Prediction of distant metastases after stereotactic body radiation therapy for early stage non-small cell lung cancer: development and external validation of a multi-institutional model

Sarah Gao, MD, Lan Jin, PhD, MSc, Hugh W. Meadows, B.A, Timothy Shafman, MD, Cary P. Gross, MD, James B. Yu, MD, MHS, Hugo JWL. Aerts, PhD, Joseph A. Miccio, MD, John M. Stahl, MD, Raymond H. Mak, MD, Roy H. Decker, MD, PhD, Benjamin H. Kann, MD

PII: S1556-0864(22)01909-8
DOI: https://doi.org/10.1016/j.jtho.2022.11.007
Reference: JTHO 2576

To appear in: Journal of Thoracic Oncology

Received Date: 24 June 2022
Revised Date: 3 November 2022
Accepted Date: 6 November 2022


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.
Prediction of distant metastases after stereotactic body radiation therapy for early stage non-small cell lung cancer: development and external validation of a multi-institutional model

Sarah Gao, MD1; Lan Jin, PhD2, MSc; Hugh W Meadows3,4, B.A.; Timothy Shafman,5 MD; Cary P. Gross, MD6; James B. Yu, MD, MHS1,6; Hugo JWL Aerts3,4,7, PhD; Joseph A Miccio, MD8; John M Stahl, MD9; Raymond H. Mak, MD3,4; Roy H. Decker, MD1, PhD; Benjamin H. Kann, MD3,4

1Department of Therapeutic Radiology, Yale School of Medicine, New Haven, CT, USA
2Department of Neurosurgery, Yale School of Medicine, New Haven, CT, USA
3Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, USA.
4Department of Radiation Oncology, Brigham and Women’s Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA.
5GenesisCare, Providence, RI, USA.
6Cancer Outcomes, Public Policy and Effectiveness Research (COPPER) Center, Yale School of Medicine, New Haven, USA.
7Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, The Netherlands.
8Department of Radiation Oncology, Penn State Milton S. Hershey Medical Center, Camp Hill, PA
9Department of Radiation Oncology, University of Alabama at Birmingham, Birmingham, Alabama

Word count: Abstract: 250; Manuscript: 3,481

Acknowledgements:

Conflicts of interest for Dr. James B. Yu:
1. Boston Scientific: Speaking and consulting fees (not relevant to current work)
2. Galera Pharmaceuticals: Advisory board (not relevant to current work)
3. Reflexion: Speaking fees (not relevant to current work)
4. Myovant/Pfizer: Speaking fees (not relevant to current work)

Conflicts of Interest for Dr. Cary P. Gross:
1. National Comprehensive Cancer Network (Pfizer/AstraZeneca): Grant funding (not relevant to current work)
2. Johnson & Johnson: Grant funding (not relevant to current work)
3. Genentech: Grant funding (not relevant to current work)
4. Flatiron: Speaking and traveling fees (not relevant to current work)

Conflicts of Interest for Dr. Raymond H. Mak:
1. AstraZeneca: Consulting
2. ViewRay: Grant funding, Lung Cancer Consortium, and Medical Advisory Board
3. Varian Medical System: Consulting

The other authors have no conflicts of interest to disclose

**Corresponding Author:** Benjamin H. Kann

**Corresponding Author email:** benjamin_kann@dfci.harvard.edu
ABSTRACT

Objectives
Distant metastases (DM) are the primary driver of mortality for patients with early stage non-small cell lung cancer (NSCLC) receiving stereotactic body radiation therapy (SBRT), yet patient-level risk is difficult to predict. We developed and validated a model to predict individualized risk of DM in this population.

Methods
We utilized a multi-institutional database of 1,280 patients with cT1-3N0M0 NSCLC treated with SBRT from 2006-2015 for model development and internal validation. A Fine and Gray (FG) regression model was built to predict 1-year DM risk and compared to a random survival forests (RSF) model. The higher performing model was evaluated on an external dataset of 130 patients from a separate institution. Discriminatory performance was evaluated using the time-dependent area under the curve (AUC). Calibration was assessed graphically and with Brier scores.

Results
The FG model yielded an AUC of 0.71 (95% CI: 0.57-0.86) compared to the RSF’s AUC of 0.69 [95% CI: 0.63-0.85] in the internal test set and was selected for further testing. On external validation, the FG model yielded an AUC 0.70 (95% CI: 0.57-0.83) with good calibration (Brier score 0.08). The model identified a high-risk patient subgroup with greater 1-year DM rates in the internal test (20.0% (3/15) v 2.9% (7/241), P=0.001) and external validation 21.4% (3/15) v 7.8% (9/116), P= 0.095). A model nomogram and online application was made available.

Conclusions
We developed and externally validated a practical model that predicts DM risk in NSCLC patients receiving SBRT that may help select patients for systemic therapy.
INTRODUCTION

Stereotactic body radiotherapy (SBRT) is standard of care for medically inoperable patients with early stage non-small cell lung cancer (NSCLC)\(^1\). Local control (LC) rates for SBRT are reported to be over 90% at 3 years.\(^2,3\) However, a primary driver of morbidity and mortality is the development of distant metastases (DM), which occur in up to 20-30% of patients.\(^4-6\) This suggests a need for systemic treatment intensification and has led to ongoing clinical trials using immunotherapy and targeted agents in this setting.\(^6-11\) Most of these trials are open to all early stage NSCLC patients despite the likelihood that there are subgroups of patients who are at higher risk than others for the development of DM.\(^12,13\) Furthermore, a significant percentage of medically inoperable patients are unable to tolerate systemic therapy due to poor performance status and other medical comorbidities.\(^6\) Previous data have shown that the indiscriminate use of adjuvant systemic therapy for NSCLC without careful patient selection is associated with worse survival outcomes.\(^12\) As such, proper identification of patients with a high risk of developing distant disease is crucial to maximizing the net benefit of adjuvant therapy strategies as well as tailoring follow-up protocols after treatment.

Several factors have been hypothesized to be associated with the increased risk of DM following SBRT, including larger tumor size, higher tumor FDG-avidity, and poor tumor differentiation.\(^13-17\) However, these factors have not been robustly validated, and it remains unclear how they should be used to guide individualized management.

By leveraging large, multi-institutional databases along with statistical and machine learning techniques, we aimed to build a model that can predict the risk of DM
following SBRT and help select patients most likely to benefit from systemic therapy or enrollment in systemic therapy trials.

METHODS

Data Sources and Study Cohorts

We received approval from the Institutional Review Boards and Human Investigation Committees of participating institutions prior to conducting this study. The report is in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement (Type 3).\(^\text{18}\) We evaluated two statistical and machine learning models to predict DM risk while accounting for the competing risk of death due to other causes, which could confound the true risk of distant metastasis.\(^\text{19}\) For model development, we utilized a de-identified multi-institutional database including patient information from five academic sites affiliated with Yale School of Medicine and 109 community sites affiliated with 21st Century Oncology (21C), a national private practice organization in the United States. Data collection for the Yale database was performed by four physician investigators (BK, JM, JS, and SG). Data collection for the 21C database was conducted by a trained data abstractor (RR) using a standardized instrument (Supplement). We included consecutive patients with clinical T1-3N0M0 (AJCC 7th edition) NSCLC without synchronous primary tumors treated with definitive SBRT from 2006 through 2015. Demographic and clinical data collected included age at diagnosis, ECOG status, tumor size (max diameter radiographically, centimeters), tumor histology (adenocarcinoma, squamous cell carcinoma, poorly differentiated NSCLC not otherwise specified
(NSCLC, NOS), and unbiopsied), smoking status, prior lung cancer, maximum tumor standardized uptake value (SUV\text{max}) on PET scan, and prescription biological equivalent dose (BED\text{10}) to 95% of the planning target volume assuming a tumor $\alpha/\beta$ of 10. The lung lobe in which the primary tumor was discovered was documented given data showing that tumor location is correlated with prognosis.\textsuperscript{20,21} Patients were excluded if they had systemic therapy before, during, or after their SBRT treatment course, or received a prescription BED\text{10} of less than 100 Gy to 95% of the planning target volume\textsuperscript{22} (Supplemental Table 1).

For external validation, we used an independent, retrospective dataset from the Brigham and Women’s Hospital/Dana-Farber Cancer Institute (BWH/DFCI) Department of Radiation Oncology, which included consecutive patients with clinical T1-3N0M0 (AJCC 7th edition) NSCLC treated with definitive SBRT from 2010 through 2016. The same exclusion criteria were applied to this patient cohort as the development set and the same demographic and clinical information as described above were abstracted (HWM/RHM). All institutions adhered to NCCN guidelines for diagnostic surveillance imaging for early stage NSCLC following SBRT.

\textit{Study Endpoint}

Patients with radiographically identified distant disease or multifocal pulmonary recurrence were documented as having DM, the primary study endpoint, and were documented as such regardless of preceding or synchronous local and/or regional failures. Months to DM and overall survival (OS) were defined as the time from the first
fraction of SBRT to the date of distant failure or death. Date of death was determined through medical record review when available or web-based search for obituary notice.

**Model Development and Statistical Analysis**

All analyses were performed in R version 4.0.0 (http://www.R-project.org; The R Foundation). We used two methods for model development: 1) Fine and Gray competing risks regression (FG) and 2) random forest survival (RFS) for competing risks. The outcome of interest was the development of DM at one year accounting for the competing risk of death. A summary of the study framework can be found in Figure 1.

The Yale-21C dataset was randomly split into a training set (80%) and a test set (20%). Each model was developed on the training set, locked, and tested on the internal test set. The highest performing model was selected for testing on the external validation set. The primary study endpoint was model discriminatory performance via the area under the curve (AUC) of the time-dependent receiver operating curve at 1-year [33,34]. While we chose a primary evaluation endpoint of 1-year DM risk partly due to the limitation of short median follow-up time in our data, we felt it was an appropriate endpoint to capture high risk patients as the majority of DM occur within one to two years after SBRT. Calibration plots were used to compare the predicted values with the observed values at 1-year, with pairs of observed and predicted values falling in line at a 45-degree angle and Brier score indicating goodness-of-fit.

Our FG model was built using the riskRegression package in R. Univariate competing risks analysis was carried out for each variable. Tumor size (max diameter)
and SUV$_{\text{max}}$ were included in the base model, given the strong evidence and rationale for their association with prognosis.$^{13,26-28}$ Tumor size was dichotomized at 2 cm, which was the median size in the training set and has been used as inclusion criteria for an ongoing systemic therapy trial.$^{11}$ Other variables were included in the final multivariate model using backwards selection if $P<0.1$.

We constructed our RFS model for competing risks using the randomForestSRC package in R (Supplement).$^{29}$ After the FG and RFS models were optimized and locked, we evaluated their discriminatory performance using the Yale-21C internal test set (20%), and the highest performing model was tested on the external validation set.

We calculated the external validation sample size needed to detect a Type I error of 5% with 80% power assuming a null hypothesis of AUC 0.50, an alternative hypothesis based on the internal test set results, and a ratio of negative to positive DM events of 5:1. Power calculations were performed using MedCalc ® v19 (MedCalc Software, Belgium).

To determine if performance was altered by imputing missing data, we conducted a sensitivity analysis using regression imputation for model development (Supplement).

RESULTS

Study Cohort Characteristics

There were 1,280 patients from the combined Yale-21C dataset that met inclusion criteria and were used as training ($n=1,024$) and internal test sets ($n=256$). There were 130 patients from the BWH/DFCI who met criteria for the external validation dataset (Supplemental Figure 1).
Median follow-up in the Yale-21C dataset was 20.7 months (IQ: 9.7-35.9) and 25.4 months (IQ: 15.9-34.4) in the BWH/DFCI dataset (P=0.098) (Table 1). There were 172 (13.4%) overall DM events in the Yale-21C dataset and 23 (17.7%) in the BWH/DFCI dataset (p=0.181). At 12 months, there were 76 (5.9%) DM events in the Yale-21C dataset and 12 (9.2%) in the BWH/DFCI dataset (p=0.139). In the Yale-21C dataset, the median age at diagnosis was higher (77 years (IQ range: 70- 83) vs 74 years (IQ range: 67- 80), P<0.001) and the percentage of tumors greater than 2 cm were higher (49.1% v 36.2%, P=0.005). There were no tumors were greater than 7 cm in either dataset. Only 1.9% of tumors in the Yale-21C dataset measured over 5 cm, while no tumors in the BWH/DFCI dataset measured over 5 cm (Supplemental Table 1). In the Yale-21C dataset, 73.4% (n=940) of patients had an ECOG performance status of <2 compared to 63.9% (n=83) of patients in the BWH/DFCI dataset (P=0.004).

Cumulative incidence curves for DM and death due to other causes are shown in

Supplemental Figure 2.

Model Development and Internal Validation

The final FG model consisted of tumor size, SUV$_{\text{max}}$, age, ECOG status, prior lung cancer, and histology (Table 2). The time-dependent AUC for the model was 0.71 [95% CI: 0.57-0.86, Brier score: 0.04] at 1-year in the internal test set (n=256) (Figure 2).

The RFS model consisted of tumor size, SUV$_{\text{max}}$, age, ECOG status, prior lung cancer, histology, lung cancer lobe location, and smoking history. The RFS model
demonstrated a time-dependent AUC of 0.69 [95% CI: 0.63-0.85, Brier score: 0.04] at 1-year in the internal test set (Supplemental Figure 4).

The time-dependent AUC for the RFS and FG models evaluated on the internal test set were plotted over a range of DM time points from 12 to 36 months (Supplemental Figure 3). Performance was comparable at all time points with overlapping 95% confidence intervals. The significant overlap between the receiver-operator curves of the RFS and FG models suggests there is no significant difference in discriminatory performance between both models. Given its simplicity, reproducibility, and interpretability, the FG model was selected for further testing. Sensitivity analysis using multiple imputation with regression to account for missing data showed no significant differences compared to the original FG model (Supplemental Tables 2 and 3).

**External Validation**

Based on an alternative hypothesis of AUC: 0.70 (based on internal test results), we calculated that an external validation sample size of at least 120 patients would be required for hypothesis testing. Following data abstraction and applying our inclusion and exclusion criteria, our final external validation dataset consisted of 130 patients. The time-dependent AUC for the FG model in the external validation dataset was 0.70 [95% CI: 0.56-0.83, Brier score 0.08] at 12 months, demonstrating good calibration over the predicted risk range (Figure 2).

The FG model was developed into a nomogram for the prediction of DM at 1-year, with values ranging from 4% to 22% (Figure 3). The FG model was also implemented as a R-based ShinyApp and publicly released online (Risk of Distant Metatasis after Lung SBRT (URL: https://predictdm.shinyapps.io/App-1/)).
Delineation of a High-Risk Patient Subgroup

Using the parameters of the nomogram, we defined a subgroup of patients at a high risk of developing DM at 1-year, including patients younger than the median age (77 years), ECOG < 2, tumor size > 2 cm, SUV\textsubscript{max} $\geq$ 2.5, and either adenocarcinoma or NSCLC NOS. The SUV\textsubscript{max} threshold of 2.5 was chosen because this value was historically used to distinguish pulmonary malignancies from benign nodules\textsuperscript{30}. Data has shown that solid malignant lesions typically have a SUV\textsubscript{max} of 2.5 or greater while ground glass lesions have lower SUV\textsubscript{max} values, suggesting that this may be a useful threshold for distinguishing radiographically distinct subsets of tumors\textsuperscript{31,32}.

High-risk patients had greater rates of DM overall and at 1-year in the training set (23.9% (17/71) v 12.5% (119/953), P=0.006 and 18.3% (13/71) v 5.6% (53/953), P<0.001, respectively), the internal test set (26.7% (4/15) v 13.3% (32/241), P=0.148 and 20.0% (3/15) v 2.9% (7/241), P=0.001, respectively), and the external validation dataset (28.6% (4/14) v 16.4% (19/116), P=0.259 and 21.4% (3/15) v 7.8% (9/116), P=0.095, respectively).

DISCUSSION

We developed and externally validated a computational model based on multi-institutional data to predict DM risk for early stage NSCLC patients treated with SBRT. It is the first externally validated model to predict distant metastases in SBRT patients and is publicly available for use online. With stable AUCs of approximately 0.70 on internal and external testing and good calibration, our model demonstrates predictive performance on par with other models that have found routine use in oncologic care.
The model relies on six commonly collected clinical and pathologic variables and, importantly, maintained stable performance (AUC: 0.70) on an external validation set that had large baseline differences in patient and tumor characteristics compared to the development set, portending strong model generalizability. Despite the consistently high rates of local control in early stage NSCLC patients treated with SBRT, distant metastasis remains the dominant reason for treatment failure. This study demonstrates that there is a wide range of DM risk for early stage NSCLC treated with SBRT and that the resulting model may be helpful in selecting patients most likely to benefit from systemic therapy and avoiding indiscriminate use in those at low risk for metastasis. Additionally, patients with a higher risk of distant disease may benefit from intensified follow-up after treatment, and our model can help clinicians determine appropriate follow-up schedules based on patient risk factors.

The National Comprehensive Cancer Network (NCCN) guidelines list adjuvant chemotherapy as a consideration for high-risk patients who receive SBRT, but this high-risk cohort is vaguely defined. As such, several retrospective studies have attempted to determine which patients may benefit from systemic therapy. Some of the most current and robust tools for the prediction of OS, progression-free survival (PFS), and time to progression (TTP) for early stage NSCLC patients are the nomograms constructed by Kang, et. al at MD Anderson Cancer Center. While these nomograms were rigorously developed, they are limited given their reliance on training data from a single institution and lack of external validation. Furthermore, they perform best on overall survival endpoints, not disease progression endpoints (AUC: 0.60 for time-to-progression). This may be related to the inclusion of certain serum inflammatory and
pulmonary markers in the model, which may be associated with overall survival, but not
disease-control or distant metastasis. Additionally, prior nomograms do not make the distinction between risk of death
due to DM or risk of death from unrelated reasons. Considering that most early stage
NSCLC patients who receive SBRT are elderly with multiple comorbidities, this
distinction is important when making decisions regarding the administration of systemic
therapy. Because our model is adjusted for competing risk of death, it may better
generalize across institutions and clinical trial settings where patient age and
performance status distribution vary. Additionally, our model distinguishes distant from
local failure, which each have different implications for treatment escalation approaches.

Several other studies have aimed to identify risk factors associated with the
development of metastatic disease in early stage NSCLC patients, outside of the SBRT
setting. Chong, et. al developed a random forest classifier that predicts the risk of
lymph node metastases in early stage lung adenocarcinoma in a surgical population
with high performance (AUC= 0.921). However, their patient sample all underwent
surgery with mediastinal and hilar lymph node dissection, and their model heavily relies
on pathologic information that is generally unavailable for patients receiving SBRT,
including lymphovascular invasion, pleural invasion, and the presence of a solid
component.

In another study, Gu et al built a logistic regression model using serum
inflammatory markers such as platelet-to-lymphocyte ratios, lactate dehydrogenase
(LDH), neural-specific enolase (NSE), carcinoembryonic antigen, and cytokeratin 19
fragments (Cyfra211) with good discriminatory ability in patients prior to receiving
therapy. This study highlights that serum laboratory markers could be correlated with increased risk of distant metastases. However, it is unclear whether the conclusions of this study can be applied to clinically node negative NSCLC patients who undergo SBRT.

Finally, Wu et. al determined imaging features on PET that were associated with the development of distant metastases in early stage NSCLC receiving surgery using a quantitative radiomic approach. Quantitative imaging features, i.e. Radiomics, represent interesting ways to better predict disease outcomes, though these findings are preliminary, have not been externally validated, and were derived from a small, single-institution cohort.

Compared to prior studies, our model provides immediate clinical relevance in that it utilizes practical clinical variables while still maintaining good discriminatory performance that was stable across different institutions. It is novel in that was developed from a large, mixed community-academic practice population, externally validated, and specific to patients receiving SBRT for ES-NSCLC.

Since more recent machine learning models have shown improved capacity to model complex data, we sought to determine whether we could build a higher performing model using a random survival forest algorithm. Notably, there was no significant difference in performance between our RFS and FG model at all time points evaluated, likely owing to the relatively low number of variables included. Additionally, the greater complexity of the RFS model came at the cost of interpretability while the simplicity of the FG model gave us greater insight into the relationships between the
model variables and the risk of DM. As such, we chose the FG model for further validation and development into a clinically usable tool.

Our FG model included size, $\text{SUV}_{\text{max}}$, age, ECOG status, prior lung cancer, and histology, which each have biologically plausible rationale as to their association with distant metastasis. The importance of tumor size has been demonstrated in multiple studies, including CALGB 9633, which showed that post-operative chemotherapy was associated with improved survival outcomes in patients with resected tumors $\geq$4 cm.$^{13}$ While this data cannot be directly extrapolated to patients who receive SBRT, it is consistent with an abundance of retrospective data that suggest tumor size is an important prognostic feature.$^{40-42}$ In addition to tumor size, $\text{SUV}_{\text{max}}$ values have been shown in numerous studies to be associated with tumor aggressiveness and risk of distant metastasis.$^{14,15,43,44}$ These findings are reflected in an ongoing trial testing the addition of atezolizumab to SBRT for early stage NSCLC patients with tumor size $\geq$2 cm and/or $\text{SUV}_{\text{max}} \geq$ 6.2.$^{11}$ Younger age and better performance status were predictive of a greater risk of metastases, which may be, in part, due to lower competing risk of death and longer follow-up associated with these features, though competing risks modeling likely diminishes this effect compared to survival models. Additionally, Suidan et. al suggested that patients younger than 50 had higher rates of driver mutations and were more likely to develop brain metastases during the course of their disease compared to their older counterparts,$^{45}$ which is consistent with the findings of our study.

There are several limitations to this study. Given the retrospective nature of the data, the results are inevitably subject to selection bias. While patient follow-up was concordant with NCCN guidelines, there may have been discrepancies in surveillance
scan intervals that could have contributed to bias in documenting DM. There were a
substantial number of patients in both datasets (22-29%) that were treated without
pathologic confirmation. While this limits the ability to model certain pathologic markers
that may be associated with distant metastasis in this subset of patients, it is also
represents a real-world inoperable population wherein SBRT sometimes is delivered
without tissue confirmation. Additionally, there may exist other risk factors for DM that
we were unable to capture given database limitations. Tumor grade was one such
variable that we did not include in our model due to the significant number of
undocumented cases. While this limits our ability to study the effects of tumor grade on
metastatic potential, it reflects real-world practice settings when grade is not always
documented on pathology report. Additionally, there is little evidence suggesting a
correlation between grade and prognosis in this setting, and $\text{SUV}_{\text{max}}$, which was robustly
captured in our study, may be a better predictor of tumor aggressiveness.\textsuperscript{46} Emerging
data suggest that tumor driver mutations are important prognostic and predictive factors
for early stage NSCLC patients.\textsuperscript{47} Future models incorporating driver mutation status
will likely further improve decision-making for these patients, though this data is not yet
routinely collected for early stage, inoperable patients. Finally, the median study follow-
up time is less than two years, which may have contributed to a lower-than-expected
incidence of DM overall. As such, our data are not mature enough to capture all future
endpoints and should only inform risk of distant metastases within the first year after
SBRT. However, the majority of distant metastases occur within one to two years of
treatment, and the primary 1-year prediction endpoint accurately identified those
patients at highest risk of dissemination.\textsuperscript{23}
In conclusion, we developed and externally validated a practical competing risks model to predict risk of distant metastasis in early stage NSCLC following SBRT. While our model has fair discriminatory ability, we recommend the model undergo independent and prospective validation given the small sample of our external validation set, and we have released a practical nomogram to facilitate this. Although we believe we have maximized the potential predictive power of routine clinicopathologic variables to predict distant metastases, there is still room for improvement. We hypothesize that genomic, radiomic, and serum-based markers could be leveraged in conjunction with more advanced machine or deep learning algorithms to generate further improvements.


35. Network NCC. Non-Small Cell Lung Cancer. 


Figure Legends

Figure 1. Study framework describing the development and validation of the RFS$^1$ and FG$^2$ models. A developmental dataset was randomly split into a training set (80% of the initial dataset) and a testing set (20% of the initial dataset). Both RFS and FG models were developed on the training set and locked for testing on the testing set. The model with the best discriminatory ability was then tested on an external validation set.

$^1$RFS = Random forest survival

$^2$FG = Fine and Gray

Figure 2. Time-dependent calibration (a) and ROC$^1$ (b) plots for the FG model for competing risks at 12 months in the internal test dataset; Time-dependent calibration (c) and ROC (d) plots for the FG model for competing risks at 12 months in the external validation dataset

$^1$ROC = receiver operating curve

Figure 3. Nomogram for predicting risk of DM$^1$ at 1 year based on Fine and Gray competing risks regression model. The model was developed on 1,024 patients from the study development set. It was validated on an internal hold-out set of 256 patients and an external cohort of 130 patients.

$^1$DM = distant metastasis
Table 1. Patient characteristics

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Yale-21st Century Oncology Lung Cancer Database n= 1280</th>
<th>BWH/DFCI Lung Cancer Database n=130</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median time to distant metastases in months, median (IQ range)</td>
<td>19.5 (8.6-34.5)</td>
<td>23.4 (12.7-32.2)</td>
<td>0.539</td>
</tr>
<tr>
<td>Median follow up time in months, median (IQ range)</td>
<td>20.7 (9.7-35.9)</td>
<td>25.4 (15.7-34.4)</td>
<td>0.098</td>
</tr>
<tr>
<td>Percentage of patients with distant metastases</td>
<td>13.4% (172/1280)</td>
<td>17.7% (23/141)</td>
<td>0.181</td>
</tr>
<tr>
<td>Tumor size in cm, median (IQ range)</td>
<td>2.0 (1.5-2.8)</td>
<td>1.8 (1.5-2.5)</td>
<td>0.004</td>
</tr>
<tr>
<td>SUV\textsubscript{max}, median (IQ range)</td>
<td>6.0 (3.5-10.0)</td>
<td>4.2 (2.4-6.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age in years, median (IQ range)</td>
<td>77 (70- 83)</td>
<td>74 (67- 80)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Median BED\textsubscript{10}, median (IQ range) (fractionation scheme)</td>
<td>132.0 (100.0-132.0)</td>
<td>151.2 (115.5-151.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tumor size in cm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤2</td>
<td>651 (50.9%)</td>
<td>83 (63.9%)</td>
<td>0.005</td>
</tr>
<tr>
<td>&gt;2</td>
<td>629 (49.1%)</td>
<td>47 (36.2%)</td>
<td></td>
</tr>
<tr>
<td>ECOG score</td>
<td></td>
<td></td>
<td>0.020</td>
</tr>
<tr>
<td>&lt;2</td>
<td>940 (73.4%)</td>
<td>83 (63.9%)</td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>340 (26.6%)</td>
<td>47 (36.2%)</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td>0.062</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>444 (34.7%)</td>
<td>51 (39.2%)</td>
<td></td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>343 (26.8%)</td>
<td>27 (20.8%)</td>
<td></td>
</tr>
<tr>
<td>NSCLC, NOS</td>
<td>210 (16.4%)</td>
<td>14 (10.8%)</td>
<td></td>
</tr>
<tr>
<td>Unbiopsied</td>
<td>283 (22.1%)</td>
<td>38 (29.2%)</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td>0.031</td>
</tr>
<tr>
<td>Well differentiated</td>
<td>76 (5.9%)</td>
<td>4 (3.08%)</td>
<td></td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>166 (13.0%)</td>
<td>14 (10.8%)</td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>199 (15.6%)</td>
<td>10 (7.7%)</td>
<td></td>
</tr>
<tr>
<td>Unknown\textsuperscript{2}</td>
<td>556 (43.4%)</td>
<td>64 (49.2%)</td>
<td></td>
</tr>
<tr>
<td>Unbiopsied</td>
<td>283 (22.1%)</td>
<td>38 (29.2%)</td>
<td></td>
</tr>
<tr>
<td>Lung lobe of primary tumor</td>
<td></td>
<td></td>
<td>0.240</td>
</tr>
<tr>
<td>Left upper lobe</td>
<td>328 (25.6%)</td>
<td>36 (27.7%)</td>
<td></td>
</tr>
<tr>
<td>Left lower lobe</td>
<td>181 (14.1%)</td>
<td>22 (16.9%)</td>
<td></td>
</tr>
<tr>
<td>Right upper lobe</td>
<td>447 (34.9%)</td>
<td>33 (25.4%)</td>
<td></td>
</tr>
<tr>
<td>Right middle lobe</td>
<td>91 (7.1%)</td>
<td>9 (6.9%)</td>
<td></td>
</tr>
<tr>
<td>Right lower lobe</td>
<td>233 (18.2%)</td>
<td>30 (23.1%)</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td>0.708</td>
</tr>
<tr>
<td>Never smoker</td>
<td>78 (6.1%)</td>
<td>9 (6.9%)</td>
<td></td>
</tr>
</tbody>
</table>
Prior or current smoker | 1202 (93.9%) | 11 (93.1%) | <0.001
Prior lung cancer | No | 1090 (85.2%) | 85 (65.4%) | Yes | 190 (14.8%) | 45 (34.6%)

1IQ= interquartile range
2Patients categorized as having unknown grade include those with tumor histologic confirmation without documentation of grade

Table 2. Fine Gray competing risks regression analysis of factors correlated with risk of distant metastasis in the internal training set (n=1024)

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Univariable Analysis</th>
<th>Multivariable Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SHR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age in years at diagnosis</td>
<td>0.98</td>
<td>0.96-0.99</td>
</tr>
<tr>
<td>SUV&lt;sub&gt;max&lt;/sub&gt;</td>
<td>1.00</td>
<td>0.98-1.03</td>
</tr>
<tr>
<td>Tumor Size</td>
<td>≤2</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>&gt;2</td>
<td>1.12</td>
</tr>
<tr>
<td>Histology</td>
<td>Adenocarcinoma</td>
<td>--</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>0.72</td>
<td>0.47-1.11</td>
</tr>
<tr>
<td>NSCLC, NOS</td>
<td>1.00</td>
<td>0.63-1.59</td>
</tr>
<tr>
<td>Unbiopsied</td>
<td>0.62</td>
<td>0.38-1.01</td>
</tr>
<tr>
<td>Lung lobe of primary tumor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left upper lobe</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Left lower lobe</td>
<td>0.96</td>
<td>0.53-1.74</td>
</tr>
<tr>
<td>Right upper lobe</td>
<td>1.21</td>
<td>0.76-1.91</td>
</tr>
<tr>
<td>Right middle lobe</td>
<td>1.68</td>
<td>0.90-3.16</td>
</tr>
<tr>
<td>Right lower lobe</td>
<td>1.30</td>
<td>0.77-2.20</td>
</tr>
<tr>
<td>ECOG</td>
<td>&lt;2</td>
<td>--</td>
</tr>
<tr>
<td>≥ 2</td>
<td>0.61</td>
<td>0.39-0.93</td>
</tr>
<tr>
<td>Smoking</td>
<td>Never Smoker</td>
<td>--</td>
</tr>
<tr>
<td>Prior or Current Smoker</td>
<td>1.40</td>
<td>0.57-3.47</td>
</tr>
<tr>
<td>Prior Lung Cancer</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>Yes</td>
<td>1.52</td>
<td>1.01-2.26</td>
</tr>
</tbody>
</table>
Development Dataset (Yale-21C Database) N=1,280

Training Dataset (80%)

Models
1) Fine and Gray
2) Random forest survival

Internal Test (20%)

External Validation
DFCI/BWH N=130

Time-dependent AUC\(^1\) for 12-month distant metastasis

Models Locked

\(^1\) AUC = Area under curve
Credit Roles:

Benjamin H. Kann and Sarah Gao: Conceptualization

Benjamin H. Kann, John M. Stahl, Joseph A Miccio, and Sarah Gao: Data curation

Benjamin H. Kann, Sarah Gao, and Lan Jin: Formal analysis

Funding acquisition: N/A

Benjamin H. Kann and Sarah Gao: Writing - original draft

Hugh W Meadows, Timothy Shafman, Cary P. Gross, James B. Yu, MD, Hugo JWL Aerts, Joseph A Miccio, John M Stahl, Raymond H. Mak, Roy H. Decker, PhD, Benjamin H. Kann, MD: Writing – Review & Editing