Biomarkers for lung SBRT

CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer

Elizabeth Huynh, Thibaud P. Coroller, Vivek Narayan, Vishesh Agrawal, Ying Hou, John Romano, Idalid Franco, Raymond H. Mak, Hugo J.W.L. Aerts

Department of Radiation Oncology; and Department of Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Article info

Article history:
Received 15 December 2015
Received in revised form 23 May 2016
Accepted 24 May 2016
Available online 10 June 2016

Keywords:
Radiomics
Imaging
Stereotactic body radiation therapy
Lung cancer

Abstract

Background: Radiomics uses a large number of quantitative imaging features that describe the tumor phenotype to develop imaging biomarkers for clinical outcomes. Radiomic analysis of pre-treatment computed-tomography (CT) scans was investigated to identify imaging predictors of clinical outcomes in early stage non-small cell lung cancer (NSCLC) patients treated with stereotactic body radiation therapy (SBRT).

Materials and methods: CT images of 113 stage I-II NSCLC patients treated with SBRT were analyzed. Twelve radiomic features were selected based on stability and variance. The association of features with clinical outcomes and their prognostic value (using the concordance index (CI)) was evaluated. Radiomic features were compared with conventional imaging metrics (tumor volume and diameter) and clinical parameters.

Results: Overall survival was associated with two conventional features (volume and diameter) and two radiomic features (LoG 3D run low gray level short run emphasis and stats median). One radiomic feature (Wavelet LLH stats range) was significantly prognostic for distant metastasis (CI = 0.67, q-value < 0.1), while none of the conventional and clinical parameters were. Three conventional and four radiomic features were prognostic for overall survival.

Conclusion: This exploratory analysis demonstrates that radiomic features have potential to be prognostic for some outcomes that conventional imaging metrics cannot predict in SBRT patients.

© 2016 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 120 (2016) 258–266

Heterogeneous disease within and among patients demands a need for an individualized approach to cancer treatment. Precision medicine aims to design treatment plans tailored to the specific disease profile of the patient to improve outcomes. However, a major challenge for individualized treatment is the inability to accurately predict how a patient's disease will behave and respond to particular therapies prior to treatment [1].

Treatment plans for cancer patients involve one or several treatment modalities involving surgery, chemotherapy and/or radiation therapy (RT). For early stage non-small cell lung cancer (NSCLC) patients, the primary treatment is surgery [2]; however, due to underlying comorbidities, medically inoperable patients are treated with hypofractionated RT, known as stereotactic body radiation therapy (SBRT), as the standard of care [3]. Compared to conventional RT, SBRT administers higher radiation doses over a hypofractionated scheme (e.g. 2 Gy/fraction over 30 fractions for conventional RT vs. 12–18 Gy/fraction over 3–5 fractions for SBRT). SBRT has demonstrated excellent local control, overall survival (OS) and cancer-specific survival (CSS). These promising treatment outcomes have motivated investigations comparing the efficacy of SBRT to surgery, as a potential alternative treatment for surgical candidates [4]. However, despite the successes of SBRT, some patients still develop distant metastases (DM) (13–23%) and local recurrence (4–14%) [5–11]. While early stage patients with larger tumors (i.e. overall stage IB-IIA) who undergo surgical resection may receive adjuvant chemotherapy [12], patients with medically inoperable disease often have comorbidities that limit their ability to tolerate systemic therapy. Therefore, systemic therapy is not a feasible global strategy for all SBRT patients and there is a need for a non-invasive patient stratification approach to identify those who are at highest risk of recurrence after SBRT. Identification of these patients prior to treatment would allow augmentation of their therapeutic approach with addition of systemic therapy and/or radiation dose intensification to reduce disease relapse rates and increase OS [13].
A novel method to classify patients could be based on their tumor phenotype derived from medical imaging. Radiomics offers a non-invasive approach to precision medicine by extracting a large number of advanced quantitative features from medical images to assess the tumor phenotype [1,14,15]. It then involves comprehensive analyses of these features with clinical outcomes as potential prognostic indicators using robust and reproducible methodology [16–18] (Fig. 1). Radiomics generates a unique imaging atlas of the tumor that is a quantification of the tumor phenotype and could provide superior prognostic power over current clinical imaging metrics (e.g. tumor diameter as a predictor of response). Radiomic features have been associated with tumor characteristics, such as genotype and protein expression [19–21], and have been prognostic of clinical outcomes [22–26].

Computed tomography (CT)-based radiomics has immense potential for developing imaging biomarkers for NSCLC patients treated with SBRT since it is the most widely used imaging modality in RT for treatment planning, guidance and follow-up. While quantitative CT imaging has been well reported for lung cancer studies on predicting outcomes in lung cancer patients undergoing SBRT [28–32]. Analysis of baseline Hounsfield Units (HU) and changes in HU or textural features after SBRT have been investigated as prognostic indicators for radiation-induced lung damage [28–30,32] and recurrence [31]. These studies have been limited in their reproducibility and prognostic power prior to SBRT, which would be important for optimizing individualized treatment plans to improve prognosis and/or prevent recurrent disease.

The aim of the current study is to apply an exploratory CT-based radiomics analysis to investigate imaging biomarkers of clinical outcomes in SBRT patients from pre-treatment images. This approach could have a large impact for precision medicine, as radiomic biomarkers are non-invasive and can be applied to imaging data that are already acquired in clinical settings.

**Materials and methods**

**Patient characteristics**

This study was Institutional Review Board (IRB) approved for analysis of non-small cell lung cancer (NSCLC) patients who underwent stereotactic body radiation therapy (SBRT) treatment at our institution between 2009 and 2014. This was a retrospective study and therefore, IRB approval was obtained for waiver of consent. The patient population was limited to patients with early stage NSCLC (overall stage I–II, N0). Patients that did not have a free breathing computed tomography (CT) scan on file (n = 10), had greater than a 1-week duration between CT image acquisition and the start of treatment (n = 2), had multiple SBRT treatments and/or multiple tumor lesions (n = 17), or received induction chemotherapy (n = 2) were excluded from this study. In addition, patients who fulfilled any of the following criteria were also excluded: had metastases to the lung from other sites of primary disease (n = 30), locally recurrent disease (n = 5), had small cell lung cancer (n = 1) or atypical carcinoid (n = 1) histology, or were overall stage III or IV (n = 1). None of the patients received additional chemotherapy after SBRT. A total of 113 patients were included in the analysis and their characteristics can be found in Table 1.

**SBRT treatment and clinical endpoints**

All patients were treated with SBRT according to institutional standards. SBRT was restricted to peripheral tumors as defined in Radiation Therapy and Oncology Group (RTOG) 0236 [9] and abdominal compression was used if tumor motion was greater than 1 cm. Treatment planning was performed on 4D CT scans where the internal target volume was defined, and a planning target volume (PTV) with a 5 mm margin with no clinical target volume margin was created. For tumors close to the chest wall, patients received a dose of 10–12 Gy/C2 fractions or 18 Gy/3 fractions for all other tumors. One patient was unable to complete the full course of treatment due to death and only received 1 fraction of 18 Gy (delivered biologically effective dose of 50.4 Gy). Exact Trac, cone-beam CT and portal imaging using a linear accelerator were used for daily setup and image-guided treatment.

Follow-up chest CT scans with contrast (unless the patient had a contraindication to contrast, e.g. renal dysfunction or allergy) were performed every three to six months after treatment based on United States national guidelines [12] to assess tumor progres-
sion. Distant metastasis (DM) and locoregional recurrence (LRR) were evaluated. The spread of disease to sites outside of the lungs (e.g., brain) was considered DM. LRR was defined as any local, lobar, and/or regional (nodal) recurrence. Recurrence within the PTV was classified as local recurrence, whereas recurrence occurring in the same lobe as the primary tumor but outside of the SBRT treatment field was considered lobar recurrence, and regional recurrence was defined as hilar, mediastinal and supraclavicular lymph node recurrence (Fig. 2).

CT image acquisition and tumor segmentation

Free breathing CT images were acquired on a GE LightSpeed RT16 CT scanner (GE Medical Systems, Milwaukee, WI, USA) according to standard clinical scanning protocols. The most common imaging slice thickness and pixel spacing was 2.5 mm and 1.27 mm by 1.27 mm, respectively. The primary tumor site was contoured on Eclipse software (Varian Medical Systems, Palo Alto, CA, USA). Tumors were manually contoured by E.H., V.A., and Y.H, and then individually verified by an expert radiation oncologist (R.H.M.).

Radiomic feature extraction

A set of 1605 radiomic features describing the tumor phenotype was extracted from the tumors in the free breathing CT images using an in-house Matlab 2013 (The Mathworks Inc., Natick, MA, USA) toolbox and 3D Slicer 4.4.0 software [33]. A bin width of 25 Hounsfield units (HU) for textural features was used to increase sensitivity relative to the raw image. This discretization step reduces image noise and normalizes intensities across all patients, which allowed for a direct comparison of all calculated textural features. All CT voxels were resampled to $1 \times 1 \times 1$ mm$^3$ using a bicubic interpolation function prior to feature extraction to standardize the voxel spacing across the cohort. These features were organized into three categories including shape, statistics and textural features. A description of the radiomics features can be found in the Supplementary material from a previous study [22]. Our feature set contained more features than Coroller et al. [22] as a result of adding textural features that had been calculated using Laplacian of Gaussian (LoG) filters.

### Table 1

<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th>Total ($n = 113$ patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>74 (47–89)</td>
</tr>
<tr>
<td>Gender</td>
<td>57/56 (50.4/49.6)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>8 (7.1)</td>
</tr>
<tr>
<td>Smoking</td>
<td>3/27/83 (2.7/23.9/73.4)</td>
</tr>
<tr>
<td>Performance status</td>
<td>50 (0.4–180.0)</td>
</tr>
<tr>
<td>T stage</td>
<td>17/51/39/6 (15.0/45.1/34.5/5.3)</td>
</tr>
<tr>
<td>Histology</td>
<td>Adenocarcinoma 48 (42.5)</td>
</tr>
<tr>
<td></td>
<td>Adenosquamous carcinoma 1 (0.9)</td>
</tr>
<tr>
<td></td>
<td>Squamous cell carcinoma 27 (23.9)</td>
</tr>
<tr>
<td></td>
<td>Undifferentiated NSCLC 17 (15.0)</td>
</tr>
<tr>
<td></td>
<td>No pathology 20 (17.7)</td>
</tr>
<tr>
<td>T stage</td>
<td>67/27/18/1 (59.3/23.9/15.9/0.9)</td>
</tr>
<tr>
<td>SBRT technique</td>
<td>3D conformal/VMAT 85/28 (75.2/24.8)</td>
</tr>
<tr>
<td>Prescribed radiation dose (Gy)</td>
<td>54 (18–60)</td>
</tr>
<tr>
<td>Radiation dose per fraction (Gy)</td>
<td>18 (10–18)</td>
</tr>
<tr>
<td>Number of radiation fractions</td>
<td>1/3/4/5</td>
</tr>
<tr>
<td>Delivered biologically effective dose (Gy)</td>
<td>151.2 (50.4–151.2)</td>
</tr>
<tr>
<td>Follow-up time (months)</td>
<td>20.8 (0.0–47.8)</td>
</tr>
<tr>
<td>Follow-up time of survivors (months)</td>
<td>25.2 (3.3–47.8)</td>
</tr>
<tr>
<td>Distant metastasis (DM)</td>
<td>90/23 (79.6/20.4)</td>
</tr>
<tr>
<td>Time to event (months)</td>
<td>10.0 (2.0–37.7)</td>
</tr>
<tr>
<td>Estimate of freedom from DM at 2 years</td>
<td>74.0%</td>
</tr>
<tr>
<td>Locoregional recurrence (LRR)</td>
<td>89/24 (78.8/21.2)</td>
</tr>
<tr>
<td>Time to event (months)</td>
<td>8.8 (2.0–26.4)</td>
</tr>
<tr>
<td>Estimate of freedom from LRR at 2 years</td>
<td>70.9%</td>
</tr>
<tr>
<td>Site of recurrence</td>
<td>Regional/lobar/local 17/13/9</td>
</tr>
<tr>
<td>Survival</td>
<td>54/59 (47.8/52.2)</td>
</tr>
<tr>
<td>Time to event (months)</td>
<td>22.5 (0.03–47.8)</td>
</tr>
<tr>
<td>Estimate of survival at 2 years</td>
<td>61.8%</td>
</tr>
<tr>
<td>Cause of death</td>
<td>Cancer/other causes/unknown cause 16/20/18</td>
</tr>
</tbody>
</table>
for tumor characteristics were also included in the study. These conventional features included tumor volume, maximum axial diameter (taken as the maximum diameter measured in a single 2D axial imaging slice) and maximum 3D diameter. The maximum axial diameter and maximum 3D diameter measurements were computed based on the 3D segmentations. The maximum 3D diameter was measured as the largest pairwise Euclidean distance between voxels on the surface of the tumor volume. The maximum axial diameter was taken as the maximum 2D tumor diameter measured on an axial slice. Therefore, the diameters were measured such that the measured segment lay entirely within the solid tumor. The conventional features were not included in the feature dimension reduction process for the radiomic features. The prognostic value of the radiomic features in predicting clinical outcomes were also compared against the performance of clinical parameters, including age, gender, performance status and overall stage.

**Radiomic feature dimension reduction**

A two-step feature dimension reduction method was used for the radiomic features. First, stable features were selected using the test–retest Reference Image Database to Evaluate Therapy Response (RIDER) dataset [34]. The RIDER dataset consists of a series of CT images from 31 NSCLC patients obtained approximately 15 min apart in a similar position. This dataset was only used to select features that were stable through the test–retest assessment with an intraclass correlation coefficient (ICC) greater than 0.8 (using the “irr” package [35]), reducing the number of features to 855 stable features.

Second, Principal Component Analysis (PCA) and factor analysis was applied on our dataset using the “FactoMineR” package [36] to further reduce the resulting high-dimensional dataset to a low-dimensional dataset while retaining most of the variation contained within the data and reducing redundancy. PCA creates a new principal component space using scores that describe the single value decomposition of the features. Scores that retained 95% of the variability from the stable features were selected. Furthermore, the features that correlated by at least 99% to the PCA scores were then selected. This feature dimension reduction process resulted in 12 radiomics features.

**Data analysis**

All statistical analyses were performed in R software version 3.2.2 [37]. Univariate analysis was conducted using the “survcomp” package [38-39] version 1.1.6 from Bioconductor [40]. We analyzed the association between features and each outcome by comparing the feature distribution at the median time of event (24 months for OS and 12 months for LRR and DM). Patients that were censored or did not have an event before the considered time point were not considered in the assessment of the differences in the feature distribution. The number of patients excluded from the association analysis was 14 patients for DM, 25 patients for LRR, 26 patients for OS and 9 patients for CSS. A two-sided Wilcoxon rank-sum test was used to assess the difference in the features between patients with or without an event. Multiple testing correction was applied by the false discovery rate (FDR) procedure introduced by Benjamini and Hochberg [41,42], where a q-value less than 0.1 was considered statistically significant. Boxplots are reported in Supplementary Figs. 1–4. Estimates of survival and event-free probabilities were determined by the Kaplan–Meier method at the median time to event using the “survival” package [43].

The prognostic value of the imaging features was evaluated by calculating the concordance index (CI) using the “survcomp” package in R [38,39]. The CI was calculated from the feature values or clinical parameters and clinical outcome, incorporating the full time data for each outcome while accounting for censoring of the data [44,45]. The CI is a generalization of the area under the receiver operating characteristic curve (AUC) (with the incorporation of time) and a measure of the probability that between two randomly drawn samples, the sample with the higher value (e.g. of an imaging feature) will have a higher likelihood of the event (e.g. DM). A CI greater than 0.5 indicates direct proportionality between the feature value and clinical outcome. Furthermore, a CI less than 0.5 indicates inverse proportionality (the lower the value, the higher the likelihood of the event). Noether’s test was used to compute the p-value to determine the significance of the CI from random (CI = 0.5). Multiple testing correction was then applied by the false discovery rate (FDR) procedure introduced by Benjamini and Hochberg [41,42], where a q-value less than 0.1 was considered statistically significant.

Multivariate models were generated using stratified cross validation by evaluating 100 iterations of each type of model (clinical, conventional, radiomic) using a partition in the dataset of 80% training and 20% validation. We used a conservative partition in the dataset such that the training and validation datasets contained roughly the same proportion of events. The reduced feature set determined by PCA was applied in each training dataset and a set of 3 features were selected only from the respective training dataset to avoid bias, based on the highest univariate performances.

![Fig. 2. Schematic of sub-sites of recurrence. Locoregional recurrence considers all recurrence in the lungs, local recurrence refers to recurrence within the SBRT field, lobar recurrence refers to recurrence occurring in the same lobe as the primary but outside the SBRT field, and regional recurrence refers to recurrence in the lymph nodes in the lungs. The blue box highlights the region of interest for the indicated pattern of recurrence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image-url)
to create each model. The radiomics model generated from the training dataset was evaluated on the validation dataset. Three features were selected for the radiomics signature because choosing 3 features would generate a similar signature for each iteration and capture the best performing features, without the addition of excessive noise. Furthermore, based on the univariate performance of the overall cohort, the top 3 performing features had CIs greater than 0.6. The \( p \)-values for these models were determined using a permutation test with 1000 iterations.

**Results**

A total of 113 early stage non-small cell lung cancer (NSCLC) patients treated with stereotactic body radiation therapy (SBRT) were included in the analysis with a median age of 74 (range: 47–89) at the initiation of treatment. The patient population was nearly equally divided by gender with 50.4% female and 49.6% male. All patients received SBRT with a median biologically effective dose of 151.2 Gy (range: 50.4–151.2 Gy). None of the patients received chemotherapy. The median follow-up time was 20.8 months (range: 0.0–47.8 months) and the median follow-up time of survivors was 25.2 months (range: 3.3–47.8 months). 20.4% of patients developed distant metastasis (DM) and 21.2% developed locoregional recurrence (LRR). The median time to DM and LRR was 10.0 months (range: 2.0–37.7 months) and 8.8 months (2.0–26.4 months), respectively. Overall survival (OS) was 52.2% with a median survival time of 22.5 months (range: 0.03–47.8). The 2-year estimates for DM, LRR and OS were 74.0%, 70.9% and 61.8%, respectively. Patient characteristics and clinical outcomes are shown in Table 1.

The full radiomic feature set was reduced to twelve radiomic features that were selected based on feature stability, preserving the variance in the data and minimizing redundancy among the features. A total of 15 imaging features (3 conventional features and 12 radiomic features) and 4 clinical parameters (age, gender, performance status, overall stage) were included in our analysis. The association between imaging features and clinical parameters with the main clinical outcomes (DM, LRR, OS, cancer-specific survival (CSS)) were evaluated at the median time of the event (Fig. 3). None of the imaging features or clinical parameters were significantly associated with DM, LRR or CSS. Four imaging features were associated with OS including two conventional features (describing tumor diameter and volume), one textural feature (describing characteristics of tumor heterogeneity) and one statistics features (describing image intensity median). One clinical parameter, performance status, was significantly associated with OS. The features and their corresponding \( q \)-values can be found in Supplementary Table 1.

The prognostic power of the imaging features was determined using the concordance index (CI) for each clinical outcome and evaluating the significance of its performance from a random guess. Conventional and radiomic features for DM and LRR are shown in Fig. 4. A statistics-based radiomic feature was prognostic for DM and none of the features were significant for LRR. Notably,

![Heatmap of the association between imaging features and clinical parameters, and clinical outcomes. Imaging features (3 conventional features and 12 radiomic features) were evaluated for distant metastasis (DM), locoregional recurrence (LRR), overall survival (OS) and cancer-specific survival (CSS) with the corresponding \( q \)-value indicated (Wilcoxon rank-sum test, FDR corrected). The considered time point for DM, LRR, OS and CSS was the median time to event (12 months, 12 months, 24 months and 24 months, respectively). \( q \)-value \(< 0.1\).](image-url)
Fig. 4. Prognostic values of imaging features for distant metastasis (DM) and locoregional recurrence (LRR) in SBRT NSCLC patients. Radiomic features are shown in red (statistics features) and blue (textural features), clinical parameters are shown in black and conventional features are shown in gray. The concordance index (CI) is plotted for each feature. “Rand.” indicates the equivalent of a random guess, “Prop.” indicates direct proportionality and “Inv. Prop.” denotes inverse proportionality. * $q$-value < 0.1 (Noether’s test, FDR corrected). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Prognostic values of imaging features for survival in SBRT NSCLC patients. The clinical outcomes evaluated were overall survival (OS) (left) and cancer-specific survival (CSS) (right). Radiomic features are shown in red (statistics features) and blue (textural features), clinical parameters are shown in black and conventional features are shown in gray. The concordance index (CI) is plotted for each feature. “Rand.” indicates the equivalent of a random guess, “Prop.” indicates direct proportionality and “Inv. Prop.” denotes inverse proportionality. * $q$-value < 0.1 (Noether’s test, FDR corrected). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Radiomic analysis of SBRT patients

Discussion

Understanding the behavior of a tumor in response to a particular therapy is crucial for precision medicine in order to design an optimized treatment plan. Medical imaging is not only used for diagnosis and monitoring of cancer patients, but also image-based tumor size metrics derived from these images are used as prognostic indicators of treatment response. However, these current clinical metrics do not describe and exploit all of the tumor information captured in these images. Radiomics assesses a large number of imaging features that characterize the tumor phenotype, using descriptors beyond tumor size, to predict clinical outcomes with increased prognostic power. This approach could be important for patients treated with radiation therapy where treatment optimization is not only based on physical parameters but also the use of biological information, such as those derived from medical images regarding the tumor phenotype, to stratify patients who are at risk of disease recurrence [46]. In particular, this is imperative for early non-small cell lung cancer (NSCLC) patients treated with stereotactic body radiation therapy (SBRT) in which 13–23% of patients develop distant metastasis (DM) and 4–14% of patients experience local recurrence after treatment [5–11].

For OS, the CIs of 7 imaging features (3 conventional, 4 radiomic) were significant from random (Fig. 5). The majority of features had Cls less than 0.5, which indicated that the feature values were inversely proportional to the probability of survival. The Cls of the conventional features ranged from 0.35 to 0.37 (q-value = 0.0024–0.0036). The best performing radiomic feature for OS (stats median) had a Cl of 0.33 (q-value = 0.0016). When the patient population was reduced to evaluate CSS, the number of prognostic features was reduced to 4 (1 conventional, 3 radiomic) (Fig. 5). The four prognostic imaging features for CSS were common with OS. The values of these features can be found in Supplementary Table 4.

The prognostic values of the clinical parameters were also assessed. None of the clinical parameters were prognostic for DM or LRR with CIs ranging from 0.48 to 0.64 (q-value = 0.72–0.81) and 0.39 to 0.68 (q-value = 0.29–0.80), respectively (Supplementary Table 5). Performance status was significantly prognostic for local recurrence (CI = 0.82, q-value = 0.00033) (Supplementary Table 6). Age (CI = 0.35, q-value = 0.076) and performance status (CI = 0.76, q-value = 0.0090) were significantly prognostic for lobular recurrence. None of the clinical parameters were prognostic for regional recurrence with CIs ranging from 0.40 to 0.62 (q-value = 0.80). Age (CI = 0.34, q-value = 0.0011) and performance status (CI = 0.33, q-value = 0.0052) were significantly prognostic for OS and age was prognostic for CSS (CI = 0.37, q-value = 0.028) (Supplementary Table 6).

Multivariate models were generated for DM based on clinical parameters, conventional imaging features and radiomic features using cross validation with 100 iterations. Radiomic multivariate models were generated by applying the reduced radiomic feature set to the training cohort alone, to avoid bias, and evaluating their univariate performances. A radiomic signature composed of the top three univariately performing features from the training dataset was then evaluated on the validation cohort for each iteration. We evaluated DM as the clinical outcome since univariate analysis identified a prognostic radiomic feature for this outcome and it had the most clinical relevance since 13–23% of patients will develop DM [5–11]. The prognostic performance of the radiomics models (median CI of 0.67) was higher than the conventional (median CI of 0.62, p-value = 0.11) and clinical (median CI of 0.62, p-value = 0.18) models, although it was not statistically significant (Supplementary Fig. 6). Within the cross validation, the most frequently occurring features in the radiomics models was Wv LLH stats range, followed by Wv LHL stats total energy (Supplementary Fig. 7). Wv LLH was included in 94 of the 100 models and Wv LHL stats totalenergy was included in 88 of the 100 models (Supplementary Table 8). These features were also the highest univariately performing features for the total cohort (Fig. 3).

Radiomic analysis of SBRT patients

none of the conventional features had significant prognostic value (Supplementary Table 2). The prognostic radiomic feature for DM described the range of voxel intensities (Wavelet LLH stats range, CI = 0.67, q-value = 0.067). Sub-sites of LRR were also evaluated, in which none of the conventional features were prognostic. However, five radiomic features were prognostic of local recurrence (two statistics features and three textural features) and three radiomic features (one statistics feature and two textural features) for lobar recurrence. None of the imaging features (conventional or radiomic) were prognostic for regional recurrence (Supplementary Fig. 5 and Supplementary Table 3).

For OS, the CIs of 7 imaging features (3 conventional, 4 radiomic) were significant from random (Fig. 5). The majority of features had Cls less than 0.5, which indicated that the feature values were inversely proportional to the probability of survival. The Cls of the conventional features ranged from 0.35 to 0.37 (q-value = 0.0024–0.0036). The best performing radiomic feature for OS (stats median) had a CI of 0.33 (q-value = 0.0016). When the patient population was reduced to evaluate CSS, the number of prognostic features was reduced to 4 (1 conventional, 3 radiomic) (Fig. 5). The four prognostic imaging features for CSS were common with OS. The values of these features can be found in Supplementary Table 4.

The prognostic values of the clinical parameters were also assessed. None of the clinical parameters were prognostic for DM or LRR with CIs ranging from 0.48 to 0.64 (q-value = 0.72–0.81) and 0.39 to 0.68 (q-value = 0.29–0.80), respectively (Supplementary Table 5). Performance status was significantly prognostic for local recurrence (CI = 0.82, q-value = 0.00033) (Supplementary Table 6). Age (CI = 0.35, q-value = 0.076) and performance status (CI = 0.76, q-value = 0.0090) were significantly prognostic for lobular recurrence. None of the clinical parameters were prognostic for regional recurrence with CIs ranging from 0.40 to 0.62 (q-value = 0.80). Age (CI = 0.34, q-value = 0.0011) and performance status (CI = 0.33, q-value = 0.0052) were significantly prognostic for OS and age was prognostic for CSS (CI = 0.37, q-value = 0.028) (Supplementary Table 6).

Multivariate models were generated for DM based on clinical parameters, conventional imaging features and radiomic features using cross validation with 100 iterations. Radiomic multivariate models were generated by applying the reduced radiomic feature set to the training cohort alone, to avoid bias, and evaluating their univariate performances. A radiomic signature composed of the top three univariately performing features from the training dataset was then evaluated on the validation cohort for each iteration. We evaluated DM as the clinical outcome since univariate analysis identified a prognostic radiomic feature for this outcome and it had the most clinical relevance since 13–23% of patients will develop DM [5–11]. The prognostic performance of the radiomics models (median CI of 0.67) was higher than the conventional (median CI of 0.62, p-value = 0.11) and clinical (median CI of 0.62, p-value = 0.18) models, although it was not statistically significant (Supplementary Fig. 6). Within the cross validation, the most frequently occurring features in the radiomics models was Wv LLH stats range, followed by Wv LHL stats total energy (Supplementary Fig. 7). Wv LLH was included in 94 of the 100 models and Wv LHL stats totalenergy was included in 88 of the 100 models (Supplementary Table 8). These features were also the highest univariately performing features for the total cohort (Fig. 3).

Discussion

Understanding the behavior of a tumor in response to a particular therapy is crucial for precision medicine in order to design an optimized treatment plan. Medical imaging is not only used for
different preventative treatments if identified prior to SBRT. For example, a high risk of local recurrence would suggest high radiation resistance of the tumor, whereas a high risk for regional recurrence would be indicative of the propensity of the tumor to spread to lymph nodes. We were able to identify radiomic features that were prognostic indicators of local and lobar recurrence, whereas none of the conventional features were able to perform significantly different than random. However, due to the relatively low number of LRR events, these findings would require further investigation and validation in a larger patient cohort.

Evaluation of conventional, clinical and radiomic multivariate models for DM demonstrated that radiomics models could potentially outperform other models, since in our dataset, the radiomics models had a median CI greater than the other models, although the difference was not statistically significant. The features that occurred most frequently in the radiomics multivariate models (Wv LLH stats range and Wv LHL stats totalenergy) may be of future interest for a potential imaging biomarker for DM. Overall, this demonstrates that radiomic models have potential for predicting DM, however, this needs to be confirmed in larger cohorts. Nonetheless, our current work is an exploratory analysis to demonstrate that radiomics can be applied for NSCLC patients treated with SBRT and that it has potential for identification of patients with a high risk of recurrence.

CT-based quantitative imaging applied for lung cancer SBRT patients has focussed primarily on patient radiosensitivity and radiation toxicity [28–32]. Recently, a study by Mattonen et al. used CT texture analysis to identify a predictor of tumor recurrence in 22 NSCLC patients treated with SBRT [31]. They reported that CT textural features extracted from post-SBRT images acquired 2–5 months after treatment could differentiate tumor recurrence from radiation induced lung injury with an area under the receiver operating characteristic curve (AUC) ranging from 0.79 to 0.81. It is important to note that this approach is distinctly different from our current study. Their purpose for differentiating tumor recurrence from lung injury after SBRT was for earlier identification of patients requiring salvage therapies, whereas, our objective is to identify patients with a high risk of recurrence before they begin SBRT treatment. Our approach would allow optimization of the patient’s treatment plan prior to the initiation of treatment to prevent recurrence.

Lasty, overall survival (OS) is an important clinical outcome that usually provides an indication of therapeutic efficacy. We investigated the association and prognostic value of the imaging features for survival. Although OS did have an association with conventional features, OS was also associated with numerous radiomic textural and statistics features. The highest ranked radiomic features had similar prognostic values with the conventional features. The contribution of therapeutic efficacy to OS may be difficult to assess in this patient population due to a median age of 74, which is reflected in the cause of death. Only 29.6% of deaths were due to the cancer, whereas other patient deaths were due to other causes, such as comorbidities, or unknown causes (Table 1). Therefore, we reduced the patient cohort to cancer-related deaths and observed a decrease in the number of prognostic imaging features for cancer-specific survival (CSS), however, the radiomic features still performed similar to the conventional features. This suggests that radiomic features could potentially provide additive information to the current clinical imaging metrics.

A limitation of many quantitative imaging studies on SBRT patients, including our study, is the patient cohort size. Over the past decade, there has been an increase in use of SBRT, especially for lung cancer patients. Pan et al. reported that nearly half of the physicians surveyed on using SBRT in the United States adopted the technique in 2008 or later [49]. The clinical implementation of a new technique or treatment often begins with limited enrollment of patients and a steady increase in treated patients with each successive year. The dataset in our current study is from patients treated with SBRT between 2009 and 2014, resulting in a cohort size of 113 patients. Therefore, due to the number of patients in our dataset, this current study is exploratory and requires further validation in a larger patient cohort in the future.

The patient cohort size also limited our feature reduction method to an unsupervised method that was not based on the patient outcomes. We chose to use principal component analysis (PCA) in order to maintain the meaning of the features (through capturing the variance in the data). The choice of PCA is arguable as it will likely not give the best features for the analysis but it gives a reduced subset of features using an unbiased selection toward the outcomes. These features can then be evaluated for their prognostic performance, as we have demonstrated univariately. The suitability of PCA for dimensionality reduction of the data in this paper is unknown since the data may lie on a nonlinear manifold.

Another shortcoming of our study is that it was limited to a single institution and tumors were delineated and verified by single individuals. Image acquisition and evaluation of clinical outcomes may differ between different institutions. Therefore, future work will be required to evaluate these radiomic features in pooled analyses of independent SBRT datasets to evaluate the generalizability of this study to all early stage NSCLC patients. In addition, the sensitivity of the prognostic accuracy of the features in this study, with respect to tumor contouring variability, is unknown and needs to be measured in a future study involving multiple observers. Lastly, since we evaluated early stage NSCLC (stage I-II) treated with SBRT, this was a limitation for extracting radiomic features due to the small size of the tumors (median 3D diameter = 2.7 cm, range: 0.85–7.1 cm) and hence, may impact the prognostic performance of the radiomic features. However, despite these limitations, we were still able to identify radiomic features that were significantly more prognostic over a random guess.

This study investigated the potential of radiomic features as prognostic indicators of clinical outcomes for early stage NSCLC patients treated with SBRT, and compared their performance with clinical metrics derived from medical images (conventional features) and clinical parameters. In our data set, radiomic features were prognostic indicators for distant metastases, whereas conventional and clinical features were not, and had similar prognostic power to conventional features for survival. This demonstrates that radiomics may have importance in precision medicine for early stage NSCLC SBRT patients by developing prognostic imaging biomarkers for clinical outcomes, although these findings require further exploration and validation in larger cohorts and independent validation datasets. Identification of patients with the highest risk of recurrent disease prior to SBRT treatment would allow clinicians to personalize their treatment plan to reduce the risks of these outcomes and improve survival.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

Authors acknowledge financial support from the National Institute of Health (NIH-USA U24CA194354, and NIH-USA U01CA190234). This project was partially funded by the Kaye Scholar Award and the Brigham and Women’s Hospital Department of Radiation Oncology Clinical Translational Grant. The study sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.