Title: Novel CT-based objective imaging biomarkers of long term radiation-induced lung damage

Short title: CT-based imaging biomarkers of RILD

Authors:

Catarina Veiga[1], PhD;

David Landau[2,3], FRCR;

Anand Devaraj[4], BSc, MRCP, FRCR, MD;

Tom Doel[1], PhD;

Jared White[1], BSc;

Yenting Ngai[5], BSc;

David J. Hawkes[1], PhD;

Jamie R. McClelland[1], PhD.

Affiliations:

[1]Centre for Medical Image Computing, Department of Medical Physics & Biomedical Engineering, University College London, London, UK

[2]Department of Oncology, Guy's & St. Thomas' NHS Trust, London, UK

[3]Department of Oncology, University College London Hospital, London, UK

[4]Department of Radiology, Royal Brompton Hospital, London, UK

[5]Cancer Research UK & UCL Cancer Trials Centre, University College London, London, UK

Keywords: lung, radiation-induced lung damage, computed tomography, fibrosis

ABSTRACT

Background and Purpose: Recent improvements in lung cancer survival have spurred an interest in understanding and minimizing long term radiation-induced lung damage (RILD). However, there is still no objective criteria to quantify RILD leading to variable reporting across centres and trials. We propose a set of objective imaging biomarkers to quantify common radiological findings observed 12-months after lung cancer radiotherapy (RT).

Material and Methods: Baseline and 12-month CT scans of 27 patients from a phase I/II clinical trial of isotoxic chemoradiation were included in this study. To detect and measure the severity of RILD, twelve quantitative imaging biomarkers were developed. These describe basic CT findings including parenchymal change, volume reduction and pleural change. The imaging biomarkers were implemented as semi-automated image analysis pipelines and assessed against visual assessment of the occurrence of each change.

Results: The majority of the biomarkers were measurable in each patient. Their continuous nature allows objective scoring of severity for each patient. For each imaging biomarker the cohort was split into two groups according to the presence or absence of the biomarker by visual assessment, testing the hypothesis that the imaging biomarkers were different in these two groups. All features were statistically significant except for rotation of the main bronchus and diaphragmatic curvature. The majority of the biomarkers were not strongly correlated with each other suggesting that each of the biomarkers is measuring a separate element of RILD pathology.

Conclusions: We developed objective CT-based imaging biomarkers that quantify the severity of radiological lung damage after RT. These biomarkers are representative of typical radiological findings of RILD.

MANUSCRIPT

1. INTRODUCTION

Radiation-induced lung damage (RILD) is an unwanted effect of radical radiotherapy (RT). RILD is a dynamic process of acute inflammation and chronic fibrosis, which leads to permanent loss of quality of life in cancer survivors[1]. Up to 17% of patients have been described to have evidence of RILD[2,3] although it is likely under-reported due to a focus on poor overall survival. RT adversely affects pulmonary function in the majority of the patients treated with curative intent[4]. As lung cancer survivorship increases it becomes more important to understand, score and minimize late treatment-related side effects[5–8].

RILD is characterized by CT imaging. The most commonly reported radiological finding is increasing intensity of the lung parenchyma[9–11]. Other anatomical distortions of the thorax are present but sparsely reported[12–18]. Radiological change is apparent in the majority of the patients treated with curative RT[19].

Long term RILD is often reported using systems developed to describe pulmonary fibrosis. The Radiation Therapy Oncology Group and European Organization for Research and Treatment of Cancer (RTOG/EORTC) late radiation morbidity scoring system score non-specified "radiological changes" as either slight, patchy or dense [20]. RTOG/EORTC criteria is subjective, with users regularly interpreting "patchy" as ground-glass opacities (GGOs) and "dense" as consolidation. The Common Terminology Criteria for Adverse Events (CTCAE) scores pulmonary fibrosis radiologically on the extent of "radiological pulmonary fibrosis" from <25% to <75% and adds "honeycombing" in grade 4[21]. The differences in RTOG/EORTC and CTCAE guidelines lead to significant variations in grading depending on the system used, since consolidation is very common but damage is localised to the treated volumes; both disregard other common anatomical changes that commonly occur[19]. Both systems are hence flawed in describing the incidence and severity of late RILD. The lack of suitable scoring criteria has led to few publications describing long term RILD[22,23] and to variable reporting of toxicity amongst trials[3]. There is therefore an urgent need to develop objective measures of RT-induced damage to better describe and predict RILD[24,25].

In a previous study, we described the common radiological findings of RILD 12-months after chemoradiation in a cohort of homogeneously treated patients[19]. Based on those findings, we propose a set of novel CT-based objective imaging biomarkers that quantify the severity of different radiological characteristics of RILD. These objective imaging biomarkers were implemented using

standardized and semi-automatic pipelines which are critically assessed in this work. This is a first step toward an objective and quantifiable description of RILD with important clinical applications, such as facilitating the correlation with RT dose and development of a scoring system.

2. METHODS AND MATERIALS

2.1 Patient population

27 patients treated with conventional chemoradiotherapy were included. This was a sub-group of XXX, a non-randomized phase I/II multicentre trial. This enrolled stage II/III non-small cell lung cancer (NSCLC) patients to receive isotoxic tumour RT doses between 63 Gy and 73 Gy in 30 fractions over 6 weeks (daily) or 63 Gy and 71 Gy in 30 fractions over 5 weeks (with one day twice daily fractionation given weekly) concurrent with two cycles of cisplatin and vinorelbine[6]. Tumour prescription was defined to achieve a mean lung dose (MLD) of 18.2 Gy (in equivalent dose in 2 Gy fractions). Each patient underwent CT imaging before treatment and 12-months after RT. Acquisition parameters and image resolution varied intra- and inter-patient, with the majority of the scans being diagnostic and acquired at breath-hold. RT volumes and doses were also available. Detailed information on patient and imaging characteristics are presented in Table 1.

2.2 Objective imaging biomarkers of radiological findings of RILD

Common radiological changes 12-months after RT were divided in three main categories: parenchymal change, volume reduction and pleural change according to our previous study[19]. Typical parenchymal change corresponds to changes in intensity and texture, with the most common pattern being consolidation[26]. Volume reduction includes apparent shrinkage in lung volume alongside anatomical changes to the thorax such as distortions of the diaphragm, central airways and anterior junction line. Pleural reactions are inflammatory responses including pleural effusion and pleural thickening.

We propose a set of twelve objective imaging biomarkers representing different manifestations of RILD. The biomarkers implemented are conceptually described below. For full description see supplementary material (Appendix A).

Normal lung volume shrinkage: The shrinkage in normal lung volume (ΔNV) is quantified as a decrease in the volume of normal lung between time points. The normal lung volume was defined

from the total lung volume by excluding parenchymal change, residual pleural reactions, large vessels and tumour masses (i.e., high-intensity pixels).

Consolidation volume: The volume of consolidation was quantified as the ratio of high-intensity volume (RV). RV is the relative proportion in high-intensity lung volume between ipsilateral and contralateral lungs, at follow-up. It is the only biomarker that does not compare two time-points as the presence of the tumour would skew the results.

Changes in lung shape: The reduction in lung width and height (ΔX and ΔZ) were quantified from change in width and height of the total lung volume. Width is defined as the maximum width in axial projection. Height is the maximum height calculated in coronal projection from the axial mid-point.

Distortions of the diaphragm: There are three common radiological findings at the diaphragm: elevation, changes in curvature and tenting. The diaphragmatic elevation (Δh) is the variation in relative height between ipsi and contralateral diaphragms in coronal view. Change in curvature (ΔC) is defined as the variation in the averaged Mean curvature of the diaphragm surface, a concept from differential geometry[27]. Finally, diaphragmatic tenting (ΔS) is identified as the variation in area of the diaphragm with abnormal curvature.

Distortion of the central airways: The major airways are often pulled either upward or downward, and/or toward the ipsilateral chest wall at 12-months. The rotation of the main bronchus ($\Delta\alpha$) is defined from the change in angle between the trachea and the ipsilateral primary bronchus on coronal view[28]. Mediastinal shift (ΔM) is the decrease in separation between the carina and the ipsilateral chest wall on coronal view.

Distortions of the anterior junction line: Rotation and thickening of the anterior junction line reflect mediastinal change due to volume loss. Rotation of the anterior junction line ($\Delta\beta$) is the rotation of the junction line toward the ipsilateral lung. The edge of the contralateral lung is used when the junction itself is not easily defined. Anterior junction thickening (Δt) was defined by the relative increase in the thickness of the junction on axial slice.

Pleural change: The severity of pleural change (ΔP) is calculated as the increase in surface of the chest wall covered by pleural reactions larger than 2 mm.

2.3 Implementation details

Figure 1 schematically describes the image analysis pipelines implemented. Full technical details of the implementation can be found in supplementary material (Appendix B). Briefly, the framework was implemented in a modular and semi-automatic fashion in MATLAB (Mathworks, Natick, MA). Baseline CT images were corrected for setup errors by putting them in a patient-specific coordinate system (supplementary material, Appendix C). Follow-up scans are rigidly co-registered to the baseline using the open-source NiftyReg (sourceforge.net/projects/niftyreg)[29]. This step maximized the alignment of the vertebrae anatomy. Segmentation of lungs and airways was performed using the open-source Pulmonary Toolkit (github.com/tomdoel/pulmonarytoolkit)[30]. Sub-volumes (i.e., normal and high-intensity lung) and surfaces (i.e., diaphragm and chest wall) of interest were generated from the automatically generated lung segmentations, followed by reviewing and (if necessary) manual editing by a radiation oncologist (XX) or medical physicist (XX) using ITK-SNAP[31]. Deformable registration was used to propagate the spinal canal segmented for RT planning to the pre- and post-RT scans[32]; this volume is required to define $\Delta\beta$ and ΔP. Skeletonisation of the airways was performed using the generalised thinning method implemented in the Pulmonary Toolkit[33]. The pipelines that calculate the imaging biomarkers from segmentations are fully automated.

The imaging biomarkers of RILD are continuous and correspond to typical anatomical changes that occur between two time-points. To quantify anatomical change, objective anatomical features are measured at each individual time-point. The comparison of the anatomical features measured between two time-points define the imaging biomarkers (with the exception of RV). The mathematical equations are defined such that positive values indicate the most common direction of changes. Some features are normalized by the corresponding feature measured from the contralateral lung. This accounts for differences in level of inhalation between scans. Namely, ΔNV, RV, ΔX, ΔZ, ΔC, ΔM and $\Delta\alpha$ are change in relative anatomical features, expressed as a percentage. For example, the biomarker "normal lung shrinkage" (ΔNV) is calculated by subtracting the anatomical feature "normal lung volume" measured at follow-up (NV_F) from the "normal lung volume" measured at baseline (NV_B), $NV_B - NV_F$, so that positive values indicate the most common change (reduction in volume). NV is normalized by the equivalent measure from the contralateral lung, and converted to a percentage. See supplementary material for full description of all biomarkers (Appendices A and B).

2.4 Critical assessment

The biomarkers were assessed against presence or absence of each type of visual change using the Wilcoxon rank sum test (10% significance level). The visual assessment was achieved by consensus by a multidisciplinary team composed of a thoracic radiologist, a clinical oncologist and a medical physicist as detailed in a previous publication [19]. Binomial classification was applied to dichotomize the continuous imaging biomarkers into presence or absence (binomial regression, threshold defined at 50% probability). This analysis makes the imaging biomarkers and visual assessment mathematically comparable. Shortcomings and advantages of the imaging markers against visual assessment were identified. The correlation between pairs of imaging biomarkers was assessed using the Pearson's correlation coefficient, with p-values adjusted using the Benjamini-Hochberg procedure (false discovery rate: 10%).

3. RESULTS

The twelve imaging biomarkers were measured for 27 subjects. For a total of five subjects, the anterior junction line position could not be unequivocally defined because tumour, consolidation or pleural thickening were present in the interface with the junction and could not be easily distinguished from the junction itself. Therefore, Δt could not be measured and the edge of the contralateral lung was used to estimate $\Delta\beta$. In four subjects, CT scans presented reconstruction errors or limited view of the diaphragm and hence ΔC could not be accurately analysed. Two subjects were excluded from $\Delta\alpha$ analysis as the bifurcation of the airway tree was narrowed and could not be segmented.

Figure 2 shows examples of anatomical changes and corresponding imaging biomarkers. Namely, the first subject (Figure2a) exhibited severe changes: a normal lung volume shrinkage of 34% (ΔNV); 4.6 times more high-intensity volume on the ipsilateral than contralateral lung (RV); a reduction in lung width and height of 24% and 11% (ΔX and ΔZ, respectively); an elevation of the diaphragm of 21 mm (Δh) with a change in curvature and surface of tenting of 30% and 333 mm^2 (ΔC and ΔS, respectively); the carina pulled toward the ipsilateral side by 31% of the corresponding distance at the contralateral side (ΔM) and main bronchus rotated by 27% ($\Delta\alpha$), the anterior junction line rotated by 14° ($\Delta\beta$) and 6 times more thickened (Δt); and an increase in the coverage of the chest wall by pleural reactions of 11% (ΔP). The biomarkers revealed significant patient to patient variation in the degree of each change.

For each imaging biomarker, the patient group was split into two subgroups according to the biomarker's presence or absence on visual assessment (Figure 3). The range of results is also shown

by the boxplots. We tested the hypothesis that the measured severity of the imaging biomarkers was different in the two subgroups. For all imaging biomarkers the difference in measured severity for the two subgroups was statistically significant ($p \leq 0.1$) except for main bronchus rotation ($p=0.71$) and diaphragmatic curvature ($p=0.13$). This analysis was repeated for a sub-group that excluded the subjects for whom the baseline CT scan was non-diagnostic (hence reducing the patients for analysis to 22 out of 27). Similar results were obtained within the subgroup. To investigate if there were systematic differences in the biomarkers depending which lung is the ipsilateral, the patient group was also split according to the ipsilateral lung. No biomarker was significantly different between the two subgroups ($p>0.1$). Data are shown in supplementary material Appendices D and E).

Pearson's linear correlation coefficient was calculated between each pair of imaging biomarkers (Figure 4). While the majority of the biomarkers were not strongly correlated, measures of normal lung volume (ΔNV), consolidation volume (RV), lung shape (ΔX) and pleural change (ΔP) were moderately to strongly correlated.

4. DISCUSSION

We propose a description of lung changes after RT using a set of imaging biomarkers representative of typical radiological findings of RILD and evaluated them on 12-month scans following radical RT for NSCLC. To the best of our knowledge, this is the first time that RILD has been objectively quantified with a broad spectrum of radiological findings. A key aspect of our study is that we propose objective and continuous features to quantify anatomical change consistently across the patient group. The continuous nature of the biomarkers allow severity to be described quantifiably rather than qualitatively.

Our imaging biomarkers are novel. Other groups have proposed methods using feature analysis and deformable image registration to score parenchymal change characteristic of acute RILD (i.e., radiation pneumonitis)[34–36]. This methodology is less applicable to chronic RILD due to the increased complexity of thoracic anatomical change. The challenges in registering serial CT scans in the presence of late RILD have been described elsewhere[37]. A study by Heo et al was the first attempt at quantifying RILD at 12-month follow-up[23]. The authors evaluated inter and intra-user variability in measuring fibrotic volumes. A similar metric had previously been used to investigate the relationship between dose and RILD[38]. Fibrotic volumes were defined as the volume of grade 1 toxicity according to RTOG/EORTC[20], which correspond to slight radiographic appearances. These studies depend on explicit segmentation of parenchymal change (which can be challenging) and disregard volume reduction and pleural change. Direct and indirect measures of volume loss and

anatomical distortion may be crucial to understand permanent decline in breathing function. Differentiation between fibrosis and inflammation is also challenging. Analysis of multiple time-points are probably necessary to distinguish the processes.

The imaging biomarkers correlated with visual assessment of radiological findings of RILD with the added benefit of objectively scoring radiological severity in a continuous scale (Figure 5a). To identify whether the range of values measured reflects the clinical severity of RILD will require correlation with RT dosimetry and lung function in larger cohorts. From dichotomisation of the biomarkers, we found good agreement when the anatomical changes were severe or mild, and more variable when changes were moderate (Figure 5b). One source of disagreement between biomarkers and visual assessment is the subjectiveness inherent to visual assessment. This subjectiveness is exacerbated when the level of inhalation varies between scans (Figure 5c). In the case of $\Delta\alpha$ (rotation of main bronchus), its poorer representation on visual assessment may result from the biomarker using the first bifurcation of the airway tree to assess distortion, whilst distortions may be apparent further down the airway tree upon visual inspection. Objective quantification improves on visual assessment particularly for mild severity changes as it provides a continuous scale and does not define subjective and arbitrary thresholds. The implemented image analysis pipelines have the potential to be used as an objective tool to help radiologists diagnose RILD and other lung diseases with similar radiological findings.

The majority of the biomarkers were not strongly correlated with each other and can each be seen as providing complementary information to fully describe RILD. Our findings suggest that each of the biomarkers is measuring separate elements of RILD pathology and that a scoring system could be proposed when investigated on larger cohorts. While each biomarker was designed to measure one particular radiological finding, some might also indirectly measure others and hence be more correlated and/or cause outliers, e.g. pleural effusions push the lung and contribute to apparent volume shrinkage and change in shape, and parenchymal change reduces the volume of residual normal lung and distorts the thoracic anatomy. The combination of multiple biomarkers will allow the definition of response sub-groups.

One limitation of the method is its dependency on the accuracy of manually edited segmentations. We minimise this issue by automatically post-processing the segmentations in the pipelines, and by not explicitly using manual segmentation of parenchymal change. Even lung segmentation is complex in the presence of toxicities, disease and image reconstruction errors. The higher correlation between RV and ΔP may be related to the difficulties in distinguishing parenchymal change and pleural thickening (Figure 5d). This leads to inaccuracies in marker measurement, generally overestimation since the

same anatomical feature is being measured simultaneously by different biomarkers. Residual tumour or recurrence cause overestimation of toxicity.

Different levels of inhalation between the pre-treatment and follow-up scans is the largest source of uncertainty in the quantification of RILD. Biomarkers of change in anatomical distortions (ΔZ, ΔC, Δt and $\Delta\alpha$) are those more likely to be affected by inhalation level. We reduced the potential impact by normalizing features by the corresponding value in the contralateral lung. The underlying assumption is that the two lungs are symmetric, changes in inhalation level affect both lungs similarly and that the contralateral lung is unchanged by therapy. These assumptions have limitations. The contralateral lung also receives dose and may be modified by the treatment. The two lungs are not symmetric and the amplitude of movement of the right lung tends to be larger than for the left lung during the breathing cycle. This could introduce systematic differences depending which lung is the ipsilateral. Our data showed no strong evidence of such effect but findings need to be investigated in larger groups. More refined corrections that account for the level of inhalation can potentially be developed if the inhalation level is known for each scan. Furthermore, CT scans from RT patients acquired at the same session but with different (and controlled) levels of inspiration of RT-treated patients are required for accurate assessment of how the anatomical features change due to inhalation level.

Other characteristics of the CT scans parameters and acquisition can also lead to errors in measuring the biomarkers. Breathing motion (or failure to hold breath for the whole acquisition) causes artifacts in images, such as replicas of the diaphragm. Biomarkers of change in lung volume (ΔNV) and anatomical shape (ΔZ, ΔC and Δh) are the most affected. Slice thickness reduces the contrast of structures and hence the detectability of diaphragmatic tenting (ΔS) and airway distortions ($\Delta\alpha$). Tenting cannot be detected properly if the peak height is of the order of the magnitude of the slice thickness. The poorer correlation with visual assessment of $\Delta\alpha$ is likely partially explained by errors in skeletonisation in coarse scans. Exact Hounsfield Units (HU) can vary due to acquisition parameters (for example, are different in low dose scans) and/or differences in inhalation levels. HU thresholds are used in the definition of $\Delta NV, RV$ and ΔP. Finally, while we account for variation in positioning, any residual patient positioning errors affect the measure of distances, as required to define of ΔX, ΔZ and Δh.

In retrospective studies, the quality of imaging data available is variable (such as in our study). High-resolution diagnostic scans are not always available at all time-points, which should not discourage quantitative analysis of RILD. Normalising to the contralateral lung to account for differences in inhalation level is a good approximation in artifact free-images. Meaningful and useful information of RILD in patient groups can still be extracted even if the level of inhalation varies (with some

biomarkers being more affected than others). Variation in inhalation level between scans impacts the uncertainty in biomarker measure with the overall effect of increasing the standard deviation of the biomarkers measured over patient groups. We recommend that prospective studies with a primary focus on RILD request the acquisition of diagnostic quality breath-hold at deep inspiration CT scans and methods to make the inhalation level reproducible at serial time-points. Devices like the active breathing control device should be explored [39] and could potentially remove the need for corrections.

The set of imaging biomarkers is representative but not comprehensive. Additional measures were investigated but those redundant and/or not diagnostic were not included. Measures of change in lung depth were discarded as there was no preferential type of change and hence are likely caused by variation in inhalation level, not RT. Lung maximum and minimum height were investigated, but were strongly correlated with height at the mid-point. There are other forms of RILD that may be investigated in the future. Radiation-induced narrowing of the airways[40] and distortion of the fissures[19] are common radiological findings that were not quantified due to coarse image resolution. The description of parenchymal change is not complete. We have not proposed methods to distinguish different sub-categories (i.e., consolidation, ground-glass opacities, reticulation and traction bronchiectasis), which may help to distinguish inflammatory from fibrotic processes. This will be explored in the future. The biomarkers are also not adequate to quantify RILD in the presence of newly developed lung collapse as this would significantly impair the accurate assessment of the biomarkers.

Our long term aim is to develop this set of biomarkers into a free open-source software that allows for automated analysis of large datasets and enables collaboration with other research groups and hospitals. This will facilitate the widespread use of the biomarkers as a new method of reporting RILD. This still requires solving the technical limitations. The current need for manual segmentations is the main limitation on the translation of the methodology to large datasets. Manual segmentation is time-consuming. In subjects with extensive parenchymal and/or pleural change it can take approximately one hour to segment the lung volumes and thoracic wall. We plan to fully automate the image segmentation steps using deep-learning methods in the future [41]. Other areas of planned work include making the pipelines compatible with high-performance computing, developing benchmarking data, and devising quality assurance protocols of the performance of the methodology in new datasets. This infrastructure will facilitate the development of a common scoring system of RILD. In the interim we are open to share implementations and benchmarking data with groups interested in replicating our methodology. Project updates will be shared at github.com/XXX/XXX/wiki.

A number of potential areas for clinical impact exist. (1) Development of a scoring system for RILD. The biomarkers will facilitate more accurate and consistent reporting across clinical trials to allow comparisons between RT and drug combination schedules. (2) Continuous end-points also open the possibility of improved models for toxicity prediction[24]. Correlation of RILD with RT dose will allow the development of dose constraints that can be used to avoid or reduce the incidence of RILD in the future. (3) Detailed analysis of RILD could be useful as an early indicator of normal tissue radiosensitisation in early phase trials of RT-drug combinations. As RILD is visible in all patients at 12 months it will reduce the number of patients required and the length of follow-up required.

The aim of the present study was to propose and demonstrate a description of lung changes that are apparent and measurable within both symptomatic and asymptomatic patients. These biomarkers will be used in a follow up publication describing in detail CT-based and respiratory function changes of RILD for the XXX study. The clinically critical RT dose-relationship and will be investigated in future research based on these novel biomarkers. Advanced 3D-based analysis methods are required to identify dosimetric predictors due to the isotoxic nature of this patient sub-group[42,43].

5. CONCLUSIONS

We propose a novel set of CT-based imaging biomarkers representative of common radiological findings of RILD. This set of features may be useful in quantifying toxicity caused by new RT techniques, RT-drug combinations and clinical trials. Correlation of RILD with RT dosimetry and respiratory function is facilitated.

FIGURE AND TABLE CAPTIONS

Table 1- Patient characteristics and image acquisition parameters.

Sex (%)			
	Male vs Female		74 vs 26
Age (y)			
	Median (range)		65 (53 – 83)
Staging (%)			
	IIB vs IIIA vs IIIB		4 vs 59 vs 37
Treatment modality (%)			
	Conformal vs IMRT		89 vs 11
Prescription dose (Gy)			
	Median (range)		69.1 (63.0 – 73.0)
Mean lung dose (Gy)			
	Median (range)		14.2 (8.8 – 20.0)
Fractionation scheme (%)			
	6 vs 5 week protocol		63 vs 37
Respiratory function tests at 12-months°			
	Lung diffusing capacity for carbon monoxide (DLCO)		
		Incidence of decline (%)	74
		Relative change (from baseline value) when function declines (Mean ±std) (%)	-20±11
	Forced expiratory volume in 1 second per unit of vital capacity^{+}		
		Incidence of decline (%)	43
		Absolute change when function declines (Mean ±std) (%)	-10±12
Incidence of radiological findings at 12-months (%)			
	Parenchymal change		100
	Volume reduction		96
	Pleural changes		82
Radiological toxicity scoring at 12-months (%)			
	ROTG late radiation morbidity scoring (lung)		
		Grade 2 vs 3	19 vs 81
	CTCAE v4.3 (pulmonary fibrosis)		
		Grade 1 vs 2	89 vs 11
Pre-RT CT characteristics			

	Scan type (no. pts)		
		Non-diagnostic vs Diagnostic*	5 vs 22
		Planning (3D free-breathing or average 4DCT) vs PET/CT (breath-hold)	2 vs 3
	Resolution (mm)		
		Median (range)	0.82×0.82×2.0 (0.61×0.61×0.80 – 1.4×1.4×5.0)
		Coarse vs fine (slice spacing >2 vs ≤2 mm)	48 vs 52%
12-month CT characteristics			
	Scan type (no. pts)		
		Non-diagnostic vs Diagnostic*	0 vs 27
	Resolution (mm)		
		Median (range)	0.83×0.83×2.5 (0.64×0.64×0.80 – 0.98×0.98×5.0)
	Resolution (%)		
		Coarse vs fine (slice spacing >2 vs ≤2 mm)	56 vs 44%
	Time from RT end to follow-up CT (days)		
		Median (range)	353 (265 – 364)
Pre-RT vs 12-month CT characteristics			
	Scan type (%)		
		Both scans diagnostic	81
		Non-diagnostic baseline scan	19
	Resolution		
		Mean difference (±std) (mm)	0.13×0.13×0.91 (±0.13×0.13×1.1)
		Both scans coarse (%)	33
		Both scans fine (%)	30
		One scan coarse, other fine (%)	37

°available for 23 out of the 27 subjects

*Breath-hold at deep inspiration

$^{+}$FEV1%=100×FEV1/FVC

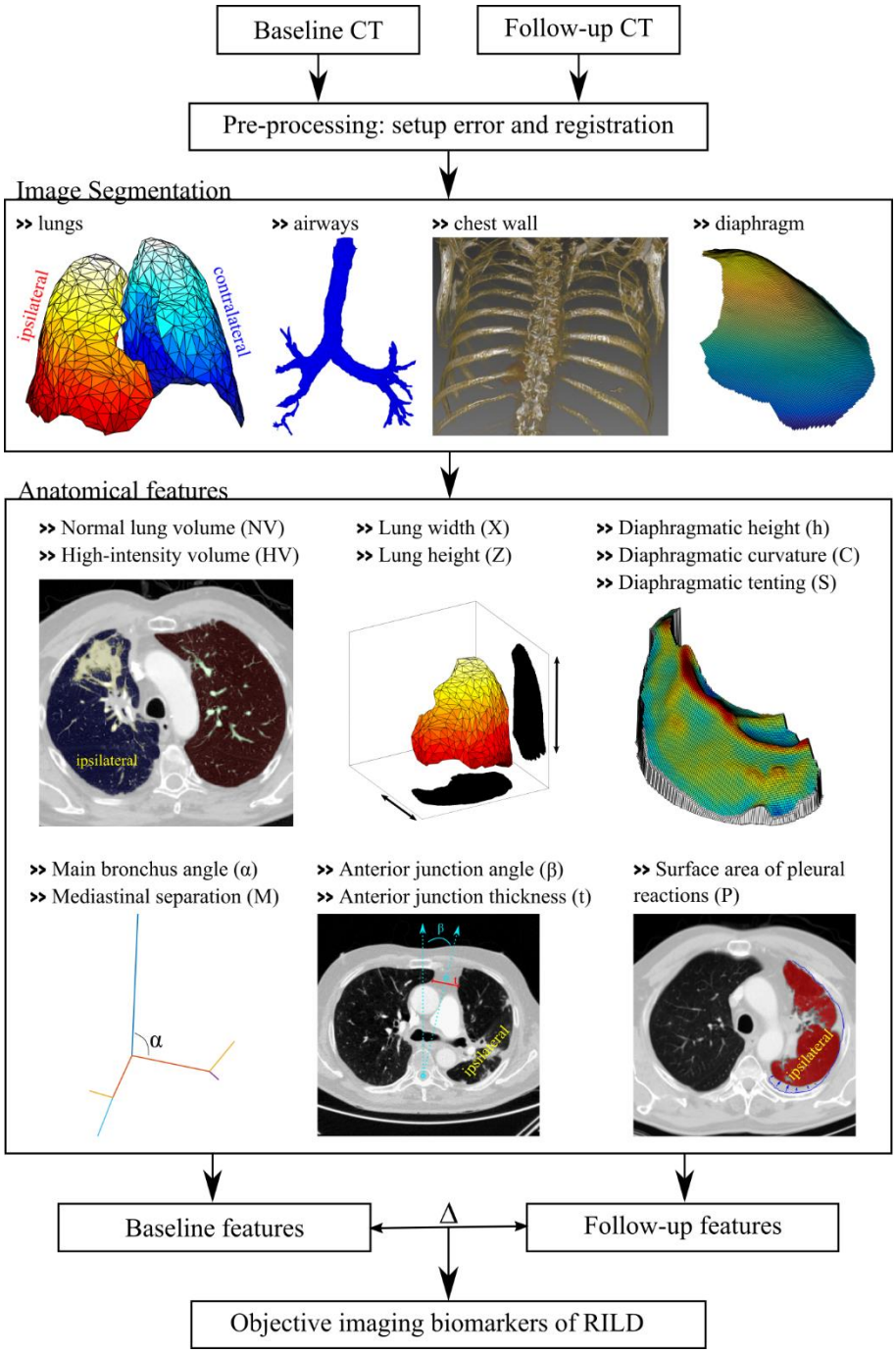

Figure 1- Generalized implementation of the imaging biomarkers.

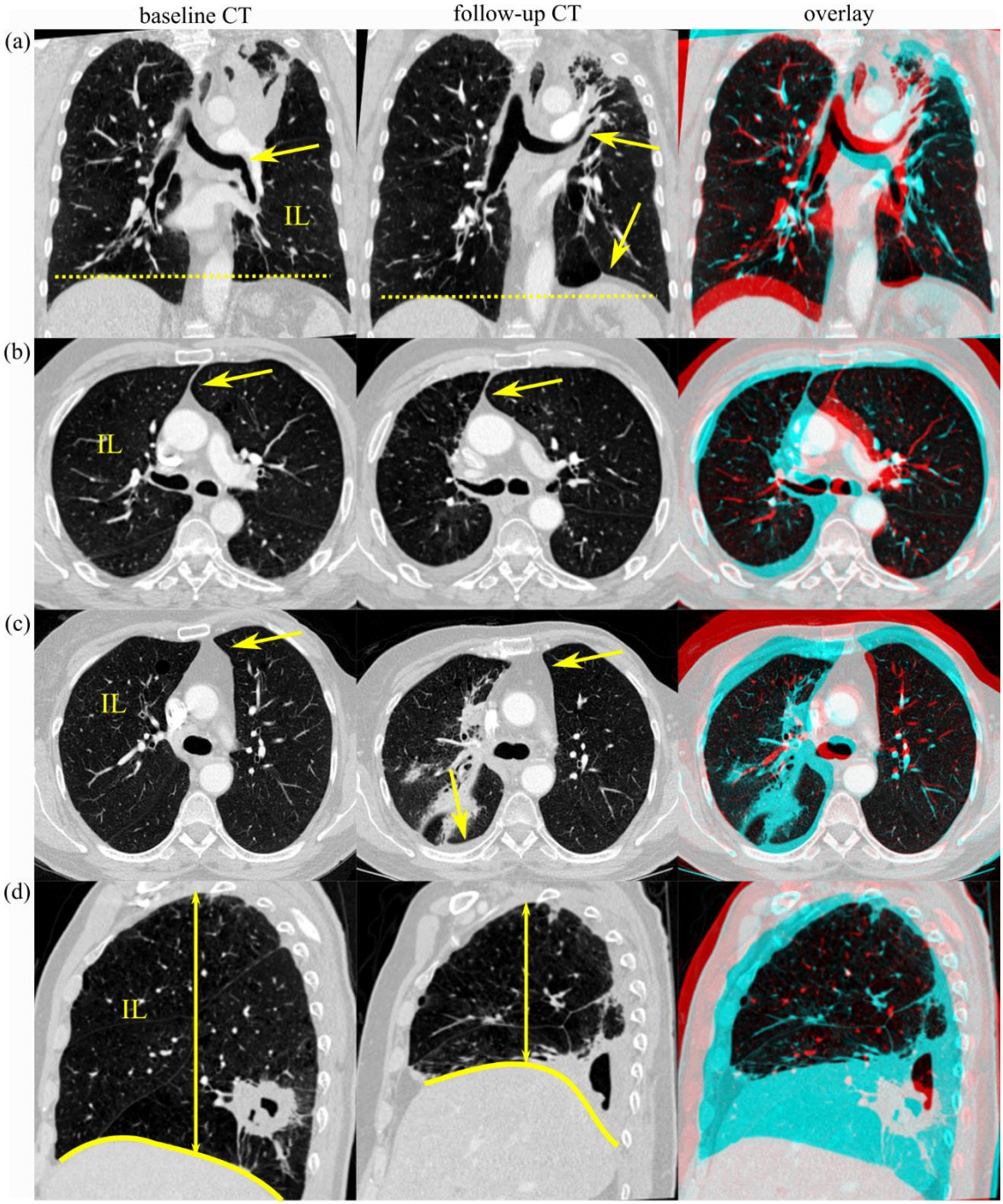

Figure 2- Imaging biomarkers in four subjects (left, baseline; middle; follow-up; right, colour overlay [baseline (red) vs follow-up (cyan)]). Ipsilateral lung (IL) indicated in baseline scan. (a) Lung volume loss with parenchymal change, mediastinal shift, elevation of the major bronchus, elevation of the diaphragm and diaphragmatic tenting ($\Delta NV = 34\%$, $RV = 4.6$, $\Delta M = 31\%$, $\Delta\alpha = 27\%$, $\Delta h = 21mm$, and $\Delta S = 333mm^2$). (b) Normal lung volume loss with reduction in lung width, rotation of the anterior junction line (no thickening), and pleural effusion ($\Delta NV = 36\%$, $RV = 2$, $\Delta X = 23\%$, $\Delta\beta = 10°$, $\Delta t = 1$, and $\Delta P = 21\%$). (c) Parenchymal change with no significant change in lung shape, exhibiting anterior junction line and pleural thickening ($\Delta NV = 17\%$, $RV = 2.5$, $\Delta X = 6\%$, $\Delta Z = 2\%$, $\Delta\beta = 0°$, $\Delta t = 2.5$,

and $\Delta P = 18\%$). (d) Lung volume loss with diaphragmatic elevation and increase in diaphragm curvature ($\Delta NV = 62\%$, $\Delta h = 45mm$, and $\Delta C = 32\%$).

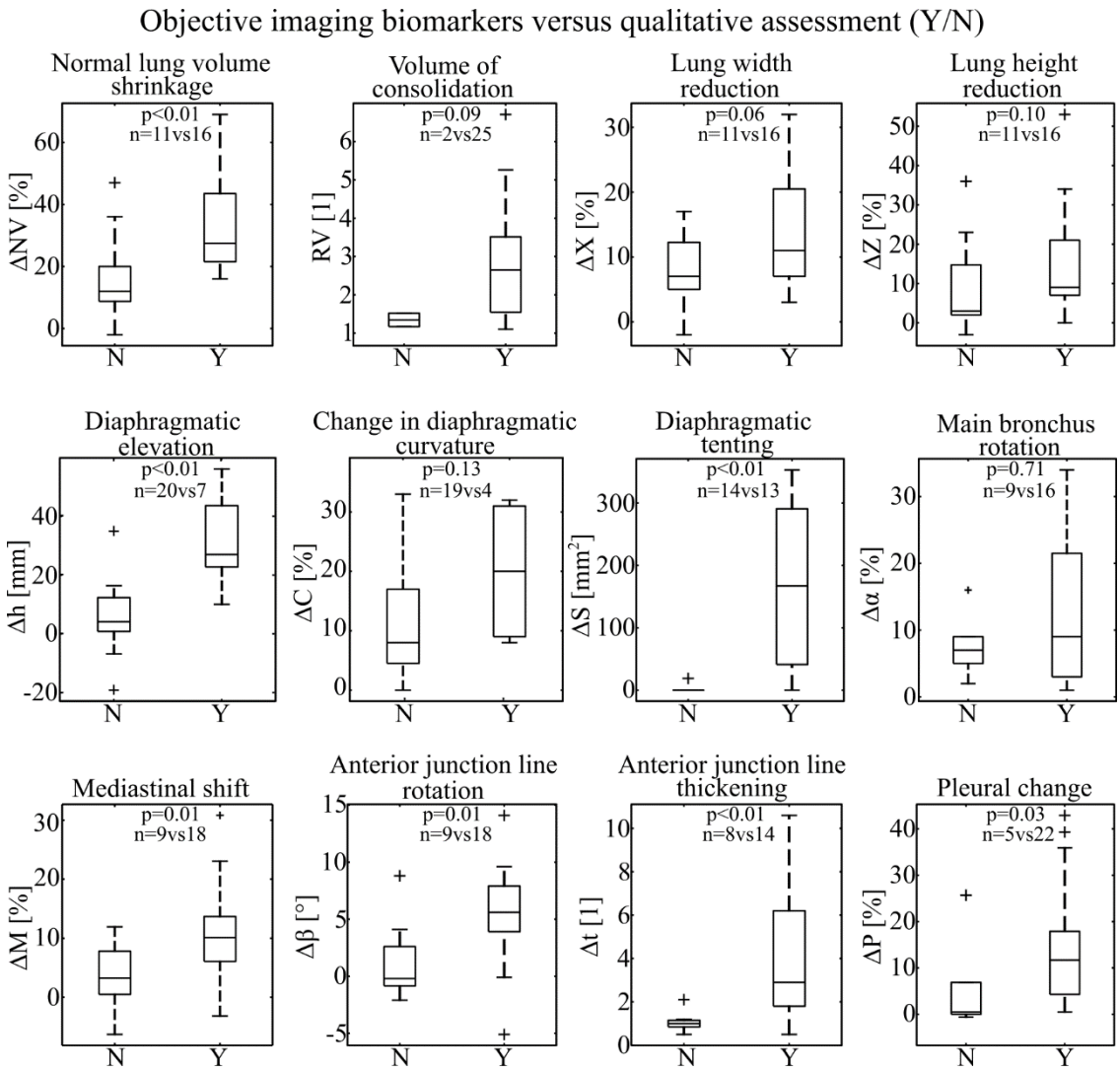

Figure 3- Boxplot of the distribution of values per imaging biomarker with the patient group split into two groups: presence (Y) or absence (N) of that radiological feature. Y/N incidence (number of subjects, n) and p-values (Wilcoxon rank sum test) detailed for each biomarker. Outliers fall outside the 99.3% coverage.

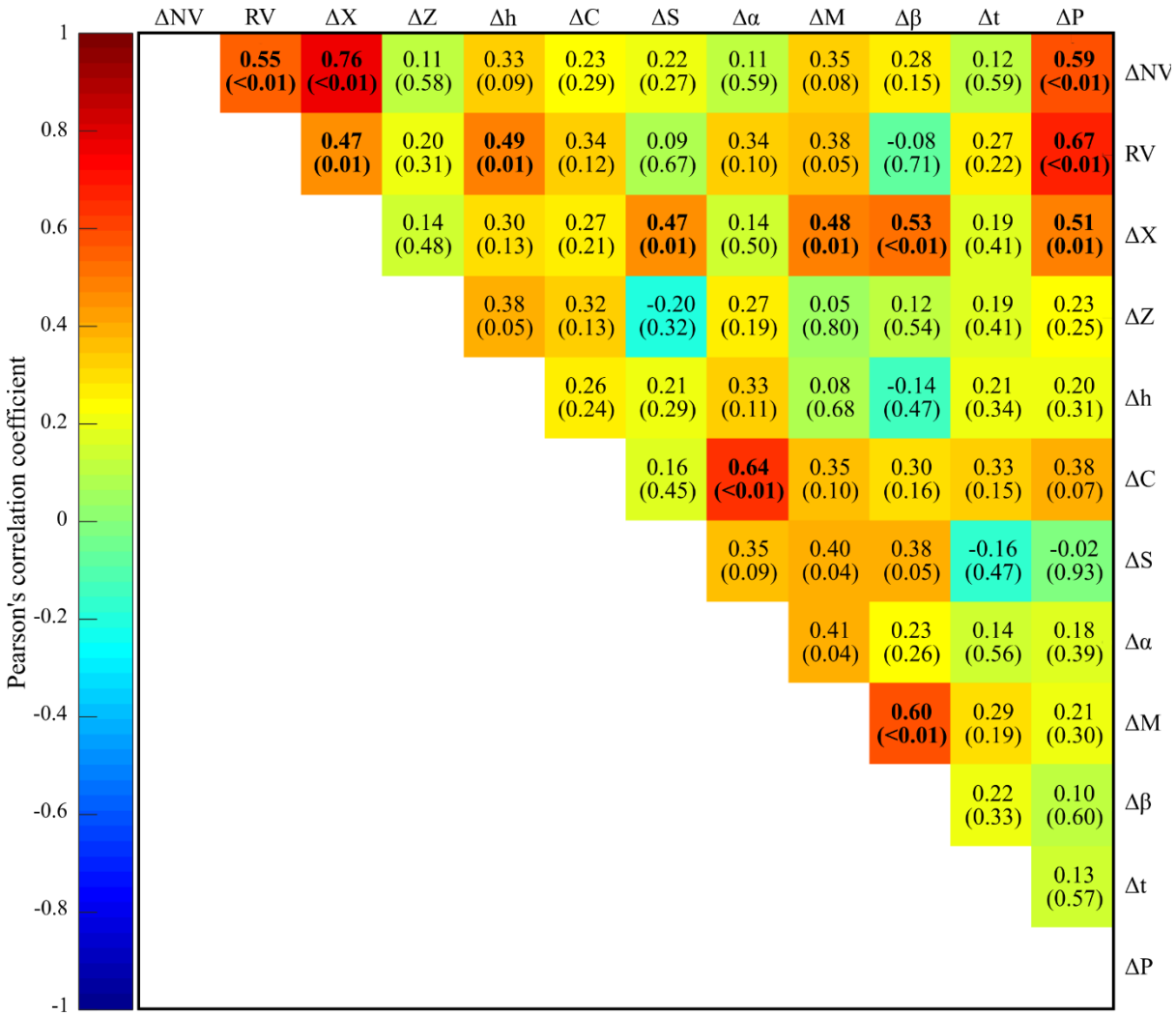

Figure 4- Heat map of Pearson's correlation coefficient between each pair of imaging biomarkers (p-values in brackets). Statistically significant results after applying the Benjamini-Hochberg procedure (10% false discovery rate) in bold (critical p-value=0.013).

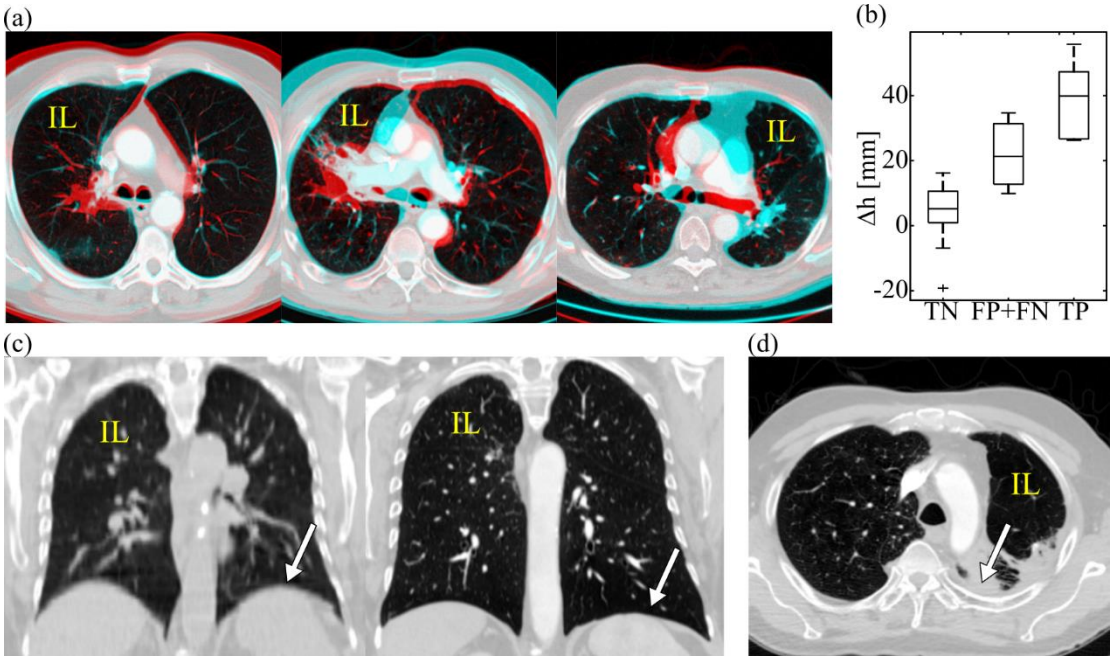

Figure 5- Advantages and short comings of the imaging biomarkers. Ipsilateral lung (IL) indicated in the CT scans. (a) Colour overlay of baseline (red) and follow-up (cyan) CT scans for increasing severity of rotation of the anterior junction line (from left to right, $\Delta\beta = \{0,3,14\}°$). (b) Boxplot of the distribution of values of diaphragmatic elevation (Δh) with the patient group split into three groups: true negative (TN), true positives (TP) and false positives and negatives (FP+FN). TN and TP correspond to cases where biomarkers and visual assessment agree; FP+FN to cases of disagreement. (c) Baseline (left) versus 12-month scan (right) where direct comparison of the curvature of the diaphragm between time-points would indicate a severe decrease in curvature. However, similar changes in curvature also occurred in the contralateral side (arrow) so it is more likely due to differences in inhalation level than RT. This was qualitatively scored as becoming flatter at 12-months while the imaging marker indicates increase in curvature ($\Delta C = 8\%$). The disagreement occurs due to the normalization to the contralateral side implemented with the imaging biomarkers. (d) The definition of the imaging biomarkers, is dependent on the accuracy of segmentations. In this 12-month CT scan, separating parenchymal change from pleural thickening is complex because of consolidation in the vicinity of the chest wall (arrow).

REFERENCES

[1] Defraene G, van Elmpt W, Crijns W, Slagmolen P, De Ruysscher D. CT characteristics allow identification of patient-specific susceptibility for radiation-induced lung damage. Radiother Oncol 2015;117:29–35. doi:10.1016/j.radonc.2015.07.033.

[2] Curran WJ, Paulus R, Langer CJ, Komaki R, Lee JS, Hauser S, et al. Sequential vs. concurrent chemoradiation for stage III non-small cell lung cancer: randomized phase III trial RTOG 9410. J Natl Cancer Inst 2011;103:1452–60. doi:10.1093/jnci/djr325.

[3] Simone CB. Thoracic Radiation Normal Tissue Injury. Semin Radiat Oncol 2017;27:370–7. doi:10.1016/j.semradonc.2017.04.009.

[4] Lopez Guerra JL, Gomez DR, Zhuang Y, Levy LB, Eapen G, Liu H, et al. Changes in Pulmonary Function after Three-Dimensional Conformal Radiation Therapy, Intensity-Modulated Radiation Therapy, or Proton Beam Therapy for Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2012;83:e537–43. doi:10.1016/j.ijrobp.2012.01.019.

[5] Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. Lancet Oncol 2015;16:187–99. doi:10.1016/S1470-2045(14)71207-0.

[6] XXX.

[7] Machtay M, Bae K, Movsas B, Paulus R, Gore EM, Komaki R, et al. Higher biologically effective dose of radiotherapy is associated with improved outcomes for locally advanced non-small cell lung carcinoma treated with chemoradiation: an analysis of the Radiation Therapy Oncology Group. Int J Radiat Oncol Biol Phys 2012;82:425–34. doi:10.1016/j.ijrobp.2010.09.004.

[8] Maguire J, Khan I, McMenemin R, O'Rourke N, McNee S, Kelly V, et al. SOCCAR: A randomised phase II trial comparing sequential versus concurrent chemotherapy and radical hypofractionated radiotherapy in patients with inoperable stage III Non-Small Cell Lung Cancer and good performance status. Eur J Cancer Oxf Engl 1990 2014;50:2939–49. doi:10.1016/j.ejca.2014.07.009.

[9] Ikezoe J, Takashima S, Morimoto S, Kadowaki K, Takeuchi N, Yamamoto T, et al. CT appearance of acute radiation-induced injury in the lung. AJR Am J Roentgenol 1988;150:765–70. doi:10.2214/ajr.150.4.765.

[10] Choi YW, Munden RF, Erasmus JJ, Park KJ, Chung WK, Jeon SC, et al. Effects of radiation therapy on the lung: radiologic appearances and differential diagnosis. Radiogr Rev Publ Radiol Soc N Am Inc 2004;24:985–997; discussion 998. doi:10.1148/rg.244035160.

[11] Linda A, Trovo M, Bradley JD. Radiation injury of the lung after stereotactic body radiation therapy (SBRT) for lung cancer: a timeline and pattern of CT changes. Eur J Radiol 2011;79:147–54. doi:10.1016/j.ejrad.2009.10.029.

[12] Davis SD, Yankelevitz DF, Wand A, Chiarella DA. Juxtaphrenic peak in upper and middle lobe volume loss: assessment with CT. Radiology 1996;198:143–9. doi:10.1148/radiology.198.1.8539368.

[13] Ghafoori P, Marks LB, Vujaskovic Z, Kelsey CR. Radiation-induced lung injury. Assessment, management, and prevention. Oncol Williston Park N 2008;22:37-47; discussion 52-53.

[14] Iyer R, Jhingran A. Radiation injury: imaging findings in the chest, abdomen and pelvis after therapeutic radiation. Cancer Imaging 2006;6:S131–9. doi:10.1102/1470-7330.2006.9095.

[15] Miller KL, Shafman TD, Anscher MS, Zhou S-M, Clough RW, Garst JL, et al. Bronchial stenosis: an underreported complication of high-dose external beam radiotherapy for lung cancer? Int J Radiat Oncol Biol Phys 2005;61:64–9. doi:10.1016/j.ijrobp.2004.02.066.

[16] Epstein DM, Littman P, Gefter WB, Miller WT, Raney RB. Radiation-induced pneumothorax. Med Pediatr Oncol 1983;11:122–4.

[17] Larici AR, del Ciello A, Maggi F, Santoro SI, Meduri B, Valentini V, et al. Lung Abnormalities at Multimodality Imaging after Radiation Therapy for Non–Small Cell Lung Cancer. RadioGraphics 2011;31:771–89. doi:10.1148/rg.313105096.

[18] Mah K, Poon PY, Van Dyk J, Keane T, Majesky IF, Rideout DF. Assessment of acute radiation-induced pulmonary changes using computed tomography. J Comput Assist Tomogr 1986;10:736–43.

[19] XXX.

[20] Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). Int J Radiat Oncol Biol Phys 1995;31:1341–6. doi:10.1016/0360-3016(95)00060-C.

[21] National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) 4.03 2009.

[22] Mazeron R, Etienne-Mastroianni B, Pérol D, Arpin D, Vincent M, Falchero L, et al. Predictive factors of late radiation fibrosis: a prospective study in non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2010;77:38–43. doi:10.1016/j.ijrobp.2009.04.019.

[23] Heo J, Cho O, Noh OK, Oh Y-T, Chun M, Kim M-H, et al. CT-based quantitative evaluation of radiation-induced lung fibrosis: a study of interobserver and intraobserver variations. Radiat Oncol J 2014;32:43–7. doi:10.3857/roj.2014.32.1.43.

[24] Niedzielski JS, Yang J, Stingo F, Martel MK, Mohan R, Gomez DR, et al. Objectively Quantifying Radiation Esophagitis With Novel Computed Tomography–Based Metrics. Int J Radiat Oncol • Biol • Phys 2016;94:385–93. doi:10.1016/j.ijrobp.2015.10.010.

[25] Jeraj R, Cao Y, Ten Haken RK, Hahn C, Marks L. Imaging for assessment of radiation-induced normal tissue effects. Int J Radiat Oncol Biol Phys 2010;76:S140-144. doi:10.1016/j.ijrobp.2009.08.077.

[26] Gotway MB, Reddy GP, Webb WR, Elicker BM, Leung JWT. High-resolution CT of the lung: patterns of disease and differential diagnoses. Radiol Clin North Am 2005;43:513–542, viii. doi:10.1016/j.rcl.2005.01.010.

[27] Abbena E, Salamon S, Gray A. Modern Differential Geometry of Curves and Surfaces with Mathematica, Third Edition. 3 edition. Boca Raton, FL: Chapman and Hall/CRC; 2006.

[28] Lee J-R. The law of cosines in a tetrahedron. Pure Appl Math 1997;4:1–6.

[29] Ourselin S, Roche A, Subsol G, Pennec X, Ayache N. Reconstructing a 3D structure from serial histological sections. Image Vis Comput 2001;19:25–31. doi:10.1016/S0262-8856(00)00052-4.

[30] Doel T. Developing clinical measures of lung function in COPD patients using medical imaging and computational modelling 2012.

[31] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 2006;31:1116–28. doi:10.1016/j.neuroimage.2006.01.015.

[32] Modat M, Daga P, Cardoso MJ, Ourselin S, Ridgway GR, Ashburner J. Parametric non-rigid registration using a stationary velocity field. 2012 IEEE Workshop Math. Methods Biomed. Image Anal. MMBIA, 2012, p. 145–50. doi:10.1109/MMBIA.2012.6164745.

[33] Palágyi K, Tschirren J, Hoffman EA, Sonka M. Quantitative analysis of pulmonary airway tree structures. Comput Biol Med 2006;36:974–96. doi:10.1016/j.compbiomed.2005.05.004.

[34] Palma DA, van Sörnsen de Koste J, Verbakel WFAR, Vincent A, Senan S. Lung Density Changes After Stereotactic Radiotherapy: A Quantitative Analysis in 50 Patients. Int J Radiat Oncol 2011;81:974–8. doi:10.1016/j.ijrobp.2010.07.025.

[35] Ruysscher DD, Sharifi H, Defraene G, Kerns SL, Christiaens M, Ruyck KD, et al. Quantification of radiation-induced lung damage with CT scans: The possible benefit for radiogenomics. Acta Oncol 2013;52:1405–10. doi:10.3109/0284186X.2013.813074.

[36] Ghobadi G, Wiegman EM, Langendijk JA, Widder J, Coppes RP, van Luijk P. A new CT-based method to quantify radiation-induced lung damage in patients. Radiother Oncol J Eur Soc Ther Radiol Oncol 2015;117:4–8. doi:10.1016/j.radonc.2015.07.017.

[37] XXX

[38] Oh Y-T, Noh OK, Jang H, Chun M, Park KJ, Park KJ, et al. The features of radiation induced lung fibrosis related with dosimetric parameters. Radiother Oncol J Eur Soc Ther Radiol Oncol 2012;102:343–6. doi:10.1016/j.radonc.2012.02.003.

[39] Kaza E, Dunlop A, Panek R, Collins D, Orton M, Symonds-Tayler R, et al. Lung volume reproducibility under ABC control and self-sustained breath-holding. J Appl Clin Med Phys 2017.;18:154–62.

[40] Kelsey CR, Kahn D, Hollis DR, Miller KL, Zhou S-M, Clough RW, et al. Radiation-induced narrowing of the tracheobronchial tree: an in-depth analysis. Lung Cancer Amst Neth 2006;52:111–6. doi:10.1016/j.lungcan.2005.11.007.

[41] Jiang F, Grigorev A, Rho S, Tian Z, Fu Y, Jifara W, et al. Medical image semantic segmentation based on deep learning. Neural Comput Appl 2017:1–9. doi:10.1007/s00521-017-3158-6.

[42] Vivekanandan S, Landau DB, Counsell N, Warren DR, Khwanda A, Rosen SD, et al. The Impact of Cardiac Radiation Dosimetry on Survival After Radiation Therapy for Non-Small Cell Lung Cancer. Int J Radiat Oncol 2017;99:51–60. doi:10.1016/j.ijrobp.2017.04.026.

[43] Chen C, Witte M, Heemsbergen W, Herk M van. Multiple comparisons permutation test for image based data mining in radiotherapy. Radiat Oncol 2013;8:293. doi:10.1186/1748-717X-8-293.