Scientific Article

Interobserver reliability in describing radiographic lung changes after stereotactic body radiation therapy

Noah S. Kalman MD, MBA [a],*, Geoffrey D. Hugo PhD [a,b], Jenna M. Kahn MD [a], Sherry S. Zhao MD [a], Nuzhat Jan MD [a], Rebecca N. Mahon MS [a], Elisabeth Weiss MD [a]

[a] Department of Radiation Oncology, Virginia Commonwealth University, Richmond, Virginia
[b] Department of Radiation Oncology, Washington University in St. Louis, St. Louis, Missouri

Abstract

Purpose: Radiographic lung changes after stereotactic body radiation therapy (SBRT) vary widely between patients. Standardized descriptions of acute ($\leq$6 months after treatment) and late (>6 months after treatment) benign lung changes have been proposed but the reliable application of these classification systems has not been demonstrated. Herein, we examine the interobserver reliability of classifying acute and late lung changes after SBRT.

Methods and materials: A total of 280 follow-up computed tomography scans at 3, 6, and 12 months post-treatment were analyzed in 100 patients undergoing thoracic SBRT. Standardized descriptions of acute lung changes (3- and 6-month scans) include diffuse consolidation, patchy consolidation and ground glass opacity (GGO), diffuse GGO, patchy GGO, and no change. Late lung change classifications (12-month scans) include modified conventional pattern, mass-like pattern, scar-like pattern, and no change. Five physicians scored the images independently in a blinded fashion. Fleiss' kappa scores quantified the interobserver agreement.

Results: The Kappa scores were 0.30 at 3 months, 0.20 at 6 months, and 0.25 at 12 months. The proportion of patients in each category at 3 and 6 months was as follows: Diffuse consolidation 11% and 21%; patchy consolidation and GGO 15% and 28%; diffuse GGO 10% and 11%; patchy GGO 15% and 15%; and no change 49% and 25%, respectively. The percentage of patients in each category at 12 months was as follows: Modified conventional 46%; mass-like 16%; scar-like 26%; and no change 12%. Uniform scoring between the observers occurred in 26, 8, and 14 cases at 3, 6, and 12 months, respectively.

Conclusions: Interobserver reliability scores indicate a fair agreement to classify radiographic lung changes after SBRT. Qualitative descriptions are insufficient to categorize these findings because

most patient scans do not fit clearly into a single classification. Categorization at 6 months may be the most difficult because late and acute lung changes can arise at that time.

Introduction

Stereotactic body radiation therapy (SBRT) for early stage lung cancer is the standard of care for medically inoperable patients.[1] The local control rates are excellent[2-4] and thoracic SBRT has found widespread application in recent years.[5] Post-treatment surveillance primarily consists of serial computed tomography (CT) scans. The accurate categorization and interpretation of post-SBRT radiographic change serves numerous purposes: Facilitation of communication between physicians, correlation between lung parenchyma changes and development of radiation-induced toxicity, and identification of tumor recurrence versus benign change.

Radiographic lung changes after SBRT vary widely between patients and can vary across different areas of the lung within individual patients, which makes the classification of post-treatment changes difficult. Pneumonitis after SBRT occurs in 11% to 20% of patients,[6-9] and early identification of high-risk patients may enable an early intervention to mitigate clinical toxicity.[10] In addition, benign inflammatory and fibrotic lung parenchyma changes can appear similar to or mask residual or recurrent disease.

The differentiation of a wide range of benign post-SBRT lung changes from local recurrence is essential to allow early salvage therapy. However, because local recurrence is rare (5%-20% of cases),[4] aggressive work up of concerning imaging findings exposes many patients to unnecessary worry and procedural risks in this commonly multi-morbid population (ie, pneumothorax with repeat biopsies or complications from surgical interventions).[11,12] Methods to identify high-risk features on

follow up CT scans have been proposed[13] but approximately 1 quarter of patients with benign post-treatment lung changes would be miscategorized using this method.[14] Indeed, no consensus exists to manage post-SBRT patients with suspicious lung changes.[15,16] Physician comfort with evaluating post-treatment imaging is critical because other accessory information such as beam arrangement, isodose distribution, and underlying patient characteristics can influence the interpretation of surveillance imaging.

To assist physicians with the evaluation of and facilitate interprovider communication with regard to post-treatment scans, standardized descriptions of acute (<6 months of treatment) and late (>6 months after treatment) benign lung changes have been proposed (Table 1).[17-20] These descriptions offer guidance to identify and categorize benign radiographic change but the reliable application of these classification systems has not been demonstrated. Prior studies have examined the interobserver agreement of grading late lung changes after SBRT[21,22] but to date, no study has evaluated the degree of interobserver reliability in the evaluation of acute lung changes after SBRT. Herein, we examine the interobserver reliability in the classification of acute and late lung changes after SBRT.

Methods and materials

Patient population

The study included 100 patients who underwent thoracic SBRT between 2007 and 2016. After approval by our institutional review board, patients' radiographic images were

Table 1 Standardized post-treatment radiographic lung change descriptions

Change categories	Definitions[17-20]
Acute (3 and 6 months)	
• Diffuse consolidation	Consolidation ≥5 cm; involved region contains more consolidation than aerated lung
• Patchy consolidation and GGO	Consolidation <5 cm and/or involved region with less consolidation than aerated lung
• Diffuse GGO	≥5 cm GGO; involved region with more GGO than aerated lung
• Patchy GGO	<5 cm GGO and/or involved region with less GGO than aerated lung
• No change	No new abnormalities or no change larger than the original tumor
Late (12 months)	
• Modified conventional pattern	Consolidation, volume loss, and bronchiectasis similar to but less extensive than conventional fibrosis
• Mass-like pattern	Focal consolidation limited around the original tumor
• Scar-like pattern	Linear opacity in the tumor regional with associated volume loss
• No change	No new abnormalities or no change larger than the original tumor

GGO, ground glass opacity.

Table 2 Patient characteristics (n = 100)

Characteristic	Median (Range) or percent present
Age, y	67 (45-91)
Female sex	54%
African-American race	52%
Stage I	86%
Planning target volume size, cc	24.4 (4.6-187.7)
Fractionation	
- 48 Gy in 4 fractions	78%
- 40-60 Gy in 5 fractions	17%
- Other	5%
Forced expiratory volume in 1 second, L[a]	1.32 (0.57-3.07)
Percent predicted diffusing capacity of the lung for carbon monoxide[b]	37 (1-130)

[a] A total of 80 patients with available data.
[b] A total of 66 patients with available data.

extracted from the medical records to create a database for analysis.

At our institution, the target volume was delineated using the maximal intensity projection images from a 4-dimensional, non-contrast, enhanced CT scan or images that were obtained with the patient in the breath-hold position. A 3 mm to 5 mm expansion was used to create the planning target volume. The most commonly prescribed fractionation regimen was 48 Gy in 4 fractions although other regimens with similar biologically equivalent doses were also used. A summary of the patient characteristics is listed in Table 2.

Imaging data

After SBRT, follow-up CT scans of the thorax were performed every 3 to 6 months. Deep-inspiration breath-hold scans that were completed 3, 6, and 12 months after treatment were included in the analysis. For patients simulated for free-breathing treatment, the 30% phase (mid-ventilation) image of the 4-dimensional treatment planning scan served as the baseline scan against which follow-up scans were compared to classify changes in lung parenchyma (Fig 1). Planning contours and plan isodose lines were available to reviewers.

Lung change scoring

The standardized lung change descriptions for acute and late changes are listed in Table 1. The 5 physician observers who all had experience in lung SBRT and the interpretation of post-SBRT follow-up imaging were provided with explanations of the lung change descriptions with example images on the basis of prior publications.[17,19,23] Subsequently, the reviewers examined de-identified patient scans
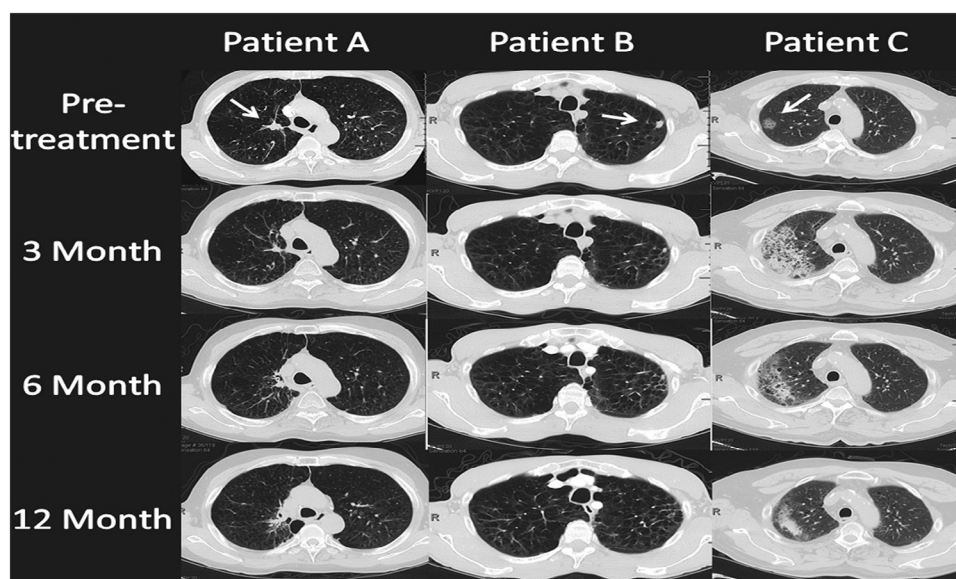

Figure 1 Examples of post-treatment radiographic lung changes. Reviewer scores (frequency) listed by patient. Follow-up scans performed at 3, 6, and 12 months after stereotactic body radiation therapy. Reviewer scores (frequency). Patient A: Uniform observations of no change at 3 months; diffuse consolidation (2) versus patchy consolidation and ground glass opacity (GGO) (1) versus patchy GGO (2) at 6 months; modified conventional pattern (4) versus scar-like pattern (1) at 12 months. Patient B: Diffuse GGO (1) versus no change (4) at 3 and 6 months; scar-like pattern (1) versus no change (4) at 12 months. Patient C: Diffuse consolidation (1) versus patchy consolidation and GGO (2) versus diffuse GGO (2) at 3 months; patchy consolidation and GGO (3) versus diffuse GGO (2) at 6 months (Of note: shape of radiographic change consistent with mass- or scar-like pattern, a late description); modified conventional pattern (1) versus mass-like pattern (3) versus scar-like pattern (1) at 12 months.

with blinding between the reviewers. Commercial software (MIM Maestro v6.6, MIM Software, Cleveland, OH) presented default lung and mediastinal contrast windows for the scans.

The baseline scans with tumor contours, beam directions, and isodose distributions were available to the reviewers. Axial, coronal, and sagittal views were evaluable for the baseline and follow-up scans. The reviewers scored patients' 3-, 6-, and 12-month scans (comparison between different time points was allowed) and completed the scoring of all scans prior to examining the next patient. The patient order was identical between reviewers to assist in evaluation for a training effect. Observers used acute lung change descriptions for the 3- and 6-month scans and the 12-month scans were graded using late lung change descriptions.

Statistical analysis

The reviewer ratings of the acute and late lung change descriptions were tabulated and summary percentages that reported the number of times a classification was used by the 5 observers, divided by the total number of reviews across all scans and patients at a given time point. Fleiss' Kappa scores were employed to quantify interobserver agreement at the 3-, 6-, and 12-month time points. Fleiss' Kappa is a statistical measure to assess the reliability of interobserver agreement for nominal data in which there are ≥2 observers.

The Fleiss' Kappa is the ratio of achieved interobserver agreement and possible interobserver agreement. Values of 0 and 1 represent no agreement above that expected by chance and complete agreement, respectively.[24] Leave-one-out analyses tested for the variability between the observers and the impact of single observers on the calculated scores. The assessments of the interobserver agreement for the first and last 20 patients were performed to evaluate for a potential effect of increasing experience with using the classification system.

Results

Lung change scoring

In 100 patients undergoing thoracic SBRT, 280 follow-up CT scans at 3, 6, and 12 months post-treatment were analyzed. All patients had evaluable 3- and 6-month scans and 80 patients had 12-month scans. The patient characteristics are provided in Table 2.

With regard to acute lung changes, the most frequently used classification at 3 months was no change (49%), followed by patchy consolidation and ground glass opacity (GGO) and patchy GGO (both 15%). At 6 months, patchy consolidation and GGO (28%) surpassed the no-change classification (25%) as the most common, followed by diffuse consolidation (21%; Table 3). The reported percentages represent the number of times a classification was used, divided by the total number of scores across all patients in a given follow-up period. Uniform scoring among the 5 observers occurred in 26 patients at 3 months and 8 patients at 6 months. In scans with uniform scoring, no change was the most common description (85%).

At 12 months, modified conventional pattern was the most commonly graded late lung change (46%), followed by scar-like pattern (26%) and mass-like pattern (16%; Table 3). Uniform scoring among the 5 raters occurred in 14 cases (18%). Modified conventional pattern was the most common description (79%) in cases with uniform rating.

Interobserver reliability

The 3- and 6-month Fleiss' Kappa scores were 0.30 and 0.20, respectively. Leave-one-out analyses demonstrated similar mean Kappa scores of 0.30 (standard deviation [SD]: 0.02) and 0.19 (SD: 0.04), respectively. Interobserver reliability improved with experience at both time points and increased from 0.20 to 0.29 between the first and last 20 patients at 3 months and from 0.03 to 0.25 at 6 months (Table 4).

Table 3   Frequency of radiographic lung change classifications

Follow-up time	Diffuse consolidation	Patchy consolidation and GGO	Diffuse GGO	Patchy GGO	No change
3 months[a]	11%	15%	10%	15%	49%
6 months[a]	21%	28%	11%	15%	25%
Follow-up time	Modified conventional		Mass-like pattern	Scar-like pattern	No change
12 months[b]	46%		16%	26%	12%

GGO, ground glass opacity.
All images were reviewed by 5 observers. The percentages present the breakdown of category use across all image reviews at a given time point.
    [a] A total of 100 patients with 498 image reviews.
    [b] A total of 80 patients with 398 image reviews.

Table 4   Interobserver reliability of lung change classifications

Follow-up time	Kappa score	Mean Kappa score (leave-one-out analysis [SD])	Kappa score (first/last 20 patients)
3 months	0.30	0.30 (0.02)	0.20/0.29
6 months	0.20	0.19 (0.04)	0.03/0.25
12 months	0.25	0.25 (0.03)	0.25/0.16

SD, standard deviation.

At 12 months, the Fleiss' Kappa score was 0.25. The leave-one-out analysis had an identical score (SD: 0.03). Interobserver agreement decreased between the first and last 20 patients with Kappa scores of 0.25 and 0.16, respectively (Table 4).

Discussion

Thoracic SBRT, which was once used only in selected medically inoperable patients, has increased in acceptance in both frail and fit patients.[3,5] This has increased the importance of detecting radiation-induced lung injury and accurately differentiating between tumor recurrence and benign post-treatment changes. Radiation oncologists, with the benefit of knowing beam arrangements and isodose distributions, are well suited to evaluate post-treatment images along with radiologists. However, the interpretation of surveillance imaging is problematic. Despite often dramatic radiographic lung changes post-treatment, the identification of radiographic changes associated with clinical toxicity and/or local treatment failure remains difficult.[14-16,25,26] Radiation oncologists and radiologists need to be familiar with typical post-SBRT lung changes and their evolution over time.

To characterize common post-treatment changes for correlation with toxicity and tumor control outcomes as well as facilitate communication between physicians, specific classifications of benign acute and late post-treatment changes have been proposed.[17-20,23] However, the consistency of these scoring systems has not been demonstrated. The purpose of our analysis was to evaluate interobserver agreement with regard to these categories of benign changes. This report represents the first study to examine the interobserver reliability of scoring acute lung changes in serial scans, which could provide the earliest indication of tumor response or relapse and patient toxicity.

Our findings show that changes in follow-up scans become more common and evolved over time. At 3 months, no change was seen in the plurality of cases and the majority of cases in which uniform scoring occurred. A lack of effect is easier to determine relative to assigning a description of change so the greatest agreement would be expected in these cases. However, the longer the follow up, the greater the incidence of radiographic changes became.

Although no change was observed in 49% of the 3-month scans, this was observed in only 25% of the 6-month scans and 12% of 12-month scans. The decrease in the number of patients with no change is consistent with findings of other studies.[21,22] At 12 months, the modified conventional pattern classification became the most common, which is again consistent with prior studies.[21]

The interpretation of Kappa scores, which assess interreviewer reliability above what is expected by chance, can be categorized as poor (0-0.20), fair (0.20-0.40), moderate (0.40-0.60), strong (0.60-0.80), and almost perfect (0.80-1) interobserver agreement.[27] Kappa scores ranged from 0.20 to 0.30, which is consistent with a fair degree of agreement. These scores are lower than those seen in studies that scored other lung conditions such as interstitial lung diseases (0.30-0.77).[28,29] Although increased category numbers and fewer reviewer and case numbers all decrease Kappa scores, these factors did not differ greatly between the cited studies and our study, which indicates a decreased interobserver agreement in our study.

This lack of reliability likely reflects the high degree of overlap between the change descriptions and the qualitative nature of the scoring. Nonetheless, our findings are consistent with a similar study of 77 patients and 6 reviewers who evaluated reliability of late (6-48 months) post-SBRT lung change classifications that reported Kappa scores that ranged from 0.17 to 0.34.[21] Another study with 8 reviewers produced similar results with a different measure of interobserver agreement using the same 77 patient cohort.[22] Leave-one-out analyses did not identify an outlier observer in the group of raters with the extent of deviation inversely proportional to the overall Kappa scores at the 3 time points.

Our findings demonstrate that interobserver variability peaked at 6 months with no dominant lung change description that emerged (Fig 1). This time point after SBRT could represent the greatest heterogeneity in patient response. Classically, radiation pneumonitis develops 2 to 6 months after SBRT.[6-9,30] At 6 months, some patients likely remain in the acute inflammatory phase of lung reaction while some have developed patterns of fibrosis that are typical of late changes. This could make lung changes at this time point extremely difficult to categorize or require the option of using late lung change descriptions if deemed appropriate. Faruqi et al.[21] reported a Kappa score of 0.34 using late change descriptions to categorize 6-month scans. Allowing the use of both acute and late change categories for 6-month scans may enable better interobserver agreement in the future. Despite that changes at 6 months have the highest interobserver variability, raters at this time point also showed the greatest training effect and the Kappa score for the final 20 patients equaled the Kappa score for the grading of the 12-month lung changes. Given that this time point appeared to give raters the most trouble, the training effect reasonably would be the strongest here.

The training effect was seen for early lung changes and consistent with that of other studies.[21] However, an inverse effect was observed for 12-month changes and interobserver agreement worsened with time. Leave-one-out analyses did not identify an outlier observer in the group. Possibly, the final scans at 12 months were especially difficult to evaluate. This warrants further evaluation in future studies.

Limitations

The qualitative nature of the lung change descriptions make interobserver reliability difficult to demonstrate. With more patients, an increased number of observers, and more rater experience, interobserver agreement could have been improved. Longer follow-up times beyond 12 months also would have allowed for a better characterization of the evolution of late lung changes. Using either acute or late change categories at a given time point may allow for more reliable classifications, especially at 6 months when the agreement was the lowest.

Alternatively, simplifying the ratings into categories such as no/minimal change, change concerning for local recurrence, and change not concerning for local recurrence could improve interobserver reliability and clinical utility. Moving forward, objective measures of lung change such as the Hounsfield Unit threshold-based delineation of hazy and dense lung parenchyma changes,[31] dose-density response curves,[32] texture features, and/or other radiomic assessments[33,34] may offer a better alternative to categorize lung changes than qualitative descriptions.

Conclusions

Interobserver reliability scores indicate a fair agreement for acute and early late lung change classifications after SBRT. Qualitative descriptions of post-SBRT radiographic lung changes are insufficient to categorize these findings because most patient scans do not fit clearly into a single classification. At 6 months post-SBRT, variability in lung changes is the greatest and most difficult to reliably categorize and patients exhibit both acute and late lung changes at that time point. Objective measures of radiographic lung changes or a combination of objective and qualitative measures may offer a better alternative to categorize these changes.

References

1. Ettinger DS, Wood DE, Aisner DL, et al. Non–small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2017;15:504-535.
2. Timmerman RD, Hu C, Michalski J, et al. Long-term results of RTOG 0236: A phase II trial of stereotactic body radiation therapy (SBRT) in the treatment of patients with medically inoperable stage I non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2014;90:S30.
3. Chang JY, Senan S, Paul MA, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: A pooled analysis of two randomised trials. *Lancet Oncol*. 2015;16:630-637.
4. Loganadane G, Martinetti F, Mercier O, et al. Stereotactic ablative radiotherapy for early stage non-small cell lung cancer: A critical literature review of predictive factors of relapse. *Cancer Treat Rev*. 2016;50:240-246.
5. Rutter CE, Corso CD, Park HS, et al. Increase in the use of lung stereotactic body radiotherapy without a preceding biopsy in the United States. *Lung Cancer*. 2014;85:390-394.
6. Baker R, Han G, Sarangkasiri S, et al. Clinical and dosimetric predictors of radiation pneumonitis in a large series of patients treated with stereotactic body radiation therapy to the lung. *Int J Radiat Oncol Biol Phys*. 2013;85:190-195.
7. Barriger RB, Forquer JA, Brabham JG, et al. A dose-volume analysis of radiation pneumonitis in non–small cell lung cancer patients treated with stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys*. 2012;82:457-462.
8. Harder EM, Park HS, Chen Z, Decker RH. Pulmonary dose-volume predictors of radiation pneumonitis following stereotactic body radiation therapy. *Pract Radiat Oncol*. 2016;6:e353-e359.
9. Matsuo Y, Shibuya K, Nakamura M, et al. Dose-volume metrics associated with radiation pneumonitis after stereotactic body radiation therapy for lung cancer. *Int J Radiat Oncol Biol Phys*. 2012;83:e545-e549.
10. Kalman NS, Zhao SS, Anscher MS, Urdaneta AI. Current status of targeted radioprotection and radiation injury mitigation and treatment agents: a critical review of the literature. *Int J Radiat Oncol Biol Phys*. 2017;98:662-682.
11. Newman B, Shah R, Weiss E. Differentiating radiation changes from local recurrence after SBRT for lung cancer: the need for better decision guidelines. *Int J Radiol Radiat Ther*. 2017;4:00106.
12. Allibhai Z, Cho BCJ, Taremi M, et al. Surgical salvage following stereotactic body radiotherapy for early-stage NSCLC. *Eur Respir J*. 2012;39:1039-1042.
13. Huang K, Dahele M, Senan S, et al. Radiographic changes after lung stereotactic ablative radiotherapy (SABR)—Can we distinguish recurrence from fibrosis? A systematic review of the literature. *Radiother Oncol*. 2012;102:335-342.
14. Ronden MI, van Sörnsen de Koste JR, Johnson C, et al. Incidence of high-risk radiologic features in patients without local recurrence after stereotactic ablative radiation therapy for early-stage non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2018;100:115-121.
15. Nguyen TK, Palma DA. Pros: After stereotactic ablative radiotherapy for a peripheral early-stage non-small cell lung cancer, radiological suspicion of a local recurrence can be sufficient indication to proceed to salvage therapy. *Transl Lung Cancer Res*. 2016;5:647-650.
16. Giuliani M, Bezjak A. Cons: After lung stereotactic ablative radiotherapy for a peripheral stage I non-small cell lung carcinoma, radiological suspicion of a local recurrence is not sufficient indication to proceed to salvage therapy. *Transl Lung Cancer Res*. 2016;5:651-654.
17. Trovo M, Linda A, El Naqa I, Javidan-Nejad C, Bradley J. Early and late lung radiographic injury following stereotactic body radiation therapy (SBRT). *Lung Cancer*. 2010;69:77-85.
18. Palma DA, van Sörnsen de Koste J, Verbakel WFAR, Vincent A, Senan S. Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients. *Int J Radiat Oncol Biol Phys*. 2011;81:974-978.
19. Dahele M, Palma D, Lagerwaard F, Slotman B, Senan S. Radiological changes after stereotactic radiotherapy for stage I lung cancer. *J Thorac Oncol*. 2011;6:1221-1228.

20. Kimura T, Matsuura K, Murakami Y, et al. CT appearance of radiation injury of the lung and clinical symptoms after stereotactic body radiation therapy (SBRT) for lung cancers: Are patients with pulmonary emphysema also candidates for SBRT for lung cancers? *Int J Radiat Oncol Biol Phys*. 2006;66:483-491.

21. Faruqi S, Giuliani ME, Raziee H, et al. Interrater reliability of the categorization of late radiographic changes after lung stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys*. 2014;89:1076-1083.

22. Raziee H, Hope A, Faruqi S, et al. Classification and reporting of late radiographic changes after lung stereotactic body radiotherapy: Proposing a new system. *Clin Lung Cancer*. 2015;16:e245-e251.

23. Linda A, Trovo M, Bradley JD. Radiation injury of the lung after stereotactic body radiation therapy (SBRT) for lung cancer: A timeline and pattern of CT changes. *Eur J Radiol*. 2011;79:147-154.

24. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.

25. Ghobadi G, Wiegman EM, Langendijk JA, Widder J, Coppes RP, van Luijk P. A new CT-based method to quantify radiation-induced lung damage in patients. *Radiother Oncol*. 2015;117:4-8.

26. Diot Q, Kavanagh B, Schefter T, Gaspar L, Stuhr K, Miften M. Regional normal lung tissue density changes in patients treated with stereotactic body radiation therapy for lung tumors. *Int J Radiat Oncol Biol Phys*. 2012;84:1024-1030.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

28. Aziz ZA, Wells AU, Hansell DM, et al. HRCT diagnosis of diffuse parenchymal lung disease: Inter-observer variation. *Thorax*. 2004;59:506-511.

29. Antunes VB, Meirelles GDSP, Jasinowodolinski D, et al. Observer agreement in the diagnosis of interstitial lung diseases based on HRCT scans. *J Bras Pneumol*. 2010;36:29-36.

30. Kanemoto A, Matsumoto Y, Sugita T. Timing and characteristics of radiation pneumonitis after stereotactic body radiotherapy for peripherally located stage I lung cancer. *Int J Clin Oncol*. 2015;20:680-685.

31. Kalman NS, Hugo GD, Mahon RN, Deng X, Mukhopadhyay ND, Weiss E. Diabetes mellitus and radiation induced lung injury after thoracic stereotactic body radiotherapy [Abstract]. *Int J Radiat Oncol Biol Phys*. 2017;99:S12.

32. Defraene G, van Elmpt W, Crijns W, Slagmolen P, De Ruysscher D. CT characteristics allow identification of patient-specific susceptibility for radiation-induced lung damage. *Radiother Oncol*. 2015;117:29-35.

33. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.

34. Mahon RN, Weiss E, Ford J, Karki K, Hugo GD. Repeatability of texture features in T1- and T2-weighted MR images [Abstract]. *Med Phys*. 2016;43:3380-3381.