independent prospective cohort (*8*).

MATERIALS AND METHODS

Study Design and Patients

This study compares two groups of patients: (a) a training cohort consisting of n=15 HNC patients who underwent dynamic 18F-FMISO PET imaging and radiotherapy treatment between 2003 and 2006 and (b) a prospective validation cohort with n=25 patients recruited 2009 - 2012 in our center. The study design is presented in detail in Figure 1. The study was carried out according to TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement (*31*).

Patient and tumor characteristics of the training cohort were reported previously (*28-30*). Patients were recruited into the validation cohort in a randomized phase II trial testing the efficacy of hypoxia imaging based dose painting. The interim analysis has been published earlier (*8*). The trial was approved by the local ethical committee as well as by the expert panel of the German Society of Radiation Oncology (DEGRO) and registered at www.clinicaltrials.gov (NCT 02352792). Written informed consent was obtained from all individual participants included in the study. Three patients received significant levels of dose escalation, i.e. 77 Gy in a volume > 5 mL and were excluded from this analysis to guarantee comparable radiation dose levels for the present analysis.

Imaging and Image Analysis

In the training as well as in the validation cohort, all patients were examined before the start of radiotherapy with [^{18}F]fluorodeoxyglucose (FDG) PET for staging purposes only and dynamic 18F-FMISO PET followed by a radiotherapy planning CT. In the validation cohort, dynamic 18F-FMISO PET/CT imaging was performed on a Biograph mCT (Siemens Healthineers, Erlangen, Germany) in radiotherapy position using thermoplastic head and shoulder masks, neck support and a flat table top. The acquisition protocol consisted of a listmode acquisition during the first 40 min after tracer injection (framing: 12x10s, 8x15s, 11x60s and 5x300s) followed by two static acquisitions 2 and 4h post injection (pi). PET data were reconstructed using an OSEM 3D algorithm with 4 iterations and 8 subsets, with 200x200 voxels per slice and a voxel

size of 4.07x4.07x5 mm^3. Corresponding CT images were acquired with 120 kV_p, with 5 mm slice thickness and an in-plane voxel size of 1.52x1.52 mm^2. 18F-FMISO PET/CT data was rigidly registered to the planning CT. Tumor volumes for radiotherapy treatment were manually defined by an experienced radiation oncologist (SW) based on the planning CT. These volumes served for PET image analysis. Firstly, a maximum tumor-to-muscle ratio (TMR_{max}) of tracer uptake in the tumor volume was derived from static 18F-FMISO PET data acquired 4h pi. Secondly, the full dynamic PET series was included in a voxel-based kinetic analysis in the tumor area using a 2-compartmental model optimized for hypoxia PET data analysis, as previously described in detail (*28*). Using this approach, for each patient a hypoxia metric M_{FMISO} was calculated from voxel-based parameters on tissue perfusion and tracer retention in the gross tumor volume (GTV). Details on the definition of M_{FMISO} can be found in the supplementary material.

Radiotherapy Treatment

In the training cohort, all patients were treated with combined radio-chemotherapy, prescribing 70, 60 and 54 Gy to the planning tumor volumes of first, second and third order, respectively. In the validation cohort, the same dose prescription was used for patients in the standard treatment arm, whereas for patients in the experimental arm a radiation dose escalation of 10% to the hypoxic part of the primary tumor was aimed for (for details, see (*8*)). However, only three patients in the validation cohort received the prescribed dose escalation because in the other patients the hypoxic volumes were too small (< 5 mL) for delivery of extra radiation dose. Patients with achieved dose escalations (n=3) were excluded from this analysis. Thus, all remaining patients (n=22) in this analysis received 70 Gy.

Tumor Control Probability (TCP) Model

Classical TCP models in radiotherapy are dose-effect-relationships, which link the radiation dose D with the expected outcome of a patient in a cohort. As a number of previously published studies report a prognostic value of hypoxia PET information for a

given dose D, we hypothesize that increased levels of hypoxia in a tumor sub-region may cause a higher level of radiation resistance and therefore counteract the dose effect. In a previous study, we developed an imaging-response-relationship in the form of a TCP model that relates tumor hypoxia measured with dynamic 18F-FMISO PET to a continuous outcome variable (*30*). For a more detailed description of this TCP model, see supplementary material. Briefly, we assume that for a constant radiation dose $D = D_0$, TCP depends mainly on the level of tumor hypoxia. This leads to the following formulation

$$-\ln TCP = A \exp(M_{FMISO})$$

where $A = -\ln TCP_0$ is a constant which refers to the TCP of a corresponding tumor, free of hypoxia. $M_{FMISO} = M_{FMISO}(H_i, P_i)$ denotes the level of tumor hypoxia, as obtained from dynamic 18F-FMISO PET where H_i and P_i are parameters related to intra-voxel hypoxia, i.e. tracer retention and perfusion derived by pharmacokinetic analysis of the dynamic 18F-FMISO PET data in the tumor region (*28*). In a previously published study, this TCP model was trained on an initial exploratory patient group consisting of 15 patients, the parameter $A = 9.9 \cdot 10^{-5}$ was derived from a fit of the model function to the observed radiotherapy outcome data (*30*).

Statistical Analysis

Imaging characteristics and general patient data of the two cohorts were compared using the Mann-Whitney-U test.

In order to test different parameters extracted from static 18F-FMISO PET acquired 4 h pi, which might be alternatives to dynamic 18F-FMISO PET that are easier to assess in clinical routine, uni-variable Cox-regression models were created based on the training data for 18F-FMISO TMR_{max}, size of the gross tumor volume V_{GTV}, as well as the hypoxia metric M_{FMISO} determined from dynamic 18F-FMISO PET and the relative hypoxic volume (rHV) associated with M_{FMISO} in order to stratify patients according to outcome after radiochemotherapy. Outcome data was available as time-to-event data for local control (LC). To evaluate the prognostic performance of the Cox-models, the concordance index (ci) was calculated. Bootstrap resampling was used to

estimate the confidence intervals (CI) for ci. Thresholds for stratification of patient subgroups were defined as median values in the training cohort. Those threshold values were then applied to the validation cohort. To compare the potential of the investigated parameters for risk group stratification, Kaplan-Meier analysis including log-rank tests were used for training and validation cohorts as well as for a merged patient group. Additionally, a logistic regression model was trained for TMR_{max} and V_{GTV} ($logit_{TMR\text{-}V}$) to check for the ability of a combined parameter to predict local failure in HNC.

To validate the TCP model, the original TCP model function (cf. equation 1) was fitted to all available data from training and validation cohorts (n=37). The original model function and the function fitted to the merged data set were compared using an ANOVA test for model comparison to assess the probability of rejecting the null hypothesis, where one model function does not fit the data better than the other one.

All statistical analyses were performed in R (version 3.1.1). In order to account for multiple testing (n=5 parameters), Bonferroni correction was used. Consequently, P-values < $0.05/n = 0.01$ were considered statistically significant. In case of statistically significant differences of local control (LC) rates, hazard ratios (HR) were calculated to estimate the risk ratio of the two groups.

RESULTS

Comparison of the two patient cohorts showed a comparable median age of patients, whereas median 18F-FMISO TMR_{max} was 2.47 in the training cohort and 1.80 in the validation cohort (p=0.0002). Similarly, differences between the two cohorts were observed in terms of tumor volumes where the median V_{GTV} was 114.7 cm^3 in the training group and 74.0 cm^3 in the validation group (p=0.027). The hypoxia metric M_{FMISO} as assessed from dynamic 18F-FMISO PET data resulted in comparable median values of 8.36 and 8.01 for training and validation cohorts, respectively (p=0.134). In contrast, the rHVs derived from M_{FMISO} differed significantly between the two groups with median values of 15.3 % and 0.9 %, respectively (p=0.0008). Further details on the two patient cohorts are presented in Table 1.

In order to establish a dedicated imaging-response-relationship for dynamic 18F-FMISO PET, a TCP model had been defined earlier based on the training cohort to link expected radiotherapy outcome to the hypoxia parameter M_{FMISO}. The TCP function defining this relationship was validated by a fit of the model to the merged patient groups. Here the parameter A was fitted as $A_{val} = (1.304 \pm 0.325) \cdot 10^{-4}$ (p=0.0003), whereas the value based the training cohort only was confirmed as $A_{tr} = (1.053 \pm 0.395) \cdot 10^{-4}$ (p=0.0183). Furthermore, the ANOVA test for model comparison resulted in a non-significant probability of $p = 0.11$ for rejecting the null hypothesis, where one model does not fit the data better than the other one. Consequently, these results confirm the initially defined TCP model based on dynamic 18F-FMISO PET. Figure 2 presents the TCP curve as a function of M_{FMISO} based on the initial training data set only (black) and the merged data groups (red) in addition to the observed clinical outcomes.

Among all uni-variable models trained for local control prediction, only the hypoxia metric M_{FMISO} showed prognostic potential in the training group (ci=0.77, p=0.001, HR=17.0). All other investigated parameters were not able to stratify patients into risk groups associated with LC. However, a logistic regression model $logit_{TMR\text{-}V}$ built from V_{GTV} and TMR_{max} yielded a ci of 0.73 (p=0.010, HR=2.4) in the training cohort. Of those two models that were associated with local control in the training cohort, only M_{FMISO} was confirmed as a prognostic parameter in the validation cohort. Consequently, the threshold defined for the hypoxia metric M_{FMISO} in the training cohort was able to

stratify patients according to outcome also in the independent validation cohort (p=0.004, HR=6.7). In contrast, the two-parameter logistic regression model could not be confirmed in the validation cohort (p=0.1). Interestingly, rHV was able to stratify the validation cohort according to outcome (p=0.001) even though in the training cohort, no significant prognostic potential had been observed. Similarly, significant patient stratification was obtained when applying the threshold values to the overall, merged cohort when using $\text{logit}_{\text{TMR-V}}$ (p=0.001) and M_{FMISO} (p<0.001, HR=9.4). However, only the hypoxia metric M_{FMISO} derived from dynamic 18F-FMISO PET was identified as prognostic parameter in the training cohort and validated in the independent validation cohort. Table 2 presents the detailed analysis of the prognostic potential of the investigated variables in terms of ci, HR and p-value of the Kaplan-Meier analysis for training, validation and merged cohorts, respectively. Figure 3 displays the Kaplan-Meier curves for local control of patients stratified according to the hypoxia metric M_{FMISO} in the training, validation and merged patient cohorts.

DISCUSSION

In this study a previously proposed TCP model using hypoxia imaging was confirmed by a prospective independent cohort of patients. The hypoxia parameter M_{FMISO} derived from dynamic 18F-FMISO PET/CT was validated as prognostic parameter for locoregional relapse after radiotherapy in HNC patients. In addition, hypoxia quantification based on dynamic 18F-FMISO was shown to be more robust compared to simple, static measures as only the hypoxia metric M_{FMISO} derived from dynamic 18F-FMISO PET was identified as prognostic parameter in the training cohort and validated in the independent validation cohort. Hence, dynamic 18F-FMISO PET data may in the future be used as a valuable tool for functional imaging based radiotherapy interventions in HNC. Of note, hypoxia dose painting might not be suitable for all tumour types as shown in a recently published trial on lung cancer (29).

The present study compared an independent validation cohort with data from an earlier training cohort. The detailed analysis showed significant differences in the two patient cohorts as several parameters such as TMR_{max}, GTV and rHV were lower in the validation cohort. This finding hints at a less hypoxic population in the validation group, which is corroborated by a lower number of observed local failures. Inherently, a difference in the overall hypoxia status of the two groups is a challenge for image biomarker validation. This issue may be a consequence of the low patient numbers in the two groups. Another reason for differences in PET data may be the fact that the training cohort was examined using a different PET scanner (Advance, GE Medical Systems, Milwaukee, US) which presented with a different overall performance in terms of hardware efficiency and image reconstruction as well as injected tracer activities. In line with our observation, biomarker studies recruiting patients over long time periods are susceptible to potential cohort effects which has been shown for example in a recently published study on 18F-FMISO PET in HNC patients (*18*). This emphasizes the need for robust and validated predictive parameters. Our TCP model appears to fulfil this requirement as it correctly predicted a lower number of local recurrences due to the lower hypoxic status in the validation. On the other side, additional knowledge may be acquired in the field during the validation phase of a model. This might potentially lead to

revision of the model itself, such as for example recent findings on hypoxia imaging using functional magnetic resonance imaging instead of PET (*32*).

The presented study is limited by the low number of patients, in training as well as in the validation phase. Furthermore, both steps were performed in a single center. A next phase of model validation would require independent, external validation.

A potentially considerable confounder of our analysis is the unknown human papilloma virus (HPV) status from the training cohort. HPV status is an established substantial prognostic factor for response to radiochemotherapy in HNC patients (*33*).

In addition, a number of aspects related to the methodology of hypoxia imaging require further discussion. TMR_{max} was used as one of the analysis parameters investigated to describe static 18F-FMISO PET data. This parameter may be subject to variation due to the noisy nature of PET data and might thus be compensated partially by using TMR_{peak} values, which has been shown to be more robust for static data analysis (*34*). Several studies have investigated the reproducibility of hypoxia PET imaging in terms of test-retest studies and found discrepant but mainly good repeatability and spatial stability of this functional imaging modality (*35-37*).

In contrast to other studies (*7,38*) our results did not identify TMR_{max} as prognostic biomarker. The small sample size during training may be a reason for this. However, the fact that dynamic 18F-FMISO was identified as powerful biomarker for LC in HNC indicates the robustness of kinetic hypoxia PET information.

In this study, we could show that the hypoxia metric M_{FMISO} derived from dynamic 18F-FMISO PET is prognostic for local control in HNC radiotherapy and thus a robust parameter to stratify patients for individualized radiotherapy approaches in the future. This finding confirms the results presented recently by other groups, which hypothesized that dynamic hypoxia PET data were more robust and reproducible in terms of hypoxia quantification in HNC (*24,25*). Dynamic data acquisition in hypoxia PET allows for the time-dependent monitoring of tracer uptake and diffusion and is therefore a very robust tool for hypoxia quantification in tumors. Nevertheless, dynamic PET scanning is very demanding of the patients and challenging in terms of scanner time and data analysis. In order to reduce examination times for patients and required scanner time, Grkovski *et*

al. propose dedicated methods for scan time reduction while maintaining the most important features of dynamic 18F-FMISO PET for hypoxia quantification (*26*). However, dynamic PET has been shown to be very reproducible in regard to the noise level of the imaging data (*27*). In contrast, static hypoxia parameters such as TMR_{max} or TMR_{peak} have been shown to be associated with only limited reproducibility due to subjective muscle activity definition for normalization (*39*). Grkovski *et al.* (*25*) report a disagreement of visual hypoxia assessment on static scans compared to pharmacokinetic modeling of dynamic data in approximately 20% of cases which may directly affect patient management in interventional trials.

A recent study identified the residual hypoxic volume after the first two weeks of radiotherapy as prognostic parameter to be used for future radiotherapy interventions to overcome hypoxia induced radiation resistance (*18*). However, also repeated 18F-FMISO PET/CT examinations before and at a second time point during radiotherapy seem to be highly challenging from a logistic point of view as well as from the patient's perspective. We hypothesize that dynamic 18F-FMISO PET information acquired before the start of radiotherapy might comprise a similar information. Dynamic hypoxia PET allows to assess the perfusion status of a tissue region as well as the active retention of hypoxia tracer in viable tumor areas (*28,40*). As such it may be the ideal tool to predict local reoxygenation during the first weeks of radiotherapy and thus substitute an examination early during treatment (*30*).

Our study confirmed that hypoxia assessment with dynamic 18F-FMISO PET can be used as input variable for a dedicated TCP model linking quantitative hypoxia information with expected outcome after HNC RT. Such a TCP model is a unique option to not only stratify patients into two binary risk groups for therapy adaptation followed by arbitrary testing of escalated radiation doses, but in contrast also allows to assess the patient's risk for relapse on a continuous scale. Assuming that radiation resistance can be counteracted by higher radiation doses, this implies that a TCP model can ultimately link the local hypoxia status to a required dose escalation level (*12,30*). Consequently, the imaging-response-relationship directly translates into a dose-response-relation and thus, the TCP model validated here inherently defines a dose prescription function for personalized, hypoxia-modification radiotherapy interventions (*5,8*). Therefore, a clinical

implementation of this model might allow not only to divide HNC patients into two risk groups, but presents a methodology to derive individual dose escalation metrics for each patient from dynamic 18F-FMISO scans. Therefore, this approach allows to personalize RT in terms of higher doses where deemed necessary and dose de-escalation in non-hypoxic patients to improve quality of life for those patients.

CONCLUSION

In this study, an imaging-response-relationship linking M_{FMISO} to the risk of local failure after radiotherapy in HNC patients was confirmed. The hypoxia parameter M_{FMISO} derived from dynamic 18F-FMISO PET was independently validated as a strong prognostic parameter. This supports the further investigation of dynamic 18F-FMISO PET examination before the start of radiotherapy for personalized hypoxia-based radiotherapy interventions.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of Interest Statement

No potential conflicts of interest relevant to this article exist.

Key Points

Question:

Within this study, the imaging-response-relationship between a hypoxia metric derived from dynamic 18F-FMISO PET and outcome to radiotherapy (RT) in head-and-neck cancer (HNC) patients was prospectively validated.

Pertinent findings:

This independent validation study showed strong prognostic value of dynamic 18F-FMISO PET for outcome to radiotherapy in HNC patients. The study confirmed a previously proposed imaging-response relationship which links dynamic 18F-FMISO PET to a continuous outcome prediction. A series of potential imaging biomarkers was tested for prognostic value, but only dynamic 18F-FMISO PET was independently validated as prognostic parameter in HNC RT.

Implications for patient care:

The results of this study are a key factor for future hypoxia imaging based radiotherapy interventions such as dose painting.

Disclosure

D. Thorwarth and D. Zips declare departmental research collaborations with Elekta, Philips and Siemens. C. la Fougère has research collaborations with Siemens Healthineers.

References

1. Nordsmark M, Overgaard M, Overgaard J. Pretreatment oxygenation predicts radiation response in advanced squamous cell carcinoma of the head and neck. *Radiother Oncol.* 1996;41:31-39.

2. Mortensen LS, Johansen J, Kallehauge J, et al. FAZA PET/CT hypoxia imaging in patients with squamous cell carcinoma of the head and neck treated with radiotherapy: results from the DAHANCA 24 trial. *Radiother Oncol.* 2012;105:14-20.

3. Toustrup K, Sorensen BS, Nordsmark M, et al. Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer. *Cancer Res.* 2011;71:5923-5931.

4. Horsman MR, Mortensen LS, Petersen JB, Busk M, Overgaard J. Imaging hypoxia to improve radiotherapy outcome. *Nat Rev Clin Oncol.* 2012;9:674-687.

5. Lee N, Schoder H, Beattie B, et al. Strategy of Using Intratreatment Hypoxia Imaging to Selectively and Safely Guide Radiation Dose De-escalation Concurrent With Chemotherapy for Locoregionally Advanced Human Papillomavirus-Related Oropharyngeal Carcinoma. *Int J Radiat Oncol Biol Phys.* 2016;96:9-17.

6. Servagi-Vernat S, Differding S, Hanin FX, et al. A prospective clinical study of (1)(8)F-FAZA PET-CT hypoxia imaging in head and neck squamous cell carcinoma before and during radiation therapy. *Eur J Nucl Med Mol Imaging.* 2014;41:1544-1552.

7. Zips D, Zophel K, Abolmaali N, et al. Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. *Radiother Oncol.* 2012;105:21-28.

8. Welz S, Monnich D, Pfannenberg C, et al. Prognostic value of dynamic hypoxia PET in head and neck cancer: Results from a planned interim analysis of a randomized phase II hypoxia-image guided dose escalation trial. *Radiother Oncol.* 2017;124:526-532.

9. Linge A, Lock S, Gudziol V, et al. Low Cancer Stem Cell Marker Expression and Low Hypoxia Identify Good Prognosis Subgroups in HPV(-) HNSCC after Postoperative Radiochemotherapy: A Multicenter Study of the DKTK-ROG. *Clin Cancer Res.* 2016;22:2639-2649.

10. Ling CC, Humm J, Larson S, et al. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys.* 2000;47:551-560.

11. Gregoire V, Chiti A. PET in radiotherapy planning: particularly exquisite test or pending and experimental tool? *Radiother Oncol.* 2010;96:275-276.

12. Thorwarth D, Eschmann SM, Paulsen F, Alber M. Hypoxia dose painting by numbers: a planning study. *Int J Radiat Oncol Biol Phys.* 2007;68:291-300.

13. Toustrup K, Sorensen BS, Lassen P, et al. Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in squamous cell carcinomas of the head and neck. *Radiother Oncol.* 2012;102:122-129.

14. Zegers CM, Hoebers FJ, van Elmpt W, et al. Evaluation of tumour hypoxia during radiotherapy using [18F]HX4 PET imaging and blood biomarkers in patients with head and neck cancer. *Eur J Nucl Med Mol Imaging.* 2016;43:2139-2146.

15. Hoogsteen IJ, Lok J, Marres HA, et al. Hypoxia in larynx carcinomas assessed by pimonidazole binding and the value of CA-IX and vascularity as surrogate markers of hypoxia. *Eur J Cancer.* 2009;45:2906-2914.

16. Wiedenmann NE, Bucher S, Hentschel M, et al. Serial [18F]-fluoromisonidazole PET during radiochemotherapy for locally advanced head and neck cancer and its correlation with outcome. *Radiother Oncol.* 2015;117:113-117.

17. Bittner MI, Wiedenmann N, Bucher S, et al. Exploratory geographical analysis of hypoxic subvolumes using 18F-MISO-PET imaging in patients with head and neck cancer in the course of primary chemoradiotherapy. *Radiother Oncol.* 2013;108:511-516.

18. Lock S, Perrin R, Seidlitz A, et al. Residual tumour hypoxia in head-and-neck cancer patients undergoing primary radiochemotherapy, final results of a prospective trial on repeat FMISO-PET imaging. *Radiother Oncol.* 2017;124:533-540.

19. Wiedenmann N, Bunea H, Rischke HC, et al. Effect of radiochemotherapy on T2* MRI in HNSCC and its relation to FMISO PET derived hypoxia and FDG PET. *Radiat Oncol.* 2018;13:159.

20. Bandurska-Luque A, Lock S, Haase R, et al. FMISO-PET-based lymph node hypoxia adds to the prognostic value of tumor only hypoxia in HNSCC patients. *Radiother Oncol.* 2019;130:97-103.

21. van Elmpt W, Zegers CM, Reymen B, et al. Multiparametric imaging of patient and tumour heterogeneity in non-small-cell lung cancer: quantification of tumour hypoxia, metabolism and perfusion. *Eur J Nucl Med Mol Imaging.* 2016;43:240-248.

22. Zegers CM, van Elmpt W, Reymen B, et al. In vivo quantification of hypoxic and metabolic status of NSCLC tumors using [18F]HX4 and [18F]FDG-PET/CT imaging. *Clin Cancer Res.* 2014;20:6389-6397.

23. Schwartz J, Grkovski M, Rimner A, et al. Pharmacokinetic Analysis of Dynamic (18)F-Fluoromisonidazole PET Data in Non-Small Cell Lung Cancer. *J Nucl Med.* 2017;58:911-919.

24. Grkovski M, Schoder H, Lee NY, et al. Multiparametric Imaging of Tumor Hypoxia and Perfusion with (18)F-Fluoromisonidazole Dynamic PET in Head and Neck Cancer. *J Nucl Med.* 2017;58:1072-1080.

25. Grkovski M, Lee NY, Schoder H, et al. Monitoring early response to chemoradiotherapy with (18)F-FMISO dynamic PET in head and neck cancer. *Eur J Nucl Med Mol Imaging.* 2017;44:1682-1691.

26. Grkovski M, Schwartz J, Gonen M, et al. Feasibility of 18F-Fluoromisonidazole Kinetic Modeling in Head and Neck Cancer Using Shortened Acquisition Times. *J Nucl Med.* 2016;57:334-341.

27. Wang W, Lee NY, Georgi JC, et al. Pharmacokinetic analysis of hypoxia (18)F-fluoromisonidazole dynamic PET in head and neck cancer. *J Nucl Med.* 2010;51:37-45.

28. Thorwarth D, Eschmann SM, Paulsen F, Alber M. A kinetic model for dynamic [18F]-Fmiso PET data to analyse tumour hypoxia. *Phys Med Biol.* 2005;50:2209-2224.

29. Thorwarth D, Eschmann SM, Scheiderbauer J, Paulsen F, Alber M. Kinetic analysis of dynamic 18F-fluoromisonidazole PET correlates with radiation treatment outcome in head-and-neck cancer. *BMC Cancer.* 2005;5:152.

30. Thorwarth D, Eschmann SM, Paulsen F, Alber M. A model of reoxygenation dynamics of head-and-neck tumors based on serial 18F-fluoromisonidazole positron emission tomography investigations. *Int J Radiat Oncol Biol Phys.* 2007;68:515-521.

31. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162:W1-73.

32. Hompland T, Hole KH, Ragnum HB, et al. Combined MR Imaging of Oxygen Consumption and Supply Reveals Tumor Hypoxia and Aggressiveness in Prostate Cancer Patients. *Cancer Res.* 2018;78:4774-4785.

33. Lassen P, Primdahl H, Johansen J, et al. Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. *Radiother Oncol.* 2014;113:310-316.

34. Laffon E, Lamare F, de Clermont H, Burger IA, Marthan R. Variability of average SUV from several hottest voxels is lower than that of SUVmax and SUVpeak. *Eur Radiol.* 2014;24:1964-1970.

35. Nehmeh SA, Lee NY, Schroder H, et al. Reproducibility of intratumor distribution of (18)F-fluoromisonidazole in head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2008;70:235-242.

36. Zegers CM, van Elmpt W, Szardenings K, et al. Repeatability of hypoxia PET imaging using [(1)(8)F]HX4 in lung and head and neck cancer patients: a prospective multicenter trial. *Eur J Nucl Med Mol Imaging.* 2015;42:1840-1849.

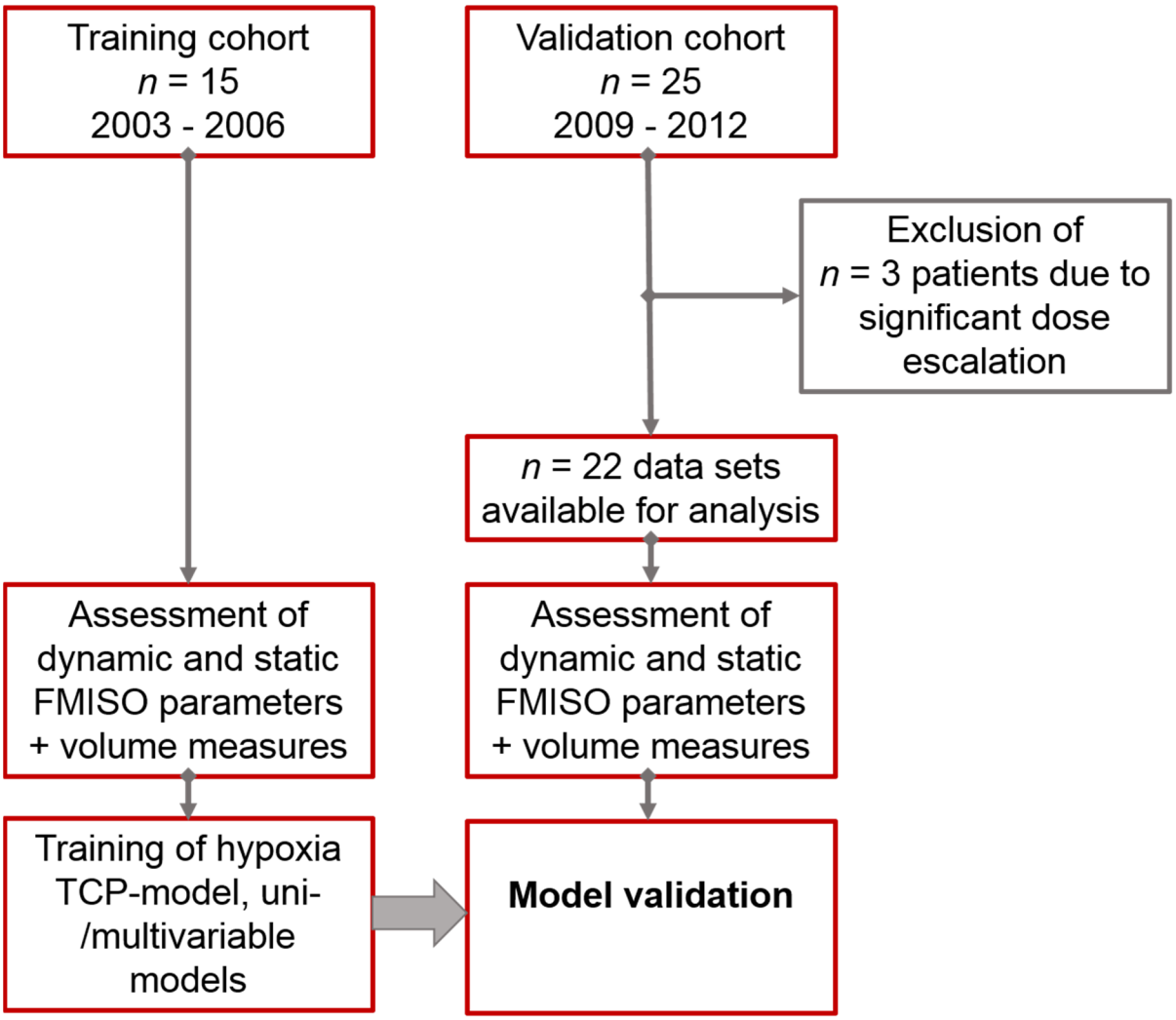
37. Silvoniemi A, Suilamo S, Laitinen T, et al. Repeatability of tumour hypoxia imaging using [(18)F]EF5 PET/CT in head and neck cancer. *Eur J Nucl Med Mol Imaging.* 2018;45:161-169.

38. Rajendran JG, Schwartz DL, O'Sullivan J, et al. Tumor hypoxia imaging with [F-18] fluoromisonidazole positron emission tomography in head and neck cancer. *Clin Cancer Res.* 2006;12:5435-5441.

39. Monnich D, Welz S, Thorwarth D, et al. Robustness of quantitative hypoxia PET image analysis for predicting local tumor control. *Acta Oncol.* 2015;54:1364-1369.

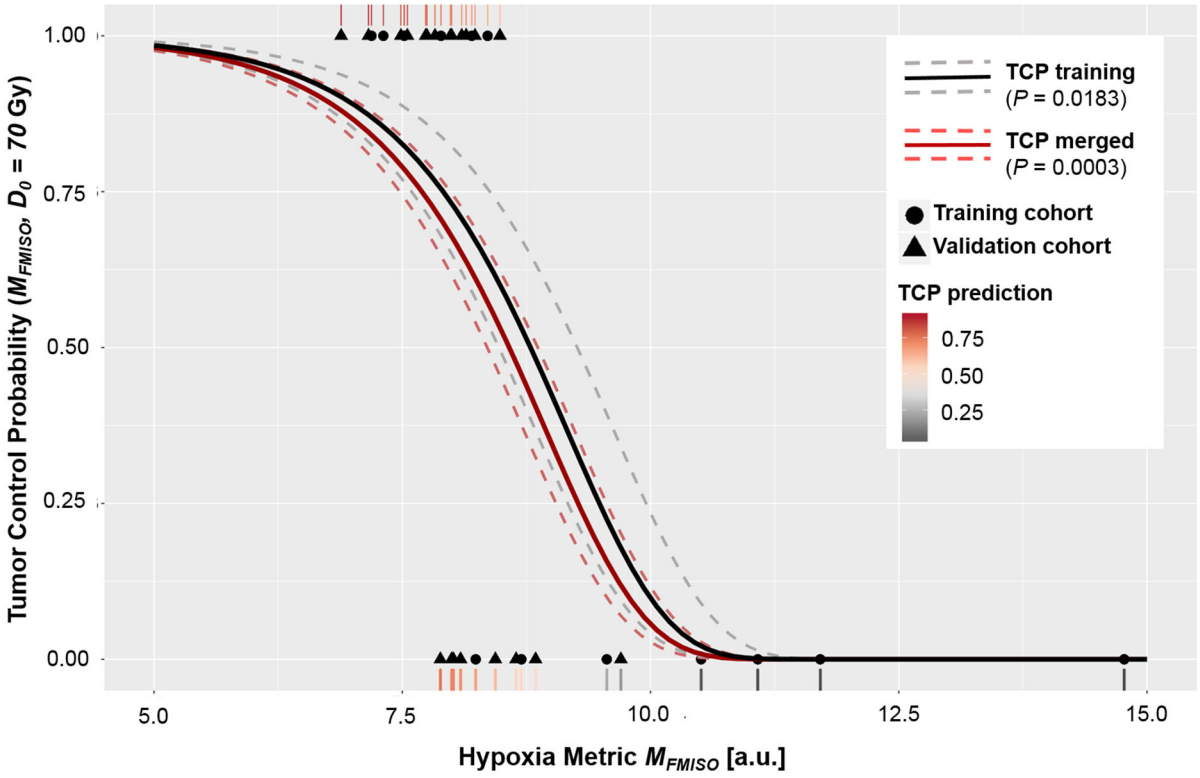
40. Grkovski M, Emmas SA, Carlin SD. (18)F-Fluoromisonidazole Kinetic Modeling for Characterization of Tumor Perfusion and Hypoxia in Response to Antiangiogenic Therapy. *J Nucl Med.* 2017;58:1567-1573.

Figure1:



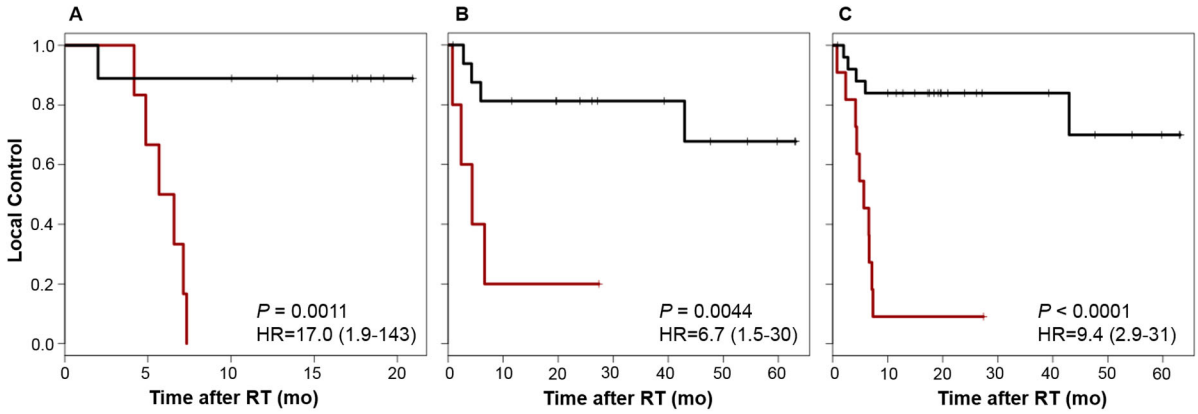
Study design.

Figure 2:



TCP model defining the expected local control rate for patients as a function of the hypoxia metric M_{FMISO} derived from dynamic 18F-FMISO PET. TCP model function defined on the training cohort initially is displayed by the black solid line (95 % confidence interval, dashed grey). The TCP function as fitted to the whole patient group, including the validation cohort is displayed in red. Observed outcomes of all individual patients (1=control, 0=recurrence) are shown for the training (circles) and the validation (triangles) cohort. The rug plot visualizes the estimated control probabilities by the TCP model for the individual patients.

Figure 3:



Kaplan-Meier curves stratified according to M_{FMISO}. (a) training cohort (n=15), (b) validation cohort (n=22), (c) merged groups (n=37). P-values are with respect to log-rank tests. Hazard ratios (HR) are given including 95% confidence intervals.

545 **Table 1:**

	Training cohort 2003 - 2006	Validation cohort 2009 - 2012	p-value
Age	59 (46 – 68)	56.5 (45 – 74)	0.698
Injected 18F-FMISO activity [MBq]	412.3 (336.6 – 473.5)	371.0 (315.0 – 444.0)	**0.005**
Tumor volume V_{GTV} [cm^3]	114.7 (32.4 – 287.6)	74.0 (23.0 – 257.0)	**0.027**
TMR_{max}*	2.47 (1.89 – 3.94)	1.80 (1.27 – 3.46)	**0.0002**
rHV† [%]	15.3 (0 – 38.8)	0.9 (0 – 26.5)	**0.0008**
M_{FMISO}	8.36 (7.19 – 14.77)	8.01 (6.88 – 9.70)	0.134

Comparison of patient and tumor characteristics of training and validation cohorts. Values are given as median (range). P-values < 0.05 are marked in bold.

Table 2:

		Training cohort (n=15)	**Validation cohort** (n=22)	**Merged groups** (n=37)
M_{FMISO} / **TCP**	ci (CI)†	0.77 (0.67 - 0.82)	0.71 (0.64 - 0.77)	0.74 (0.71 - 0.77)
	p-value*	**0.001**	**0.004**	**< 0.0001**
	HR (CI)‡	17.0 (1.9 – 143)	6.7 (1.5 – 30)	9.4 (2.9 – 31)
rHV	ci (CI)†	0.57 (0.36 - 0.57)	0.63 (0.58 - 0.67)	0.61 (0.57 - 0.67)
	p-value*	0.269	**0.001**	0.020
	HR (CI)‡	-	8.6 (0.22 – 331)	-
TMR_{max}	ci (CI)†	0.58 (0.40 - 0.62)	0.60 (0.47 - 0.61)	0.61 (0.54 - 0.68)
	p-value*	0.435	0.028	0.043
	HR (CI)‡	-	-	-
V_{GTV}	ci (CI)†	0.67 (0.62 - 0.77)	0.53 (- §)	0.60 (0.52 - 0.66)
	p-value*	0.083	0.317	0.031
	HR (CI)‡	-	-	-
$Logit_{TMR\text{-}V}$	ci (CI)†	0.73 (0.70 - 0.85)	0.58 (0.53 - 0.62)	0.65 (0.62 - 0.70)
	p-value*	**0.010**	0.1	**0.001**
	HR (CI)‡	2.4 (0.56 – 11)	-	3.1 (0.91 – 11)

Detailed analysis of the prognostic potential of different models trained in the training cohort and evaluated in the validation as well as in the merged groups, respectively. *p-value of log-rank test regarding Kaplan-Meier analysis. P-values < 0.01 are printed in bold. †Concordance index (ci) of the uni-variable cox model and ‡hazard ratio (HR) including 95% confidence intervals (CI). §CI estimation not possible in this case.