ORIGINAL ARTICLE

Prediction of Distant Metastases After Stereotactic Body Radiation Therapy for Early Stage NSCLC: Development and External Validation of a Multi-Institutional Model

Sarah J. Gao, MD,[a] Lan Jin, PhD, MSc,[b] Hugh W. Meadows, BA,[c,d] Timothy D. Shafman, MD,[e] Cary P. Gross, MD,[f] James B. Yu, MD, MHS,[a,f] Hugo J. W. L. Aerts, PhD,[c,d,g] Joseph A. Miccio, MD,[h] John M. Stahl, MD,[i] Raymond H. Mak, MD,[c,d] Roy H. Decker, MD, PhD,[a] Benjamin H. Kann, MD[c,d,*]

[a]*Department of Therapeutic Radiology, Yale School of Medicine, New Haven, Connecticut*
[b]*Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut*
[c]*Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, Massachusetts*
[d]*Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts*
[e]*GenesisCare, Providence, Rhode Island*
[f]*Cancer Outcomes, Public Policy and Effectiveness Research (COPPER) Center, Yale School of Medicine, New Haven, Connecticut*
[g]*Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, the Netherlands*
[h]*Department of Radiation Oncology, Penn State Milton S. Hershey Medical Center, Camp Hill, Pennsylvania*
[i]*Department of Radiation Oncology, University of Alabama at Birmingham, Birmingham, Alabama*

Received 24 June 2022; revised 3 November 2022; accepted 6 November 2022
Available online - XXX

ABSTRACT

Introduction: Distant metastases (DMs) are the primary driver of mortality for patients with early stage 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 used a multi-institutional database of 1280 patients with cT1-3N0M0 NSCLC treated with SBRT from 2006 to 2015 for model development and internal validation. A Fine and Gray (FG) regression model was built to predict 1-year DM risk and compared with a random survival forests model. The higher performing model was evaluated on an external data set 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% confidence interval [CI]: 0.57–0.86) compared with the AUC of random survival forest at 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 of 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 of 15] versus 2.9% [7 of 241], $p = 0.001$) and external validation (21.4% [3 of 15] versus 7.8% [9 of 116], $p = 0.095$). A model nomogram and online application was made available.

*Corresponding author.

Disclosure: Dr. Yu reports receiving speaking and consulting fees (not relevant to current work) from Boston Scientific; serving on the advisory board (not relevant to current work) of Galera Pharmaceuticals; and receiving speaking fees (not relevant to current work) from RefleXion and Myovant/Pfizer. Dr. Gross reports receiving grant funding (not relevant to current work) from the National Comprehensive Cancer Network, Pfizer, AstraZeneca, Johnson & Johnson, and Genentech. Dr. Mak reports having consulting for AstraZeneca; receiving grant funding, being a part of lung cancer consortium, and serving on the medical advisory board of ViewRay; and having consulting for Varian Medical System. The remaining authors declare no conflict of interest.

Address for correspondence: Benjamin H. Kann, MD, Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts. E-mail: benjamin_kann@dfci.harvard.edu

ISSN: 1556-0864
https://doi.org/10.1016/j.jtho.2022.11.007

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

Keywords: NSCLC; SBRT; Distant metastases; Systemic therapy; Nomogram

Introduction

Stereotactic body radiotherapy (SBRT) is standard of care for medically inoperable patients with early stage NSCLC.[1] Local control rates for SBRT are reported to be more than 90% at 3 years.[2,3] Nevertheless, a primary driver of morbidity and mortality is the development of distant metastases (DMs), which occur in up to 20% to 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 patients with early stage NSCLC 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 considerable percentage of medically inoperable patients are unable to tolerate systemic therapy owing to poor performance status and other medical comorbidities.[6] Previous data have revealed 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 and tailoring follow-up protocols after treatment.

Several factors have been hypothesized to be associated with the increased risk of DM after SBRT, including larger tumor size, higher tumor fluorodeoxyglucose avidity, and poor tumor differentiation.[13–17] Nevertheless, 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 after SBRT and help select patients most likely to benefit from systemic therapy or enrollment in systemic therapy trials.

Materials and Methods

Data Sources and Study Cohorts

We received approval from the Institutional Review Boards and Human Investigation Committees of participating institutions before conducting this study. The report is in accordance with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement (type 3).[18] We evaluated two statistical and machine learning models to predict DM risk while accounting for the competing risk of death owing to other causes, which could confound the true risk of DM.[19] For model development, we used a deidentified 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 (American Joint Committee on Cancer seventh edition) NSCLC without synchronous primary tumors treated with definitive SBRT from 2006 to 2015. Demographic and clinical data collected included age at diagnosis, Eastern Cooperative Oncology Group (ECOG) status, tumor size (maximum diameter radiographically, in centimeters), tumor histology (adenocarcinoma, squamous cell carcinoma, poorly differentiated NSCLC not otherwise specified, and unbiopsied), smoking status, prior lung cancer, maximum tumor standardized uptake value (SUV_{max}) on positron emission tomography scan, and prescription biological equivalent dose to 95% of the planning target volume assuming a tumor alpha-beta ratio of 10. The lung lobe in which the primary tumor was discovered was documented given data revealing that tumor location is correlated with prognosis.[20,21]

Patients were excluded if they had systemic therapy before, during, or after their SBRT treatment course or received a prescription biologically effective dose (BED_{10}) of less than 100 Gy to 95% of the planning target volume[22] (Supplementary Table 1).

For external validation, we used an independent, retrospective data set 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 (American Joint Committee on Cancer seventh edition) NSCLC treated with definitive SBRT from 2010 to 2016. The same exclusion criteria were applied to this patient cohort as the development set and the same demographic and clinical information as described previously were abstracted. All institutions adhered to the National Comprehensive Cancer Network (NCCN) guidelines for diagnostic surveillance imaging for early stage NSCLC after SBRT.

Study End Point

Patients with radiographically identified distant disease or multifocal pulmonary recurrence were

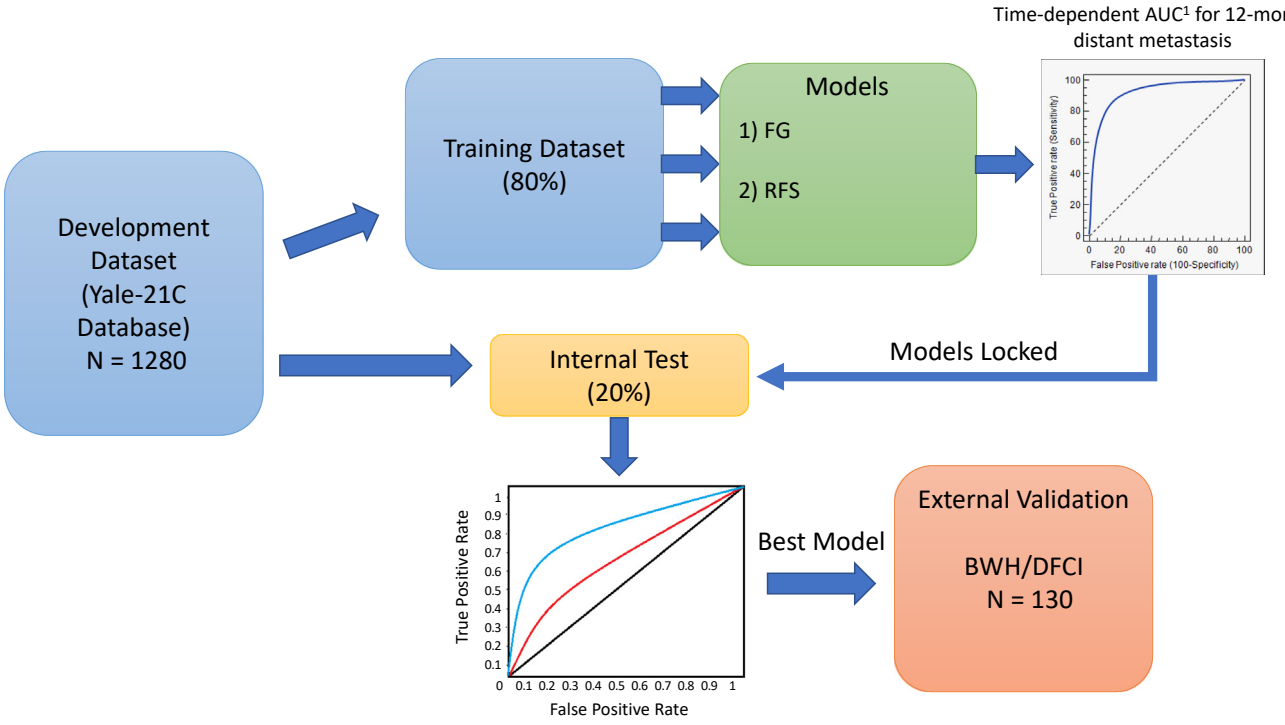

Figure 1. Study framework describing the development and validation of the RFS and FG models. A developmental data set was randomly split into a training set (80% of the initial data set) and a testing set (20% of the initial data set). Both RFS and FG models were developed on the training set and locked for testing on the testing set. The model with the best discriminatory ability was then tested on an external validation set. AUC, area under the curve; BWH/DFCI, Brigham and Women's Hospital/Dana-Farber Cancer Institute; FG, Fine and Gray; RFS, random forest survival.

documented as having DM, the primary study end point, and were documented as such regardless of preceding or synchronous local 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 the following two methods for model development: (1) Fine and Gray (FG) competing risk regression and (2) random forest survival (RFS) for competing risks. The outcome of interest was the development of DM at 1 year accounting for the competing risk of death. A summary of the study framework can be found in Figure 1.

The Yale-21C data set 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 end point was model discriminatory performance by means of the area under the curve (AUC) of the time-dependent receiver operating curve at 1 year. Although we chose a primary evaluation end point of 1-year DM risk partly owing to the limitation of short median follow-up time in our data, we felt that it was an appropriate end point to capture high-risk patients as most DM occur within 1 to 2 years after SBRT.[23] 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° angle and Brier score indicating goodness of fit.

Our FG model was built using the riskRegression package in R.[24,25] Univariate competing risk analysis was carried out for each variable. Tumor size (maximum diameter) and SUV_{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 backward selection if p value is less than 0.1.

We constructed our RFS model for competing risks using the randomForestSRC package in R (Supplementary Materials).[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 on the basis of 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 (Supplementary Materials).

Results

Study Cohort Characteristics

There were 1280 patients from the combined Yale-21C data set who met the inclusion criteria and were used as training (n = 1024) and internal (n = 256) test sets. There were 130 patients from the BWH/DFCI who met the criteria for the external validation data set (Supplementary Fig. 1).

Median follow-up in the Yale-21C data set was 20.7 months (interquartile range [IQR]: 9.7–35.9) and 25.4 months (IQR: 15.9–34.4) in the BWH/DFCI data set ($p = 0.098$) (Table 1). There were 172 (13.4%) overall DM events in the Yale-21C data set and 23 (17.7%) in the BWH/DFCI data set ($p = 0.181$). At 12 months, there were 76 (5.9%) DM events in the Yale-21C data set and 12 (9.2%) in the BWH/DFCI data set ($p = 0.139$). In the Yale-21C data set, the median age at diagnosis was higher (77 y [IQR: 70–83] versus 74 y [IQR: 67–80], $p < 0.001$) and the percentage of tumors greater than 2 cm was higher (49.1% versus 36.2%, $p = 0.005$). There were no tumors greater than 7 cm in either data set. Only 1.9% of tumors in the Yale-21C data set measured more than 5 cm, whereas no tumors in the BWH/DFCI data set measured more than 5 cm (Supplementary Table 1). In the Yale-21C data set, 73.4% (n = 940) of patients had an ECOG performance status of less than 2 compared with 63.9% (n = 83) of patients in the BWH/DFCI data set ($p = 0.004$). Cumulative incidence curves for DM and death owing to other causes are found in Supplementary Figure 2*A* and *B*.

Model Development and Internal Validation

The final FG model consisted of tumor size, SUV_{max}, age, ECOG status, prior lung cancer, and histology (Table 2). The time-dependent AUC for the model was 0.71 (95% confidence interval [CI]: 0.57–0.86, Brier score: 0.04) at 1 year in the internal test set (n = 256) (Fig. 2*A* and *B*).

The RFS model consisted of tumor size, SUV_{max}, age, ECOG status, prior lung cancer, histology, lung cancer lobe location, and smoking history. The RFS model revealed 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 (Supplementary Fig. 3).

The time-dependent AUC for the RFS and FG models evaluated on the internal test set was plotted over a range of DM time points from 12 to 36 months (Supplementary Fig. 4). Performance was comparable at all time points with overlapping 95% CIs. The significant overlap between the receiver-operator curves of the RFS and FG models suggests that 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 revealed no significant differences compared with the original FG model (Supplementary Tables 2 and 3).

External Validation

On the basis of an alternative hypothesis of AUC at 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 data set consisted of 130 patients. The time-dependent AUC for the FG model in the external validation data set was 0.70 (95% CI: 0.56–0.83, Brier score: 0.08) at 12 months, revealing good calibration over the predicted risk range (Fig. 2*C* and *D*).

The FG model was developed into a nomogram for the prediction of DM at 1 year, with values ranging from 4% to 22% (Fig. 3). The FG model was also implemented as an 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 y), ECOG less than 2, tumor size greater than 2 cm, SUV_{max} greater than or equal to 2.5, and either adenocarcinoma or NSCLC not otherwise specified. The SUV_{max} threshold of 2.5 was chosen because this value was

Table 1. Patient Characteristics

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

[a]Patients categorized as having unknown grade include those with tumor histologic confirmation without documentation of grade.
BED_{10}, biologically effective dose; BWH/DFCI, Brigham and Women's Hospital/Dana-Farber Cancer Institute; DM, distant metastasis; ECOG, Eastern Cooperative Oncology Group; IQR, interquartile range; NOS, not otherwise specified; SUV_{max}, maximum tumor standardized uptake value.

historically used to distinguish pulmonary malignancies from benign nodules.[30] Data have revealed that solid malignant lesions typically have a SUV_{max} of 2.5 or greater whereas ground-glass lesions have lower SUV_{max} values, suggesting this may be a useful threshold for distinguishing radiographically distinct subsets of tumors.[31,32]

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

Discussion

We developed and externally validated a computational model on the basis of multi-institutional data to predict DM risk for patients with early stage NSCLC treated with SBRT. It is the first externally validated model to predict DMs 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 reveals predictive performance on par with other models that have found routine use in oncologic care.[33,34] The model relies on six often

Table 2. Fine and Gray Competing Risk Regression Analysis of Factors Correlated With Risk of Distant Metastasis in the Internal Training Set (N = 1024)

	Univariable Analysis			Multivariable Analysis		
Patient Characteristics	SHR	95% CI	*p*	SHR	95% CI	*p*
Age in y at diagnosis	0.98	0.96-0.99	0.010	0.98	0.96-0.99	0.019
SUV_{max}	1.00	0.98-1.03	0.808	1.01	0.98-1.04	0.678
Tumor size						
≤2	–	–	–	–	–	–
>2	1.12	0.80-1.57	0.499	1.20	0.84-1.73	0.323
Histology						
Adenocarcinoma	–	–	–	–	–	–
Squamous cell carcinoma	0.72	0.47-1.11	0.141	0.71	0.45-1.12	0.145
NSCLC, NOS	1.00	0.63-1.59	0.989	0.98	0.61-1.56	0.918
Unbiopsied	0.62	0.38-1.01	0.057	0.63	0.39-1.03	0.063
Lung lobe of primary tumor						
Left upper lobe	–	–	–			
Left lower lobe	0.96	0.53-1.74	0.900			
Right upper lobe	1.21	0.76-1.91	0.418			
Right middle lobe	1.68	0.90-3.16	0.102			
Right lower lobe	1.30	0.77-2.20	0.325			
ECOG						
<2	–	–	–	–	–	–
≥2	0.61	0.39-0.93	0.021	0.62	0.40-0.95	0.027
Smoking						
Never smoker	–	–	–			
Prior or current smoker	1.40	0.57-3.47	0.461			
Prior lung cancer						
No	–	–	–	–	–	–
Yes	1.52	1.01-2.26	0.045	1.47	0.96-2.27	0.080

CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; NOS, not otherwise specified; SHR, subdistribution hazard ratio; SUV_{max}, maximum tumor standardized uptake value.

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 with the development set, portending strong model generalizability. Despite the consistently high rates of local control in patients with early stage NSCLC treated with SBRT, DM remains the dominant reason for treatment failure. This study reveals 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. In addition, 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 on the basis of patient risk factors.

The NCCN guidelines list adjuvant chemotherapy as a consideration for high-risk patients who receive SBRT, but this high-risk cohort is vaguely defined.[35] 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, and time to progression for patients with early stage NSCLC are the nomograms constructed by Kang et al.[16] at MD Anderson Cancer Center. Although 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 OS end points, not disease progression end points (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 OS, but not disease control or DM.[36]

In addition, prior nomograms do not make the distinction between risk of death owing to DM or risk of death from unrelated reasons. Considering that most patients with early stage NSCLC 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

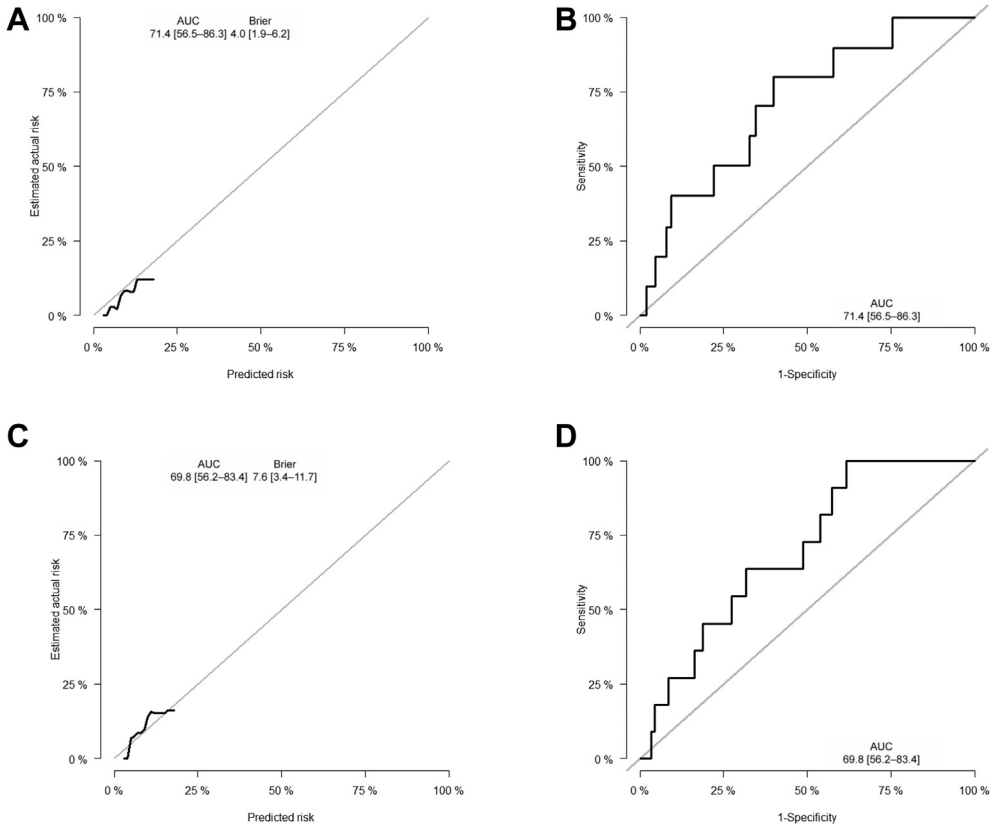

Figure 2. Time-dependent calibration (*A*) and ROC (*B*) plots for the FG model for competing risks at 12 months in the internal test data set. Time-dependent calibration (*C*) and ROC (*D*) plots for the FG model for competing risks at 12 months in the external validation data set. AUC, area under the curve; FG, Fine and Gray; RFS, random forest survival.

performance status distribution vary. In addition, our model distinguishes distant from local failure, which each has different implications for treatment escalation approaches.

Several other studies have aimed to identify risk factors associated with the development of metastatic disease in patients with early stage NSCLC, outside of the SBRT setting. Chong et al.[37] 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).[37] Nevertheless, 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.[38] built a logistic regression model using serum inflammatory markers such as platelet-to-lymphocyte ratios, lactate dehydrogenase, neural-specific enolase, carcinoembryonic antigen, and cytokeratin 19 fragments with good discriminatory ability in patients before receiving therapy. This study highlights that serum laboratory markers could be correlated with increased risk of DMs. Nevertheless, it is unclear whether the conclusions of this study can be applied to clinically node-negative patients with NSCLC who undergo SBRT.

Finally, Wu et al.[14] determined imaging features on positron emission tomography that were associated with the development of DMs in early stage NSCLC receiving surgery using a quantitative radiomic approach. Quantitative imaging features, that is, *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 with prior studies, our model provides immediate clinical relevance in that it uses 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 extensive-stage NSCLC.

Because more recent machine learning models were found to have improved capacity to model complex data, we sought to determine whether we could build a higher

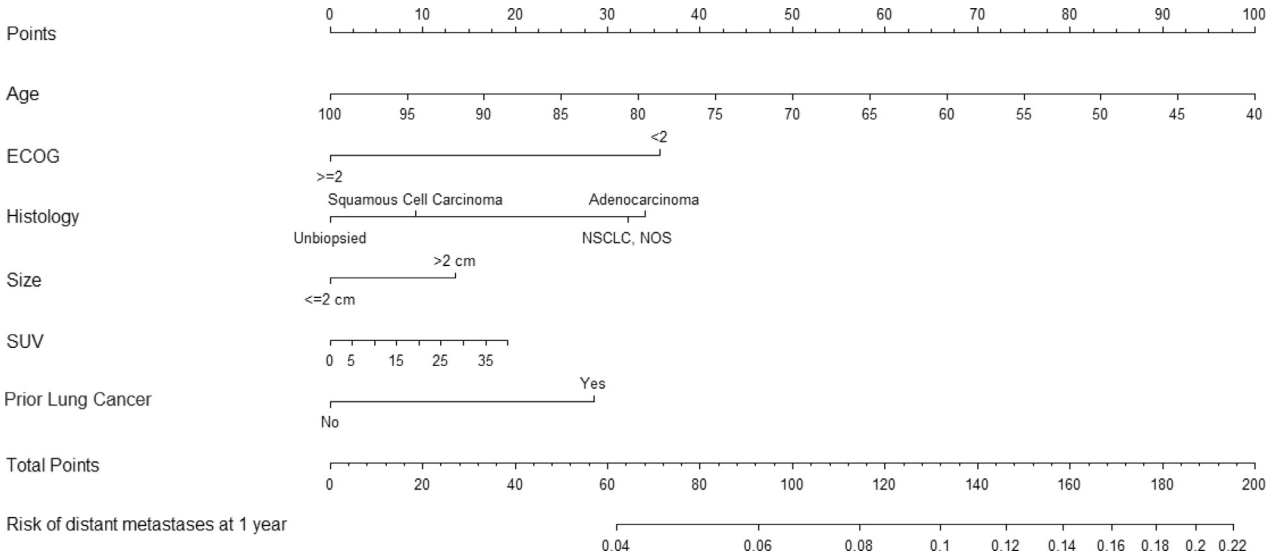

Figure 3. Nomogram for predicting risk of DM at 1 year on the basis of Fine and Gray competing risk regression model. The model was developed on 1024 patients from the study development set. It was validated on an internal hold-out set of 256 patients and an external cohort of 130 patients. DM, distant metastasis; ECOG, Eastern Cooperative Oncology Group; NOS, not otherwise specified; SUV, standardized uptake value.

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. In addition, the greater complexity of the RFS model came at the cost of interpretability whereas 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, SUV_{max}, age, ECOG status, prior lung cancer, and histology, which each has biologically plausible rationale as to their association with DM. The importance of tumor size has been found in multiple studies, including CALGB 9633, which revealed that postoperative chemotherapy was associated with improved survival outcomes in patients with resected tumors greater than or equal to 4 cm.[13] Although these 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.[39–41] In addition to tumor size, SUV_{max} values have been found in numerous studies to be associated with tumor aggressiveness and risk of DM.[14,15,42,43] These findings are reflected in an ongoing trial testing the addition of atezolizumab to SBRT for patients with early stage NSCLC with tumor size greater than or equal to 2 cm and/or SUV_{max} greater than or equal to 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 risk modeling likely diminishes this effect compared with survival models. In addition, Suidan et al.[44] suggested that patients younger than 50 years had higher rates of driver mutations and were more likely to develop brain metastases during the course of their disease compared with their older counterparts, 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. Although patient follow-up was concordant with the 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 data sets (22%–29%) who were treated without pathologic confirmation. Although this limits the ability to model certain pathologic markers that may be associated with DM in this subset of patients, it also represents a real-world inoperable population wherein SBRT sometimes is delivered without tissue confirmation. In addition, 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 owing to the significant number of undocumented cases. Although 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 surgical pathology report. In addition, there is little evidence suggesting a correlation between grade and prognosis in this setting,

and SUV_{max}, which was robustly captured in our study, may be a better predictor of tumor aggressiveness.[45] Emerging data suggest that tumor driver mutations are important prognostic and predictive factors for patients with early stage NSCLC.[46] Future models incorporating driver mutation status will likely further improve decision-making for these patients, though these data are not yet routinely collected for early stage, inoperable patients. Finally, the median study follow-up time is less than 2 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 end points and should only inform risk of DMs within the first year after SBRT. Nevertheless, most DMs occur within 1 to 2 years of treatment, and the primary 1-year prediction end point accurately identified those patients at highest risk of dissemination.[23]

In conclusion, we developed and externally validated a practical competing risk model to predict risk of DM in early stage NSCLC after SBRT. Although 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 that we have maximized the potential predictive power of routine clinicopathologic variables to predict DMs, 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.

CRediT Authorship Contribution Statement

Benjamin H. Kann, Sarah J. Gao: Conceptualization.

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

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

Benjamin H. Kann, Sarah J. Gao: Writing—original draft.

Hugh W. Meadows, Timothy D. Shafman, Cary P. Gross, James B. Yu, MD, Hugo J. W. L. Aerts, Joseph A. Miccio, John M. Stahl, Raymond H. Mak, Roy H. Decker, Benjamin H. Kann: Writing—review and editing.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at https://doi.org/10.1016/j.jtho.2022.11.007.

References

1. Videtic GMM, Donington J, Giuliani M, et al. Stereotactic body radiation therapy for early-stage non-small cell lung cancer: executive summary of an ASTRO evidence-based guideline. *Pract Radiat Oncol*. 2017;7:295-301.
2. Baumann P, Nyman J, Hoyer M, et al. Outcome in a prospective phase II trial of medically inoperable stage I non-small-cell lung cancer patients treated with stereotactic body radiotherapy. *J Clin Oncol*. 2009;27:3290-3296.
3. Wegner RE, Ahmed N, Hasan S, Schumacher LY, Van Deusen M, Colonias A. SBRT for early stage lung cancer: outcomes from biopsy-proven and empirically treated lesions. *Lung Cancer Manag*. 2018;7:LMT01.
4. Cushman TR, Gomez D, Kumar R, et al. Combining radiation plus immunotherapy to improve systemic immune response. *J Thorac Dis*. 2018;10(suppl 3):S468-S479.
5. Lin AJ, Roach M, Bradley J, Robinson C. Combining stereotactic body radiation therapy with immunotherapy: current data and future directions. *Transl Lung Cancer Res*. 2019;8:107-115.
6. Kann BH, Miccio JA, Stahl JM, et al. Stereotactic body radiotherapy with adjuvant systemic therapy for early-stage non-small cell lung carcinoma: a multi-institutional analysis. *Radiother Oncol*. 2019;132:188-196.
7. Ernani V, Appiah AK, Marr A, Marr A, et al. Adjuvant systemic therapy in patients with early-stage NSCLC treated with stereotactic body radiation therapy. *J Thorac Oncol*. 2019;14:475-481.
8. ClinicalTrials.gov. SBRT with immunotherapy in early stage non-small cell lung cancer: tolerability and lung effects. https://clinicaltrials.gov/ct2/show/NCT03383302. Accessed March 22, 2022.
9. ClinicalTrials.gov. Pembrolizumab after lung SBRT for medically inoperable early stage non-small cell lung cancer. https://clinicaltrials.gov/ct2/show/NCT03574220. Accessed March 22, 2022.
10. ClinicalTrials.gov. Stereotactic body radiation therapy (SBRT) combined with avelumab (anti-PD-L1) for management of early stage non-small cell lung cancer (NSCLC). https://www.clinicaltrials.gov/ct2/show/NCT03050554. Accessed March 22, 2022.
11. ClinicalTrials.gov. Testing the addition of the drug atezolizumab to the usual radiation treatment for patients with early non-small cell lung cancer. https://clinicaltrials.gov/ct2/show/NCT04214262. Accessed March 22, 2022.
12. Foster CC, Rusthoven CG, Sher DJ, et al. Adjuvant chemotherapy following stereotactic body radiotherapy for early stage non-small-cell lung cancer is associated with lower overall survival: a National Cancer Database analysis. *Lung Cancer*. 2019;130:162-168.
13. Strauss GM, Herndon JE 2nd, Maddaus MA, et al. Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B. *J Clin Oncol*. 2008;26:5043-5051.
14. Wu J, Aguilera T, Shultz D, et al. Early-stage non-small cell lung cancer: quantitative imaging characteristics

of 18F fluorodeoxyglucose PET/CT allow prediction of distant metastasis. *Radiology*. 2016;281:270–278.
15. Clarke K, Taremi M, Dahele M, et al. Stereotactic body radiotherapy (SBRT) for non-small cell lung cancer (NSCLC): is FDG-PET a predictor of outcome? *Radiother Oncol*. 2012;104:62–66.
16. Kang J, Ning MS, Feng H, et al. Predicting 5-year progression and survival outcomes for early stage non-small cell lung cancer treated with stereotactic ablative radiation therapy: development and validation of robust prognostic nomograms. *Int J Radiat Oncol Biol Phys*. 2020;106:90–99.
17. Wood DE. National Comprehensive Cancer Network (NCCN) clinical practice guidelines for lung cancer screening. *Thorac Surg Clin*. 2015;25:185–197.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg*. 2015;102:148–158.
19. Basak R, Mistry H, Chen RC. Understanding competing risks. *Int J Radiat Oncol Biol Phys*. 2021;110:636–640.
20. Lv X, Cao J, Dai X, Rusidanmu A. Survival rates after lobectomy versus sublobar resection for early-stage right middle lobe non-small cell lung cancer. *Thorac Cancer*. 2018;9:1026–1031.
21. Peleg H, Antkowiak JG, Lane WW, Regal AM, Takita H. Prognosis after resection of non-small cell lung cancer of the right middle lobe. *J Surg Oncol*. 1987;35:230–234.
22. Onishi H, Shirato H, Nagata Y, et al. Hypofractionated stereotactic radiotherapy (HypoFXSRT) for stage I non-small cell lung cancer: updated results of 257 patients in a Japanese multi-institutional study. *J Thorac Oncol*. 2007;2(suppl 3):S94–S100.
23. Pfannschmidt J. Editorial on "Long-term survival outcome after postoperative recurrence of non-small cell lung cancer: who is 'cured' from postoperative recurrence?". *J Thorac Dis*. 2018;10:610–613.
24. Gerds TA, Blanche P, Mortensen R, et al. riskRegression, R Package version 3.5.0. 2021. https://cran.r-project.org/web/packages/riskRegression/riskRegression.pdf; 2021. Accessed March 22, 2022
25. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. *Stat Med*. 2014;33:3191–3203.
26. Pathak R, Goldberg SB, Canavan M, et al. Association of survival with adjuvant chemotherapy among patients with early-stage non-small cell lung cancer with vs without high-risk clinicopathologic features. *JAMA Oncol*. 2020;6:1741–1750.
27. Hanin FX, Lonneux M, Cornet J, et al. Prognostic value of FDG uptake in early stage non-small cell lung cancer. *Eur J Cardio Thorac Surg*. 2008;33:819–823.
28. Horne ZD, Clump DA, Vargo JA, et al. Pretreatment SUVmax predicts progression-free survival in early-stage non-small cell lung cancer treated with stereotactic body radiation therapy. *Radiat Oncol*. 2014;9:41.
29. Ishwaran H. Fast unified random forests for survival, regression, and classification (RF-SRC), R package version 2.9.3. Vol. 2020. https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf. Accessed March 22, 2022.
30. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA*. 2001;285:914–924.
31. Heyneman LE, Patz EF. PET imaging in patients with bronchioloalveolar cell carcinoma. *Lung Cancer*. 2002;38:261–266.
32. Nomori H, Watanabe K, Ohtsuka T, Naruke T, Suemasu K, Uno K. Evaluation of F-18 fluorodeoxyglucose (FDG) PET scanning for pulmonary nodules less than 3 cm in diameter, with special reference to the CT images. *Lung Cancer*. 2004;45:19–27.
33. Tendulkar RD, Agrawal S, Gao T, et al. Contemporary update of a multi-institutional predictive nomogram for salvage radiotherapy after radical prostatectomy. *J Clin Oncol*. 2016;34:3648–3654.
34. Creutzberg CL, van Stiphout RG, Nout RA, et al. Nomograms for prediction of outcome with or without adjuvant radiation therapy for patients with endometrial cancer: a pooled analysis of PORTEC-1 and PORTEC-2 trials. *Int J Radiat Oncol Biol Phys*. 2015;91:530–539.
35. National Comprehensive Cancer Network. Non-small cell lung cancer. https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf. Accessed March 22, 2022.
36. Tong YS, Tan J, Zhou XL, Song YQ, Song YJ. Systemic immune-inflammation index predicting chemoradiation resistance and poor outcome in patients with stage III non-small cell lung cancer. *J Transl Med*. 2017;15:221.
37. Chong Y, Wu Y, Liu J, et al. Clinicopathological models for predicting lymph node metastasis in patients with early-stage lung adenocarcinoma: the application of machine learning algorithms. *J Thorac Dis*. 2021;13:4033–4042.
38. Gu W, Hu M, Wang W, Shi C, Mei J. Development and validation of a novel nomogram for predicting tumor-distant-metastasis in patients with early T1-2 stage lung adenocarcinoma. *Ther Clin Risk Manag*. 2020;16:1213–1225.
39. Dubben HH, Thames HD, Beck-Bornholdt HP. Tumor volume: a basic and specific response predictor in radiotherapy. *Radiother Oncol*. 1998;47:167–174.
40. Dehing-Oberije C, De Ruysscher D, van der Weide H, et al. Tumor volume combined with number of positive lymph node stations is a more important prognostic factor than TNM stage for survival of non-small-cell lung cancer patients treated with (chemo)radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;70:1039–1044.
41. Shan Q, Fan Y, Guo J, Han X, Wang H, Wang Z. Relationship between tumor size and metastatic site in patients with stage IV non-small cell lung cancer: a large SEER-based study. *PeerJ*. 2019;7:e7822.
42. Cerfolio RJ, Bryant AS, Ohja B, Bartolucci AA. The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival. *J Thorac Cardiovasc Surg*. 2005;130:151–159.
43. Yildirim F, Yurdakul AS, Türk M, Akdemir ÜÖ, Öztürk C. The affect of the metabolic activity of the primary tumor to lymph node and distant metastasis in non-small cell lung cancer. *Eur Respir J*. 2014;44(suppl 58):344.

44. Suidan AM, Roisman L, Belilovski Rozenblum A, et al. Lung cancer in young patients: higher rate of driver mutations and brain involvement, but better survival. *J Glob Oncol*. 2019;5:1–8.
45. Duan XY, Wang W, Li M, Li Y, Guo YM. Predictive significance of standardized uptake value parameters of FDG-PET in patients with non-small cell lung carcinoma. *Braz J Med Biol Res*. 2015;48:267–272.
46. Chevallier M, Borgeaud M, Addeo A, Friedlaender A. Oncogenic driver mutations in non-small cell lung cancer: past, present and future. *World J Clin Oncol*. 2021;12:217–237.