**Oregon Health & Science University**

**School of Medicine-Graduate Studes**

# Longitudinal Analysis of Functional Effort Correction to Assess Lung Damage after Radiation Therapy

By

Justin Louis Reel

A Thesis

Presented to the Department of Medical Physics
and the Oregon Health and Science University
School of Medicine
in partial fulfillment of the
requirements for the degree of

Masters of Science

June 2025

# Contents

# Acknowledgements

I must acknowledge my boundless gratitude to my dear Shannon, my partner in all things, who suffers my quixotic ambitions with patience and grace.

My thanks to Dr. Lindsay DeWeese, Dr. Anna Mench and Dr. Malcom Heard who excel at balancing a thoughtful prodding with a kind word.

This project wouldn't have been possible without Will Bethard, the man with the plan. His guidance was essential, and his passion inspiring.

Finally, my thanks and appreciation to Dr. John Bayouth. The important work he does to improve the lives of lung cancer patients is deeply meaningful. He is generous with his time and advice, motivating in lectures, and serves as a guide to my future endeavors.

# List of Abbreviations

| | |
|---|---|
| 4DCT | Four-Dimensional Computed Tomography |
| B-spline | Basis Spline (used in deformable image registration) |
| CT | Computed Tomography |
| DIR | Deformable Image Registration |
| DICOM | Digital Imaging and Communications in Medicine |
| ELV | Equivalent Lung Volume |
| EPS | Expanded Polystyrene |
| ETV | Equivalent Tidal Volume |
| FBP | Filtered Back Projection |
| FEC | Functional Effort Correction |
| FEC1 | Functional Effort Correction using fixed 0.75 L extrapolation |
| FEC2 | Functional Effort Correction using 40% lobe tidal volume |
| FLART | Functional Lung Avoidance Radiation Therapy |
| HFL | High-Functional Lung |
| Hp-MRI | Hyperpolarized Magnetic Resonance Imaging |
| IN | Inhale (phase label, e.g., 100IN) |
| Jacobian | Jacobian Determinant (of the deformation field) |
| LERN | Local Expansion Ratio over N phases |
| LFL | Low-Functional Lung |
| LLL | Left Lower Lobe |
| LUL | Left Upper Lobe |
| MRI | Magnetic Resonance Imaging |
| MVV | Maximal Voluntary Ventilation |
| NSCLC | Non-Small Cell Lung Cancer |
| PET | Positron Emission Tomography |
| PFT | Pulmonary Function Test |

| | |
|---|---|
| RILI | Radiation-Induced Lung Injury |
| RLL | Right Lower Lobe |
| RML | Right Middle Lobe |
| RUL | Right Upper Lobe |
| SPECT | Single-Photon Emission Computed Tomography |
| TV | Tidal Volume |
| TPS | Treatment Planning System |
| V/Q | Ventilation/Perfusion |
| VIDA | VIDA Diagnostics (software/tool for lung segmentation) |

# Abstract

This work investigates the application and refinement of Functional Effort Correction (FEC) methods in the longitudinal assessment of radiation-induced lung damage using four-dimensional computed tomography (4DCT) images. Traditional ventilation imaging approaches such as Equivalent Tidal Volume (ETV) and Equivalent Lung Volume (ELV) attempt to correct for inter-scan variability in respiratory effort, but may also normalize damaged regions, masking radiation induced changes. FEC aims to overcome this limitation by identifying a stable, minimally irradiated and unaffected lung subregion and using it as a physiologic reference point for normalization.

The work compares multiple implementations of FEC (FEC1 and FEC2) with ETV and ELV. The FEC1 approaches applies a fixed extrapolation volume of 0.75 L scaled to lung-wide volume ratios, while FEC2 calculates effort variation using 40% of the tidal volume of the selected lobe. Both approaches aim to ensure that pre- and post-treatment Jacobian maps represent comparable physiologic states, enhancing sensitivity to real changes in lung function.

The study uses LERN (Local Expansion Ratio over N phases) as a voxel-level metric of ventilation derived from Jacobian determinants over the full respiratory cycle. Lung damage is assessed based on LERN ratios and thresholds (e.g., $< 0.94$), allowing for spatial and temporal quantification. The dataset includes multiple timepoints, pre-treatment, and 3, 6, and 12 months post-treatment, and was drawn from an earlier clinical trial. Rigorous resampling, deformable image registration, and dose alignment were used to ensure voxel consistency across methods.

Key findings include that FEC2 yields lower variability and smaller estimates of damage compared to FEC1, likely due to its more conservative interpolation limits. Statistically significant differences were observed between FEC1 and both ELV and FEC2, particularly in the most irradiated lobe. However, FEC2 showed no significant difference from ETV or ELV, suggesting that it may be more consistent but less sensitive. No consistent temporal pattern in damage progression was observed across the 3, 6, and 12 month timepoints, possibly due to segmentation issues, imaging artifacts, or insufficient statistical power.

Despite lacking conclusive superiority for FEC, the study demonstrates the importance of effort correction in ventilation imaging and provides a framework for future research. Custom Python scripts and a modular data pipeline were developed for image processing and analysis. Limitations include variability in segmentation accuracy, image quality in 4DCT vs. diagnostic CT, and exclusion of subjects due to artifacts. The work concludes that further refinement of segmentation techniques, broader subject inclusion, and investigation of sublobar normalization may enhance the utility of FEC in functional imaging-guided radiotherapy.

# 1. INTRODUCTION

Lung cancer impacts thousands every year and remains one of the most lethal forms of cancer. Radiation therapy is often an effective treatment, but it can lead to serious side effects such as radiation pneumonitis and pulmonary fibrosis. Efforts to mitigate these toxicities include directing radiation away from high-functioning areas of the lung, a strategy known as functional lung avoidance. To improve such approaches, it is necessary to accurately identify and quantify radiation-induced lung injury.

Four-Dimensional Computed Tomography (4DCT) imaging holds promise in this regard through the use of ventilation biomarkers derived from deformable image registration. However, patients exhibit natural variability in breathing effort, which introduces uncertainty into ventilation measurements and can obscure true functional changes. Several methods have been proposed to correct for this variability, including Equivalent Tidal Volume (ETV), Equivalent Lung Volume (ELV), and more recently, Functional Effort Correction (FEC).

This work examines the FEC method under different parameterizations and compares its performance over time using data from a prior clinical trial. By analyzing its limits and potential, this study aims to improve the assessment of lung function after radiation and ultimately inform treatment planning to reduce pulmonary toxicity.

# 2. BACKGROUND

## 2.1 Introduction to Lung Function and Injury in Radiation Therapy

Lung cancer remains the leading cause of cancer related death worldwide, accounting for approximately 1.8 million deaths annually according to the World Health Organization.[1] In the United States alone, over 230,000 new cases of lung cancer are diagnosed each year, with non-small cell lung cancer (NSCLC) comprising about 85% of cases. [2] Local tumor control through radiation is an essential component for many patients with inoperable diseases. Treatment, however, is complicated by the risk of damaging healthy lung tissue, which may result in debilitating or even fatal toxicity. As survivorship improves and more patients undergo curative radiation treatment, the need to preserve pulmonary function remains a priority in thoracic oncology.

Radiation therapy is a cornerstone in the treatment of non-small cell lung cancer (NSCLC), but its curative potential is often constrained by radiation-induced lung injury (RILI). Most significant among these injuries are radiation pneumonitis and pulmonary fibrosis. [3] Historically, the lungs were modeled as functionally uniform in treatment planning, despite evidence of substantial regional variability in ventilation and perfusion. [4] By considering variable functions of lung components treatment planners can utilize functional lung avoidance strategies, in which high functioning regions are preferentially spared from high radiation doses. This preserves more lung function by reducing radiation damage to those areas.

Advancements in imaging technology allow for improved identification of functional region and radiation damage. In particular, four dimensional computed tomography (4DCT) enables imaging of lung function via deformable image registration (DIR), wherein voxel-level expansion can be quantified using the Jacobian determinant. The Jacobian determinant is a scalar value that represents the amount of compression or expansion for a particular volume element in the lung. This approach has proven to be robust and clinically promising for defining ventilation maps that inform functional dose metrics such as functional mean lung dose. [5]

Nevertheless, longitudinal use of these biomarkers is limited by inter-scan variability in breathing effort. To address this, Bethard introduced a functional effort correction (FEC) method, which normalizes voxel-level Jacobian values using sublobar regression models fit to normal expansion curves, improving consistency and enhancing sensitivity to therapy-induced changes. [4] The FEC approach is being tested and revised and is the primary subject of this work. By correcting for breathing effort a clearer indication of radiation damage can be drawn and a better treatment planning steps can be made to minimize RILI.

## 2.2 Pulmonary Anatomy and Functional Heterogeneity

The lungs are anatomically divided into five lobes, three in the right lung and two in the left. Figure 1 illustrates these lobes, separated by visible fissures, and major branches of the airway tree. Fissures are composed of elastin and collagen fibers. Each lobe may be further subdivided into bronchopulmonary segments. These segments are defined by the tertiary bronchi and represent functionally discrete units used in both surgical planning and physiologic modeling. [6-7]

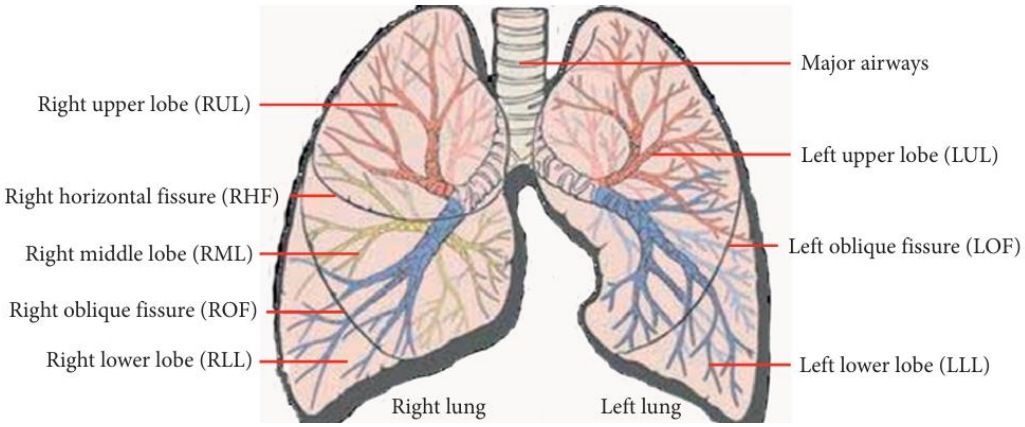


**Figure 1**: General pulmonary anatomy identifying lobes and fissures with airway distribution. Reprinted with permission from Flakus. [3]

These anatomical boundaries are not merely structural, they map onto differences in regional compliance, airway resistance, and ventilation.[8] Throughout this work, the term sublobe is used to describe tertiary bronchial segments. Twenty sublobe designations are shown in Figure 2 in mode lateral and medial views.

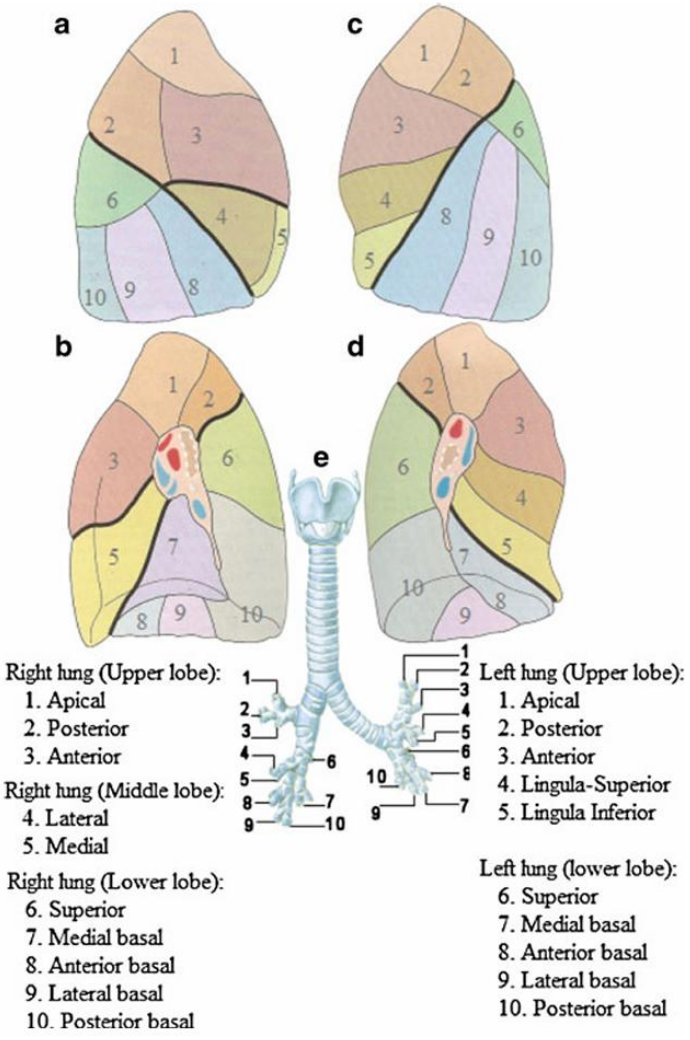


**Figure 2:** Diagram of lung airways and sublobe. Kabir [9]

Lung function is highly heterogeneous, influenced by patient posture, gravitational gradients, airway tree architecture, and underlying pathology. Gravity-dependent perfusion and ventilation gradients result in higher function in dependent (posterior and basal) regions when supine [8]. Disease processes such as fibrosis or emphysema further exacerbate spatial variation, often disproportionately affecting specific lobes or segments [10]. These considerations challenge the traditional paradigm of treating the lung as a homogeneous organ in radiation therapy.

To resolve this, recent approaches incorporate anatomical segmentation with functional imaging. Bethard, for example, used VIDA Diagnostics tools and a combination of virtual bronchoscopy and manual editing to segment sublobes based on third-generation bronchi, an approach grounded in the structure described by Weibel and others. These sublobes serve as the spatial framework for functional effort correction and longitudinal comparison. Similarly, Flakus emphasized the role of regional anatomy in refining dose-function models and improving predictive accuracy of functional decline. The convergence of anatomical and functional models offers a foundation for personalized radiotherapy, where both dose and risk are allocated based on regional function.

## 2.3 Quantifying Lung Function

Quantitative assessment of lung function is essential in the diagnosis and treatment planning for pulmonary diseases, particularly when balancing tumor control with preservation of healthy lung tissue. Lung disease is often diagnosed with pulmonary function tests(PFT) that include spirometry, and assessments of Total Lung Capacity (TLC), Residual Volumes (RV), diffusing capacity, and Maximal Voluntary Ventilation (MVV).  among others.  These provide informative metrics but lack spatial resolution and sensitivity to regional damage.[11] To address this, several imaging modalities have been developed that enable visualization and quantification of regional lung function.

Single-photon emission computed tomography (SPECT) and positron emission tomography (PET) offer functional imaging based on radiotracer distribution. These modalities can map regional ventilation and perfusion and are widely used for detecting ventilation-perfusion mismatch, especially in chronic obstructive pulmonary disease (COPD) and pulmonary

embolism[12]. Hyperpolarized gas MRI (Hp-MRI), using helium-3 or xenon-129, provides high-resolution, non-ionizing functional imaging of ventilation patterns, though it remains largely limited to research settings due to equipment and isotope availability. [13] Ultrasound is a common imaging modality used for pleural assessments of effusions or pneumothorax but is poorly suited to measure internal parenchymal function. Computed Tomography (CT) uses multiple X-ray projections to reconstruct a three-dimensional image with high spatial resolution, surpassing many other imaging modalities. However, imaging of the lungs is complicated by respiratory motion, which introduces artifacts and blurring. This limitation is addressed in four-dimensional CT (4DCT), which acquires images over time throughout the breathing cycle. The resulting data are retrospectively sorted into respiratory phases, allowing reconstruction of motion-resolved image sets. These phase-specific volumes enable visualization and quantification of lung function across the respiratory cycle.

CT creates cross-sectional images of the body by rotating an x-ray source and detector array around the patient, collecting multiple projection images at different angles. These projections reflect x-ray attenuation, which arises primarily from Compton scattering and photoelectric absorption in tissue. A sinogram is generated from the set of projections, and cross-sectional images are reconstructed using algorithms such as filtered back projection or iterative reconstruction. The resulting volumes have excellent spatial resolution, making CT a mainstay of thoracic imaging and radiation therapy planning.

During a 4DCT scan, the patient breathes freely while respiratory motion is tracked using an external surrogate, such as infrared markers, pressure belts, spirometers, or surface-tracking systems. The CT data are retrospectively sorted into discrete respiratory phases, typically ten evenly spaced bins ranging from full exhale (0EX) to full inhale (100IN). Each phase is

reconstructed into its own 3D volume, yielding a time-resolved set of images that captures anatomical deformation throughout the breathing cycle. Figure 3 depicts aspects of tracking a breathing cycle, visualizing changes in chest movements and scan progression through regions of the lung.

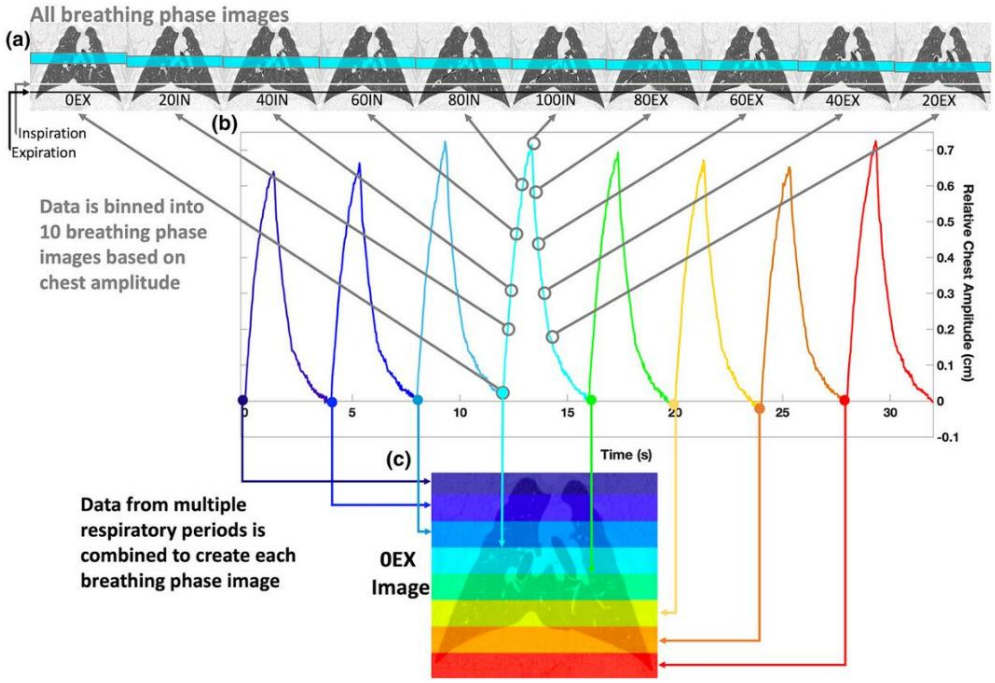


**Figure 3:** Helical 4DCT acquisition. (A) Ten reconstructed breathing phase images are shown with white and black lines across all images identifying full exhale (0EX) and inhale (100IN) diaphragm positions, respectively. (B) A respiratory trace with 8 periods (each a unique color) is shown. Due to helical acquisition, the corresponding anatomical region is slightly inferior in each subsequent phase image. (C) Individual breathing phase images are generated from data acquired over multiple respiratory periods. Reprinted with permission from [3].

This dataset enables the calculation of regional lung ventilation using deformable image registration (DIR). [4] DIR computes a voxel-wise displacement field between two respiratory phases, most commonly from 0EX to a selected inhale phase. A B-spline transformation model is typically used, optimized via similarity metrics that incorporate intensity differences and spatial regularization. [14]From this deformation field, the Jacobian determinant is derived at each voxel, representing local volume change. A Jacobian of 1 indicates no change, values >1 indicate expansion, and values <1 indicate contraction. The Jacobian thus serves as a surrogate for ventilation, forming the basis of the Local Expansion Ratio (LER) maps, discussed below.

Although Jacobian ventilation mapping provides high-resolution spatial information, it is sensitive to variability in respiratory effort between imaging sessions. If the tidal volume (TV)—the difference in lung volume between exhale and inhale—varies across longitudinal scans, the resulting ventilation maps may reflect effort changes rather than physiological alterations. To address this, an Equivalent Tidal Volume (ETV) matching approach is employed. [18] This involves identifying inhale phases in both scans that produce similar global TV when paired with a common exhale phase (e.g., 0EX). Selecting these matched pairs improves the repeatability of Jacobian-based ventilation estimates and enhances the detection of true biological changes. [4]

The major advantage of 4DCT-based ventilation imaging is its integration into existing clinical workflows. Because 4DCT is often acquired for radiation treatment planning, it enables functional imaging without additional scanning, cost, or patient dose. This makes it an attractive modality for developing functionally guided radiotherapy strategies, such as dose painting or functional avoidance, where spatial knowledge of lung function can inform treatment decisions.

## 2.4 Local Expansion Ratio

In 4D computed tomography (4DCT) ventilation imaging, the Jacobian determinant is a widely recognized estimate for regional lung ventilation. It is derived by performing deformable image registration (DIR) between two phases of the breathing cycle. The Jacobian determinant is a scalar-valued function of the Jacobian matrix. The vector matrix is comprised of first order partial differential equations. The term "Jacobian" is often used to refer both to the matrix and to its determinant. Equation 1 describes the Jacobian matrix in the present context of partial differential equations for physical space.

$$J = \begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} & \frac{\partial x'}{\partial z} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} & \frac{\partial y'}{\partial z} \\ \frac{\partial z'}{\partial x} & \frac{\partial z'}{\partial y} & \frac{\partial z'}{\partial z} \end{bmatrix} = \begin{bmatrix} 1 + \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & 1 + \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & 1 + \frac{\partial u_z}{\partial z} \end{bmatrix} \tag{1}$$

The determinant of this provides then a scalar value representing the amount of local expansion or contraction, at the voxel level, between lung scans that are registered to each other. A mapping of these values provides a way to assess lung damage. Impaired lung tissue generally shows a loss of elasticity which negatively impacts ventilation and gas oxygenation. The expression for Jacobian determinant is given by

$$det(J) = (1 + u\_\{x,x\})((1 + u\_\{y,y\})(1 + u\_\{z,z\}) - u\_\{y,z\}\, u\_\{z,y\})$$

$$- u\_\{x,y\}(u\_\{y,x\}(1 + u\_\{z,z\}) - u\_\{z,x\}\, u\_\{y,z\}) \quad (2)$$

$$+ u\_\{x,z\}(u\_\{y,x\}\, u\_\{z,y\} - (1 + u\_\{y,y\})\, u\_\{z,x\})$$

This technique enables voxel-level quantification of tissue deformation and has become a common surrogate for ventilation imaging in radiation therapy planning and functional lung assessment. [15]

Two phase Jacobian methods assume that each region of the lung reaches maximum expansion and minimum contraction at the same global time points, an assumption often violated due to regional ventilation heterogeneity and asynchronous breathing mechanics. [16] As a result, errors in ventilation estimation can arise, particularly in diseased lungs or in patients with irregular respiratory patterns, where different regions may reach their peak expansion at different moments in the cycle [16].

To address these limitations, the Local Expansion Ratio over N phases (LERN) approach was developed. [15] LERN builds upon Jacobian-based ventilation by incorporating data from all respiratory phases within a 4DCT acquisition, rather than relying on a single inhale and exhale pair. At each voxel, the Jacobian determinant is computed for every respiratory phase and the maximum and minimum values across the cycle are used to calculate a ratio.

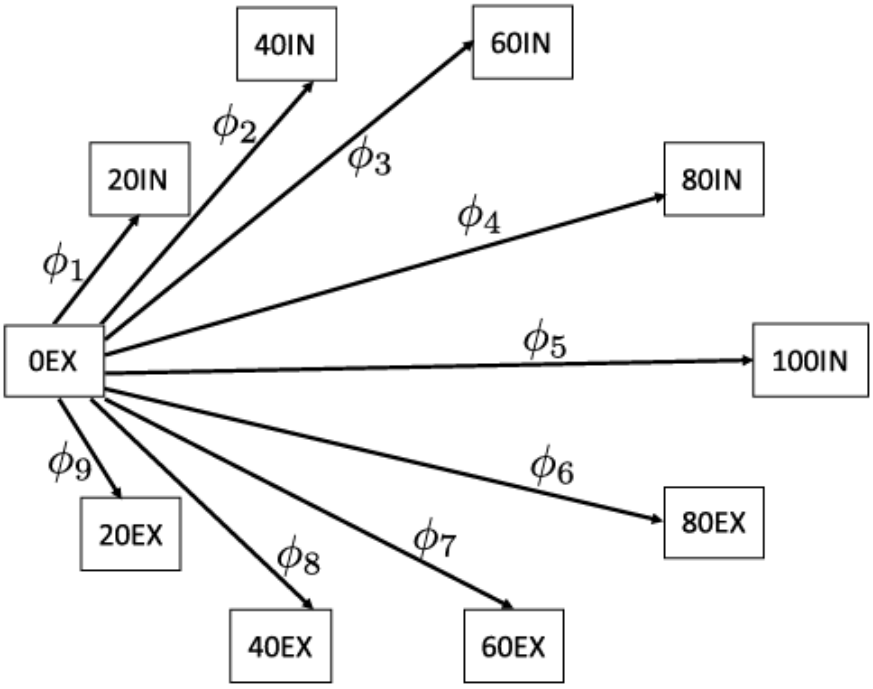


**Figure 4:** A depiction of LER-N method developed by Shao. Each breathing phase is represented and a Jacobian determinant is calculated for each phase.

This ratio, the LERN value, reflects the full extent of local volumetric change over time and is mathematically defined as $LER_n(x) = \max_j J_j(x) / \min_j J_j(x)$, where $J_j(x)$ is the Jacobian determinant at voxel x during phase j. [16] By considering the entire respiratory cycle, LERN captures regions that expand asynchronously with global lung motion and is more robust to errors caused by suboptimal phase selection.

The benefits of LERN have been demonstrated in both simulation and clinical settings. Shao showed that LERN reduces sensitivity to phase selection bias, improves inter-scan reproducibility, and better detects localized ventilation defects compared to conventional two-

phase Jacobian analysis. [16] In particular, LERN avoids underestimating expansion in regions that may not reach their maximum volume during the end of the inhale phase, improving the physiological accuracy of ventilation maps. Additionally, LERN avoids the need for phase pair selection heuristics, which are often unstable in longitudinal imaging where breathing effort may vary between sessions [16]. By anchoring the ventilation calculation to the actual temporal behavior of each voxel, LERN supports more consistent voxel-level comparisons across timepoints. The LERN approach won the AAPM Grand Challenge in 2019 for accurately estimating regional lung ventilation.

LERN is typically applied in conjunction with standard DIR workflows, requiring only that multiple respiratory phases be available. It has been successfully integrated with B-spline algorithms, and the resulting maps have shown promise in studies of lung function after radiation therapy [16]. This multiphase framework provides an important step toward physiologically meaningful ventilation imaging and may complement or enhance newer correction approaches such as Functional Effort Correction.

## 2.5 Functional Effort Correction

Jacobian ventilation maps derived from 4DCT are sensitive to variability in respiratory effort between imaging sessions. Small differences in tidal volume, caused by inconsistent breathing patterns, can lead to large changes in computed ventilation, even in the absence of true physiological change. This presents a major challenge in longitudinal studies, where the goal is to quantify regional damage or functional loss due to therapy.

Several strategies have been proposed to correct for this variability. Guerrero introduced a method to derive ventilation images form 4DCT using deformable image registration. They

then calculated regional lung variation at the voxel level by using Hounsfield units to estimate the fraction of air present in a region. [17] A more systematic solution was proposed by Du, et. al, who introduced what would later be called Equivalent Tidal Volume (ETV) and Equivalent Lung Volume (ELV) methods. These techniques identify inhale phases in each scan that produce similar total lung volumes when paired with the same exhale phase. When applied to Jacobian-based ventilation mapping, ETV and ELV showed markedly improved reproducibility, with correlation coefficients approaching 0.97 in repeat scans. [18] Bethard proposed the idea of using stable, uninvolved lobes as internal references, assuming that function in those lobes remained constant between timepoints. [4]

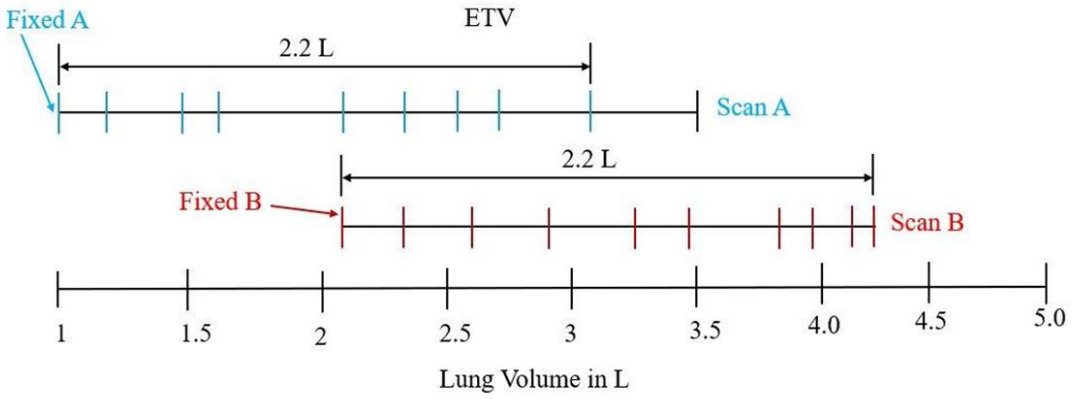


**Figure 5:** Illustration of Equivalent Tidal Volume (ETV) correction applied to two 4DCT scans with equal tidal volumes (2.2 L), but different starting lung volumes due to variable respiratory effort. ETV aligns the dynamic phases across both scans to a common tidal volume range, enabling consistent voxelwise comparison. [4]

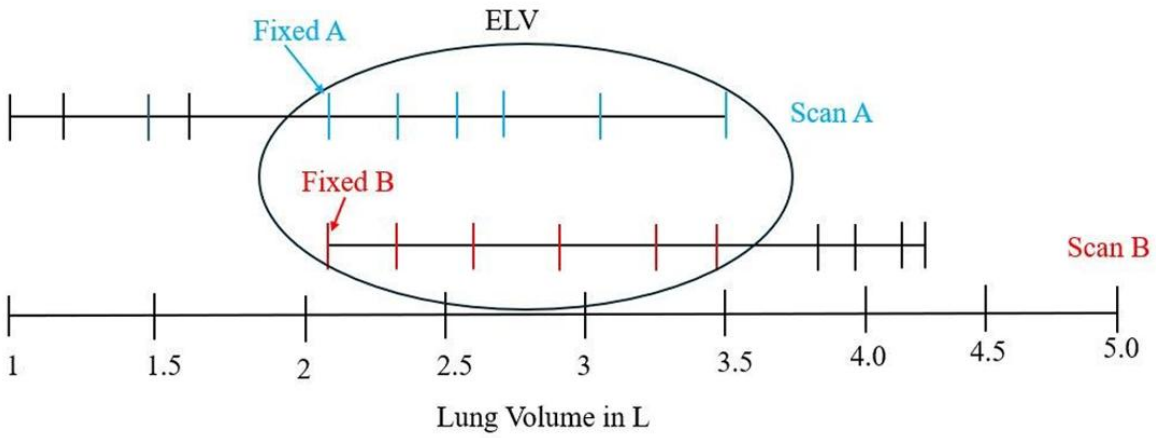


**Figure 6:** Illustration of Equivalent Lung Volume (ELV) correction applied to 4DCT scans with differing respiratory efforts. Although Scan A (blue) and Scan B (red) begin at different fixed phases, in ELV the Jacobian is calculated using only those phases that overlap in absolute lung volume across both scans. This ensures matched anatomical inflation, improving deformable registration fidelity. [4]

Global normalization approaches, such as ETV and ELV, are limited because they effectively mask RILI by normalizing on regions that have received high dose. To address this, regional effort correction methods have been proposed. One such approach involves identifying anatomically distinct, low dose sublobes that exhibit consistent expansion between scans, and using these as local volume references. The objective is to isolate and preserve true functional decline while correcting for global breathing differences.

It has been shown that Jacobian based ventilation estimates are highly sensitive to small uncertainties in deformable image registration. This supports the idea that both deformable registration and functional effort normalization should be anchored to stable anatomical substructures, such as the central airways or diaphragm, rather than relying solely on global tidal

volume. [19] Such spatially localized control regions may enhance reproducibility and improve sensitivity to real physiological changes. To enable robust effort correction without obscuring true regional damage, Bethard's Functional Effort Correction (FEC) method begins by selecting a stable sublobe as a normalization reference. The process identifies the lowest-dose sublobe (typically receiving <5 Gy) and evaluates its suitability using two criteria: (1) the sublobe's volume between scans must remain within ±15% (i.e., volume ratio between 0.85 and 1.15), and (2) the correlation between sublobe volume and total lung volume across breathing phases must exceed 0.9. If these conditions are not met, the next lowest dose sublobe is evaluated, continuing iteratively until a qualifying region is found. This ensures that the selected reference region is minimally affected by therapy and reliably reflects physiological breathing effort.

Figure 7 provides an example of FEC using an unaffected sublobe region as a basis for normalizing changes due to effort. An extrapolation region is computed for expanding the volumes and therefore breathing phases that are available for proper comparison for damage calculation.

Once a suitable sublobe is identified, Jacobian values are computed for all candidate inhale phases registered to a common exhale phase (0EX). The average Jacobian value within the sublobe is plotted against sublobe volume, producing a volume–ventilation curve for each timepoint. ΔV refers to the difference in the selected inhale-phase volumes between two timepoints, specifically within the reference sublobe. After fitting the volume–ventilation curves at each scan, the FEC method selects inhale phases whose sublobe volumes are as close as possible. The difference in those volumes is ΔV.

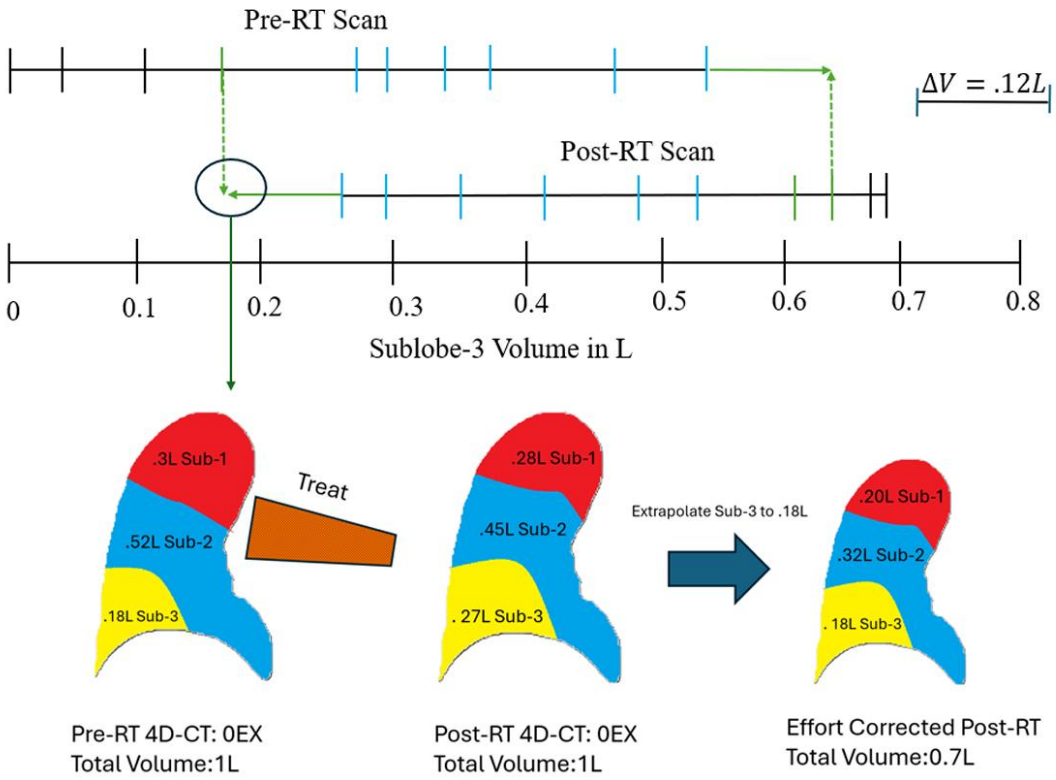


**Figure 7**: Illustration of Functional Effort Correction (FEC) applied to sublobe ventilation analysis. The upper panel shows the volume-time curves for a representative sublobe (Sub-3) from pre- and post-RT 4DCT scans, with the post-RT curve demonstrating reduced effort. The lower panel depicts anatomical sublobes at end-expiration (0EX), with the post-RT scan showing reduced Sub-3 volume due to both treatment and decreased effort. FEC extrapolates post-RT Sub-3 to its pre-treatment volume (0.18 L), enabling voxelwise comparisons at matched effort. Reprinted with permission from Bethard.

Keeping $\Delta V \leq 0.83$ L ensures that the resulting Jacobian comparison occurs between anatomically equivalent states, minimizing interpolation error when estimating functional

change. Bethard empirically demonstrated that maintaining $\Delta V \leq 0.83$ L keeps Jacobian interpolation error below 1%. Using this strategy, FEC revealed significantly greater functional loss in 13 out of 18 subjects compared to uncorrected methods, indicating enhanced sensitivity to radiation-induced damage while controlling for inter-scan variability.

## 2.6 Functional Lung Avoidance

Functional lung avoidance radiation therapy (FLART) is an advanced radiotherapy approach that integrates functional imaging data into treatment planning to minimize radiation-induced lung injury (RILI).[20] Unlike conventional radiotherapy, which treats the lung as a homogeneous organ, FLART leverages imaging modalities such as 4DCT, SPECT, and MRI to identify and spare regions of the lung with higher functional capacity, thereby preserving pulmonary function and reducing toxicity. [21]

The rationale for FLART stems from the recognition that lung function is heterogeneously distributed and that damage to highly functional regions can lead to significant morbidity. By incorporating functional imaging into the planning process, clinicians can tailor radiation doses to avoid these critical areas. For instance, studies have demonstrated that FLART can significantly reduce the mean dose to high-functioning lung tissue without compromising tumor control. [22]

Recent advancements have focused on automating the FLART planning process. Xiong (2024) developed an automatic planning framework that integrates function-guided beam angle selection and plan optimization. This approach effectively redirects doses from high-functional lung (HFL) to low-functional lung (LFL), achieving significant reductions in HFL mean dose and volume receiving 20 Gy (V20) and 5 Gy (V5).[22]

Clinical trials have also explored the benefits of FLART. The FLAIR trial investigated whether functional lung avoidance based on He3- MRI could improve quality of life for patients undergoing chemoradiotherapy. Although the trial did not reach full accrual, it provided insights into the feasibility and potential benefits of incorporating functional imaging into radiotherapy planning. [23] Moreover, FLART has been applied to esophageal cancer treatment. Integrating 4DCT ventilation function images into radiotherapy planning could effectively reduce the dose to functional lung regions while maintaining target coverage. [22]

FLART is a significant advancement in personalized radiotherapy, offering the potential to reduce pulmonary toxicity and preserve lung function by incorporating functional imaging into treatment planning.

## 2.7 Motivation

The primary objective of this study was to evaluate the effectiveness of Functional Effort Correction (FEC) compared to Equivalent Tidal Volume (ETV) and Equivalent Lung Volume (ELV) methods in quantifying radiation-induced lung damage over time. Assessments were conducted at 3, 6, and 12 months post-radiation therapy to determine if FEC provides improved sensitivity and specificity in detecting functional impairments.

To quantify lung damage, two primary metrics were employed: the LERN ratio and the percentage of lung voxels exhibiting Jacobian values below a threshold (e.g., 0.94). These metrics aimed to capture both the extent and severity of ventilation deficits. By analyzing these parameters across different time points, the study sought to identify patterns of injury progression or resolution.

Furthermore, the study aimed to investigate indirect effects of radiation therapy on lung function. Emerging evidence suggests that radiation can impair perfusion in lung regions not directly exposed to high doses but supplied by irradiated vasculature. This phenomenon indicates that vascular injury can lead to downstream functional deficits, emphasizing the need to consider vascular pathways in treatment planning. [24]

An additional focus was to observe temporal patterns in patients exhibiting clinical symptoms such as radiation pneumonitis or fibrosis. For instance, some patients may demonstrate increased lung damage at 3 months due to acute inflammatory responses, which could subsequently recede by 6 months. Conversely, fibrotic changes might not be apparent at early time points but could manifest at 12 months post-treatment. Understanding these patterns is crucial for timely intervention and management of radiation-induced lung injuries (RILI). [24]

# 3. METHODS

The effort correction methods described above were applied to a set of scans obtained from a prior clinical trial. This chapter describes aspects of that trial and the image processing workflows necessary for analysis. Modifications to the FEC method led to two different variations of that model that tested, which are also described, in addition to approaches to identify an appropriate damage metric and potential sources of uncertainty.

## 3.1 Clinical Trial Overview and Data Utilization

This study utilized imaging and dosimetric data from a clinical trial under a Data Sharing Agreement from with University of Iowa, focusing on patients undergoing thoracic radiation therapy for non-small cell lung cancer (NSCLC). [25]The trial aimed to investigate the relationship between radiation dose and subsequent changes in lung function, employing 4DCT imaging to capture these dynamics.

Participants were selected based on specific inclusion criteria, including a confirmed diagnosis of NSCLC, eligibility for curative-intent radiation therapy, and the ability to undergo serial imaging studies. Exclusion criteria encompassed prior thoracic radiation, significant comorbidities affecting lung function, and contraindications to imaging procedures.

Imaging data were collected at multiple time points: pre-treatment, and at 3, 6, and 12 months post-treatment. Each imaging session included 4DCT scans to assess lung ventilation and perfusion. Patients were scanned twice at each timepoint, within approximately five minutes. This provides a way to test variability of the measurements, as well as giving an additional measurement to mitigate the occurrence of artifacts. Patients were given breathing prompts in an

attempt to maintain a consistent breathing pattern.  Radiation dose distributions were obtained from treatment planning systems, providing spatial maps of delivered dose across the lung parenchyma.

## 3.2 Image Processing and Analysis Workflow

The analytical pipeline developed for this study processes four primary input files per patient to assess lung damage following radiation therapy. These files include the Local Expansion Ratio over N phases (LERN) map, the warped post-treatment image, the lobe mask, and the radiation dose distribution.

Initially, 4DCT scans were acquired and stored in Digital Imaging and Communications in Medicine (DICOM) format and then converted to Neuroimaging Informatics Technology Initiative (NIFTI) format. This conversion is useful because NIFTI files are more compatible with various image processing libraries and facilitate easier manipulation and analysis within Python-based environments. DICOM metadata also encompasses unwanted and unneeded meta data such as personally identifiable patient information. Prior to conversion the image phase information was extracted from the DICOM metadata and used to sort the images into their respective phases.

The LERN map is generated through deformable image registration (DIR), aligning the post-treatment scan to the pre-treatment scan. This process computes the Jacobian determinant of the deformation field obtained from DIR, resulting in a map that quantifies voxel-wise local volume changes. The warped post-treatment image is created by applying the deformation field to the post-treatment scan, transforming it to align with the pre-treatment scan. This alignment allows for direct voxel-wise comparison between pre- and post-treatment scans.

Lung lobe segmentation is then applied to the aligned scan using methods developed by Dr. Sarah Gerard's lab at the University of Iowa. Their approach employs deep learning techniques, such as FissureNet and LobeNet, to detect pulmonary fissures and segment the lungs into lobes, even in cases with incomplete fissures or diseased lungs. This method has demonstrated robustness and accuracy across various pulmonary diseases, including COPD, IPF, lung cancer, and COVID-19.[26, 27] The resultant lung mask is then aligned with the image scan so particular voxels can be associated with individual lobes. NIFTI files assign lobes with a bitmask identifier using powers of two. Conventionally these are 8, 16, 32, 64, 128. Table 1 lists the particular values associated with individual lobes in our work.

**Table 1:** Identification of NIFTI imaging data tags with particulars lobes used in this study.

| Identifier | Lobe code | Lobe description |
|---|---|---|
| 8 | LUL | Left Upper Lobe |
| 16 | LLL | Left Lower Lobe |
| 32 | RUL | Right Upper Lobe |
| 64 | RML | Right Middle Lobe |
| 128 | RLL | Right Lower Lobe |

The radiation dose distribution is extracted from the Treatment Planning System (TPS) as a DICOM RT Dose file. This file contains the radiation dose delivered to lung tissues and is converted and resampled to align with the coordinate space of the LERN map and lobe mask.

Resampling is a critical step to ensure all input files have consistent voxel spacing and dimensions for accurate voxel-wise analysis. In this context, resampling involves interpolating

image data to a new grid that matches a reference image's spatial parameters. Third-order spline (cubic) interpolation is used for this purpose, balancing computational efficiency with smoothness and preserving anatomical details. This process is vital when combining images from different sources or modalities, as it standardizes the data for accurate analysis and comparison.

Custom scripts, developed in Python, computes the ratio of LERN values between post-treatment and pre-treatment scans for each voxel. Table 2 describes the particular software packages used to create different data structures used in the analysis. This LERN ratio indicates relative functional changes over time. Voxels with LERN ratios below a certain threshold are identified as indicative of potential damage. Paired Student's t-tests are performed to compare LERN ratios and the percentage of damaged voxels across different time points (e.g., 3, 6, and 12 months post-treatment).

**Table 2:** Software and packages used

| | |
|---|---|
| Python | 3.11.5 |
| Matplotlib | 3.8.3 |
| Nibabel | 5.2.0 |
| Numba | 0.59.0 |
| NumPy | 1.26.4 |
| Pandas | 2.2.0 |
| SciPy | 1.12.0 |
| SimpleITK | 2.3.1 |

The output of the script includes a structured array or DataFrame containing voxel-wise information, including lobe designation, dose, LERN ratio, and damage status. Arrays for each subject, each time point, and each correction model (FEC, ETV, ELV) were created. Subjects were filtered out and excluded because some patients did not complete the study, the presence of severe artifacts, and the absence of proper dose information. Plots are generated using these arrays to visualize the lobar distribution of lung damage and its correlation with radiation dose.

## 3.3 Functional Effort Correction (FEC) Methodology

The original Functional Effort Correction (FEC) method was applied to account for variations in patient effort during imaging. This approach involved normalizing ventilation measurements to a standard effort level. Modifications to the original FEC method were implemented to improve accuracy, including adjustments for baseline ventilation heterogeneity.

Reference volumes were originally tertiary bronchi, or sublobes, but uncertainty around segmentation methods, specially around the use arterial tree data, prompted a revised approach. Identified lobes were used in place of sublobes as the fissures separating lobes promote increased certainty in region identification. First, the lowest irradiated lobe is selected as a candidate reference lobe. Second, the volume of the lobe is tracked through all phases of the scan (0EX to 100IN) and plotted against the overall increase in lung volume. Through a linear regression analysis the correlation coefficient is determined. The criterion is $R^2$<.6m or a negative slope results in the rejection of the candidate. If a lobe is not inflating and deflating similarly enough with the lung there is little confidence in the segmentation and therefore the candidate is unsuitable. Additional filtering is done by comparing the absolute difference in lobe volume pre- and post-radiation across all phases, expressed as a percentage. If this average percent deviation

exceeds 1.2%, the lobe is also excluded. Among the remaining lobes, the one with the lowest average dose is selected as the FEC lobe.

Once the FEC lobe is identified, the next step is to calculate the lobe-specific ΔV, which defines the maximum allowed interpolation range. This is computed by first averaging the lung volume across all phases, then dividing it by the average volume of the selected FEC lobe across the same phases. This was done in two different methods described as FEC1 and FEC2. The first, FEC1, was as Bethard originally described, taking the ratio and multiplying by a constant lung ΔV (0.75 L). It has been shown that a extrapolation volume of up to 0.83 L was possible without sacrificing accuracy due to noise. [4]. After reviewing outlier data and accounting for unrealistic LERN values, a more conservative scaling factor of 0.75 L was selected.

The FEC model contemplates sublobe regions, defined around tertiary bronchi, with the goal of achieving finer resolution for the reference region, and thereby increasing the likelihood of scaling on an undamaged region. A sublobe segmentation model using AirQuant open source tool was trained on a set of CT images. However, the ability to distinguish proper sublobe regions on our 4DCT data set was limited. Table 3 highlights the relative loss of airway segmentations that were unusable when switching from CT training data to the 4DCT used in the clinical trial. 4DCT data provides the temporal resolution necessary for respiration imaging but comes with the tradeoffs of motion artifacts, lower SNR, and lower spatial resolution as compared to standard CT.

**Table 3:** Comparison in airway segmentations between CT training data and 4DCT data. Statistically fewer airways can be identified on 4DCT data

| Metric | CT (n = 47) | 4D-CT (n = 119, subjects = 42) | t-statistic | p-value |
|---|---|---|---|---|
| Avg. terminal generation | 7.41 ± 1.59 | 6.67 ± 0.77 | 4.028 | $8.58 \times 10^{-5}$ |
| Total airways | 208.0 ± 54.6 | 151.6 ± 57.5 | 5.740 | $4.48 \times 10^{-8}$ |
| Max generation | 11.3 ± 2.6 | 10.3 ± 1.8 | 2.797 | $5.78 \times 10^{-3}$ |

4DCT images suffer from lower spatial resolution than standard diagnostic CT. There are fewer acquisitions and faster acquisitions leading to reduced signal to noise. Fewer visible airways result in fewer control point to determine sublobar volumes with accuracy. In particular the issue is most pronounced in the LLL where multiple artifacts are often present. Furthermore, there is a lack of arterial tree knowledge in the segmentation model, making sublobe identification more difficult. For these reasons the FEC reference region chosen to be at the lobar level. Anatomically distinct and separated by fissures, lobes provide more robust and reproducible segmentation, enabling better consistency in volume tracking across phases.

The FEC lobe volumes from both pre- and post-treatment scans are then passed to the Equivalent Volume Matching function. This function identifies overlapping phases between the two scans. To match beyond this overlap, the algorithm attempts to predict expansion toward a larger volume. It evaluates whether additional phases can be included within the computed lobe $\Delta V$ and selects the phase in the opposing scan that results in the largest possible expansion. The difference between the largest original volume and this new candidate phase defines the "expansion volume larger." This expansion is then translated from lobe volume to lung volume using a fitted conversion function, another known error source. The same process is repeated in the direction of contraction to predict toward a smaller volume.

After determining the matched phases and predicted lung volumes, these are used to guide image registration. The lowest-volume lung phase serves as the fixed image, and all other

phases become moving images. Deformable image registration is performed, and Jacobian determinant maps are generated for each transformation.

In the scan being expanded to a larger volume, the Jacobian map from the phase with the largest lung volume is selected. This Jacobian is scaled by the ratio of the predicted expansion volume to the lung volume at that phase, yielding a predicted Jacobian at the target volume. The same process is carried out in the reverse direction for contraction.

To generate the final LERN maps, the maximum and minimum Jacobian values are extracted voxel-wise across all respiratory phases and the predicted expansion/contraction volumes. The LERN is computed as the voxel-wise ratio of maximum to minimum Jacobian values. To standardize across comparisons, all Jacobians are renormalized by dividing by the predicted Jacobian from the lowest-volume image.

The final output has two LERN maps, one from post-RT and one from pre-RT. They are compared via DIR to generate a transformation from the postRT- to the pre-RT space. The Post-RT map is then warped using this transformation to the pre-RT scan. The ratio of these two LERN maps, now combined into a single frame, provides the measure of lung damage on a voxel basis.

The Functional Effort Correction (FEC) method was originally developed to account for variability in patient respiratory effort during 4DCT imaging. By normalizing ventilation measurements to a standardized volume state, the method enables meaningful comparison of lung function across timepoints and treatment conditions. Over time, the implementation of FEC has undergone several important revisions, driven by practical constraints in segmentation accuracy and the desire to reduce interpolation error.

The FEC lobe is selected through a multi-step filtering process. First, the lobe receiving the least about of dose is chosen as a candidate lobe. It is tested for linear inflation behavior by plotting their volume across all phases (from 0EX to 100IN) and fitting a linear regression. Lobes with a poor correlation ($R^2 < 0.6$) or negative slope are excluded. Additional filtering eliminates lobes whose average absolute volume deviation between pre- and post-treatment exceeds 1.2% . If the candidate fails, the lobe with the next least amount of dose is selected and the process is repeated until the FEC lobe is identified.

As described earlier, FEC1 scales based on a set factor of .75L. A second approach, called FEC2 was also defined. Rather than scaling a fixed lung $\Delta V$, this approach calculates $\Delta V$ as 40% of the tidal volume of the lobe itself. This threshold was determined empirically. First, the top 15% of observed lobe tidal volume changes using paired scans at the same time point were analyzed, outliers excluded, and the resulting values were found to range from 38% to the low 50s. A conservative 40% limit was chosen to reflect plausible effort-related variation due to unconscious changes in breathing depth. This method avoids the indirect assumptions involved in translating lung-to-lobe volume ratios and bases the interpolation threshold directly in lobe-specific physiology.

Once the $\Delta V$ is defined, the pre- and post-treatment lobe volumes are passed to the Equivalent Volume Matching function. This identifies overlapping phases and determines whether expansion toward larger or smaller predicted volumes is possible within the allowed $\Delta V$. Expansion predictions are converted from lobe volume to lung volume using a fitted phase-wise lung-to-lobe mapping function. The resulting target volumes are used to guide deformable image registration. The lowest-volume phase serves as the fixed image, and others become moving phases. Jacobian determinant maps are computed for each registration. In scans requiring

extrapolation to a larger volume, the Jacobian from the phase with the largest lung volume is scaled by the ratio of predicted volume to that phase's actual lung volume. The same logic is applied for contraction to a smaller volume. Finally, LERN maps are generated as voxel-wise ratios of maximum to minimum Jacobians across all observed and predicted phases.

Additionally, the predicted Jacobians are now computed without converting lobe volumes into lung volumes. Instead, once a predicted lobe volume is selected, the algorithm simply computes the ratio of the predicted volume to the end-of-range volume in the scan. This ratio is applied directly to the Jacobian map from the endpoint phase, scaling it to the target predicted state. This change eliminates the uncertainty introduced by phase-wise lung-to-lobe fitting models and simplifies the process.

The LERN combined with dose arrays described in section 3.2 were generated with both FEC1 and FEC2, and then compared. Analysis was restricted to those subjects for which a valid reference lobe could be identified for both methods. Two measures of damage, the LERN ratio, and the number of voxels below a threshold amount of damaged, are used to compare effort correction methods with ETV, ELV, FEC1 and FEC2 at different timepoints.

## 3.4 Data Structuring and Statistical Analysis

Processed data were organized into structured NumPy arrays (.npy files), with each voxel annotated with corresponding lobe designation, radiation dose, and Local Expansion Ratio over N phases (LERN) values. This structured format facilitated efficient analysis of regional lung function changes and their association with radiation doses.

Statistical comparisons were conducted using Stata/BE v. 18 (StataCorp, College Station, Texas), Python scipy package, and Microsoft Excel. Employing paired Student's t-tests to

evaluate differences in LERN ratios and the percentage of lung voxels exhibiting Jacobian values below a variety of LERN thresholds. Testing for potential bimodal distribution was done using the Bimodality Coefficient (BC) defined in Equation 3.

$$BC = \frac{g^2+1}{k} \quad (3)$$

where g is skewness and k is kurtosis.

Once a damage score was determined, a longitudinal analysis was performed to compare the effort correction methods at 3, 6, and 12 months. Statistical differences were evaluated using paired Student's t-tests.

## 3.5 Damage Score

The LERN ratio serves as the primary metric to define lung damage but defining a precise range with that metric presents a challenge. There is no ground truth to validate the extent to which damage has occurred. A LERN value of less than one indicates a lung region is ventilating less than it did prior to radiation therapy serving a surrogate for damage. Developing a clear picture of damage requires examining the entire distribution of LERN ratios.

Two scans were taken, five minutes apart, prior to treatment. It is assumed there is no appreciable change in lung function over so short of time. Differences in LERN between these scans therefore are attributable only to measurement uncertainty and random errors. The distribution of change in LERN, when computed as a ratio of values, can then be considered as Gaussian with a central peak of LERN ratio = 1.000, or no change. It has been found that the standard deviation in this comparison is ±6%, indicating that LERN ratio values of < 0.94 are more likely due to actual damage. [4]

It was unknown whether the LERN distribution would appear to be a Gaussian distribution when comparing post-RT scans to pre-RT scans, wherein we expect radiation damage. Overall damage is most accurately described as the summation of integral damage of all the voxels revealing an LERN ratio < 1.000. It was hypothesized that there might be a bimodal distribution of LERN ratios, with a peak in the damaged range. If true, employing a simple threshold may obscure meaningful information about the damage and an integral damage score would provide insight.

To evaluate damage score for effort correction methods, the distributions of LERN were examined by constructing histograms from the multidimensional NumPy arrays. A distribution was examined for each subject, lobe, effort correction method, and time point. The Bimodality Coefficient was calculated and tabulated using Python scipy package.

## 3.6 Measuring with and without Interscan Registration

The multidimensional NumPy arrays are designed for analysis of lung damage on a voxelwise basis. All the images, lobe masks and dose distributions are aligned in the same coordinate system and resampled to allow for this voxelwise analysis. By warping one image or LERN map from one timepoint to another, the scans must be registered. Any registration process will introduce a degree of uncertainty in the matched image alignment.

To assess the degree of uncertainty introduced the FEC2 method was studied with and without the interscan registration process. Instead of using the NumPy arrays, which include a single lobe mask for the chosen coordinate system, two separate masks are needed, one from pre-RT and one for post-RT. Each LERN map is then paired with the appropriate mask. An average LERN for each lobe is then easily calculated. The two averages, one pre-RT LERN and the other

post-RT LERN, and the calculated as a ratio, provided the LERN ratio average, and thus measured damage, for the lobe. Equations 4 and 5 contrast the two approaches. The interscan registration process is an average of the LERN ratio values, based on a per voxel approach. To avoid registration uncertainty it must be a ratio of LERN averages.

$$\text{Lobe-Averaged Ratio} = \frac{1}{N_{\text{lobe}}} \sum_{i \in \text{lobe}} \frac{\text{LERN}_{\text{Post-RT},i}}{\text{LERN}_{\text{Pre-RT},i}} \tag{4}$$

$$\text{Lobe-Averaged Ratio} = \frac{\frac{1}{N_{\text{lobe}}^{\text{Post}}} \sum_{i \in \text{lobe}} \text{LERN}_{\text{Post-RT},i}}{\frac{1}{N_{\text{lobe}}^{\text{Pre}}} \sum_{j \in \text{lobe}} \text{LERN}_{\text{Pre-RT},j}} \tag{5}$$

These two difference approaches were used to calculate average LERN ratio values for each subject, lobe and time point for the FEC2 method in order to assess the effect of registration.

# 4. RESULTS

## 4.1 Damage Distribution

Each effort correction methodology determines whether a subject can be included or excluded, meaning that a proper set of pre-RT and post-RT LERN maps can be generated and warped through DIR into a common coordinate system. To achieve the most comprehensive

comparison, the subjects analyzed were limited to those subjects for which LERN maps could be generated for all effort correction methods. This limited the number to 24 for the 3-month time period, 19 for the 6-months, and 32 for the 12-month period, as shown in Table 4.

**Table 4:** Change in qualifying subjects in various FEC modes

| | Number of subjects | | |
|---|---|---|---|
| | **FEC1** | **FEC2** | **Change** |
| **3 months** | 43 | 24 | -44% |
| **6 months** | 37 | 19 | -49% |
| **12 months** | 38 | 32 | -16% |

The LERN ratio of 0.94 or serves as a threshold for determining when a particular volume of the lung is considered damaged. In assessing overall damage to a larger volume such as a lobe, a distribution of voxel based LERN ratios must be considered. Figure 8 shows the correction effort methods (ETV, ELV, FEC1, FEC2) and the extent of damage revealed over a range of thresholds.

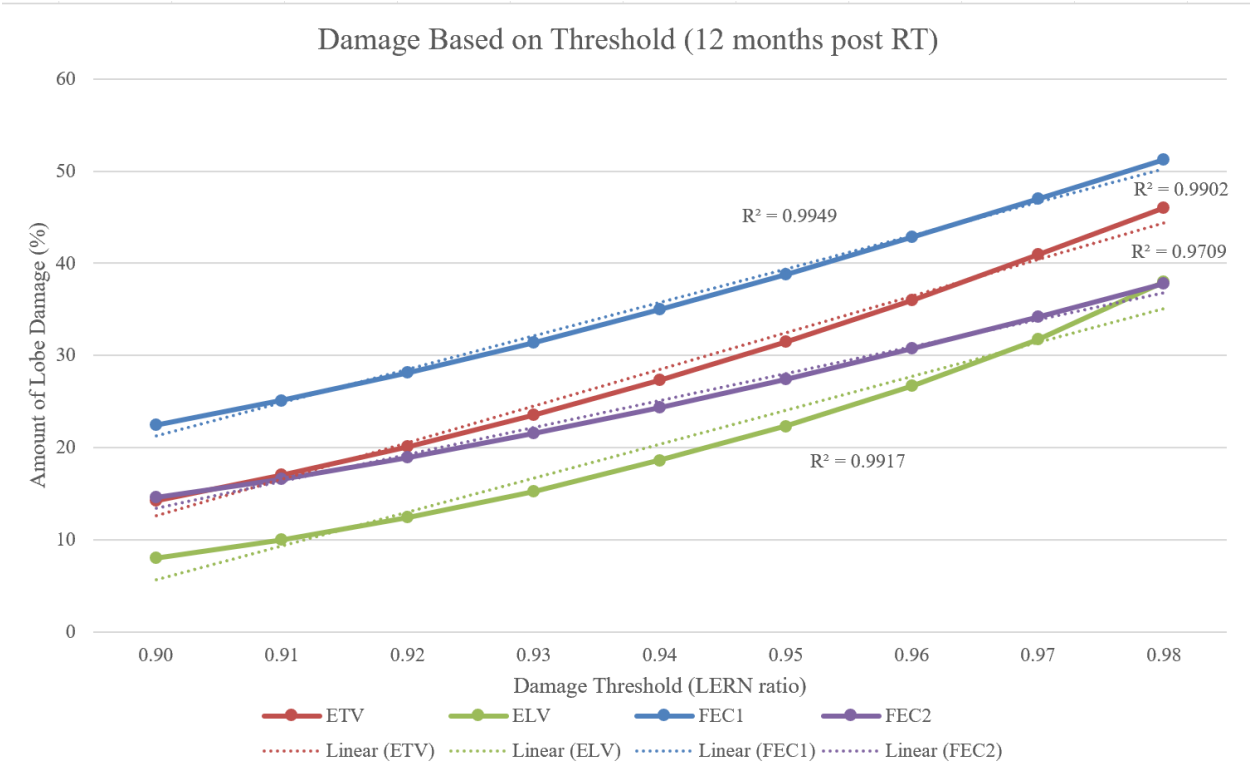


**Figure 8:** The relationship between a damage threshold and the amount of lobe that meets that threshold for four different effort correction approaches, along with linear fit lines.

Below .90 LERN the amount of damage is within a standard deviation of zero. Although it was initially hypothesized that the distribution might exhibit bimodal characteristics or secondary peaks near the tail, the observed data revealed a strongly unimodal, centrally peaked distribution. The region under investigation demonstrated a nearly linear ascent toward the central maximum, fitting well to a linear model. The results of the Bimodal Coefficient revealed an average value of $0.34 \pm 0.07$. This is approximately three standard deviations below the unimodal threshold value of $BC > .555$, indicating a strong unimodularity. [28] Of all the lobes only 1.3% were above the threshold that might indicate bimodality. As a result, there is no compelling need to apply continuity corrections to the metric. Within this region, the area under

the curve closely approximates the average value, supporting the use of straightforward averaging without further transformation.

**Table 5:** Damage amounts based on different thresholds at the 12-month time period. (n=32) Data is for the most irradiated lobe for each subject

| LERN Ratio | ETV (% damage) | ELV (% damage) | FEC1 (% damage) | FEC2 (% damage) | FEC1 vs. ETV | FEC1 vs. ELV | FEC2 vs. ETV | FEC2 vs. ELV | FEC1 vs. FEC2 |
|---|---|---|---|---|---|---|---|---|---|
| < 0.90 | 14.2 ± 10.8 | 8 ± 7.9 | 22.4 ± 25.4 | 14.6 ± 20.3 | 0.046 | 0.003 | 0.893 | 0.057 | 0.004 |
| < 0.91 | 17 ± 12 | 10 ± 9.2 | 25.1 ± 26.4 | 16.6 ± 21.6 | 0.052 | 0.002 | 0.916 | 0.073 | 0.003 |
| < 0.92 | 20.1 ± 13.4 | 12.4 ± 10.8 | 28.1 ± 27.3 | 18.9 ± 22.9 | 0.060 | 0.002 | 0.724 | 0.097 | 0.003 |
| < 0.93 | 23.5 ± 14.7 | 15.2 ± 12.6 | 31.4 ± 27.9 | 21.5 ± 24 | 0.069 | 0.002 | 0.548 | 0.133 | 0.002 |
| < 0.94 | 27.3 ± 16 | 18.6 ± 14.4 | 35 ± 28.5 | 24.3 ± 25.2 | 0.082 | 0.003 | 0.392 | 0.188 | 0.002 |
| < 0.95 | 31.5 ± 17.1 | 22.3 ± 16.4 | 38.8 ± 28.9 | 27.4 ± 26.2 | 0.099 | 0.003 | 0.256 | 0.272 | 0.003 |
| < 0.96 | 36 ± 18.1 | 26.7 ± 18.3 | 42.8 ± 29.2 | 30.7 ± 27 | 0.128 | 0.005 | 0.146 | 0.400 | 0.003 |
| < 0.97 | 40.9 ± 18.8 | 31.7 ± 19.9 | 47 ± 29.3 | 34.2 ± 27.7 | 0.170 | 0.007 | 0.074 | 0.606 | 0.004 |
| < 0.98 | 46 ± 19.1 | 38 ± 21.3 | 51.2 ± 29.4 | 37.8 ± 28.1 | 0.231 | 0.017 | 0.033 | 0.974 | 0.006 |

. **Table 6:** Damage amounts based on different thresholds at the 12 month time period. (n=31) Data is for the reference lobe for each subject, typically the lobe that receives the least amount of dose.

| LERN Ratio | ETV (% damage) | ELV (% damage) | FEC1 (% damage) | FEC2 (% damage) | FEC1 vs. ETV | FEC1 vs. ELV | FEC2 vs. ETV | FEC2 vs. ELV | FEC1 vs. FEC2 |
|---|---|---|---|---|---|---|---|---|---|
| < 0.90 | 8.7 ± 10.3 | 7.5 ± 8.5 | 16.9 ± 19.3 | 7.6 ± 11.6 | 0.027 | 0.014 | 0.637 | 0.974 | 0.000 |
| < 0.91 | 10.5 ± 11.4 | 9.4 ± 10.1 | 19.9 ± 20.8 | 9.3 ± 12.9 | 0.019 | 0.013 | 0.634 | 0.947 | 0.000 |
| < 0.92 | 12.7 ± 12.7 | 11.7 ± 11.8 | 23.1 ± 22.3 | 11.2 ± 14.3 | 0.014 | 0.012 | 0.626 | 0.872 | 0.000 |
| < 0.93 | 15.2 ± 14.1 | 14.5 ± 13.5 | 26.7 ± 23.8 | 13.5 ± 15.7 | 0.012 | 0.013 | 0.600 | 0.788 | 0.000 |
| < 0.94 | 18.1 ± 15.3 | 17.6 ± 15.3 | 30.4 ± 25.2 | 16.1 ± 17.2 | 0.012 | 0.015 | 0.561 | 0.696 | 0.000 |
| < 0.95 | 21.5 ± 16.6 | 21.3 ± 17 | 34.2 ± 26.5 | 19 ± 18.7 | 0.012 | 0.019 | 0.509 | 0.592 | 0.000 |
| < 0.96 | 25.4 ± 17.8 | 25.4 ± 18.6 | 38.3 ± 27.6 | 22.1 ± 20.1 | 0.015 | 0.025 | 0.448 | 0.490 | 0.000 |
| < 0.97 | 29.6 ± 18.8 | 30 ± 20 | 42.3 ± 28.5 | 25.6 ± 21.3 | 0.020 | 0.035 | 0.379 | 0.388 | 0.000 |
| < 0.98 | 34.3 ± 19.6 | 35 ± 21.1 | 46.5 ± 29.1 | 29.3 ± 22.3 | 0.029 | 0.053 | 0.305 | 0.286 | 0.000 |

Tables 5 and 6 presents the amount of damaged voxels in a lobe, either the highest dosed, or the reference lobe for effort correction. A Student-T test results are displayed to compare the four methods (ETV, ELV, FEC1, and FEC2). A statistically significant difference at the $\alpha = 0.05$ level is found between FEC1 and ELV as well as FEC1 and FEC2. At all thresholds FEC2 reveals less damage than FEC1.

## 4.2 Damage over Time with FEC and ETV

For the first set of results the FEC1 and FEC2 are compared with ETV. Each plot includes timepoints of 3, 6, and 12 months. Two sets of comparison are made. The first compares the most irradiated lobe for all the subjects. The most irradiated lobe varies between subject but the point of comparison is that these are the lobe that received the most amount or dose and therefore should suffer the greatest amount of radiation induced injury. The second comparison is with the reference lobe. This is typically, but not always the lobe receiving the least amount of radiation. This is particularly meaningful for the FEC method since it is based on examining a relatively unaffected region of the lung to account for change related to effort. Furthermore, both change in LERN and amount of volume of lung damage (defined as % of voxels where LERN ratio < 0.94) are compared.

**FEC v. ETV ( most irradiated lobe)**

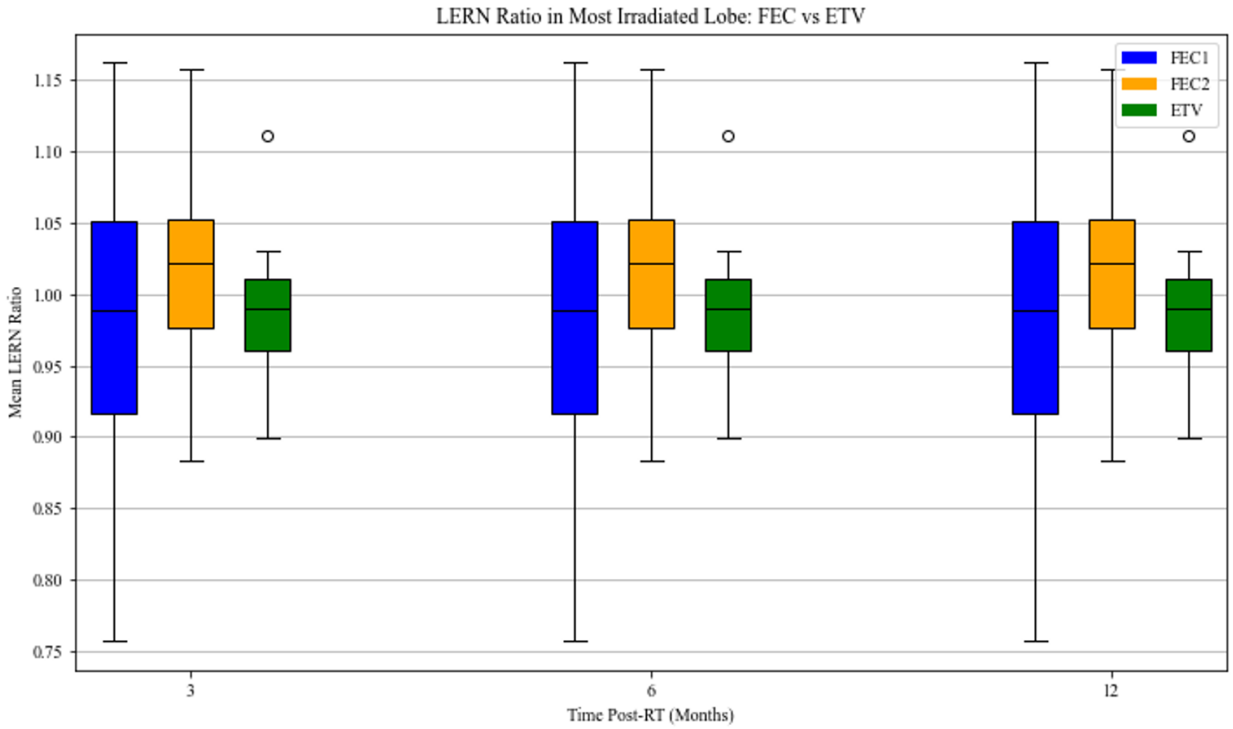


**Figure 9:** Box plot showing the distribution of LERN ratio at different time points. FEC1, FEC2 and ETV are compared for the most irradiated lobe, where damage is expected to be greatest.

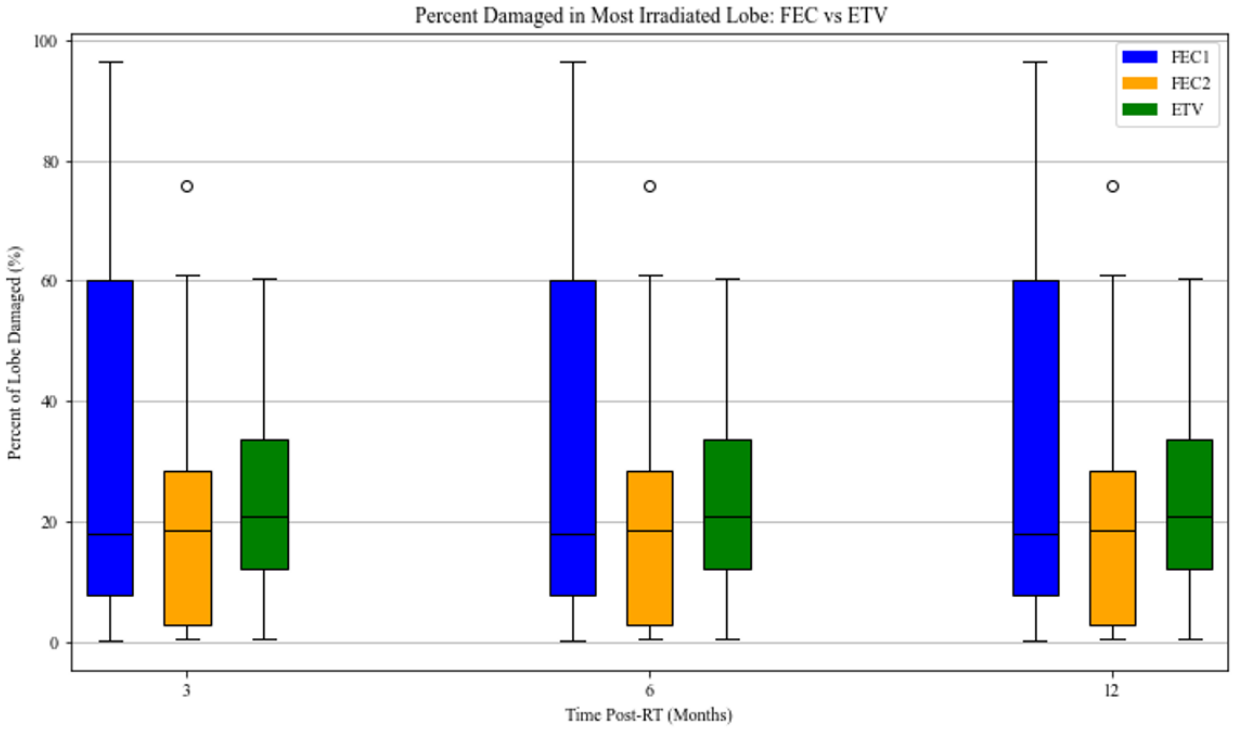


**Figure 10:** Box plot showing the distribution of amount of lobe damage at different time points. FEC1, FEC2 and ETV are compared for the most irradiated lobe, where damage is expected to be greatest.

**Table 7:** Comparison at LERN ratio at 3 months between FEC1, FEC2 and ETV for most irradiated lobe

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN vs ETV (α = .05) | p-val Damage vs ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.986 | 0.099 | 31.7 | 31.8 | 0.998 | 0.043 |
| FEC2 | 1.011 | 0.059 | 16.8 | 18.3 | 0.020 | 0.524 |
| ETV | 0.986 | 0.029 | 18.8 | 14.8 | | |

**Table 8:** Comparison at LERN ratio at 6 months between FEC1, FEC2 and ETV for most irradiated lobe

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN vs ETV (α = .05) | p-val Damage vs ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.971 | 0.091 | 39.0 | 32.9 | 0.305 | 0.006 |
| FEC2 | 1.018 | 0.069 | 20.1 | 22.2 | 0.054 | 0.873 |
| ETV | 0.993 | 0.037 | 19.5 | 13.0 | | |

**Table 9:** Comparison at LERN ratio at 12 months between FEC1, FEC2 and ETV for most irradiated lobe

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ETV (α = .05) | p-val Damage v. ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.987 | 0.095 | 35.0 | 28.5 | 0.567 | 0.012 |
| FEC2 | 1.011 | 0.075 | 21.5 | 24.0 | 0.870 | 0.548 |
| ETV | 0.987 | 0.040 | 23.5 | 14.7 | | |

For the most irradiated lobe there is a statistically significant difference between FEC1 and ETV at all time points. There is no difference between FEC2 and ETV at any of the points. There is a significant decrease in damage between FEC1 and FEC2. Overall there is no noticeable trend over the three time points.

**FEC v. ETV ( Reference lobe)**

In the following plots (Figures 11 and 12) and tables (Tables 10-12) the comparison between FEC methods and ELV is done by examining the reference lobe. The particular lobe may vary between subjects but they are selected because it is believe they have suffered the least amount of radiation damage and are large enough to make a good comparison.

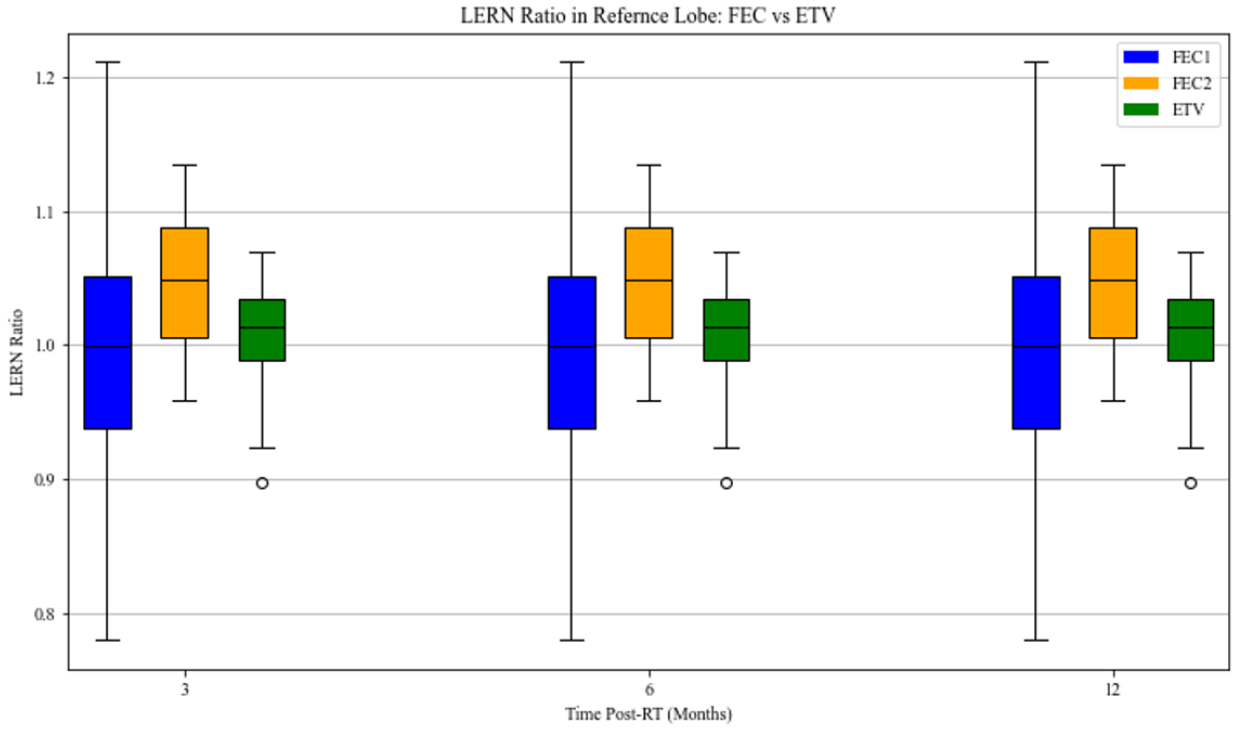


**Figure 11:** Box plot showing the distribution of LERN ratio at different time points. FEC1, FEC2 and ETV are compared for the reference lobe, where damage is expected to be least.

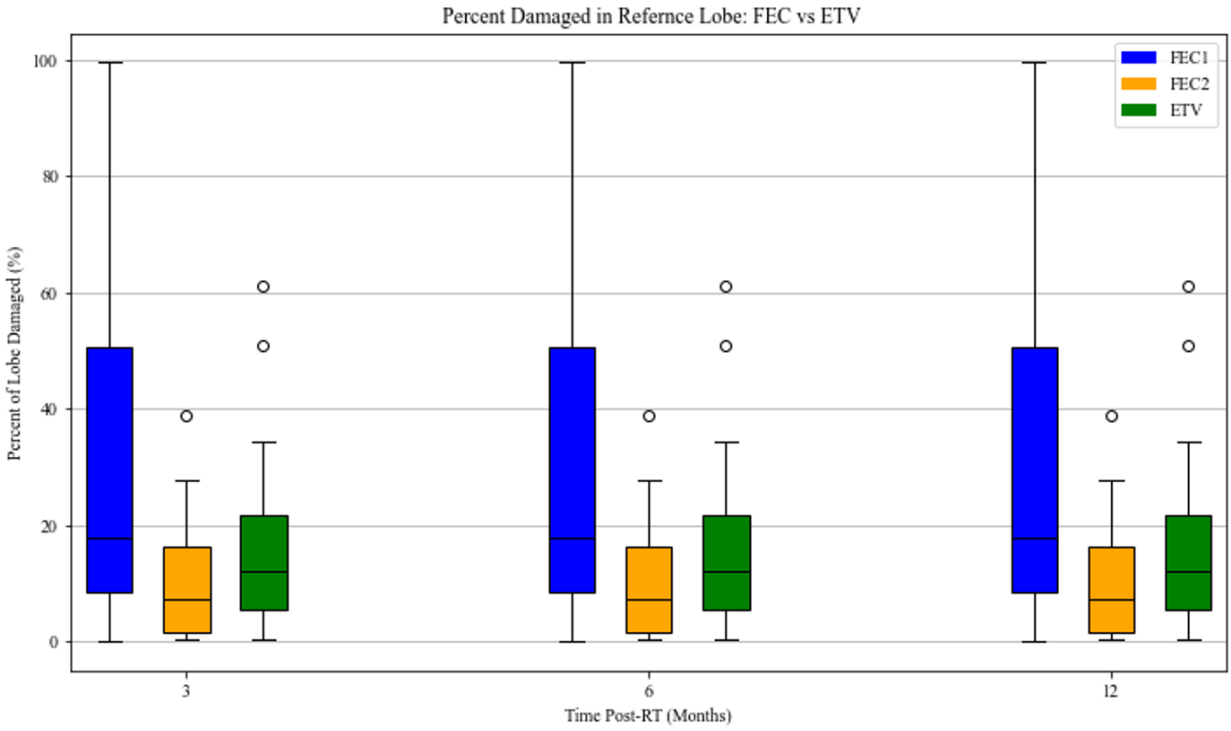


**Figure 12:** Box plot showing the distribution of amount of lobe damage at different time points. FEC1, FEC2 and ETV are compared for the reference lobe, where damage is expected to be least.

**Table 10:** Comparison at LERN ratio at 3 months between FEC1, FEC2 and ETV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ETV (α = .05) | p-val Damage v. ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.993 | 0.100 | 31.2 | 31.4 | 0.773 | 0.024 |
| FEC2 | 1.030 | 0.061 | 12.6 | 18.2 | 0.041 | 0.516 |
| ETV | 1.000 | 0.036 | 15.0 | 11.3 | | |

**Table 11:** Comparison at LERN ratio at 6 months between FEC1, FEC2 and ETV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ETV (α = .05) | p-val Damage v. ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.991 | 0.086 | 31.2 | 25.4 | 0.608 | 0.028 |
| FEC2 | 1.046 | 0.054 | 10.9 | 11.3 | 0.031 | 0.096 |
| ETV | 1.004 | 0.046 | 16.4 | 13.4 | | |

**Table 12:** Comparison at LERN ratio at 12 months between FEC1, FEC2 and ETV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ETV (α = .05) | p-val Damage v. ETV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 1.004 | 0.097 | 30.4 | 25.2 | 0.518 | 0.002 |
| FEC2 | 1.033 | 0.062 | 13.5 | 15.7 | 0.791 | 0.600 |
| ETV | 1.008 | 0.039 | 15.2 | 14.1 | | |

For the reference lobe there is a statistically significant difference between FEC1 and ETV at all time points. There is no difference between FEC2 and ETV at any of the points. There is a significant decrease in damage between FEC1 and FEC2. Overall there is no noticeable trend over the three time points.

## 4.3 Damage over Time with FEC and ELV

For the first set of results the FEC1 and FEC2 are compared with ELV. Each plot includes time points of 3, 6, and 12 months. Two sets of comparison are made. The first compare the most irradiated lobe for all the subjects. The most irradiated lobe varies between subject but the point of comparison is that these are the lobe that received the most amount or dose and therefore should suffer the greatest amount of radiation induced injury. The second comparison is with the reference lobe. This is typically, but not always the lobe receiving the least amount of radiation. This is particularly meaningful for the FEC method since it is based

on examining a relatively unaffected region of the lung to account for change related to effort. Furthermore, both change in LERN and amount of volume of lung damage (defined as % of voxels where LERN ratio < 0.94) are compared.

**FEC v. ELV ( most irradiated lobe)**

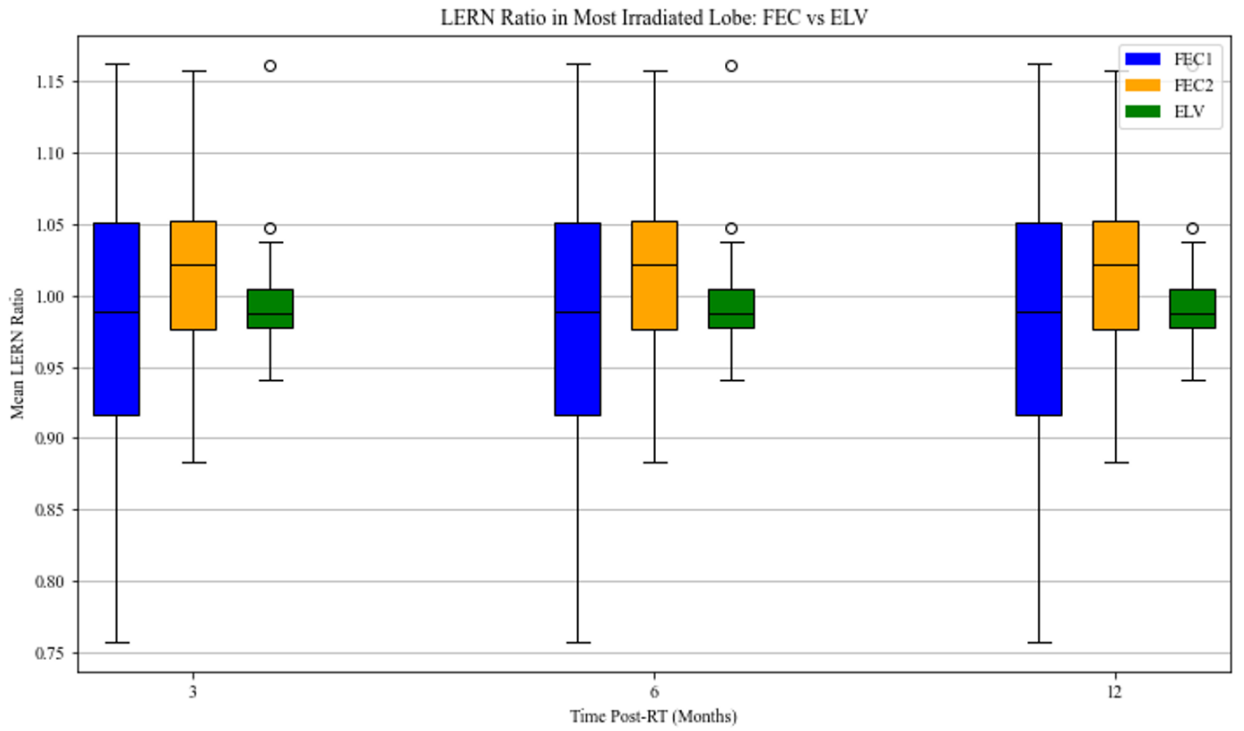


**Figure 13:** Box plot showing the distribution of LERN ratio at different time points. FEC1, FEC2 and ELV are compared for the most irradiated lobe, where damage is expected to be greatest.

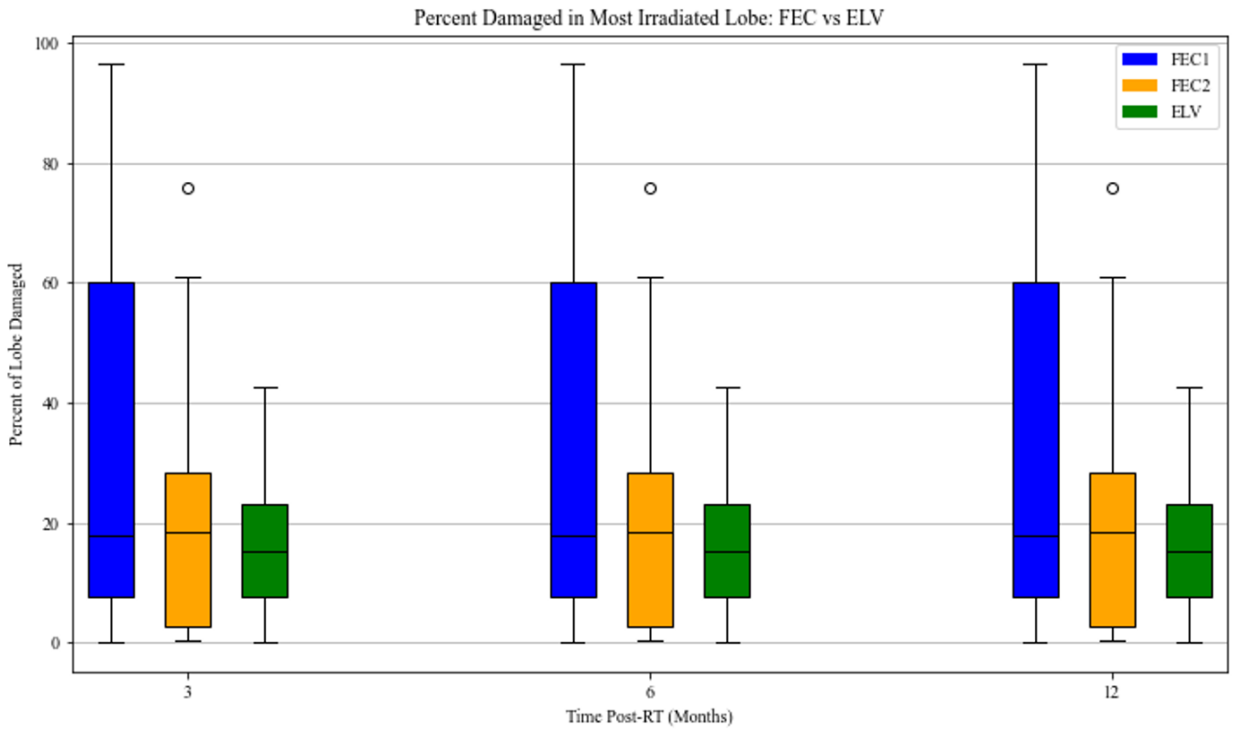


**Figure 14:** Box plot showing the distribution of amount of lobe damage at different time points. FEC1, FEC2 and ELV are compared for the most irradiated lobe, where damage is expected to be greatest.

**Table 13:** Comparison at LERN ratio at 3 months between FEC1, FEC2 and ELV for most irradiated lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.986 | 0.099 | 31.7 | 31.8 | 0.864 | 0.019 |
| FEC2 | 1.011 | 0.059 | 16.8 | 18.3 | 0.039 | 0.874 |
| ELV | 0.989 | 0.032 | 17.3 | 15.1 | | |

**Table 14:** Comparison at LERN ratio at 6 months between FEC1, FEC2 and ELV for most irradiated lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.971 | 0.091 | 39.0 | 32.9 | 0.379 | 0.006 |
| FEC2 | 1.018 | 0.069 | 20.1 | 22.2 | 0.119 | 0.485 |
| ELV | 0.990 | 0.039 | 15.1 | 17.3 | | |

**Table 15:** Comparison at LERN ratio at 12 months between FEC1, FEC2 and ELV for most irradiated lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.987 | 0.095 | 35.0 | 28.5 | 0.684 | 0.000 |
| FEC2 | 1.011 | 0.075 | 21.5 | 24.0 | 0.602 | 0.171 |
| ELV | 0.997 | 0.039 | 15.4 | 10.9 | | |

For the most irradiated lobe there is a statistically significant difference between FEC1 and ELVV at all time points. There is no difference between FEC2 and ETV at any of the points.

There is a significant decrease in damage between FEC1 and FEC2. Overall there is no noticeable trend over the three time points.

**FEC v. ELV ( Reference lobe)**

In the following plots (Figures 15 and 16) and tables (Tables 16-18) the comparison between FEC methods and ELV is done by examining the reference lobe. The lobe may vary between subjects but they are selected because it is believe they have suffered the least amount of radiation damage and are large enough to make a good comparison.

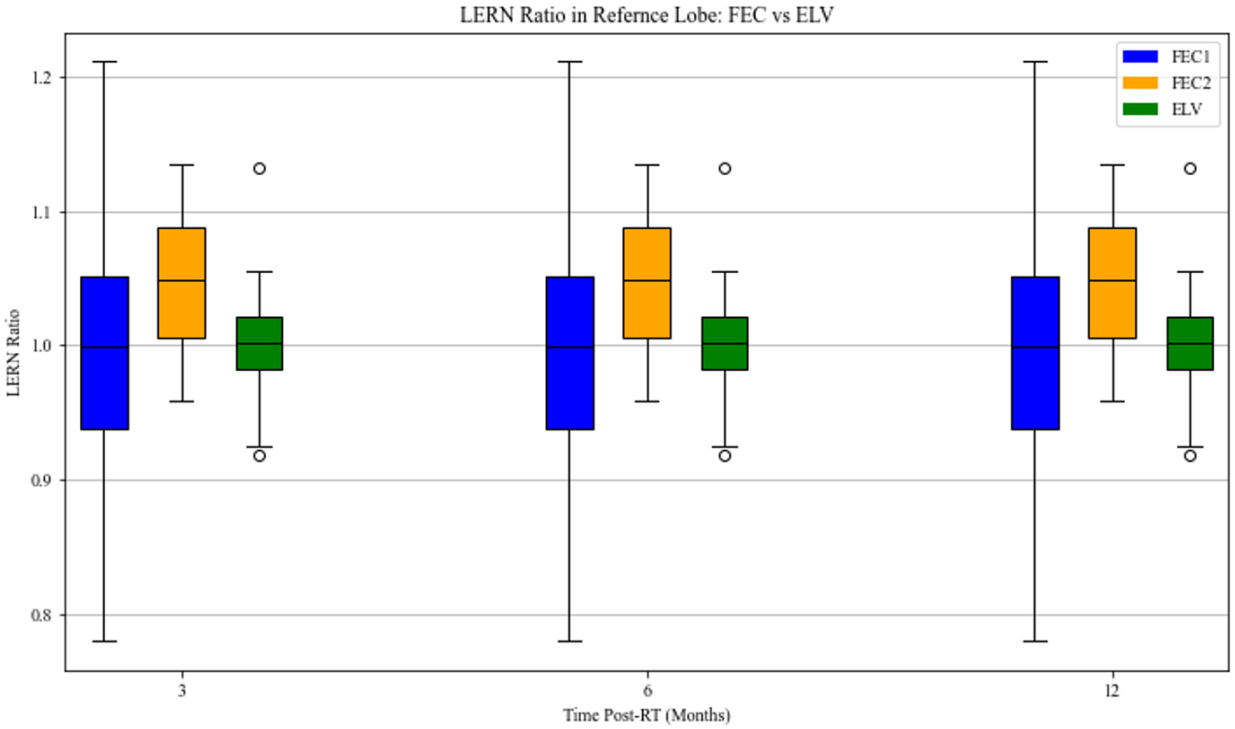


**Figure 15:** Box plot showing the distribution of LERN ratio at different time points. FEC1, FEC2 and ELV are compared for the reference lobe, where damage is expected to be least.

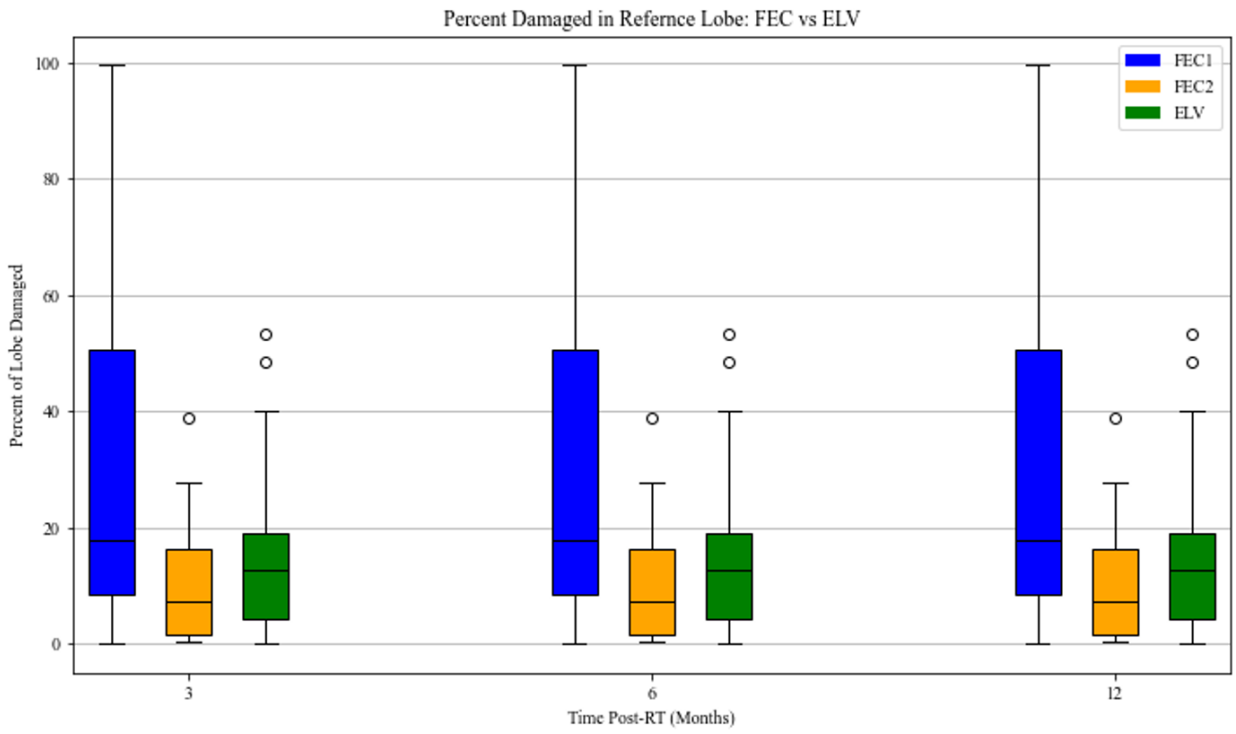


**Figure 16:** Box plot showing the distribution of amount of lobe damage at different time points. FEC1, FEC2 and ELV are compared for the reference lobe, where damage is expected to be least.

**Table 16:** Comparison at LERN ratio at 3 months between FEC1, FEC2 and ELV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.993 | 0.100 | 31.2 | 31.4 | 0.775 | 0.015 |
| FEC2 | 1.030 | 0.061 | 12.6 | 18.2 | 0.048 | 0.812 |
| ELV | 1.000 | 0.032 | 13.6 | 10.4 | | |

**Table 17:** Comparison at LERN ratio at 6 months between FEC1, FEC2 and ELV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 0.991 | 0.086 | 31.2 | 25.4 | 0.812 | 0.077 |
| FEC2 | 1.046 | 0.054 | 10.9 | 11.3 | 0.019 | 0.191 |
| ELV | 0.997 | 0.043 | 17.2 | 17.7 | | |

**Table 18:** Comparison at LERN ratio at 12 months between FEC1, FEC2 and ELV for reference lobe.

| | Mean LERN | SD LERN | Mean % Damaged | SD % Damaged | p-val LERN v. ELV (α = .05) | p-val Damage v. ELV (α = .05) |
|---|---|---|---|---|---|---|
| FEC1 | 1.004 | 0.097 | 30.4 | 25.2 | 0.619 | 0.003 |
| FEC2 | 1.033 | 0.062 | 13.5 | 15.7 | 0.976 | 0.748 |
| ELV | 1.003 | 0.041 | 14.7 | 13.3 | | |

For the reference lobe there is a statistically significant difference between FEC1 and ELV at all time points. There is no difference between FEC2 and ELV at any of the points. There is a significant decrease in damage between FEC1 and FEC2. Overall there is no noticeable trend over the three time points.

## 4.4 Results of Avoiding Interscan Registration

The results of processing the FEC2 dataset by applying separate pre-RT and post-RT lobe masks and then calculating lobe average LERN ratios is displayed in Figure 18 with the tabulated averages in Table 20. These results are also compared to the FEC1 voxelwise results. This information provides a direct contrast based on the removal of registration uncertainty. The data is based on the reference lobe. We expect to see LERN ratios near unity for the reference lobe since it selected precisely because it is not expected to have undergone radiation injury.

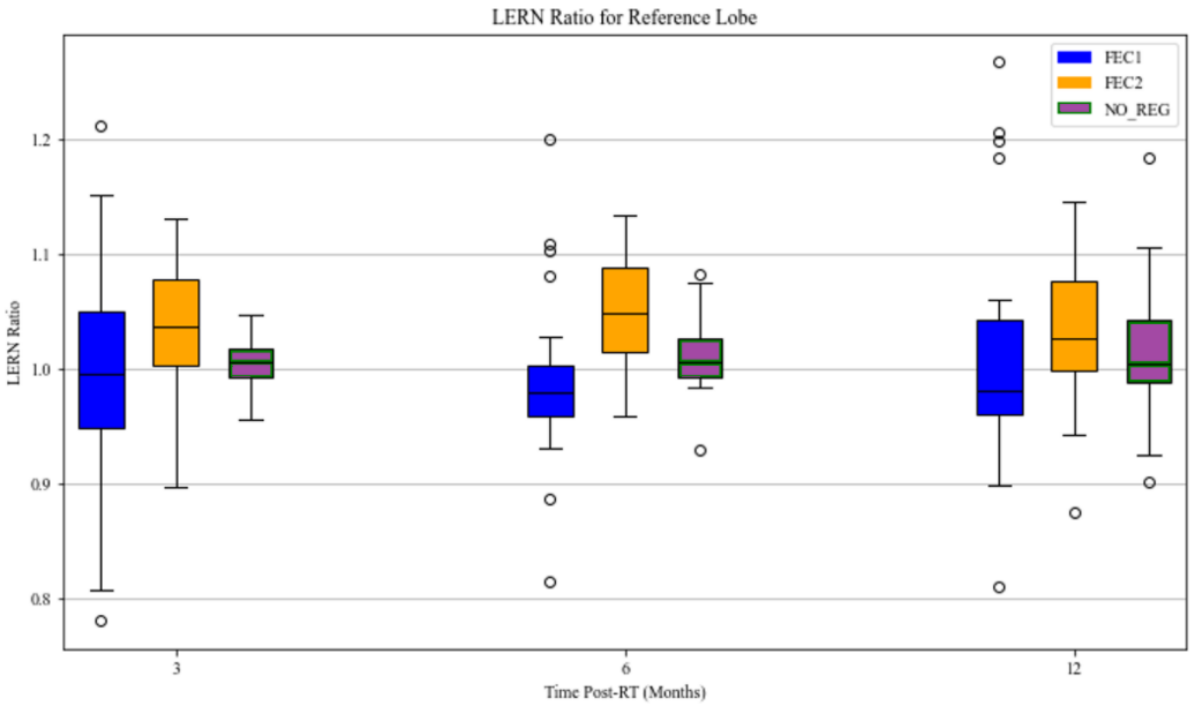


**Figure 18.** LERN ratio for FEC1, FEC2 and FEC2 (No Interscan Registration). The most relevant comparison is between the purple and yellow bars. Without registration uncertainty the same base scans reveal a tighter grouping and LERN ratio values closer to 1.000.

**Table 20:** Show the average, standard deviation, and median LERN ratio for lobes based on FEC1, FEC2 and FEC2 (No Interscan Registration). Both average and median values for FEC2 (No Interscan Registration) are closer to 1.000 than FEC2 using Interscan Registration.

| 3 months (n=24) | FEC1 | FEC2 | No_Reg |
|---|---|---|---|
| Avg | 0.993 | 1.030 | 1.005 |
| SD | 0.100 | 0.061 | 0.025 |
| Median | 0.999 | 1.037 | 1.005 |

| 6 months (n=19) | FEC1 | FEC2 | No_Reg |
|---|---|---|---|
| Avg | 0.991 | 1.046 | 1.013 |
| SD | 0.086 | 0.054 | 0.037 |
| Median | 0.979 | 1.049 | 1.006 |

| 12 months (n=32) | FEC1 | FEC2 | No_Reg |
|---|---|---|---|
| Avg | 1.004 | 1.033 | 1.016 |
| SD | 0.097 | 0.062 | 0.057 |
| Median | 0.981 | 1.027 | 1.005 |

# 5. DISCUSSION

The primary objective of this study is to evaluate FEC at different time periods post-RT and to determine if it is more effective at revealing radiation induced lung damage than prior methods of ETV and ELV. Different parameters of the FEC model were altered (FEC1 and FEC2) allowing for a comparison between approaches. The Python scripts developed provide a compact and easily analyzed set of data structures that can be employed to test different FEC algorithms with dose at the lobar level.

First a set damage threshold of LERN ratio < 0.94 was established based on prior studies and an analysis of LERN distributions. The value of 0.94 is calculated from the variance in nearly contemporaneous pre-RT scans, revealing a 6% standard deviation from 1.000. Based on a Gaussian distribution, approximately 68% of LERN ratio measurements would be expected to fall within that 6% standard deviation. Values below 0.94 would only occur through random uncertainty about 16% of the time. Most values below this point could be attributed to radiation damage.

Prior to this study it was uncertain whether an FEC approach would result in a unimodal distribution of damage. If the distribution matched a non-Gaussian or multi-modal profile, then using a simple threshold may not be as revealing as other measures. A damage score metric was proposed that might be more informative, based on an integral approach. An accumulated damage score may indicate a greater extent of damage than a simple count of damaged voxels based on a single threshold.

The results, however, indicate that the distributions are unimodal and that the amount of damage fits with a linear profile over LERN ratios less than unity. Figure 8 shows a clear linear pattern without bimodalism. This profile could mask a uniform distribution of bimodal peaks through the damage range of 0.90-0.98 LERN, so it was necessary to analyze distributions individually. Employing the statistical approach of Bimodal Coefficients (BC), the average value of BC was 0.34 ± .07, well below an expected unimodal value of approximately 0.555. A bimodal distribution is only suspected above that value. 98.3% of all lobes analyzed had a value below this, strongly indicating unimodal behavior

With a linear relationship between LERN ratio and damage, and no indications of bimodalism, the utility of an integrated damage score is obviated. A straightforward inequality metric suffices to inform the amount of damage being measured. There is no ground truth for damage with this data, making an accurate selection of a threshold an uncertain process. Based on the previous study on pre-RT reproducibility the value of 0.94 was selected for the rest of the analysis in this work. The multidimensional histograms used, however, require no alterations based on the threshold selection. It is a simple matter to generate results based on any threshold desired which will support future investigations.

FEC2 relies on a smaller extrapolation volume and results in a reduced amount of damage found. This is a useful result and can guide further refinements of the algorithm. Noticeably, the change in methods resulted in a change from FEC showing significantly higher damaged volume in FEC1, to showing statistically higher LERN ratio (less damage) in FEC2. FEC hypothesizes that damage is masked in ETV and ELV due to scaling on regions of the lung based on damaged regions of the lung. Finding less damage with FEC2 as compared to FEC1 is somewhat unexpected but it is revealing as to the sensitivity of the FEC approach. FEC2 uses

smaller extrapolation volumes than FEC1 pointing to a strong correlation between that volume and overall measured damage. However, more scrutiny is needed in assessing these outcomes.

The range and standard deviations for both LERN and percentage of lobe damage are very large in the case of FEC. 7 of the 32 subjects with data at the 12 month timepoint had an average LERN of greater than or less 10% for the reference lobe. By definition, the reference lobe is expected to normalize to near 1.000 LERN. A systematic review of the subject and sources of uncertainty may provide justification for exclusion of these subjects.

Throughout the course of the FEC reference region identification, extrapolation calculations, image registration, and warping of transformation, uncertainty is introduced. Although lobe segmentation is more certain than sublobe segmentation, it is still infallible. Fissures are not always clear or captured. In the original FEC, $\Delta V$ is scaled for the lobe based on lung-wide changes. This assumes that each lobe expands or shrinks proportionally to its volume fraction, but real lung may ventilate unequally due to disease, anatomy, or positioning. FEC2 reduces this uncertainty by using actual tidal volume, but the functional threshold value of 40% is a conservative estimate and has not been thoroughly tested. Finally, in the LERN warping process there is always a degree of uncertainty that stems from deformable registration inaccuracies, interpolation artifacts, and the misalignment of functionally significant regions.

The comparisons between FEC2 and FEC2 without interscan registrations show a clear effect of this introduced uncertainty. Any registration process introduced uncertainty into the results. Interscan registration is inherently more uncertain than intrascan registration. Interscan occurs months after the subject has undergone radiation therapy. Anatomy will change from RILI, tumor control, or other non-related factors. The voxel intensities will likely always be less similar than a intrascan differences.

Table 20 shows that the reference lobe for FEC2 is showing higher LERN ratios while using interscan registration, approximately 3-4% above unity. By removing interscan registration uncertainty the medians and average are much closer to unity, around 1%. That is as expected since the FEC reference lobe selection process is designed to identify regions that are unaffected by radiation. Using this approach, however, precludes a voxel level analysis. One area for further study would be to use the no interscan registration procedure on the FEC1 dataset and compare.

Another topic to consider for future work is trying to explain the relatively higher LERN ratio values seen with FEC2 by considering the nature of the metric in the presence of unchanging tidal volume with decreases lung volumes. Typically, lung volumes decline after radiation therapy and that has been observed with this dataset. Recent investigations also reveal that the tidal volumes are relatively consistent over time even with shrinking volumes. This may result in an augmentation of the LERN ratio without a true reflection of physiological behavior. There may be an unaccounted-for modification of the workflow to account for this scenario and should be comprehensively examined.

Quantitative analysis of uncertainties may lead to refinement of the FEC approach. A greater number of subjects may also reduce the standard deviation of results and more useful conclusions. The alteration of the FEC method led to the exclusion of a significant number of subjects. It's possible that the tighter constraints of FEC2 properly excluded subjects that were viable under FEC1's approach and leads to more certain results. However, a detailed investigation may provide justification for the reintroduction of more patients and thus provide more subjects for comparison. Furthermore, finding additional subjects through the accounting of improper segmentation or image artifacts may improve the model's predictive power.

The original FEC method contemplated sublobe identification for a reference region. Current techniques of segmentation down to that level of functional region are inadequate. Further development however, may yield more precision in the results as lobar complexities may confound the benefits provided by the non-irradiated reference region.

No trends over time were observed. At each time point damage was registered but no pattern emerged as to difference between 3, 6, and 12 months. Lung toxicities such as radiation pneumonitis and pulmonary fibrosis manifest and subside on different time scales. Additionally, indirect damage further complicates the analysis. Radiation induced damage of an airway may yield to volume changes and implied damage in a different section of the lung. When examining how damage changed over time there were no discernable trend over time. The spread of damage measured is large and may be refined with more subjects that could tease out time-related effects. A further study might be possible if enough subjects with known clinical outcomes come be identified and teased out of the larger dataset.

There are opportunities for future work to further develop and test FEC. Improved segmentation techniques at both the lobar and sublobar level may yield more certain results. A systemic review of segmentation through direct visualization and examination may lead to a rejection or reassessment of candidates that may serve to reduce the range of result, producing greater certainty and confidence. Once there is more confidence in effort correction an analysis of airways may lead to better modelling of indirect damage.

# 5. CONCLUSION

This study served as a functional stress test for successive refinements to the FEC method and its comparison with existing effort correction techniques such as ETV and ELV. Although the current results do not demonstrate a consistent or statistically significant advantage for FEC in detecting greater lung damage, they highlight key areas of methodological sensitivity and variability. Importantly, the development of flexible Python tools and a structured data output format enables streamlined analysis and testing of future algorithmic improvements. These resources position the FEC framework for ongoing refinement, deeper integration with dose analysis at the lobar level, and potential clinical application as segmentation, registration, and subject characterization continue to advance.

# REFERENCES


1. **World Health Organization. (2021).** *Global cancer observatory: Cancer today*. International Agency for Research on Cancer. Retrieved from https://gco.iarc.fr/

2. **American Cancer Society. (2024).** *Cancer facts & figures 2024*. American Cancer Society.

3. **Flakus, M.** (2020). *Advancing CT-ventilation biomarker utility in lung functional avoidance radiation therapy* (Doctoral dissertation, University of Wisconsin). University of Wisconsin Institutional Repository.

4. **Bethard, W.** (2024). *Functional effort correction of the lungs in radiation therapy longitudinal studies* (Master's thesis, Oregon Health & Science University). Oregon Health & Science University Institutional Repository.

5. **Simon, B. A., Christensen, G. E., Low, D. A., & Reinhardt, J. M**. (2005). *Computed tomography studies of lung mechanics*. Proceedings of the American Thoracic Society, 2(6), 517–521. https://doi.org/10.1513/pats.200507-076DS

6. **Weibel, E. R.** (1963). *Morphometry of the human lung*. Springer.

7. **Hegedüs, C., Näppi, J. J., Yoshida, H. (**2014). Segmentation of pulmonary lobes using an iterative algorithm with dynamic 3D shape models. *Medical Image Analysis*, 18(7), 1087–1096.

8. **Lutfi, M. F. (2017).** ***The physiological basis and clinical significance of lung volume measurements*****.** ***Multidisciplinary Respiratory Medicine*****, 12, 3. https://doi.org/10.1186/s40248-017-0084-5**9. **Kabir, N. A., Harvey, S. B., & Rosenfeld, A.** (2010). Design and fabrication of pulmonary embolism phantom for planar and SPECT V/Q imaging quality assurance.

*Australasian Physical & Engineering Sciences in Medicine*, *33*(3), 271–277. https://doi.org/10.1007/s13246-010-0025-5

10. **Clark, K., Vendt, B., Smith, K.,** (2015). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 28(1), 104–113.

11. **Miller, R., Castillo, R., Castillo, E., Jones, B. L., Miften, M., Kavanagh, B., Lu, B., Werner-Wasik, M., Ghassemi, N., Lombardo, J., Barta, J. A., Grills, I., Rusthoven, C. G., Guerrero, T., & Vinogradskiy, Y.** (2022). *Characterizing pulmonary function test changes for patients with lung cancer treated on a 2-institution, 4-dimensional computed tomography-ventilation functional avoidance prospective clinical trial*. *Advances in Radiation Oncology*, 8(2), 101133. https://doi.org/10.1016/j.adro.2022.101133

12. **Kupitz, D., Wissel, H., Wuestemann, J., Bluemel, S., Pech, M., Amthauer, H., Kreissl, M. C., & Grosser, O. S**. (2021). *Optimization of SPECT/CT imaging protocols for quantitative and qualitative 99mTc SPECT*. *EJNMMI Physics*, 8(1), 57. https://doi.org/10.1186/s40658-021-00405-3

13. **Zhang, Z., Li, H., Xiao, S., Zhou, Q., Liu, S., Zhou, X., & Fan, L.** (2024). *Hyperpolarized gas imaging in lung diseases: Functional and artificial intelligence perspective*. *Academic Radiology, 31*(10), 4203–4216. **https://doi.org/10.1016/j.acra.2024.01.014**

14. **Chen, Y., Heidari Pahlavian, S., Jacobs, P., Neupane, T., Forghani-Arani, F., Castillo, E., Castillo, R., & Vinogradskiy, Y.** (2024). *Systematic Evaluation of the Impact of Lung Segmentation Methods on 4-Dimensional Computed Tomography Ventilation Imaging Using a Large Patient Database*. International Journal of Radiation Oncology, Biology, Physics, 118(1), 242–252. https://doi.org/10.1016/j.ijrobp.2023.08.017

15. **Shao, W., Wallat, E., Wuschner, A., Flakus, M. J., & Bayouth, J. E.** (2020). N-phase local expansion ratio for characterizing out-of-phase lung ventilation. *Medical Physics*, *47*(4), 1635–1645. https://doi.org/10.1002/mp.14023

16. **Gorbunova, et al.** (2012). Mass preserving image registration for lung CT. *Medical Image Analysis*, *16*(1), 157–169. https://doi.org/10.1016/j.media.2011.06.006

17. **Guerrero, T., Sanders, K., Castillo, E.,** (2005). Dynamic ventilation imaging from four-dimensional computed tomography. *Physics in Medicine and Biology*, 50(16), 3767–3780. https://doi.org/10.1088/0031-9155/50/16/018

18. **Du K, Reinhardt JM, Christensen GE, Ding K, Bayouth JE.** Respiratory effort correction strategies to improve the reproducibility of lung expansion measurements. Med Phys. 2013 Dec;40(12):123504. doi: 10.1118/1.4829519**.**

19. **Yim, Y., Hong, H., Shin, Y.,** (2010). Deformable lung resigrations between exhale and inhale CT scans using active cells in a bombined gradient force approach. *Mecical Physics,* 37 (8), 4307-17. https://doi.org/10.1118/1.3460316

20. **Farr, K. P., West, K., Yeghiaian-Alvandi, R., Farlow, D., Stensmyr, R., Chicco, A., & Hau, E. (2019).** *Functional perfusion image guided radiation treatment planning for locally advanced lung cancer*. Physics and Imaging in Radiation Oncology, 11, 76–81. https://doi.org/10.1016/j.phro.2019.08.007

21. **Yamamoto, T., Kabus, S., von Berg, J., Lorenz, C., & Keall, P. J**. (2011). Reproducibility of Four-dimentional Compute Tomography-based Lung Ventilation Imaging. *International Journal of Radiation Oncology, Biology, Physics*, 79(1), 279–288. https://doi.org/10.1016/j.ijrobp.2010.02.008

22**. Zhou, P.-X., Wang, R.-H., Yu, H., Zhang, Y., Zhang, G.-Q., & Zhang, S.-X.** (2022). Different functional lung-sparing strategies and radiotherapy techniques for patients with esophageal cancer. *Frontiers in Oncology*, 12, 898141.

23. **Hoover, D. A., Capaldi, D. P., Sheikh, K., Palma, D. A., Rodrigues, G. B., Dar, A. R., Yu, E., Dingle, B., Landis, M., Kocha, W., Sanatani, M., Vincent, M., Younus, J., Kuruvilla, S., Gaede, S., Parraga, G., & Yaremko, B. P.** (2022). Functional Lung Avoidance for Individualized Radiotherapy (FLAIR): Results of a Double-Blind Randomized Controlled Trial. *International Journal of Radiation Oncology, Biology, Physics*, 112(3), 532–539. https://doi.org/10.1016/j.ijrobp.2022.04.047PubMed

24. **Wuschner, A. E., Flakus, M. J., Wallat, E. M., Reinhardt, J. M., Shanmuganayagam, D., Christensen, G. E., & Bayouth, J. E.** (2022). Measuring indirect radiation-induced perfusion change in fed vasculature using dynamic contrast CT. *Journal of Personalized Medicine*, *12*(8), 1254. https://doi.org/10.3390/jpm12081254

25. **Baschnagel, A. M., Flakus, M. J., Wallat, E. M., Wuschner, A. E., Chappell, R. J., Bayliss, R. A., Kimple, R. J., Christensen, G. E., Reinhardt, J. M., Bassetti, M. F., & Bayouth, J. E.** (2024). A Phase 2 Randomized Clinical Trial Evaluating 4-Dimensional Computed Tomography Ventilation-Based Functional Lung Avoidance Radiation Therapy for Non-Small Cell Lung Cancer. *International Journal of Radiation Oncology, Biology, Physics*, *119*(5), 1393–1402. https://doi.org/10.1016/j.ijrobp.2024.02.019

26. **Gerard, S. E., Patton, T. J., Christensen, G. E., Bayouth, J. E., & Reinhardt, J. M.** (2019). FissureNet: A deep learning approach for pulmonary fissure detection in CT images.

*IEEE Transactions on Medical Imaging, 38*(1), 1565–1576.

https://doi.org/10.1109/TMI.2018.2858202

27. **Gerard, S.E., Herrmann, J., Xin, Y. *et al.*** CT image segmentation for inflamed and fibrotic lungs using a multi-resolution convolutional neural network. *Sci Rep* **11**, 1455 (2021).

https://doi.org/10.1038/s41598-020-80936-4

28. **Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B.** (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology, 4*, 700.

https://doi.org/10.3389/fpsyg.2013.00700

# Appendix

Script Availability: Scripts for all portions of our work are available upon request from Bayouth Lab.