



## OPEN ACCESS

## EDITED BY

Sharon R Pine,  
University of Colorado, United States

## REVIEWED BY

Francesco Ricchetti,  
Sacro Cuore Don Calabria Hospital  
(IRCCS), Italy  
Nicola Simoni,  
University Hospital of Parma, Italy

## \*CORRESPONDENCE

Angela Davey  
✉ [angela.davey@manchester.ac.uk](mailto:angela.davey@manchester.ac.uk)

RECEIVED 01 February 2023

ACCEPTED 27 June 2023

PUBLISHED 12 July 2023

## CITATION

Davey A, Thor M, van Herk M, Faivre-Finn C,  
Rimmer A, Deasy JO and McWilliam A (2023)  
Predicting cancer relapse following lung  
stereotactic radiotherapy: an external  
validation study using real-world evidence.  
*Front. Oncol.* 13:1156389.  
doi: 10.3389/fonc.2023.1156389

## COPYRIGHT

© 2023 Davey, Thor, van Herk, Faivre-Finn,  
Rimmer, Deasy and McWilliam. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Predicting cancer relapse following lung stereotactic radiotherapy: an external validation study using real-world evidence

Angela Davey<sup>1\*</sup>, Maria Thor<sup>2</sup>, Marcel van Herk<sup>1</sup>,  
Corinne Faivre-Finn<sup>1,3</sup>, Andreas Rimmer<sup>4</sup>, Joseph O. Deasy<sup>2</sup>  
and Alan McWilliam<sup>1</sup>

<sup>1</sup>Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom, <sup>2</sup>Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States, <sup>3</sup>Department of Clinical Oncology, The Christie NHS Foundation Trust, Manchester, United Kingdom, <sup>4</sup>Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, United States

**Purpose:** For patients receiving lung stereotactic ablative radiotherapy (SABR), evidence suggests that high peritumor density predicts an increased risk of microscopic disease (MDE) and local-regional failure, but only if there is low or heterogenous *incidental* dose surrounding the tumor (GTV). A data-mining method (*Cox-per-radius*) has been developed to investigate this dose-density interaction. We apply the method to predict local relapse (LR) and regional failure (RF) in patients with non-small cell lung cancer.

**Methods:** 199 patients treated in a routine setting were collated from a single institution for training, and 76 patients from an external institution for validation. Three density metrics (mean, 90<sup>th</sup> percentile, standard deviation (SD)) were studied in 1mm annuli between 0.5cm inside and 2cm outside the GTV boundary. Dose SD and fraction of volume receiving less than 30Gy were studied in annuli 0.5-2cm outside the GTV to describe *incidental* MDE dosage. Heat-maps were created that correlate with changes in LR and RF rates due to the interaction between dose heterogeneity and density at each distance combination. Regions of significant improvement were studied in Cox proportional hazards models, and explored with and without re-fitting in external data. Correlations between the dose component of the interaction and common dose metrics were reported.

**Results:** Local relapse occurred at a rate of 6.5% in the training cohort, and 18% in the validation cohort, which included larger and more centrally located tumors. High peritumor density in combination with high dose variability (0.5 - 1.6cm) predicts LR. No interactions predicted RF. The LR interaction improved the predictive ability compared to using clinical variables alone (optimism-adjusted C-index; 0.82 vs 0.76). Re-fitting model coefficients in external data confirmed the importance of this interaction (C-index; 0.86 vs 0.76). Dose variability in the 0.5-1.6 cm annular region strongly correlates with heterogeneity inside the target volume (SD;  $\rho = 0.53$  training,  $\rho = 0.65$  validation).

**Conclusion:** In these real-world cohorts, the combination of relatively high peritumor density and high dose variability predicts increase in LR, but not RF, following lung SABR. This external validation justifies potential use of the model to increase low-dose CTV margins for high-risk patients.

#### KEYWORDS

image-based data mining, real world data, biomarker-by-treatment interactions, local relapse, NSCLC, stereotactic ablative body radiotherapy (SABR), personalized medicine, external validation

## 1 Introduction

Patients with early-stage lung cancer who are medically inoperable or refuse surgery will receive stereotactic body radiotherapy (SABR) as standard of care (1). It is a well-tolerated and successful treatment, with five-year local relapse (LR), regional failure (RF), and distant metastasis (DM) rates at ranges between 8-11%, 10-13%, and 11-22% respectively (2–5). SABR is characterized by high dose radiation delivered by multiple conformal beams to precisely target the tumor and avoid surrounding tissue. To maintain a conformal dose distribution, dose heterogeneity inside the planning target volume (PTV) is permitted (6). It is still unknown what part of the tumoral dose distribution is affecting tumor control the most (7), and further considerable institutional differences exist in treatment planning approaches (8). Better understanding of the level of tumor dose homogeneity/heterogeneity could lead to changes in radiotherapy planning to improve patient outcomes. In some situations, data from real-world settings allows us to test hypotheses on the impact of changes that can be made to the treatment planning process as an alternative to costly clinical trials. Real world data also has the advantage of being more inclusive of a general population. This is particularly relevant in the context of patients with lung cancer treated with SABR as most of them are elderly, frail and with multiple comorbidities. These patients are typically excluded or underrepresented in clinical trials (9).

The association between dose and LR has been well investigated through the link between the prescription dose and tumor control probability. Such studies report dose associations with the isodose surface encompassing the PTV (10, 11), the isocenter (12) and the average of the two (13). A study on 1500 patients emphasized the importance of ensuring high doses within the gross tumor volume (GTV) to promote local control (13). Reports have also detailed the importance of high dose *outside* the GTV as means of treating microscopic disease (MDE) and nodal micro-metastases that could be responsible for treatment failure (14, 15). Further studying this effect may identify associations that can be utilized to personalize treatments for patients at *high-risk* of failure.

Assessing pre-treatment imaging biomarkers is a non-invasive approach to identify high-risk patients who could be candidates for treatment adaptation. Salguero et al. demonstrated that CT-based

GTV circularity and surface density predict high-risk of MDE (14). High-risk of MDE also translated into an increased risk of local-regional failure if patients had low dose within 1.5cm from the GTV (16). Previous efforts have, however, focused on using tumor dose parameterizations, histological, or clinical characteristics alone to study and stratify risk of MDE (10–13, 17).

Ignoring risk stratification can lead to incorrectly claiming a lack of association (18). In our previous work, we developed a *Cox-per-radius* method to investigate the interaction between imaging density biomarkers and dose in independent annuli surrounding the GTV (19). In that study we demonstrated that the interaction between CT-based imaging biomarkers and dose far outside the tumor is linked to DM. In this study, we will use our previously developed method to assess whether a similar associations can be identified for LR and RF using two real-world patients cohorts. We will utilize real-world data from geographically separate institutions to explore the generalizability of any identified patterns.

## 2 Methods

### 2.1 Clinical data and patient follow-up

#### 2.1.1 Training data

Data was available for 195 patients with T1-2 N0M0 non-small cell lung cancer (confirmed histologically or suspected based on radiology) who were treated with SABR for primary lung cancer during 2011-2017 at The Christie NHS Foundation Trust, with 60 Gy in 5 fractions on consecutive weekdays. Institutional approval was granted to collect and analyse this data (REC reference: 17/NW/0060). All patients were staged with both a CT and 18F-FDG positron emission tomography (PET) scan, but did not always receive a histological diagnosis. Four-dimensional computed tomography scans (4D-CT) and three-dimensional dose distributions were available, as described previously and in [Supplementary Material \(SM\)](#), Section 1 (19). Clinical data was retrospectively collected from structured e-forms completed in routine practice. Clinical data was available on tumour lobe location, age, sex, Eastern Cooperative Oncology Group (ECOG) performance status (functional ability), ACE27 comorbidity score (describing the presence and severity of existing medical conditions), and histological sub-type. Where

available, histological sub-type diagnosis was classified as ‘adenocarcinoma not otherwise specified (ADC NOS)’, ‘squamous cell carcinoma (SCC)’, ‘carcinoma NOS’ or other. Information on tumor centrality was not available.

In routine practice, patients were followed-up every three months for the first year, and six monthly thereafter. At the discretion of the clinician, a free-breathing CT is performed, PET or biopsy recommended when treatment failure is suspected. Recorded data on treatment failure was retrospectively collected from electronic records. Local relapse (LR) was defined as progression in or adjacent to the original treatment volume, based on clinical interpretation of ‘adjacent’ following reported definitions as guidance (20). Regional failure (RF) was defined as recurrence in regional lymph nodes (hilar or mediastinal). Time to failure was recorded from the start of radiotherapy to the date of the first scan that showed progression. Patients were censored at the most recent follow-up in the absence of failure.

### 2.1.2 Validation data

For validation, data were available for 139 patients with T1-2 NOM0 early-stage NSCLC treated with a range of fractionation regimes (see SM, Table 1) treated at Memorial Sloan Kettering Cancer Center (MSKCC) between 2014 and 2017. There was no agreement in treatment schedules between the training and validation data. For validation, we limited selection to patients treated with 50Gy in 5 fractions treated on consecutive weekdays, as the most common treatment schedule, and included all such patients treated with that regime during this time frame. All patients included had been staged with a CT and PET scan. A data-sharing agreement was in place and a study analysis plan made available online at the start of the collaboration (21). Clinical data was collected retrospectively from structured e-forms completed in routine practice. Clinical data was available on tumor lobe location, age, sex, Karnofsky performance status (a rating of 0-100 measuring a patient’s ability to perform daily tasks), and histological sub-type (SCC, ADC, or other). Information on

TABLE 1 A table to demonstrate patient demographic differences in training and validation data.

Characteristic	Training, N = 195 <sup>1</sup>	Validation, N = 76 <sup>1</sup>	p-value <sup>2</sup>
<b>Tumor volume (cc)</b>	4 (0 - 31)	14 (0 - 158)	<0.001
<b>Tumor motion (cm)</b>	0.56 (0.00 - 3.43)	0.36 (0.00 - 2.73)	0.016
<b>Tumor lobe location</b>			0.7
Lower	69 (35%)	29 (38%)	
Upper	126 (65%)	47 (62%)	
<b>Age (years)</b>	75 (45 - 92)	78 (52 - 92)	0.036
<b>Biological sex</b>			0.3
Female	99 (51%)	44 (58%)	
Male	96 (49%)	32 (42%)	
<b>Histological subtype</b>			<0.001
Adenocarcinoma, NOS	35 (37%)	45 (59%)	
Squamous cell carcinoma	37 (39%)	24 (32%)	
Carcinoma, NOS	14 (15%)	0 (0%)	
Other	8 (8.5%)	7 (9.2%)	
Unknown	101	0	
<b>Performance status (ECOG)</b>			<0.001
0	3 (1.8%)	29 (38%)	
1	64 (37%)	43 (57%)	
2	83 (49%)	4 (5.3%)	
3	21 (12%)	0 (0%)	
Unknown	24	0	
<b>Local relapse</b>	13 (6.7%)	12 (16%)	0.020
<b>Regional failure</b>	15 (7.7%)	5 (6.6%)	0.8

<sup>1</sup>Statistics presented: n (%); Median (range).

<sup>2</sup>Wilcoxon rank sum test; Pearson’s Chi-squared test; Fisher’s exact test.

Tumours were larger and less mobile in the validation data compared to training. Worse performance status was reported in the training set. Differences were also identified in histological sub-type with a larger proportion of adenocarcinoma in the validation data, but this could be influenced by the missing data reported for training. In the validation set there were significantly more local relapses but a similar percentage of regional failures.

comorbidity score or tumor centrality was not available. The Karnofsky performance status was scaled to the ECOG gradings to match the training cohort using a published guide (22).

Data was also collected on recurrences for this cohort. Local relapse was defined as new tumor growth at the site of prior CT; and recurrences were typically confirmed by PET (to demonstrate local FDG avidity) and/or biopsy. Regional failure was defined as recurrence in regional lymph nodes.

## 2.2 Imaging and dosimetric data

Both training and validation cohorts had treatment plans that included a 'motion-adapted' GTV (iGTV) which incorporates the GTV combined across all respiratory phases. In both institutions, this is outlined on the maximum intensity projection (MIP) and edited on individual respiratory phases. The PTV was also recorded, which represents the iGTV plus a 5mm expansion (with a 2-3mm additional CTV in the validation cohort). To extract the tumor volume (GTV) for every phase, an in-house process was applied to estimate and remove the motion adaptation using rigid tumor registration across 4D phases (23). As a result, a GTV contour was available for every phase and two additional clinical variables recorded: tumor volume and tumor motion amplitude. Results of the registration were assessed visually on a movie-loop of registered phases, and all GTV contours were visually approved by a single observer.

Three-dimensional dose distributions were available for all patients and were converted to biologically equivalent doses in 2Gy fractions (EQD2) using  $\alpha/\beta=10$ . Independently, the same set of distributions were blurred according to respiratory motion (derived from the registration above) and then converted to EQD2 – which represents the *blurred* dose represented on a reference phase (for which we used the middle, i.e., 50% phase). The mean, maximum, minimum, and standard deviation (SD) of the dose inside the PTV was calculated in both cohorts based on the planned dose. The *blurred* dose provides an indication of the *planned* tumor dose over the respiratory cycle. From this distribution, the mean, maximum, minimum, and SD dose on the generated GTV was calculated.

### 2.2.1 Density metrics

The radial histogram framework described in our previous work was implemented to measure density and dose at radial distances from the generated GTV for all patients (19). A 2D cross-histogram of density vs distance in bins of 1mm annuli from -0.5 to 2cm from the GTV was formed for each 4D phase considering lung tissue only. Only lung tissue was considered for density metrics so that higher density is not just a surrogate for nearby organs-at-risk. For each patient, a single phase was selected as the most stable compared to neighboring phases and used for density analysis in the remaining analysis (as detailed in previous work) (19, 24). Summary curves were then extracted from the 2D histogram for each patient to describe the mean, 90<sup>th</sup> percentile, and SD density at the defined distances from the GTV. The summary curves were

smoothed with Gaussian smoothing ( $\sigma = 1.5\text{mm}$ ) (as annuli thickness is less than slice thickness) and stored for each patient.

### 2.2.2 Dose metrics

To extract dosimetric information, scans were cropped based on the body contour, and dose-volume histograms were extracted in 1mm annuli from the *blurred* dose distribution in the region 0.5-4cm radially from the GTV on the reference phase (19). Smoothed curves were summarized by dose SD, and the fraction of volume in each 1mm rim that receives EQD2 of less than 30Gy, which is a dose threshold reported for controlling MDE (25). An additional curve of mean dose was extracted for visualization only.

### 2.2.3 Exploratory comparison

The average curve across patients for each metric as function of radius was calculated in training and validation. The curves did not inform selection of which variables to use in the remaining analysis but assist in interpreting the results. We visually assessed the images for all patients to note any qualitative characteristics that may lead to differences between the two groups.

## 2.3 Model development

### 2.3.1 Clinical model and 'standard' dose metrics

For training, a reference Cox model was derived for both LR and RF including clinical variables with complete data for all patients – which were tumor lobe location, age, sex, tumor motion amplitude, and tumor volume. The concordance index (C-index), Akaike Information Criterion (AIC), and any variables that statistically significantly predicted LR and/or RF were recorded. Each GTV and PTV dose-related parameter extracted in Section 2.2 was then included in the Cox model individually to determine whether it was associated with RF and LR, and a likelihood-ratio test was performed for models with and without the dose parameter to determine if there was a significant improvement in model performance.

### 2.3.2 Interaction maps

The *Cox-per-radius* method was applied for each outcome separately (LR and RF); full details reported in previous work (19). Briefly, for each combination of dose and density in independent annuli, the density feature, dose metric, and interaction term (*density\*dose*) were added to the clinical model. A likelihood-ratio test was performed for models with and without the interaction to produce a heat-map of p-values describing the benefit of the interaction for prediction in each dose-density annulus combination. Regions of statistical significance (defined by  $p<0.05$ ) were highlighted on the heat-map.

Region size post-processing was then used as multiple-testing correction on the heat-maps to ensure regions were truly not less than 3mm thickness in either the dose or density scale (likely representing spurious associations) (26, 27). The average height and width of each region defined boxes on a heat-map that represent the significant distances for dose and density independently. From the

defined independent annuli, the dose and density values in the identified regions were extracted and included in overall Cox models. Internal validation was performed over 500 bootstrap resamples to estimate the interaction coefficient stability and model performance. A secondary post-processing step was then performed, which removed unstable interaction coefficients on internal validation or interactions that did not improve the performance of the clinical model (C-index). Unstable coefficients are those that show both a negative and positive effect within the 95% confidence interval across resamples.

### 2.3.3 Final models, internal validation, and interpretation

Coefficients and p-values were reported for models with and without the interaction term. To interpret the direction and size of the association between density and outcome, *contrast plots* were created with log(hazard ratio) on the y-axis and different density values on the x-axis for the 10<sup>th</sup> percentile, median, and 90<sup>th</sup> percentile value of the dose parameter in the relevant region (28).

Based on the results from the internal validation, the concordance-index was adjusted for optimism. For each bootstrapped model we calculated an estimate of the C-index and calculated the difference between the C-index of the model in the bootstrapped data and in the original data. To calculate the optimism-adjusted C-index, the median difference across all resamples was subtracted from the original C-index. This internal validation was performed for the clinical model, and the new models with and without an interaction term for comparison. For all models, the median and 95% confidence interval of the C-index across the bootstrap resamples were recorded.

For interpretation of the identified dose location, correlations were investigated between the relevant dose parameter and the metrics extracted in Section 2.2.

## 2.4 External validation

No formal advice is currently reported for external ‘validation’ in an image-based data mining framework. Following the stricter advice of Royston et al. on validation in prediction modelling (29), the model coefficients developed on the training data were applied in the new data-set to build a validation Cox model. The C-index of the validation model was recorded and compared to that identified from the internal validation. This process was performed for the clinical model and repeated for the new models with and without the interaction terms.

As a second comparison (as opposed to strict prediction-model validation) we re-fit the model coefficients in the new data set to

determine if the significance of the interaction would still be identified. Contrast plots were created for this interaction, and the same correlations between other dose metrics performed for interpretation. As a further validation step, we also studied the stratification of patients with Kaplan-Meier plots in both training and validation using an arbitrary cut off for the median peritumour density, and the lowest and highest third of the identified dose metrics.

## 3 Results

### 3.1 Clinical, imaging and dosimetric data

For training, 195 patients with complete clinical, image and dose data were available (19). Thirteen patients had LR (6.5%) and 15 patients (8%) had RF. Complete clinical data was available on tumor lobe location, age, sex, tumor motion amplitude, and tumor volume. In the training data, we report 52% missing data on histological sub-type and 12% on performance status – hence these were excluded from further analysis.

In the validation set, we first selected patients who were treated with the same number of fractions as the training data-set, this left 95 patients treated with 50Gy in 5 fractions. Out of the remaining patients ten more were excluded due to: missing radiotherapy data (n = 1), missing 4D-CT phases (n = 5), alternative planning approach (breath-hold; n = 2), or no iGTV (n = 2). On visual assessment of the GTV generation, we observed seven failures related to registration and three related to segmentation. Overall, 76 patients were available for analysis with complete data, and the difference between the training and the validation cohort are shown in Table 1. Twelve patients had LR (16%) and five RF (6.6%). In the validation cohort, the tumours were larger with a mean tumour volume of 14cc compared to 4cc, and a maximum volume of 158cc compared to 31cc. The tumours in the validation cohort also were more centrally located on visual assessment, which contributed to a significantly lower tumour motion amplitude. In the validation data, a larger distribution of adenocarcinoma compared to squamous cell carcinoma was recorded, and patients had better performance status.

### 3.2 Clinical models for training data

Firstly, using the training data, multivariable Cox models were built to predict LR and RF containing all clinical variables in Table 1, and the full tables are reported in SM Table 2. The summary of these models in Table 2 demonstrates that only

TABLE 2 Clinical model for local and regional failure in training data with significant variables, C-index, and Akaike Information Criterion (AIC) reported. Performance metrics were evaluated on the full data-set.

Outcome	Multivariable prognostic factors (p<=0.05)	C-index	AIC
Regional	None	0.74	142.3
Local	ln(Tumour volume) (HR = 3.05, p = 0.006)	0.79	100.0

tumor volume is a prognostic factor for LR, and there are no prognostic factors for RF. Including ‘standard’ dose metrics in each model did not significantly improve model performance (SM, Table 3).

### 3.3 Radial dose and density metrics

Three density metrics (mean, 90<sup>th</sup> percentile, SD) and three dose metrics (mean, SD, fraction volume receiving less than 30Gy) were extracted from both training and validation cohorts. These metrics were averaged across all patients and the differences visualised between cohorts as shown in Figure 1. An increased density heterogeneity inside and outside the tumour in the validation data was observed. The mean dose is lower in the validation cohort as the total prescription was 50Gy in 5 fractions compared to 60Gy in 5 fractions delivered in the training cohort. Differences in the other curves (Figure 1) include variation in both dose and density heterogeneity which likely represent institutional planning differences.

### 3.4 Dose-density interaction maps

Using the dose and density curves extracted for all patients, six *Cox-per-radius* maps were produced from combinations of the three density metrics and two dose metrics (SD, fraction volume receiving less than 30Gy) extracted for this stage of the analysis.

In the analysis of LR, post-processing 1-40% of significant pixels were removed (SM, Figure 1 and 2). After this, up to nine candidate regions were identified on each map. Of these nine regions, three were below the size threshold, one had unstable coefficients on internal validation, and four did not improve the clinical model performance (SM, Table 3), therefore only one consistently identified region remained (Figure 2).

This region indicates that 90<sup>th</sup> percentile density at the peritumor border (-0.1 to 0.3cm) interacts with the dose SD 0.5-1.6cm and jointly predict LR. The interaction term was significant in 74% of bootstrap resamples and the interaction model performance had a median of 0.81 (95% confidence interval: 0.70 - 0.85) on internal validation. The larger region (light pink border; Figure 2) was not considered important as it did not improve the clinical model performance with a median C-index of 0.79.

For RF, the maps before and after post-processing are shown in SM, Figures 3 and 4 where between 25-100% of spurious significant pixels were removed across maps. Two regions remained (SM, Table 4), but neither met the 3mm threshold for annuli size and hence no regions were considered for further analysis of regional failure.

### 3.5 Final training and validation models

The first column of Table 3 reports the median and 95% confidence interval of the training C-index for the clinical LR model, the model with additional dosimetric and density information, and the final model with an interaction term. In the second column, we record the optimism-adjusted estimates of these values. In both cases, the interaction model out-performs the other models. The strict prediction-model validation results (without refitting) demonstrate that the interaction model does not directly translate to the external data, but the dose and density terms alone do improve on the clinical model in the validation data-set. However, when model coefficients were re-fitted, the importance of the interaction term was re-established.

### 3.6 Model interpretation

The model coefficients for the interaction model in the training data and for the re-fitted model in the validation data are shown in Table 4. The HRs for the interaction term are reported at the 90<sup>th</sup> percentile of SD dose (11.7Gy) and peritumor density (-107.4HU) as this is a multiplicative model, the hazard can only be interpreted as a single value at specified values of the interacting variables.

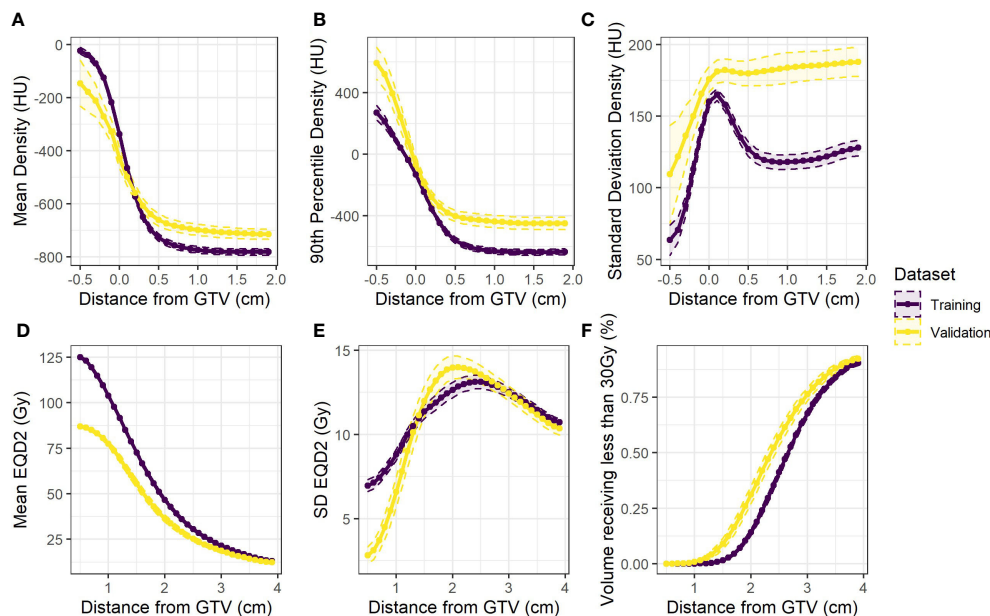
For interpretation on the direction of effect we plot the log (hazard ratio) for peritumor 90<sup>th</sup> percentile density as a continuous variable at three reference values of dose SD in Figure 3. The plot demonstrates that high peritumor density and high dose SD in the identified region is linked to increased risk of LR. At other values of dose SD there is no significant association between density and LR.

An example stratification with these models is demonstrated in Figure 4, where it can be seen that only at high dose SD, peritumour density is linked to worse outcome in training and validation cohorts.

TABLE 3 Concordance index of models on training and validation data.

Model C-index	Training	Optimism-adjusted	Validation (without refitting)	Validation (with refitting)
<i>Clinical</i>	0.81 (0.66-0.94)	0.76 (0.61-0.80)	0.58 (0.37-0.76)	0.76 (0.58-0.91)
<i>Clinical + SD dose + 90<sup>th</sup> percentile density</i>	0.84 (0.71-0.95)	0.77 (0.61-0.82)	0.64 (0.37-0.80)	0.81 (0.66-0.99)
<i>Clinical + SD dose + 90<sup>th</sup> percentile density + SD dose*90<sup>th</sup> percentile density</i>	0.88 (0.79-0.99)	0.82 (0.69-0.93)	0.56 (0.41-0.70)	0.86 (0.71-1.00)

Optimism-adjusted results represent the internally validated models on the training data. Validation without re-fitting represents a strict prediction model validation procedure, whereas with re-fitting is building a new model with all the identified variables in the training data for comparison.



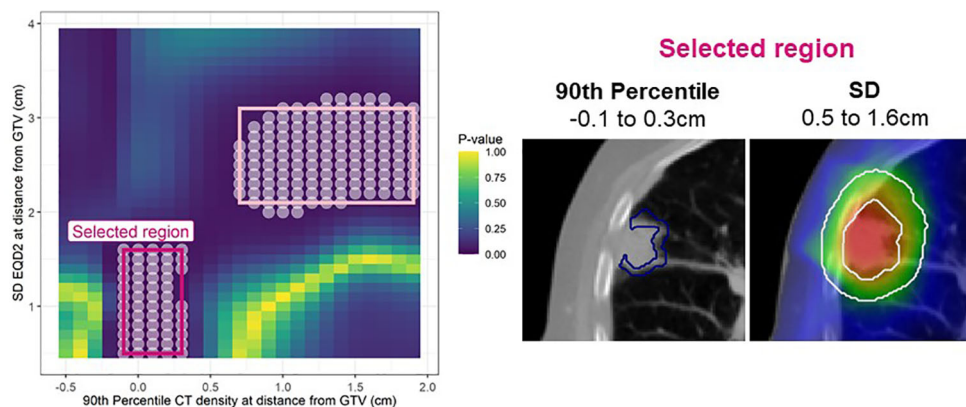
**FIGURE 1**  
Population averages (dashed 95% confidence intervals) for the extracted variables against distance from the GTV for the training data (purple) and the validation data (yellow) (A) mean (B) 90<sup>th</sup> percentile and (C) standard deviation of density. Also (D) mean and (E) standard deviation of equivalent dose, and (F) the fraction of volume in each rim receiving less than 30Gy.

### 3.7 Dosimetric correlations

The correlations between dose in the PTV and GTV and the SD dose in the identified region was studied to investigate the relationship between *incidental* dose and more ‘standard’ dose metrics. The correlations among PTV dose metrics are shown in Figure 5, and among GTV metrics in SM, Figure 5. Dose heterogeneity (SD) in the identified region (0.5 – 1.6cm from GTV) was positively correlated with mean PTV and mean GTV dose in training data, but the correlation was weaker in the validation data. Similar observations are seen for SD, max and min dose to PTV and GTV.

### 4 Discussion

In this study based on real world datasets, as illustrated by the median age of patients and high performance status, the ‘*Cox-per-radius*’ method developed in previous work was applied to investigate spatial interactions of density and dose for the purpose of better understanding local relapse (LR) and regional failure (RF) among early-stage lung cancer patients treated with SABR (19). We identified that higher peritumor density is significantly associated with increased chance of LR for patients who have high dose variability 0.5-1.6cm outside of the GTV. Higher dose variability



**FIGURE 2**  
Left: Cox per radius significance map of interaction between 90<sup>th</sup> percentile density vs SD EQD2 at distance from the GTV. The p-value reflects a likelihood-ratio test of improvement in model performance due to inclusion of the interaction between dose and density at each location. All significant points are shown with a white circle and the regions extracted for assessment in bootstrap are highlighted in pink. After post-processing only one region remained (bright pink). Right: The annuli volumes defined by the selected regions overlaid on an example patient.

90th Percentile Density: -0.1 to 0.3cm | Dose EQD2 (Gy) SD: 0.5 to 1.6cm

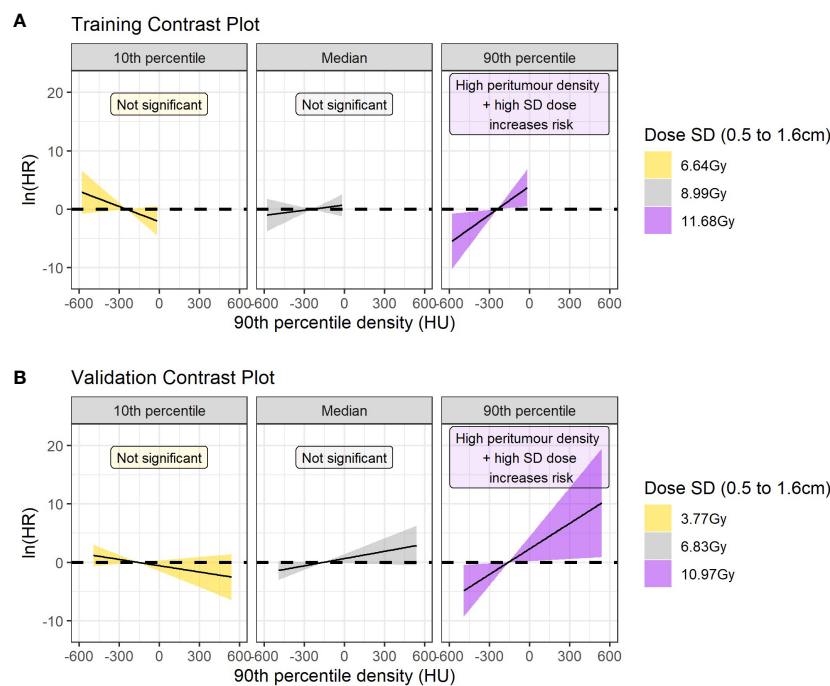


FIGURE 3

Contrast plot displaying the log(hazard ratio) versus 90<sup>th</sup> percentile peritumour density at different values of dose SD (standard deviation). Significant association between density and dose is only detected at high dose variability. In this case, higher peritumour density is associated with increased risk. (A) Shows the results in the training data, and (B) shows the validation results.

in this region is likely driven by increased heterogeneity in the PTV leading to steeper dose fall-off, as shown by correlations between the PTV dose and dose in this region further out. Similarly, Salguero et al. (14) found that patients with high risk of MDE (based on higher GTV surface density and more complex shape) had increased risk of local-regional failure if receiving a low minimum dose up to 1.5cm from the GTV. Both results suggest that microscopic disease coverage is important to prevent LR and could potentially be customized according to peritumor density. Using a second real-world validation data-set, we confirmed the importance of the interaction between peritumor density and dose variability in the same location despite large demographic differences observed in the two cohorts.

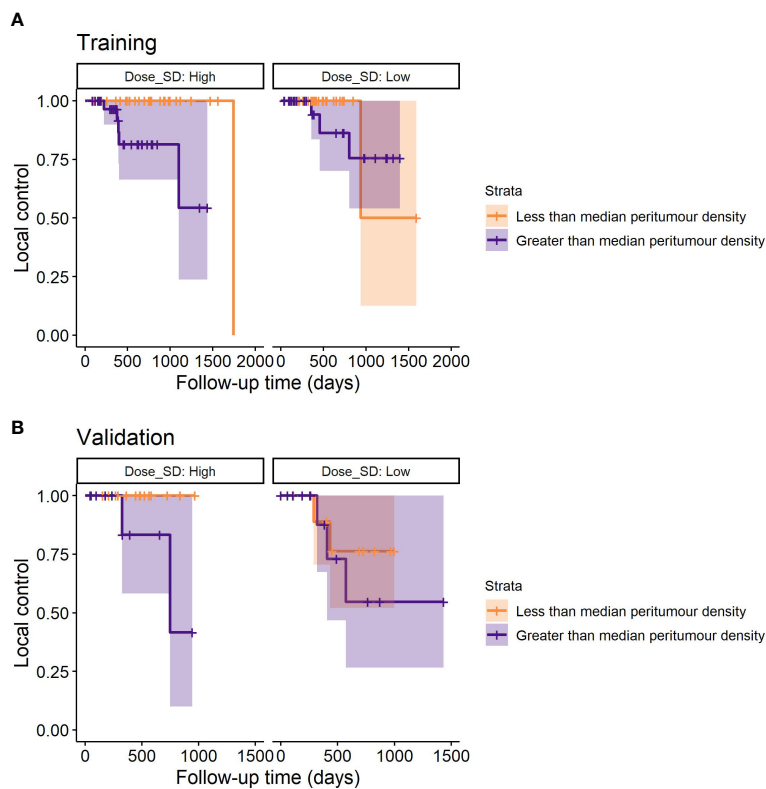
Interestingly, the <1.6cm region identified is considerably narrower compared to the corresponding region discovered by our group for DM, in which dose variability and underdosage ~3cm was predominant (19). The direct mechanisms to DM, LR, and RF are not fully understood, but the idea of these being assigned to different locations of importance is supported.

No dose-density interactions were found to provide predictive ability for RF. In other work, reduced risk of RF has been linked to higher *incidental* dose to the ipsilateral hilum, which suggests RF may be a result of microscopic disease presence at the site of failure or undiagnosed nodal metastases (15). This finding could not be confirmed in this work as the radial approach removes information of proximity to specific anatomy. However, regions were identified

prior to post-processing, so sensitivity testing is required to ensure we are not removing insightful information in this process. As the work of Salguero et al. reports only on local-regional failure it is impossible to determine whether it is the RF or LR that dominate their result or whether it is a combination thereof (14). The results of our study suggest considering LR and RF independently. To better understand RF, anatomical information could be included by combining density interactions with a voxel-based dose analysis (30), but this was beyond the scope of this work.

It is promising that the results of our study generalize to an external validation cohort despite differences in CT acquisition parameters. We did not apply any correction for scanner differences, since inter-scanner variability is typically small compared to inter-patient variability for density metrics (otherwise known as first-order radiomic features) (31). Differences between cohorts may require controlling for more complex texture features, such as, the Grey Level Run Length Matrix features. In this study, the focus on density features was motivated by interpretability and preliminary radiomic analysis that shown relation between 90<sup>th</sup> percentile density and metastasis prediction (24). In addition to other radiomic features, different imaging modalities could be considered to improve prediction of recurrence, e.g. 18-FDG PET (32). Imaging that tracks changes during treatment (e.g., cone-beam CT) could also be utilized to consider tumor reduction (33), and the location of the surrounding microscopic disease (34).





**FIGURE 4** Example stratification of LR based on the median peritumour density for the lowest and highest third of dose SD (discarding uninformative middle values). For low dose SD (assumed adequate MDE coverage), peritumour density no longer stratifies patients for LR – suggesting that peritumour density is a potential predictor for MDE. (A) Shows the survival plots on the training data, and (B) shows the validation data.

A limitation of the methodology presented in this work is the potential risk of spurious correlations between density at the border of an automatically generated GTV and the spatially offset dose annuli. In particular, density at the peritumor border is closely linked to the ability to accurately define the GTV. Higher peritumor density could be representative of high density spiculations that

may or may not be included in the GTV or demonstrate an under-sampled iGTV for tumors with large motion amplitude (35, 36). Such contour variation could link to under-treatment of the GTV as opposed to microscopic disease. In addition, an organ-at-risk abutting the GTV could be associated with an increase in peritumor density. Although care was taken to visually check all

**TABLE 4** The hazard ratios (HR) and p-values for the final models built in the training and validation data for predicting LR.

	Training		Validation	
	HR	P-value	HR	P-value
ln(Tumour volume)	4.29	<b>0.008</b>	0.53	0.095
Tumour motion [cm]	1.68	0.334	1.22	0.784
Tumour location (lower lobe reference)	10.3	<b>0.047</b>	0.52	0.454
Age [years]	1.01	0.885	0.97	0.552
Sex (female reference)	1.4	0.619	8.36	0.059
90th percentile density -0.1 to 0.3cm from GTV [unit: 100HU]	0.02	<b>0.013</b>	0.27	<b>0.057</b>
Dose variability 0.5 to 1.6cm from GTV [Gy]	2.90	<b>0.02</b>	1.20	0.184
90th percentile density * Dose variability [100HU * Gy]	1.64	<b>0.009</b>	1.29	<b>0.035</b>

The bold values are those that are significant predictors in the model (p<0.05).

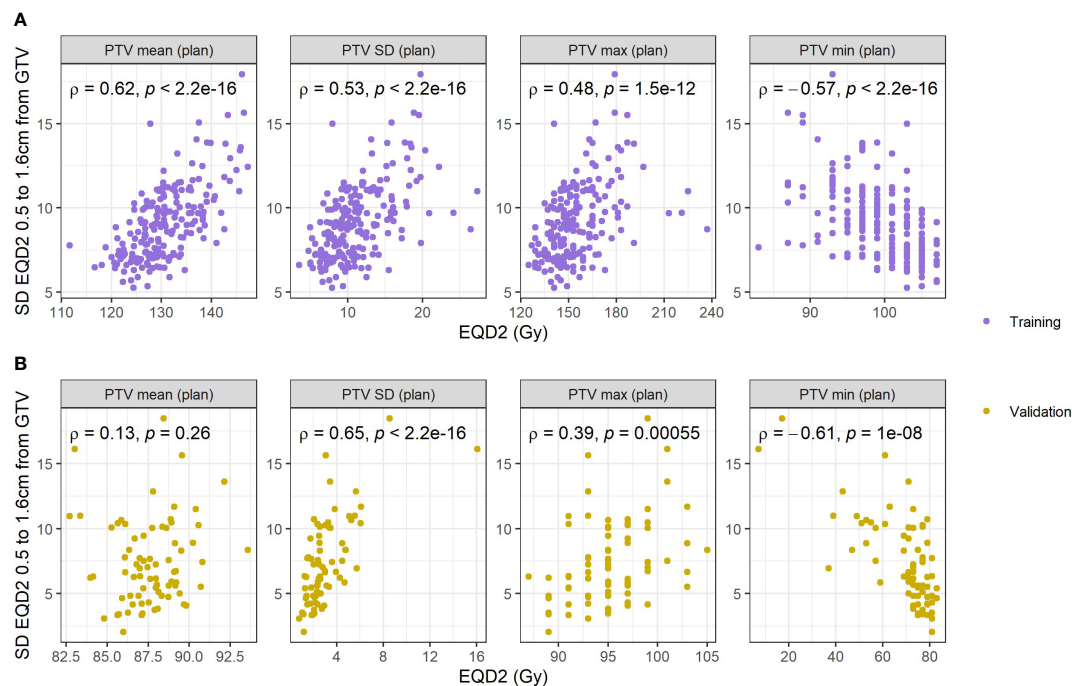


FIGURE 5

Correlations with standard deviation of the respiratory blurred dose extracted from the region identified and different dose metrics extracted from the PTV on the planned dose distribution. (A) Reports correlations in the training data, and (B) reports correlations in the validation data.

contours this cannot be fully excluded as a potential confounding factor. The influence of these factors on the association between high peritumor density and LR could be further studied by exploring the impact of contour variation on LR (37). Further, adopting a physics-approach of annuli at set distances regardless of surrounding anatomy means the dose annuli assessed in this study are not restricted to specific anatomical locations. The annuli can, therefore, include regions for some patients where microscopic disease is unlikely to be a biologically plausible route to LR (i.e., chest-wall) (14). To investigate the sensitivity of the method, a further improvement could include using dose information sampled from the lungs only (similar to density biomarkers), but one would have to be cautious as this could also lead to bias due to loss of different amounts of data when performing lung cropping for peripheral tumors or those close to surrounding organs. The radial data-mining technique demonstrated may provide information on routes to LR, but it would be important to assess histological characteristics of the tumor when making biological conclusions. This was limited by access to such information in routine practice.

There was also limited access to information on baseline pathologies that are common in lung cancer patients (e.g., chronic obstructive pulmonary disease (COPD) and emphysema) that have CT density presentations (e.g., fibrosis and bullae) that could make SABR planning more difficult. Such pathologies could influence the dose-density analysis, so care was taken to review images, dose, and density curves to detect possible outliers. In future study, this information such be incorporated in the analysis to understand causal links behind the associations identified in this

work. Another potential avenue of further investigation could involve the underlying reason of the importance of dose alone (13, 38), while in this study and others the dose and density interaction is required to identify a similar association with outcomes (14, 19). This may in part be due to cohort differences, including tumor size or planning techniques combined with typically small cohort sizes and low number of events all contributing to limited power to determine true effects. Despite the low event rates, we were able to both produce a predictive model for LR with an optimism-adjusted performance of 0.82 in two independent data sets. In particular, the two cohorts included difference in demographics (i.e., performance status), tumor volume, and tumor location, which is representative of the challenges faced using real-world data as opposed to carefully selected patients enrolled on a clinical trial. Despite these challenges, we identified significant support towards the peritumor density as a predictive image biomarker, and increased dose variability up to 1.6cm for high-risk patients leads to worse clinical outcomes. The larger SD in density outside of the GTV in the validation cohort suggests that GTV delineations are tighter, which is consistent with the higher observed LR rate.

Furthermore, the higher LR rate observed in the validation data could also be explained by the lower overall dose (50Gy in 5 compared to 60Gy in 5). Here it is relevant to note that the model of Jeong et al., available online at <https://tcp4rt.info>, which predicts local failure rates quite close to those observed (39). Namely: predicted failure rates: 6.3% (12Gy x 5), 13% (10Gy x 5); observed rates: 6.7% (12Gy x 5), 16% (10Gy x 5). The Jeong fit to early-stage lung cancer SBRT saturates at a failure rate of 5%, even

for very high biological effective dose. It seems reasonable to hypothesize that the mechanism identified in this paper, namely, underdosing of peripheral local disease, is partly responsible for these commonly observed high-dose failures.

The interaction between peritumoral density and dose variability to the identified region (0.5 – 1.6cm from GTV) maintained significance in external validation despite differences in planning approaches implemented at both institutions. The difference in planning approaches was demonstrated by difference in dosimetric correlations between the two cohorts. In training data, a positive correlation was found between standard deviation of dose to identified region, and mean PTV and mean GTV dose. This correlation was weaker in the validation data. In future, it would be worth comparing planning approaches to determine how to best homogenize dose to the identified region.

The inclusion of interaction terms in analysis of real-world data-sets has so far been under investigated. Whilst the current gold-standard evidence for a predictive biomarker is a significant ‘interaction test’ in a randomized controlled trial (40), we have demonstrated inclusion of dose-density interactions in retrospective analyses could allow us to explore and hypothesize on the impact of personalized radiotherapy using real-world data-sets. As randomized controlled trials have long timescales and are limited to specific populations, this complementary method is beneficial to assess the impact of smaller changes to clinical practice in a real-world cohort (41). The discovery and validation of a significant interaction suggests that patients could potentially be stratified based on risk of local relapse pre-treatment which could lead to changes in radiotherapy delivery (e.g., increased margins or dose intensification) or selection for immunotherapy to improve patient outcome.

## 5 Conclusion

In summary, we have applied a previously described data-mining technique to predict LR and RF following lung SABR. High peritumor density was found to interact with dose variability up to 1.6cm outside the GTV and they jointly predicted LR in two independent data-sets from different institutions. No direct association with clinical outcome was found in GTV and PTV dose metrics, but correlations demonstrated that within PTV heterogeneity may be chiefly responsible for this interaction. Overall, external confirmation of the model supports the use of density biomarkers to predict risk of microscopic disease extension, and that adequate dose coverage outside the intended treatment region of the tumor could reduce the risk of local failure. The proposed density x dose biomarker could be investigated in the future to personalize SABR dose distributions and possibly to reduce the rate of residual failures observed even at high doses.

## Data availability statement

The datasets presented in this article are not readily available because ethical permission was not granted for general publication

of the dataset. Requests to access the datasets should be directed to Dr Alan McWilliam, alan.mcwilliam@manchester.ac.uk and Dr Maria Thor, thorm@mskcc.org.

## Ethics statement

Our retrospective analysis of anonymized routine data was approved by institutional information governance and research ethics committee (The Christie NHS Foundation Trust and Caldicott Committee). The research was carried out according to a protocol approved by the Caldicott committee. A data-sharing agreement was in-place for remote analysis of the data from Memorial Sloan Kettering Cancer Center.

## Author contributions

AD developed the methodology, collated the data, performed primary analysis, and wrote the manuscript. MT collated the validation data and was involved in detailed discussions about the planned validation work. AM and MH provided expert insight at all stages of the experiment. CF-F provided expert clinical insight to the project. AR and JD provided insight into the presentation of results and supported the validation work. All authors contributed to the article and approved the submitted version.

## Funding

AD, AM, CF-F and MH are supported by Cancer Research UK via the funding to Cancer Research Manchester Centre [C147/A25254]. AM, MH, and CF-F are also supported by the National Institute for Health and Care Research Manchester Biomedical Research Centre. CF-F and MH are also supported by Cancer Research UK RadNet Manchester [C1994/A28701]. Open access publication fees were provided by The University of Manchester library.

## Acknowledgments

The authors would like to thank Jason Kennedy, Helen Craggs, and Sean Brown for the collection of clinical data to aid this analysis.

## Conflict of interest

AR: Research funding from AstraZeneca, Merck, Pfizer, Boehringer Ingelheim, and Varian Medical Systems. Consulting fees/advisory boards for AstraZeneca, Merck and MoreHealth.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1156389/full#supplementary-material>

## References

- Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2017) 28 (Supp 4):iv1–21. doi: 10.1093/annonc/mdx222
- Senthi S, Lagerwaard FJ, Haasbeek CJA, Slotman BJ, Senan S. Patterns of disease recurrence after stereotactic ablative radiotherapy for early stage non-small-cell lung cancer: a retrospective analysis. *Lancet Oncol* (2012) 13(8):802–9. doi: 10.1016/S1473-2045(12)70242-5
- Sun B, Brooks ED, Komaki RU, Liao Z, Jeter MD, McAleer MF, et al. 7-year follow-up after stereotactic ablative radiotherapy for patients with stage I non-small cell lung cancer: results of a phase 2 clinical trial. *Cancer*. (2017) 123(16):3031–9. doi: 10.1002/cncr.30693
- Dess RT, Sun Y, Matuszak MM, Sun G, Soni PD, Bazzi L, et al. Cardiac events after radiation therapy: combined analysis of prospective multicenter trials for locally advanced non-small-cell lung cancer. *J Clin Oncol* (2017) 35(13):1395–402. doi: 10.1200/JCO.2016.71.6142
- Spratt DE, Wu AJ, Adeseye V, Din SU, Shaikh F, Woo KM, et al. Recurrence patterns and second primary lung cancers after stereotactic body radiation therapy for early-stage non-Small-Cell lung cancer: implications for surveillance. *Clin Lung Canc* (2016) 17(3):177–183.e2. doi: 10.1016/j.clcc.2015.09.006
- Benedict SH, Yenice KM, Followill D, Galvin JM, Hinson W, Kavanagh B, et al. Stereotactic body radiation therapy: the report of AAPM task group 101. *Med Phys* (2010) 37(8):4078–101. doi: 10.1118/1.3438081
- Wilke L, Andratschke N, Blanck O, Brunner TB, Combs SE, Grosu AL, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams: statement from the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery. *Strahlentherapie und Onkol* (2019) 195(3):193–8. doi: 10.1007/s00066-018-1416-x
- Giglioli FR, Strigari L, Ragona R, Borzi GR, Cagni E, Carbonini C, et al. Lung stereotactic ablative body radiotherapy: a large scale multi-institutional planning comparison for interpreting results of multi-institutional studies. *Phys Medica* (2016) 32(4):600–6. doi: 10.1016/j.ejmp.2016.03.015
- Price G, van Herk M, Faivre-Finn C. Data mining in oncology: the uKCAT project and the practicalities of working with routine patient data. *Clin Oncol* (2017) 29 (12):814–7. doi: 10.1016/j.clon.2017.07.011
- Loganadane G, Martinetti F, Mercier O, Krhili S, Riet FG, Mbagui R, et al. Stereotactic ablative radiotherapy for early stage non-small cell lung cancer: a critical literature review of predictive factors of relapse. *Cancer Treat Rev* (2016) 50(1):240–6. doi: 10.1016/j.ctrv.2016.10.002
- Ohri N, Werner-Wasik M, Grills IS, Belderbos J, Hope A, Yan D, et al. Modeling local control after hypofractionated stereotactic body radiation therapy for stage I non-small cell lung cancer: a report from the Elekta collaborative lung research group. *Int J Radiat Oncol Biol Phys* (2012) 84(3):e379–84. doi: 10.1016/j.ijrobp.2012.04.040
- Guckenberger M, Klement RJ, Allgauer M, Appold S, Dieckmann K, Ernst I, et al. Applicability of the linear-quadratic formalism for modeling local tumor control probability in high dose per fraction stereotactic body radiotherapy for early stage non-small cell lung cancer. *Radiother Oncol* (2013) 109(1):13–20. doi: 10.1016/j.radonc.2013.09.005
- Klement RJ, Sonke JJ, Allgauer M, Andratschke N, Appold S, Belderbos J, et al. Correlating dose variables with local tumor control in stereotactic body radiation therapy for early-stage non-small cell lung cancer: a modeling study on 1500 individual treatments. *Int J Radiat Oncol* (2020) 107(3):579–86. doi: 10.1016/j.ijrobp.2020.03.005
- Salguero FJ, Belderbos JSA, Rossi MMG, Blaauwgeers JLG, Stroom J, Sonke JJ. Microscopic disease extensions as a risk factor for loco-regional recurrence of NSCLC after SBRT. *Radiother Oncol* (2013) 109(1):26–31. doi: 10.1016/j.radonc.2013.08.028
- Lao L, Hope AJ, Maganti M, Brade A, Bezjak A, Saibishkumar EP, et al. Incidental prophylactic nodal irradiation and patterns of nodal relapse in inoperable early stage NSCLC patients treated with SBRT: a case-matched analysis. *Int J Radiat Oncol Biol Phys* (2014) 90(1):209–15. doi: 10.1016/j.ijrobp.2014.05.006
- van Loon J, Siedschlag C, Stroom J, Blaauwgeers H, van Suylen R-JJ, Kneijens 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. doi: 10.1016/j.ijrobp.2010.09.001
- Leeman JE, Rimner A, Montecalvo J, Hsu M, Zhang Z, von Reibnitz D, et al. Histologic subtype in core lung biopsies of early-stage lung adenocarcinoma is a prognostic factor for treatment response and failure patterns after stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys* (2017) 97(1):138. doi: 10.1016/j.ijrobp.2016.09.037
- Betensky RA, Louis DN, Cairncross JG. Influence of unrecognized molecular heterogeneity on randomized clinical trials. *J Clin Oncol* (2002) 20(10):2495–9. doi: 10.1200/JCO.2002.06.140
- Davey A, van Herk M, Faivre-Finn C, McWilliam A. Radial data mining to identify density-dose interactions that predict distant failure following SABR. *Front Oncol* (2022) 0:790. doi: 10.3389/fonc.2022.838155
- Varlotto JM, Recht A, Flickinger JC, Medford-Davis LN, Dyer A-M, DeCamp MM. Varying recurrence rates and risk factors associated with different definitions of local recurrence in patients with surgically resected, stage I nonsmall cell lung cancer. *Cancer*. (2010) 116(10):2390–400. doi: 10.1002/cncr.25047
- Davey A, Thor M, van HM, Faivre-Finn C, Rimner A, Deasy J, et al. Identifying predictive factors for treatment failure in early-stage non-small cell lung cancer patients treated with stereotactic body radiotherapy. *Open Sci Framework OSF* (2021). doi: 10.17605/OSF.IO/TRZBY
- Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern cooperative oncology group. *Am J Clin Oncol* (1982) 5(6):649–55. doi: 10.1097/00000421-198212000-00014
- Davey A, van Herk M, Faivre-Finn C, Brown S, McWilliam A. Automated gross tumor volume contour generation for large-scale analysis of early-stage lung cancer patients planned with 4D-CT. *Med Phys* (2021) 48(2):724–32. doi: 10.1002/mp.14644
- Davey A, van Herk M, Faivre-Finn C, Brown S, McWilliam A. Optimising use of 4D-CT phase information for radiomics analysis in lung cancer patients treated with stereotactic body radiotherapy. *Phys Med Biol* (2021) 66(1):115012. doi: 10.1088/1361-6560/abfa34
- Marks LB. A standard dose of radiation for “microscopic disease” is not appropriate. *Cancer*. (1990) 66(12):2498–502. doi: 10.1002/1097-0142(19901215)66:12<2498::AID-CNCR2820661209>3.0.CO;2-X
- Palorini F, Cozzarini C, Gianolini S, Botti A, Carillo V, Iotti C, et al. First application of a pixel-wise analysis on bladder dose-surface maps in prostate cancer radiotherapy. *Radiother Oncol* (2016) 119(1):123–8. doi: 10.1016/j.radonc.2016.02.025
- Shelley LEA, Sutcliffe MPF, Thomas SJ, Noble DJ, Romanchikova M, Harrison K, et al. Associations between voxel-level accumulated dose and rectal toxicity in prostate radiotherapy. *Phys Imaging Radiat Oncol* (2020) 14(1):87–94. doi: 10.1016/j.phro.2020.05.006
- Breheny P, Burchett W. Visualization of regression models using visreg. *RJ*. (2017) 9(2):56–71. doi: 10.32614/RJ-2017-046
- Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol* (2013) 13(1):33. doi: 10.1186/1471-2288-13-33
- Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: a methodological cookbook. *Phys Medica* (2020) 69(1):192–204. doi: 10.1016/j.ejmp.2019.12.013
- Ger RB, Zhou S, Chi PCM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies. *Sci Rep* (2018) 8(1):1–14. doi: 10.1038/s41598-018-31509-z

32. Mazzola R, Fiorentino A, Di Paola G, Giaj Levra N, Ricchetti F, Fersino S, et al. Stereotactic ablative radiation therapy for lung oligometastases: predictive parameters of early response by 18FDG-PET/CT. *J Thorac Oncol* (2017) 12(3):547–55. doi: 10.1016/j.jtho.2016.11.2234
33. Mazzola R, Fiorentino A, Ricchetti F, Niccolò NN, Levra G, Fersino S, et al. Cone-beam computed tomography in lung stereotactic ablative radiation therapy: predictive parameters of early response. *Br J Radiol* (2016) 89(1064):20160146. doi: 10.1259/bjr.20160146
34. Amugongo LM, Osorio EV, Green A, Cobben D, van HM, McWilliam A. Identification of patterns of tumour change measured on CBCT images in NSCLC patients during radiotherapy. *Phys Med Biol* (2020) 65(21):215001. doi: 10.1088/1361-6560/aba7d3
35. Peulen H, Belderbos J, Guckenberger M, Hope A, Grills I, van Herk M, et al. Target delineation variability and corresponding margins of peripheral early stage NSCLC treated with stereotactic body radiotherapy. *Radiother Oncol* (2015) 114(3):361–6. doi: 10.1016/j.radonc.2015.02.011
36. Peulen H, Belderbos J, Rossi M, Sonke JJ. Mid-ventilation based PTV margins in stereotactic body radiotherapy (SBRT): a clinical evaluation. *Radiother Oncol* (2014) 110(3):511–6. doi: 10.1016/j.radonc.2014.01.010
37. Jenkins A, Soares Mullen T, Johnson-Hart C, Green A, McWilliam A, Aznar M, et al. Novel methodology to assess the effect of contouring variation on treatment outcome. *Med Phys* (2021) 48(6):3234–42. doi: 10.1002/mp.14865
38. Diamant A, Chatterjee A, Faria S, El Naqa I, Bahig H, Filion E, et al. Can dose outside the PTV influence the risk of distant metastases in stage I lung cancer patients treated with stereotactic body radiotherapy (SBRT)? *Radiother Oncol* (2018) 128(3):513–9. doi: 10.1016/j.radonc.2018.05.012
39. Jeong J, Oh JH, Sonke JJ, Belderbos J, Bradley JD, Fontanella AN, et al. Modeling the cellular response of lung cancer to radiation therapy for a broad range of fractionation schedules. *Clin Cancer Res* (2017) 23(18):5469–79. doi: 10.1158/1078-0432.CCR-16-3277
40. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol* (2009) 27(24):4027. doi: 10.1200/JCO.2009.22.3701
41. Van Loon J, Grutters J, Macbeth F. Evaluation of novel radiotherapy technologies: what evidence is needed to assess their clinical and cost effectiveness, and how should we get it? *Lancet Oncol* (2012) 13(1):169–77. doi: 10.1016/S1470-2045(11)70379-5