Radiotherapy and Oncology 136 (2019) 78–85

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Longitudinal radiomics of cone-beam CT images from non-small cell lung cancer patients: Evaluation of the added prognostic value for overall survival and locoregional recurrence

Janna E. van Timmeren [a,*], Wouter van Elmpt [b], Ralph T.H. Leijenaar [a], Bart Reymen [b], René Monshouwer [c], Johan Bussink [c], Leen Paelinck [d], Evelien Bogaert [d], Carlos De Wagter [d], Elamin Elhaseen [d], Yolande Lievens [d], Olfred Hansen [e,g], Carsten Brink [e,f], Philippe Lambin [a]

[a] The D-Lab: Decision Support for Precision Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+; [b] Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre (MUMC), Maastricht; [c] Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, the Netherlands; [d] Ghent University Hospital and Ghent University, Ghent, Belgium; [e] Institute of Clinical Research, University of Southern Denmark; [f] Laboratory of Radiation Physics, Odense University Hospital; and [g] Department of Oncology, Odense University Hospital, Odense, Denmark

ARTICLE INFO

Article history:
Received 4 December 2018
Received in revised form 7 March 2019
Accepted 29 March 2019
Available online 11 April 2019

Keywords:
Non-small cell lung cancer
Radiomics
Cone-beam CT
Longitudinal
Overall survival

ABSTRACT

Background and purpose: The prognostic value of radiomics for non-small cell lung cancer (NSCLC) patients has been investigated for images acquired prior to treatment, but no prognostic model has been developed that includes the change of radiomic features during treatment. Therefore, the aim of this study was to investigate the potential added prognostic value of a longitudinal radiomics approach using cone-beam computed tomography (CBCT) for NSCLC patients.

Materials and methods: This retrospective study includes a training dataset of 141 stage I–IV NSCLC patients and three external validation datasets of 94, 61 and 41 patients, all treated with curative intended (chemo)radiotherapy. The change of radiomic features extracted from CBCT images was summarized as the slope of a linear regression. The CBCT slope-features and CT-extracted features were used as input for a Cox proportional hazards model. Moreover, prognostic performance of clinical parameters was investigated for overall survival and locoregional recurrence. Model performances were assessed using the Kaplan–Meier curves and c-index.

Results: The radiomics model contained only CT-derived features and reached a c-index of 0.63 for overall survival and could be validated on the first validation dataset. No model for locoregional recurrence could be developed that validated on the validation datasets. The clinical parameters model could not be validated for either overall survival or locoregional recurrence.

Conclusion: In this study we could not confirm our hypothesis that longitudinal CBCT-extracted radiomic features contribute to improved prognostic information. Moreover, performance of baseline radiomic features or clinical parameters was poor, probably affected by heterogeneity within and between datasets.

Quantitative image analysis has been widely applied using treatment planning computed tomography (CT) images acquired at baseline prior to radiation treatment [1]. Investigation of the prognostic and predictive performance of large amounts of quantitative image features (i.e. radiomics) could eventually lead to the discovery of biomarkers that can contribute to the development of precision medicine [2–4]. The prognostic value of radiomic features extracted from CT images has already been shown for non-small cell lung cancer (NSCLC) [5–7]. Besides that, the potential added value of CT imaging during treatment has been explored for outcome prediction for both lung cancer and head-and-neck cancer [8,9]. These during-treatment images were acquired in a research setting and are generally not included in the clinical workflow. Nonetheless, a large number of cone-beam CT (CBCT) images are usually available, since these are acquired in clinical practice prior to radiation treatment fractions for positioning verification purposes. This means that additional potential prognostic information could be derived without extra patient burden. CBCT images have previously been proposed to monitor tumor response during treatment, either using radiomics [10–12] or simpler metrics like volume, density and mean Hounsfield units (HU)

* Corresponding author at: The D-Lab: Decision Support for Precision Medicine, Maastricht University, Universiteitssingel 40, 6229ER Maastricht, the Netherlands.
E-mail address: j.vantimmeren@maastrichtuniversity.nl (J.E. van Timmeren).

https://doi.org/10.1016/j.radonc.2019.03.032
0167-8140/

[13–16]. Early prognostic information could aid in clinical decision making regarding treatment adaption.

We hypothesize that the change of radiomic features might contain additional information over baseline radiomic information, since it would be possibly able to capture an early treatment response. A previous study showed that radiomic features extracted from cone-beam CT images acquired at the first radiotherapy fraction contains similar prognostic performance for overall survival as baseline CT measurements [10]. Moreover, a recent study showed that radiomic features acquired from CBCT images do change more than the day-to-day variability during the course of treatment [11]. Nevertheless, the prognostic value of these changes in feature values has not yet been explored. Therefore, in the current study we investigated the potential added prognostic value of a longitudinal radiomics approach using CBCT images, over the prognostic value of CT-extracted radiomic features, for non-small cell lung cancer (NSCLC) patients.

Methods

Patients

Four patient cohorts from four different institutes were collected retrospectively, consisting of 141 (Dataset 1), 94 (Dataset 2), 61 (Dataset 3) and 41 patients (Dataset 4). Two clinical endpoints were considered in this study: overall survival and locoregional recurrence, both measured from the start of radiotherapy for all cohorts. Additionally, clinical parameters including age, gender, TNM-stage, overall stage and histology were collected. Detailed information about data collection, patients and exclusion criteria can be found in Supplementary Material. The study was approved by Maastro Clinic's Institutional Review Board. Data collection was approved by each institutional ethics committee.

Imaging

All patients in the cohort received a treatment planning CT (pCT) image and at least four CBCT images acquired during the course of radiotherapy. Moreover, the first available CBCT image was acquired prior to the first (FX1) or second (FX2) radiotherapy fraction. Patients were excluded from the study in case the field-of-view (FOV) of one or multiple CBCT image(s) did not comprise the entire primary tumor volume, which was the case for 19 patients in Dataset 1. Also if the required minimum number of 4 CBCT images was not available, patients were excluded. This was the case for 5 patients of Dataset 1 and 1 patient of Dataset 3. Technical details on the image acquisitions are presented in the Supplementary Material, as well as a histogram visualizing the timing of all 2154 analyzed CBCT images.

For all CBCT images, an intensity value correction was applied before feature extraction, using a region of interest (ROI) in the heart and a reference value from literature of 50 HU [17]. This is identical to the first step of the correction procedure described in more detail in [10].

Tumor segmentation

The treatment planning delineations of the primary gross tumor volume (GTV) on the pCT images were used to define corresponding tumor segmentations on all subsequent cone-beam CT images with the deformable registration algorithm 'Morphons' of the open source software REGGUI (http://openreggui.org). For Datasets 3 and 4, deformable registration in MIM (MIM Software Inc, Cleveland, OH, USA) was used. The use of MIM reduced the workload by being able to register the images and adjust the contours in one-go. We experienced similar performance for the deformable registrations of both software packages and believe this did not introduce a bias. Moreover, deformable registration was followed by manual adjustments to improve the delineations by removing spikes, loose pixels and irregularities or in case the registration procedure was not able to properly register the images. All tumor segmentations were checked and if needed manually adjusted by experienced radiation oncologists. In Dataset 1, for two patients the original GTV delineation contained two separate volumes and it was not possible to define which one was the primary tumor. Moreover, for 12 patients in Dataset 1, the tumor was indefinable on the CBCT images due to infiltration, atelectasis or image blurring. These 14 patients were therefore not included, resulting in the final cohort of 141 patients. For Dataset 4, the tumor was indefinable on the CBCT images for 4 patients.

Feature extraction

A total of 2317 radiomic features were extracted from all pCT and CBCT images included in the analysis, using in-house developed software programmed in Matlab 2014a (Mathworks, Natick, MA). Additional information about the features and image preprocessing can be found in Supplementary Material.

Cone-beam CT radiomics

Features were extracted from weekly or bimonthly CBCT scans with at least four time points available for each patient throughout the entire treatment. Due to the limited number of data points, linear regression was performed to estimate the development of the radiomic features over time. Features values were first divided by the initial value to obtain relative numbers to also include baseline information. Then, a linear regression was performed and the slope parameter of the linear regression was then used as a new CBCT feature that captures the longitudinal information into one single feature. The coefficient of determination (R^2) was evaluated to verify the goodness of fit.

Feature selection

Feature selection was performed on the training dataset: Dataset 1. The process consisted of several steps: (1) Feature clean-up, (2) Selecting only changing CBCT features and calculate linear regression slope, (3) Combine CBCT-slope feature set with pCT feature set and (4) Remove correlated features OR calculate principal components. The steps are explained in further detail in Supplementary Material and the resulting feature selection numbers are provided in the Results section.

Model development and validation

As agreed upon prior to the analysis, three models were developed using only Dataset 1 as training dataset. The first model was based on radiomics features itself (CBCT-slope and pCT). The second model was based on the calculated principal components (PCs) from the radiomic features. To investigate the clinical relevance, the prognostic performance of radiomics was also compared to those of clinical parameters: the input of the third model included all available clinical parameters without radiomic features. Interaction terms were not addressed in this study.

Penalized Cox proportional hazard models were developed using a Least Absolute Shrinkage and Selection Operator (LASSO) procedure [18]. A 10-fold cross validation procedure was repeated 200 times to stabilize the method and to find the optimal lambda value with the smallest error. The performance of this Cox model was evaluated without retraining the coefficients. The prognostic index (PI), defined as $\sum_i \beta_i x_i$, was calculated for each patient. The

distribution of the PIs on each dataset was evaluated to visualize as suggested by Royston and Altman [19]. Moreover, the model fit was evaluated by calculating estimating the regression coefficient ('calibration slope') on the PI of the validation datasets, where a slope of 1 would indicate that the discrimination in the validation datasets is similar. Furthermore, the model misclassification was tested by offsetting the original PI and running a Cox regression on the covariates in the validation datasets. A significant p-value of the joint test on all regression coefficients would indicate a lack of fit of one or more variables.

The performance of Cox model was evaluated using the Kaplan–Meier curves, split between a low-risk and high-risk group using the median PI-value of all patients. A log-rank (LR) test was performed to evaluate the significance of the split. Moreover, Harrell's concordance index (c-index) was calculated to assess the discriminative power of the model, which is a measure that ranges from 0.5 (no discrimination) to 1.0 (perfect discrimination) [20].

The developed models on Dataset 1 were validated on Datasets 2, 3 and 4 or a combination of all validation datasets. The Kaplan–Meier curves were evaluated as well as the c-index.

All statistics and analyses were performed using R (version 3.4.3).

Evaluation of the study design

The study design was evaluated using two methods: the Radiomics Quality Score (available online on www.radiomics.world) and the TRIPOD classification [21,22]. The TRIPOD checklist for "Prediction Model Development and Validation" was used to guide the study and the "TRIPOD Adherence data extraction checklist" was used to calculate a score (0–100%) for the final study (both available online on www.tripod-statement.org).

Results

The patient characteristics of the final cohorts are shown in Table 1. Missing data in Dataset 1 was corrected using mean imputation: for 1 patient the WHO status was unknown, for 7 patients the smoking status was unknown and for 28 patients the FEV_1 was unknown. A comparison of the overall survival and locoregional recurrence is shown in Fig. 1.

A summary of the feature selection workflow is provided in Fig. 2. Feature selection resulted in a final set of 283 radiomic features, consisting of 88 CBCT-slope features and 195 CT features. A total of 50 principal components described 95% of the variance in the data and were subsequently used as input for the modeling. The median of the mean R^2 values calculated per feature by averaging over all patients, showing the goodness of fit of the linear regression for acquiring the CBCT-slope features, was 0.40 ± 0.05 (median ± SD).

Three models were developed for both overall survival and locoregional recurrence, using different input variables: (1) radiomic features (models 1.1 and 1.2), (2) principal components (models 2.1 and 2.2) and (3) clinical parameters (models 3.1 and 3.2). Model 1.1 was developed to predict overall survival and resulted in a three-feature model containing only CT features. Model 1.2 for locoregional recurrence resulted in a six-feature model, consisting of three CBCT-slope features and three CT features. Models 2.1 and 2.2 could not be build, as no prognostic features could be identified for either overall survival or locoregional recurrence using the principal components as input. Model 3.1 included six clinical features to predict overall survival. Model 3.2 is a three-feature clinical model to predict locoregional recurrence.

All models with the feature's respective β-coefficients, plus the median and range of the feature values in Dataset 1 (which was used for training), are summarized in Table 2. Explanations of all features in these models can be found in Supplementary Material. The distribution of the PIs, the calibration slope of the PI and the joint test on all regression coefficients are also shown in Supplementary Material. The C-indices of all identified prognostic models are shown in Fig. 3. The clinical model contained parameters which were not available for Dataset 4 (WHO performance status), so this model was only validated in Datasets 2 and 3. The Kaplan–Meier curves for the radiomics model predicting overall survival (model 1.1) are shown in Fig. 4. The Kaplan–Meier curves for model 1.2, for predicting locoregional recurrence, are shown in Supplementary Material.

The study design was evaluated using the Radiomics Quality Score (RQS) and the TRIPOD statement. We achieved an RQS of 53.8% as feature reduction was applied, three validation datasets were available and a clinical model was investigated. On the other hand, there was no phantom study involved, the correlation with underlying biology was not investigated, the data were not available from a prospective study registered in a trial database and no cost-effective analysis was performed. On the other hand, feature reduction was performed and three validation datasets were available to validate the model without retraining and without adaption of the cut-off value. Moreover, the radiomics model performance was compared to that of clinical parameters, including TNM-stage and volume, which are commonly seen as the gold standard. The TRIPOD checklist for Prediction Model Development and Validation Scoring [21,22], using the TRIPOD items for 'Development and external validation (of the same model)', reached a score of 81%. The latter can be found in Supplementary Material.

Discussion

The aim of this study was to investigate the complementary prognostic value of CBCT radiomic features for overall survival and locoregional recurrence of NSCLC patients. We hypothesized that the change of radiomic features contains additional information over baseline radiomic information. To this extent, models were developed using CBCT-slope and planning CT-derived radiomic features. Model performances were compared to that of a model including solely clinical parameters.

For the first models, selected CBCT-slope and pCT features were used as input for LASSO. A prognostic model was identified for overall survival, which contains three CT based radiomic features. The performance of the model was average (c-index 0.63 [95% C.I. 0.57–0.69]), but it could be validated on Dataset 2 (c-index 0.59 [95% C.I. 0.54–0.65]) and also on the combined set of Datasets 2, 3 and 4. A model developed by Fave et al. [9] containing clinical parameters, baseline CT-radiomic features and delta-CT radiomic features found a c-index 0.675, but no external validation was present and weekly CT imaging acquired in a research setting was used for this study. For locoregional recurrence, six radiomic features were selected including three CBCT-slope features and three CT features. This model did not reach significance in any of the validation datasets. Also in the study of Fave et al. [9], no delta CT features were significant predictors for locoregional recurrence.

For the second models, principal components were used as input for LASSO. For both overall survival and locoregional recurrence, no prognostic models could be identified.

The third and last models, containing clinical parameters, achieved a c-index of 0.62 for overall survival and 0.69 for locoregional recurrence on Dataset 1. Both models did not validate on either Dataset 2, Dataset 3 or a combination of both.

Previous studies have shown that a four-feature radiomics model derived from CT images has prognostic value for NSCLC patients. Also, a previous study showed that this model could also

Table 1
Patient characteristics for all datasets. The dashed numbers indicate to which dataset the respective variable was significantly different ($p < 0.05$).

		Dataset 1 (n = 141)	Dataset 2 (n = 94)	Dataset 3 (n = 61)	Dataset 4 (n = 41)
Age		*3,4*	*3*	*1*	*1,2*
	Range (median)	45–86 (70)	42–83 (68)	39–83 (65)	41–85 (62)
	Mean ± SD	68.6 ± 9.6	67.0 ± 8.5	64.2 ± 9.3	62.3 ± 10.7
Gender		*4*	*4*	*1,2*	
	Male	86 (60.6%)	45 (47.9%)	48 (78.7%)	26 (63.4%)
	Female	56 (39.4%)	49 (52.1%)	13 (21.3%)	15 (36.6%)
WHO performance status		*2,3*	*1,3*	*1,2*	
	0	16 (11.3%)	27 (28.7%)	42 (68.8%)	
	1	96 (68.1%)	53 (56.4%)	15 (24.6%)	
	2	24 (17.0%)	14 (14.9%)	4 (6.6%)	
	3	4 (2.8%)	0 (0%)	0 (0%)	
Stage					
	I	12 (8.5%)	4 (4.3%)	0 (0%)	2 (4.9%)
	II	15 (10.6%)	7 (7.4%)	3 (4.9%)	4 (9.7%)
	IIIa	44 (31.2%)	50 (53.2%)	37 (60.7%)	25 (61.0%)
	IIIb	55 (39.0%)	33 (35.1%)	21 (34.4%)	10 (24.4%)
	IV	15 (10.6%)*	0 (0%)	0 (0%)	0 (0%)
Histology†					
	Adenocarcinoma	37 (26.2%)	34 (36.2%)	34 (55.7%)	14 (34.1%)
	Squamous cell carcinoma	60 (42.6%)	42 (44.7%)	22 (36.1%)	15 (36.6%)
	Large cell carcinoma	5 (3.5%)	5 (5.3%)	3 (4.9%)	3 (7.3%)
	Undifferentiated	0 (0%)	6 (6.4%)	0 (0%)	0 (0%)
	Not otherwise specified	39 (27.7%)	7 (7.4%)	2 (3.3%)	9 (22.0%)
GTV (cm^3)					
	Range (median)	0.61–341 (38)	2.1–397 (38)	1.5–425 (36)	1.7–415 (55)
	Mean ± SD	62.3 ± 70.6	70.2 ± 74.6	72.3 ± 93.0	90.1 ± 72.3
Concurrent chemotherapy					
	Yes	90 (63.8%)	63 (67.0%)	35 (57.4%)	
	No	51 (36.2%)	31 (33.0%)	26 (42.6%)	
Interval CT–RT (Days)		*2,3,4*	*1,3*	*1,2*	*1*
	Range (median)	3–16 (7)	5–21 (11)	6–39 (13)	6–21 (13)
	Mean ± SD	7.2 ± 1.6	10.9 ± 2.4	15.6 ± 7.7	12.7 ± 3.9
Received RT dose (Gy)		*2,3,4*	*1,3,4*	*1,2,4*	*1,2,3*
	Range (median)	45–76 (69)	60–66 (66)	60–70 (70)	52–67.5 (66)
	Mean ± SD	66.4 ± 5.6	64.3 ± 2.7	68.5 ± 3.1	65.2 ± 3.2
Radiotherapy schedule		*2,3,4*	*1*	*1*	*1*
	30–35 * 2 Gy (daily)	0 (0%)	94 (100%)	61 (100%)	41 (100%)
	30 × 1.5 Gy (twice daily) + 5–12 × 2 Gy (daily)	78 (54.9%)	0 (0%)	0 (0%)	0 (0%)
	23–24 × 2.75 Gy (daily)	28 (19.7%)	0 (0%)	0 (0%)	0 (0%)
	38–42 × 1.8 Gy (daily)	26 (18.3%)	0 (0%)	0 (0%)	0 (0%)
	Other	10 (7.0%)	0 (0%)	0 (0%)	0 (0%)
Overall survival (year)					
	Median [range]	2.0 [0.1–4.8]	1.7 [0.1–9.8]	2.8 [0.4–7.7]	1.8 [0.4–6.5]
Events at time of analysis					
	Survival	91	84	32	31
Follow-up (year)		*2,3*	*1,3,4*	*2,4*	*1,2,3*
	Median [range]	3.0 [0.2–4.8]	8.6 [6.6–9.8]	3.4 [1.1–7.7]	5.7 [5.3–6.5]

* This group of patients is treated with curative intent and has similar prognosis and the stage III patients.
† For histology, it was investigated whether the number of adenocarcinoma and squamous cell carcinoma patients was significantly different between the datasets.

be validated using CBCT images acquired prior to the first radiotherapy fraction [10]. Therefore, we also investigated the performance of a previously developed radiomic signature [5] during treatment. The c-index was 0.59 for all datasets combined, which was not significantly different compared to the validation performance in other studies [5,10,23]. The performance of the signature on CBCT images acquired during radiation treatment shows that the signature is robust over time, but the prognostic value does not improve at later time points (data not shown).

Complementary prognostic value of longitudinal CBCT radiomic features could not be shown in this study. No CBCT-slope features were selected by LASSO to predict overall survival. For locoregional recurrence, CBCT-slope features were selected, but the model did not validate on any of the validation datasets. The lack of prognostic value of CBCT-slope features can be caused by several factors. First of all, the slope surrogate might not be the best method to use for representing CBCT features changing over time. Only a few data points per patient were included to estimate the slope, since the GTV segmentations on the CBCT were time-consuming as they needed manual verification and adjustment. Due to this limited number of data points available, the slope is sensitive to outliers. Potentially, automatic segmentations, for instance atlas-based methods or deep learning, could improve this methodology [24,25]. Besides the uncertainty in the slope parameters due to the limited number of data points, it is possible that the development of CBCT extracted radiomic features over time is not linear. A more advanced model to represent the behavior of radiomic features over time might be more appropriate (e.g. a quadratic model or

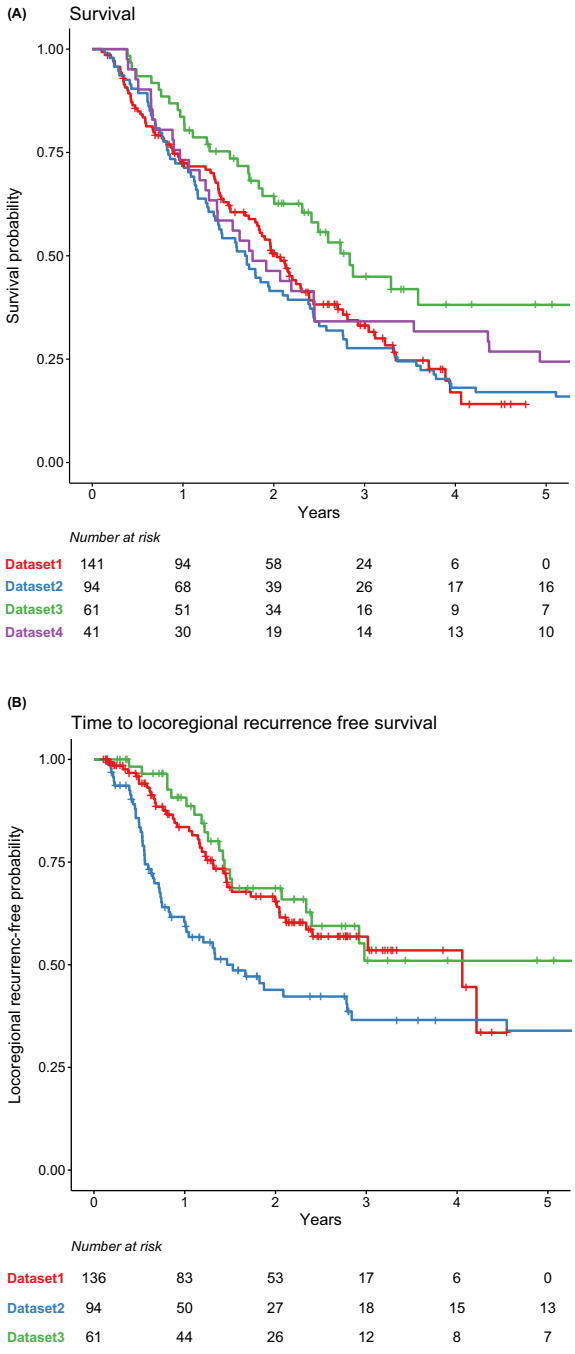

Fig. 1. Kaplan Meier curves showing the overall survival (A) and locoregional recurrence (B) of all datasets.

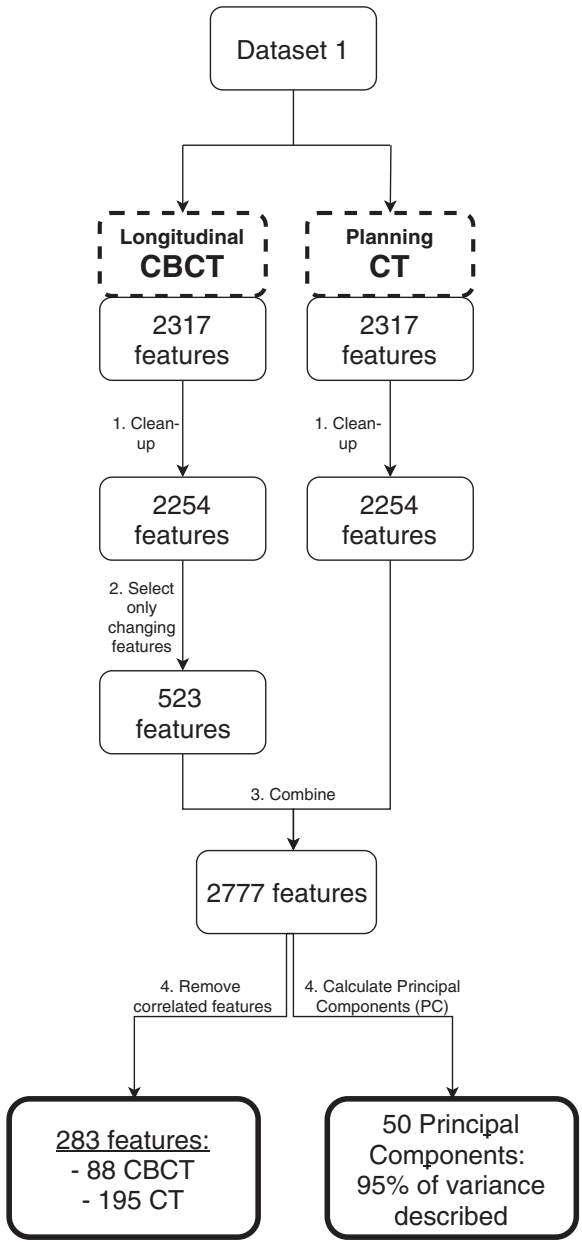

Fig. 2. Schematic overview of the feature selection process applied to Dataset 1. The initial 2317 features were initially reduced to 2254 features by a clean-up (see text) and the CBCT features were successively reduced based on time variance (see text). Finally the combined set of features (2777) are modeled either after removable based on correlations or by use of PCA in order to investigate model stabilities.

polynomial). However, the limited number of available data points in combination with a model with more degrees of freedom would have resulted in overfitting and most likely a false-positive result. Furthermore, it is also possible that the slope features contain no prognostic information, whereas the absolute difference between the feature values of two CBCT scans does have prognostic value. In a recent study it was shown that a large number of features change more than expected by chance in week 3 and week 4 of treatment [11], but this was not yet related to outcome information. The lack of prognostic value of current CBCT features may also be influenced by the low image quality of cone-beam CT. Although a previous study showed that a batch of radiomic features are interchangeable between planning CT and cone-beam CT, other features might be subject by high levels of noise. Several efforts are being made for improving the image quality of CBCT images, e.g. using artifact reduction techniques, hardware developments, improvement of 4D imaging or the implementation of iterative reconstruction methods [26–28]. In this study three different CBCT systems were involved: Varian Truebeam (Dataset 1), Elekta Synergy (Datasets 2 and 4) and Varian Clinac (Dataset 3). The image quality of these three systems differs substantially, which could have affected the lack of validation performance for the model containing CBCT slope features. Furthermore, the CBCT images analyzed in the current study are free-breathing, which for lung tumors causes a blurring of the region of interest. Considering all these factors, it is likely that (some of) these factors hamper the discovery of prognostic value in CBCT-extracted radiomic features. Once the image quality of CBCT is improved and automatic

Table 2
Model features with corresponding model coefficients and indication the median and range of the feature values in training Dataset 1. Model 1.1 and model 3.1 were developed to predict overall survival and model 1.2 and model 3.2 were developed to predict locoregional recurrence. For models 2.1 and 2.2, no prognostic features were identified. Feature definitions are presented in Supplementary Material.

Model	β-coefficients	Feature names	Median feature value [range]
Models 1.1 and 1.2: radiomics features for overall survival and locoregional recurrence, respectively			
Model 1.1	$-4.25 \cdot 10^{-3}$	CT_LoG_sigma_2_5_mm_3D_pos_IH_mode	1 [1–20]
	$9.66 \cdot 10^{-4}$	CT_LocInt_peakLocal	882.4 [414.5–1442.7]
	$1.96 \cdot 10^{-4}$	CT_Wavelet_LLL_GLCM_inverseVar	0.30 [0.13–0.47]
Model 1.2	$3.12 \cdot 10^{-3}$	CT_LoG_sigma_2_5_mm_3D_pos_IH_maxGradI	12 [1–29]
	0.0255	CT_LoG_sigma_5_5_mm_3D_IH_p10	11 [2.8–22]
	−0.665	CT_Wavelet_HHH_Stats_median	0.011 [−1.28 to 0.84]
	−11.0	CBCT-slope_Wavelet_HHL_GLCM_maxCorr	0.0033 [−0.0088 to 0.021]
	−2.42	CBCT-slope_Wavelet_HLH_GLCM_infoCorr2	0.0035 [−0.0070 to 0.022]
	3.46	CBCT-slope_Wavelet_HLL_GLCM_inverseVar	−0.0022 [−0.020 to 0.016]
Models 2.1 and 2.2: principal components for overall survival and locoregional recurrence, respectively			
No models could be identified			
Models 3.1 and 3.2: clinical variables for overall survival and locoregional recurrence, respectively			
Model 3.1	0.424	WHO 2	*Dummy variables*
	0.0567	Concomitant	
	0.453	Stage II	
	0.165	Stage IIIa	
	−0.121	Histology (other)	
	0.178	Histology (squamous cell)	
Model 3.2	0.122	WHO 1	*Dummy variables*
	0.223	Stage IIIa	
	0.558	Histology (squamous cell)	

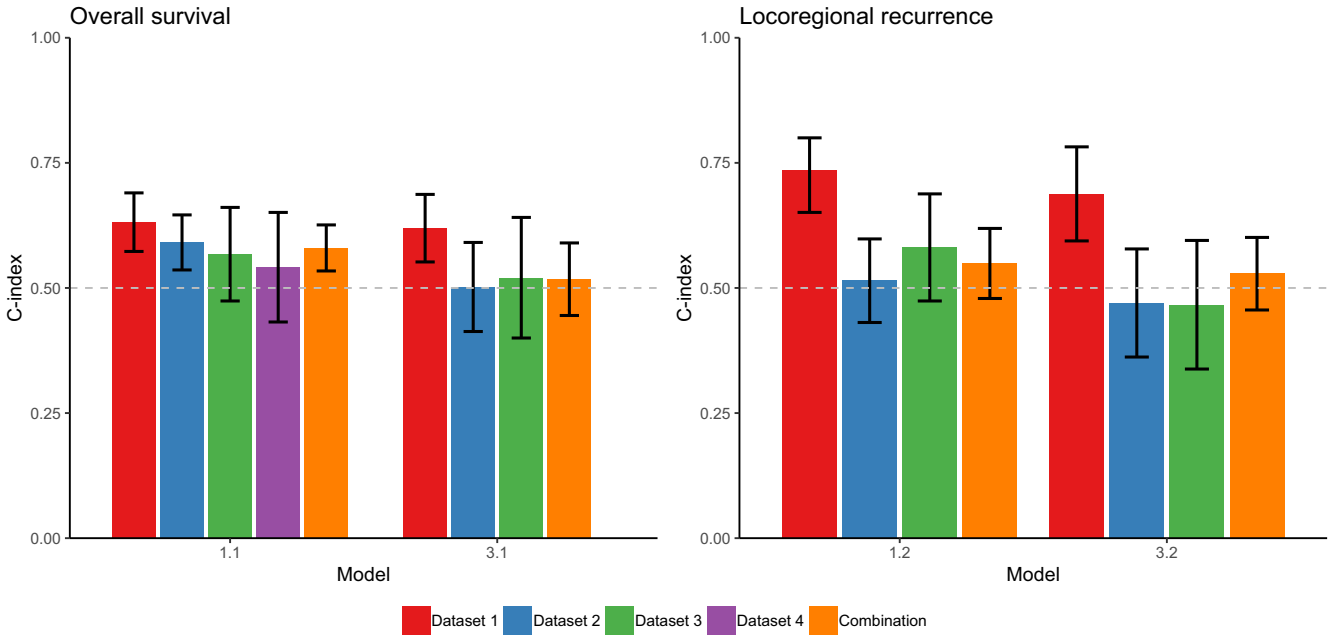

Fig. 3. C-indices of the prognostic models identified for overall survival (left) and locoregional recurrence (right). The combination was either Datasets 2, 3 and 4 (model 1.1) or Datasets 2 and 3 (models 1.2, 3.1 and 3.3). Models 1.1 and 1.2 contain only radiomic features and models 3.1 and 3.2 contain only clinical parameters. For models 2.1 and 2.2 no prognostic features could not be identified.

segmentation methods are developed, it would be interesting to redo the analysis.

A factor that could have influenced the validation performance of model 1.2 (containing CBCT-slope features to predict locoregional recurrence) is the software package used to perform the deformable registration to transfer the GTV contour of the pCT onto the CBCT images: a different software package was used in Datasets 1 and 2 than in Datasets 3 and 4. Although we have not experienced a difference between the two methods and the radiation oncologist manually verified all contours, this could potentially have resulted in a systematic bias in the data. Nevertheless, this model also did not validate on Dataset 2, for which the same software package was used as in training Dataset 1. Moreover, this does not explain the lack of validation performance of the models containing only pCT-features or clinical parameters.

In general, models based on (pre)treatment imaging or parameters do most likely not contain enough information to more accurately predict a future event like death. Treatment changes, and in particular second-line or third-line therapies, have an enormous effect on overall survival. Another explanation for the poor validation performance could be the intra- and inter-dataset heterogeneity and the limited sample size of the validation datasets, which are

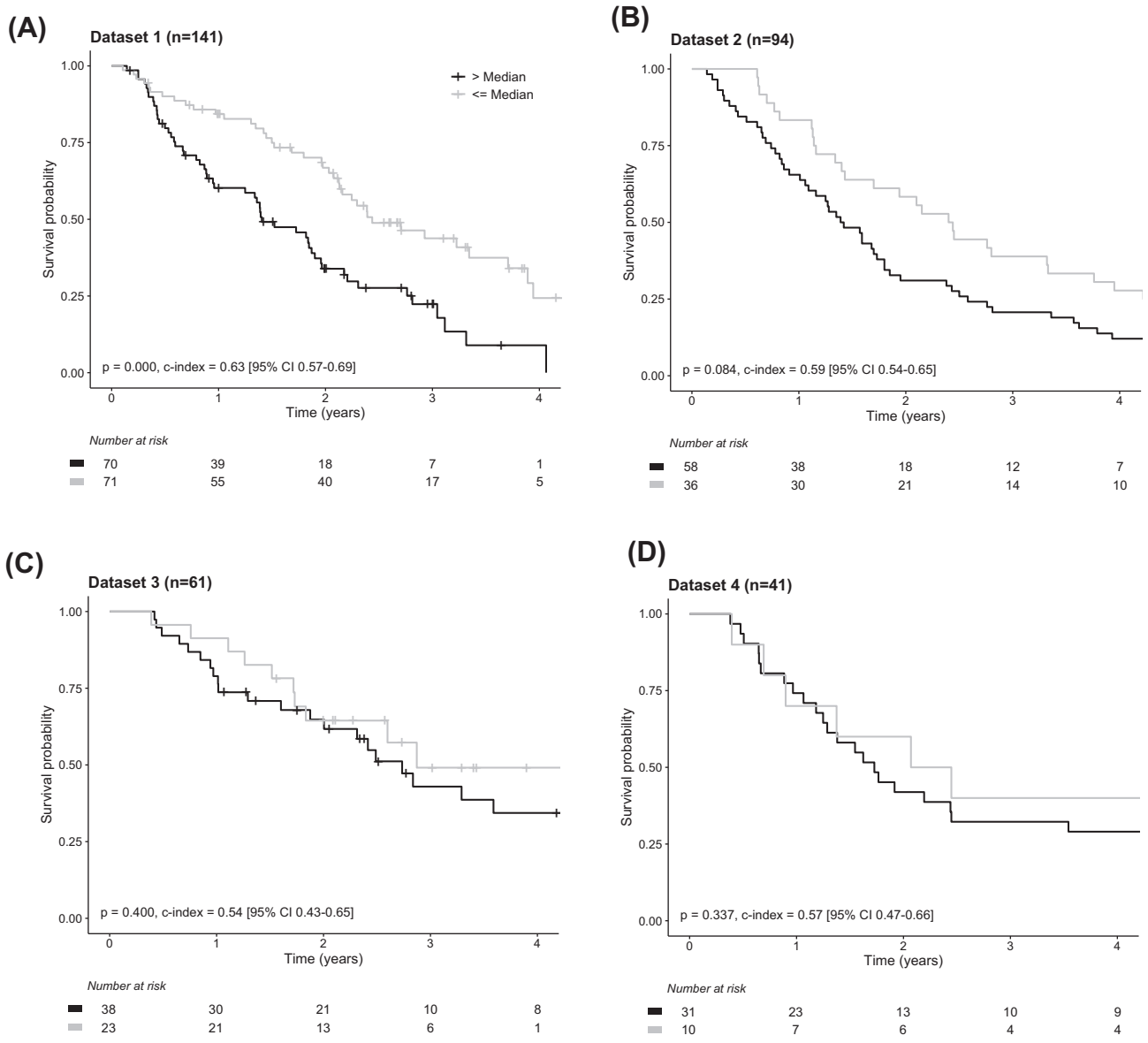

Fig. 4. Kaplan Meier curves of model 1.1 containing three CT features to predict overall survival. A) Training: Dataset 1, B) Validation 1: Dataset 2, C) Validation 2: Dataset 3, D) Validation 3: Dataset 4. For Datasets 2 and 4, there were no censored data up to 4 years of follow-up.

common phenomena of retrospective (multicentric) studies. Comparing patient characteristics and scanner parameters show that the datasets are not much alike, which makes it difficult for a model to perform well on the external dataset. Locoregional recurrence free survival in Dataset 2 is worse compared to the other datasets, which might be caused by an earlier detection of tumor recurrence because of the use of PET in the follow-up program in this institution [29]. Overall survival is better in Dataset 3, in which a larger number of patients have WHO performance status 0 compared to the other datasets. Moreover, radiotherapy schedule, age and imaging protocols differed between datasets.

Several validation measures were evaluated to investigate the potential cause for the fact that the developed models could not be validated on the validation datasets. The distribution of the prognostic indices for the radiomics model which was developed for overall survival is comparable for all datasets. Moreover, the calibration slope of the PI was not significantly different from 1 for any of the validation datasets. Furthermore, the joint test on all coefficients was not significant, showing that there is no evidence for a lack of fit. These results show that model covariates and coefficients are comparable in the validation datasets, which indicates that it is most likely that the lack of validation is a result of the small cohort sizes. We recommend to perform future studies on larger, more homogeneous datasets.

In this study we could not confirm our initial hypothesis that longitudinal radiomic features extracted from CBCT images contribute to improved prognostic information. Nevertheless, the model developed for overall survival containing only CT features, could be validated in one external dataset and also on a combination of all three external validation datasets. On the other side, validation of the models containing only clinical parameters showed poor performance, which indicates that there is still a lack of reproducible prognostic models for outcome prediction in NSCLC.

Competing interests

Author RL is CTO and shareholder of Oncoradiomics SA and co-inventor of a patent related to Radiomics.

Author PL is member of the advisory board, shareholder of Oncoradiomics SA and coinventor of two licensed & approved patents on Radiomics.

Acknowledgements

Philippe Lambin acknowledges financial support from ERC advanced grant (ERC-ADG-2015, n° 694812 – Hypoximmuno), ERC-2018-PoC (n° 81320 – CL-IO). This research is also supported by the Dutch Technology Foundation STW (grant n° 10696 DuCAT & n° P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs. Authors also acknowledge financial support from SME Phase 2 (RAIL – n°673780), EUROSTARS (DART, DECIDE, COMPACT), the European Program H2020-2015-17 (BD2Decide – PHC30-689715, ImmunoSABR – n° 733008, PREDICT – ITN – n° 766276), TRANSCAN Joint Transnational Call 2016 (JTC2016 "CLEARLY" – n° UM 2017-8295), Interreg V-A Euregio Meuse-Rhine ("Euradiomics"), Kankeronderzoekfonds Limburg from the Health Foundation Limburg and the Dutch Cancer Society.

Olfred Hansen and Carsten Brink acknowledge support from AgeCare (Academy of Geriatric Research at Odense University Hospital).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2019.03.032.

References

[1] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.
[2] O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169–86.
[3] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[4] Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. Physica Med 2017;38:122–39.
[5] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[6] Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. Invest Radiol 2015;50:719–25.
[7] Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345–50.
[8] Leger S, Zwanenburg A, Pilz K, Zschaeck S, Zophel K, Kotzerke J, et al. CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. Radiother Oncol 2019;130:10–7.
[9] Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep 2017;7:588.
[10] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol 2017;123:363–9.
[11] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Lambin P. Feature selection methodology for longitudinal cone-beam CT radiomics. Acta Oncol 2017;1–7.
[12] Fave X, Mackin D, Yang J, Zhang J, Fried D, Balter P, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? Med Phys 2015;42:6784.
[13] Bernchou U, Hansen O, Schytte T, Bertelsen A, Hope A, Moseley D, et al. Prediction of lung density changes after radiotherapy by cone beam computed tomography response markers and pre-treatment factors for non-small cell lung cancer patients. Radiother Oncol 2015;117:17–22.
[14] Bertelsen A, Schytte T, Bentzen SM, Hansen O, Nielsen M, Brink C. Radiation dose response of normal lung assessed by Cone Beam CT – a potential tool for biologically adaptive radiation therapy. Radiother Oncol 2011;100:351–5.
[15] Rosen BS, Hawkins PG, Polan DF, Balter JM, Brock KK, Kamp JD, et al. Early changes in serial CBCT-measured parotid gland biomarkers predict chronic xerostomia after head and neck radiotherapy. Int J Radiat Oncol Biol Phys 2018.
[16] Brink C, Bernchou U, Bertelsen A, Hansen O, Schytte T, Bentzen SM. Locoregional control of non-small cell lung cancer in relation to automated early assessment of tumor regression on cone beam computed tomography. Int J Radiat Oncol Biol Phys 2014;89:916–23.
[17] Cardiac CT imaging: diagnosis of cardiovascular disease. New York, NY: Springer Berlin Heidelberg; 2016.
[18] Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc: Ser B (Methodol) 1996:267–88.
[19] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33.
[20] Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–6.
[21] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63.
[22] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.
[23] de Jong EEC, van Elmpt W, Rizzo S, Colarieti A, Spitaleri G, Leijenaar RTH, et al. Applicability of a prognostic CT-based radiomic signature model trained on stage I–III non-small cell lung cancer in stage IV non-small cell lung cancer. Lung Cancer (Amsterdam, Netherlands) 2018;124:6–11.
[24] Schipaanboord B, Boukerroui D, Peressutti D, van Soest J, Lustberg T, Kadir T, et al. Can Atlas-based Auto-segmentation Ever be Perfect? Insights from Extreme Value Theory. IEEE Transactions on Medical Imaging; 2018.
[25] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 2018;126:312–7.
[26] Star-Lack J, Sun M, Oelhafen M, Berkus T, Pavkovich J, Brehm M, et al. A modified McKinnon-Bates (MKB) algorithm for improved 4D cone-beam computed tomography (CBCT) of the lung. Med Phys 2018.
[27] Martin R, Ahmad M, Hugo G, Pan T. Iterative volume of interest based 4D cone-beam CT. Med Phys 2017;44:6515–28.
[28] Zhu L, Xie Y, Wang J, Xing L. Scatter correction for cone-beam CT in radiation therapy. Med Phys 2009;36:2258–68.
[29] Pan Y, Brink C, Schytte T, Petersen H, Wu YL, Hansen O. Planned FDG PET-CT scan in follow-up detects disease progression in patients with locally advanced NSCLC receiving curative chemoradiotherapy earlier than standard CT. Medicine (Baltimore) 2015;94:e1863.