International Journal of Radiation Oncology biology • physics

www.redjournal.org

Clinical Investigation

Predicting Overall Survival After Stereotactic Ablative Radiation Therapy in Early-Stage Lung Cancer: Development and External Validation of the Amsterdam Prognostic Model

Alexander V. Louie, MD, MSc,[*,†,‡] Cornelis J.A. Haasbeek, MD, PhD,[*] Sahar Mokhles, BSc,[§] George B. Rodrigues, MD, PhD,[†] Kevin L. Stephans, MD,[‖] Frank J. Lagerwaard, MD, PhD,[*] David A. Palma, MD, PhD,[†] Gregory M.M. Videtic, MD, CM,[‖] Andrew Warner, MSc,[†] Johanna J.M. Takkenberg, MD, PhD,[§] Chandana A. Reddy, MS,[‖] Alex P.W.M. Maat, MD,[§] Neil M. Woody, MD,[‖] Ben J. Slotman, MD, PhD,[*] and Suresh Senan, MRCP, FRCR, PhD[*]

Department of Radiation Oncology, VU University Medical Center, Amsterdam, the Netherlands; [†]Department of Radiation Oncology, London Regional Cancer Program, University of Western Ontario, London, Ontario, Canada; [‡]Department of Epidemiology, Harvard School of Public Health, Harvard University, Boston, Massachusetts; [§]Department of Cardio-Thoracic Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands; and [‖]Department of Radiation Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio

Received Apr 1, 2015, and in revised form Apr 30, 2015. Accepted for publication May 1, 2015.

Summary

Survival after stereotactic ablative radiation therapy (SABR) for early-stage non-small cell lung cancer (ES-NSCLC) patients is variable, owing to various tumor, treatment, and patient

Purpose: A prognostic model for 5-year overall survival (OS), consisting of recursive partitioning analysis (RPA) and a nomogram, was developed for patients with early-stage non-small cell lung cancer (ES-NSCLC) treated with stereotactic ablative radiation therapy (SABR).

Methods and Materials: A primary dataset of 703 ES-NSCLC SABR patients was randomly divided into a training (67%) and an internal validation (33%) dataset. In the former group, 21 unique parameters consisting of patient, treatment, and tumor factors were entered into an RPA model to predict OS. Univariate and multivariate models were constructed for RPA-selected factors to evaluate their relationship with

Reprint requests to: Dr Alexander V. Louie, MD, MSc, Department of Radiation Oncology, London Regional Cancer Program, 790 Commissioners Rd E, N6A 4L6 London, ON, Canada. Tel: (519) 685-8500; ext. 53025; ext. 58627; E-mail: Dr.alexlouie@gmail.com

This work was supported by a 2013 CARO-Elekta Research Fellowship, a Western University Resident Research Career Development Program Award, and a 2014 Detweiler Travelling Fellowship, Royal College of Physicians and Surgeons of Canada (A.V.L.).

Conflict of interest: F.J.L., S.S., and B.J.S. have received honoraria and travel support from Varian Medical Systems. B.J.S. has received honoraria and travel support from BrainLAB AG. VU University Medical Center has a research agreement with Varian Medical Systems.

Supplementary material for this article can be found at www.redjournal.org.

http://dx.doi.org/10.1016/j.ijrobp.2015.05.003

factors. Unlike ES-NSCLC patients who undergo surgery, validated prognostic models are lacking for those treated with SABR. To enable clinicians to more precisely estimate the survival of individual ES-NSCLC patients after SABR, we developed and validated the Amsterdam prognostic model, a tool that has been made available in an online application.

OS. A nomogram for OS was constructed based on factors significant in multivariate modeling and validated with calibration plots. Both the RPA and the nomogram were externally validated in independent surgical (n = 193) and SABR (n = 543) datasets.

Results: RPA identified 2 distinct risk classes based on tumor diameter, age, World Health Organization performance status (PS) and Charlson comorbidity index. This RPA had moderate discrimination in SABR datasets (c-index range: 0.52-0.60) but was of limited value in the surgical validation cohort. The nomogram predicting OS included smoking history in addition to RPA-identified factors. In contrast to RPA, validation of the nomogram performed well in internal validation ($r^2 = 0.97$) and external SABR ($r^2 = 0.79$) and surgical cohorts ($r^2 = 0.91$).

Conclusions: The Amsterdam prognostic model is the first externally validated prognostication tool for OS in ES-NSCLC treated with SABR available to individualize patient decision making. The nomogram retained strong performance across surgical and SABR external validation datasets. RPA performance was poor in surgical patients, suggesting that 2 different distinct patient populations are being treated with these 2 effective modalities.

Introduction

The advent of stereotactic ablative radiation therapy (SABR) is a major advance in curative treatment for early-stage non-small cell lung cancer (ES-NSCLC) (1). In the Netherlands, the introduction of SABR for ES-NSCLC correlated with a population improvement in overall survival (OS) in elderly patients (>75 years), primarily due to increased use of SABR in medically inoperable patients who would otherwise have been left untreated (2). Despite this population-level benefit, survival of patients after lung SABR is variable (3).

A major challenge for the judicious use of curative treatment in ES-NSCLC is the ability to accurately predict life expectancy. This issue is compounded by the difficulty in determining each factor's predictive magnitude when combined in individual patients (4). In ES-NSCLC patients who are being considered for surgery, externally validated tools to predict perioperative mortality and OS can be used to guide clinical decision making (5-7). Just as surgeons should be careful in operating on patients with medical comorbidities (8), radiation oncologists must be careful when selecting patients most likely to benefit from SABR (9). To our knowledge, currently no validated instrument exists to assist in determining the prognosis of ES-NSCLC patients treated with SABR (10), which would be useful in maximizing benefit to the broader community and minimizing undertreatment (11).

Two types of prognostic models are commonly used in this situation. Recursive partitioning analysis (RPA) creates decision trees that stratify members of a population into different groups based on dichotomous covariates. Nomograms allow for prognostication at the individual level. In this study, we developed both an RPA model and a nomogram for OS by using a large single-institution cohort of ES-NSCLC SABR patients. In developing the Amsterdam prognostic model (APM) for ES-NSCLC SABR, we measured its performance in 2 independent datasets composed of surgical and SABR patients.

Methods and Materials

Primary dataset for RPA and creation of the nomogram

The VU University Medical Center (VUMC) maintains a database of ES-NSCLC patients treated with SABR. All patient cases are discussed in a multidisciplinary tumor board before being accepted for treatment, and when no pathology is available, patients are treated in accordance with guidelines of the European Society for Medical Oncology (12). Details for baseline characteristics, treatment, and follow-up information are prospectively entered. SABR was delivered using a risk-adapted scheme of 54 Gy in 3 fractions, 55 Gy in 5 fractions, or 60 Gy in 8 fractions, all based on tumor size and location. Treatment planning and follow-up details have been described previously (13).

A total of 1136 patients were identified from the VUMC database between January 1, 2003, and December 31, 2012. The following patients were excluded from analysis: any diagnosis of malignancy (except for basal cell cancer of the skin) within 2 years of ES-NSCLC, metastatic lung tumors, multiple lung tumors, and small-cell lung cancer diagnosis. After excluding ineligible patients, we selected the remaining 703 patients for the primary dataset and randomly divided them into a training (n = 469 [67%]) and a validation (n = 234 [33%]) dataset.

External validation of RPA and the nomogram

Two independent datasets consisting of clinically staged surgical and SABR ES-NSCLC patients were used for external validation of the derived models. Diagnostic and treatment details for these patients from the Erasmus Medical Center (EMC) (surgery, n = 196) and Cleveland Clinic (CC) (SABR, n = 543) are summarized in Supplemental File E1 (available online at www.redjournal.com) and also have been described previously in more detail (14, 15). In both

external validation datasets, descriptive statistics were generated and compared with VUMC patients, using χ^2, Fisher exact, or 2-sample *t* test, as appropriate (Supplemental Table E1; available online at www.redjournal.com).

Medical ethics review for this study was not obtained for the VUMC and EMC datasets, because in the Netherlands, retrospective studies of patient records, as in the present study, do not fall under the scope of the Medical Research Involving Human Subjects Act. Institutional Review Board approval was obtained to use the CC dataset for the purposes of this study.

Model creation

Using a random number generator, patients from the primary (VUMC) dataset were dichotomized into a training set (two-thirds) and an internal validation set (one-third) without stratification. Descriptive statistics were generated for baseline patient (age, sex, World Health Organization [WHO] performance status [PS], smoking status, chronic obstructive pulmonary disease Global Initiative for Chronic Obstructive Lung Disease [GOLD] score, Charlson comorbidity index [CCI], previous malignancy, time from previous malignancy, previous lung cancer), time from previous lung cancer, tumor (maximum tumor diameter, T stage, cancer stage, laterality, use of positron emission tomography [PET], pathology-proven histology, location, lobe), and treatment (number of fractions, biological equivalent dose [BED_{10}]) characteristics for all patients, and were compared between training and validating sets. BED_{10} was calculated assuming that the tumor alpha:beta ratio was 10 Gy (16). The primary endpoint was OS at 5 years and was calculated using the Kaplan-Meier method from the time of treatment initiation to the date of last follow-up or death. Follow-up was calculated using the reverse Kaplan-Meier method. Patients alive as of May 29, 2014, were censored as of that date. To examine potential discrepancies related to possible inaccurately coded deaths or death dates, patient death status was verified using the Dutch national death registry.

Statistical analysis

RPA was performed using the training dataset to predict the primary endpoint based on the inventory of 21 factors. In the RPA procedure, R software (version 3.0.3, open source; www.r-project.org) default settings were used, where a minimum number of 20 observations in a node were required to enable further splitting, followed by trimming of less important downstream branches as needed. A default of a minimum of 7 observations was required for a terminal node. Rounding of cutpoints to the nearest significant digit was performed to increase clinical utility. OS rates were compared between the training and validating dataset RPA risk groups, using the log-rank test.

Each RPA risk group was evaluated and compared between training and validating (internal and external) sets, using the log-rank test. Univariate Cox regression was performed to evaluate RPA risk group in terms of its ability to predict OS, separately for the training and validation datasets. Univariate Cox regression was also performed using RPA-selected factors to identify significant predictors of OS. Covariates with *P* values of $<.05$ for the training set were entered into a multivariate model to confirm significant predictors of OS. The final multivariate model obtained based on the training dataset was also assessed using the validation dataset.

A nomogram based on the final multivariate Cox regression model for OS (using the training dataset only) was generated to calculate individual patient-level probability estimates for 5-year OS according to each patient's unique combination of baseline characteristics. Nomogram equations were created to first calculate and then assign a total number of points per patient based on known baseline characteristics. Internal and external validations of the nomogram were performed via calibration plots of Kaplan-Meier—observed estimates versus nomogram-predicted probability for 5-year OS (17). The predictive accuracy and discriminative ability of the models were measured using the concordance index (C-index) and goodness-of-fit (r^2). After creation and assessment of the RPA model and nomogram on the primary datasets, they were separately evaluated using the 2 external validation datasets.

All statistical analyses were performed using SAS version 9.3 software (SAS institute, Cary, USA) and R software, using 2-sided statistical testing at the 0.05 significance level.

Results

Patient demographics

The distribution of baseline patient, tumor, and treatment characteristics for RPA-selected variables stratified between training and internal/external validation datasets is summarized in Table 1. The remaining stratified factors that were entered as covariates in the RPA are summarized in Supplemental Table E1 (available online at www.redjournal.com). Compared to VUMC patients, CC patients were more likely to be female and have higher rates of pathologic confirmation, better GOLD score and T-stage, and less use of PET staging (all $P<.001$). EMC patients were more commonly T1 stage, and tended to be more fit and have lower CCI, better GOLD scores, and improved PS (all $P<.001$). EMC patients were also younger and had lower rates of previous malignancy and higher rates of pathologic confirmation of malignancy and lower lobe location of disease (all $P<.001$).

Survival results and prognostic factors

The median follow-up periods for VUMC, EMC, and CC patients were 64.2, 63.0, and 34.1 months, respectively. The median OS periods for VUMC and CC patients were 40.2 and 31.2 months, respectively, and was not reached for

Table 1 Baseline tumor, patient, and treatment characteristics of all patients, training set, and validation set for RPA/nomogram selected factors

Characteristic	VUMC	Training set	Validation set	EMC	CC
Age					
Mean ± SD	72.9 ± 8.8	73.3 ± 8.7	71.9 ± 8.9	64.6 ± 9.3	73.2 ± 9.7
Median	74.4	74.5	73.4	65.2	74.0
Age distribution n (%)*					
<75 (y)	378 (53.8)	249 (53.1)	129 (55.1)	167 (85.2)	279 (51.4)
≥75 (y)	325 (46.2)	220 (46.9)	105 (44.9)	29 (14.8)	264 (48.6)
WHO performance status no. (%)					
0	84 (12.0)	52 (11.1)	32 (13.7)	131 (66.8)	12 (2.2)
1	354 (50.4)	240 (51.2)	114 (48.9)	56 (28.6)	332 (61.4)
2	229 (32.6)	152 (32.4)	77 (33.1)	8 (4.1)	190 (35.1)
3	35 (5.0)	25 (5.3)	10 (4.3)	1 (0.5)	7 (1.3)
Charlson comorbidity index					
Mean ± SD index	2.5 ± 1.7	2.5 ± 1.7	2.5 ± 1.9	1.6 ± 1.0	2.6 ± 1.6
Distribution of Charlson index no. (%)*					
0-2	418 (59.5)	273 (58.2)	145 (62.0)	162 (82.7)	279 (51.4)
≥3	285 (40.5)	196 (41.8)	89 (38.0)	34 (17.4)	264 (48.6)
No. with GOLD score shown (%)					
0	147 (21.3)	103 (22.3)	44 (19.3)	91 (46.4)	99 (20.8)
1	95 (13.8)	67 (14.5)	28 (12.3)	105 (53.6)	206 (43.3)
2	219 (31.7)	139 (30.0)	80 (35.1)	-	122 (25.6)
3	172 (24.9)	121 (26.1)	51 (22.4)	-	49 (10.3)
4	58 (8.4)	33 (7.1)	25 (11.0)	-	
Mean ± SD diameter (mm)	28.9 ± 12.0	29.3 ± 12.2	28.1 ± 11.5	30.6 ± 16.1	26.0 ± 13.0
No. with diameter (mm) shown (%)*					
<20	174 (24.8)	112 (23.9)	62 (26.5)	50 (25.5)	186 (34.3)
≥20	529 (75.3)	357 (76.1)	172 (73.5)	146 (74.5)	356 (65.6)
Smoker	667 (97.2)	445 (97.2)	222 (97.4)	179 (91.3)	517 (95.2)

Abbreviations: CC = Cleveland Clinic; EMC = Erasmus Medical Center; GOLD = Global Initiative for Chronic Obstructive Lung Disease; RPA = recursive partitioning analysis; VUMC = VU University Medical Center; WHO = World Health Organization.
* RPA selected cutoff point.

EMC patients. In contrast to the SABR validation datasets, EMC patients had improved OS compared to the training dataset ($P<.001$).

RPA: Model development and VUMC internal validation

Initial RPA using the VUMC training dataset resulted in a 3-class stratification; class 1 tumor diameter was <23 mm and age <75 years; class 2 had tumor diameter of ≥23 mm, WHO PS of 0-1, and CCI of 0-2; and class 3 had tumor diameter of <23 mm and were ≥75 years of age or tumor diameter of ≥23 mm, a WHO PS of 0-1, and CCI of ≥3, or a tumor diameter of ≥23 mm and WHO PS of 2-3. A sensitivity analysis of the RPA-selected tumor diameter was performed using cutpoints of 20 and 25 mm. As these did not significantly alter differences in OS for the 3 RPA stratification classes, 20 mm was selected as the tumor diameter consensus, consistent with the T1a and T1b inflection points using the American Joint Committee on Cancer (AJCC) staging system. Finally, only a trend in OS difference was noted between classes 1 and 2 in the training dataset ($P=.059$), a finding that was not demonstrated in the internal validation dataset ($P=.684$). Accordingly, classes 1 and 2 were collapsed into a single class. In this final recursive partitioning model for OS (Fig. 1), there were significant differences in OS between RPA class 1 and 2 patients on univariate Cox analysis in the training (hazard ratio [HR, 95% confidence interval]: 1.86 [1.44-2.38], $P<.001$) and internal validation (HR: 1.95 [1.34-2.84], $P<.001$) datasets. The C-index used to quantify RPA stratification discrimination demonstrated moderate performance in the training (0.58) and internal validation (0.58) datasets.

RPA external validation

OS differences among RPA classes remained significant in the CC dataset (HR: 1.39 [1.09-1.75], $P=.007$). The RPA classes, however, failed to demonstrate significant differences for the EMC dataset (HR: 1.16 [0.70-1.92], $P=.577$). Actuarial survival estimates for OS for all datasets stratified by RPA class are shown in Figure 2. The C-indexes for CC (0.55) and EMC (0.52) datasets also demonstrated discrimination only marginally better than chance.

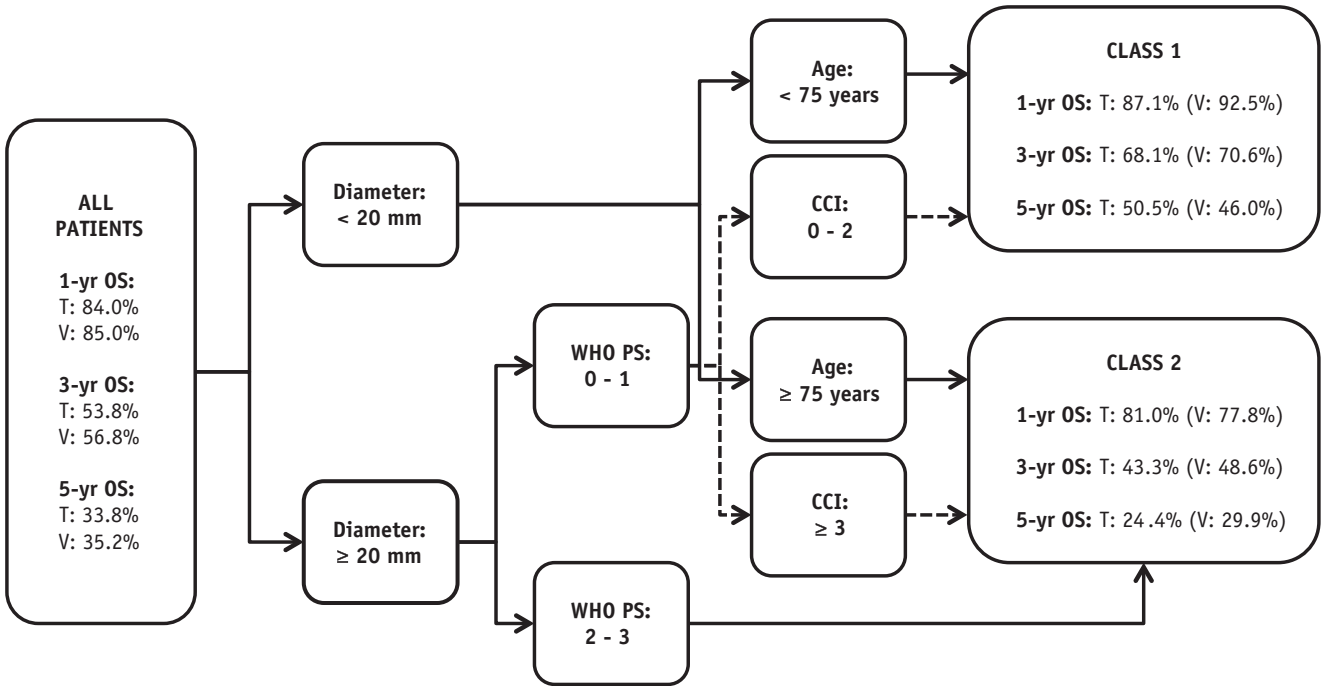

Fig. 1. Two-class RPA stratification of early lung cancer patients treated with SABR. CCI = Charlson comorbidity index; OS = overall survival; T = training dataset; V = internal validating dataset.

Nomogram development

Results from the univariate and multivariate analyses of the training and internal validation dataset are shown in Table 2. In addition to RPA-selected factors, smoking status (HR: 4.56 [1.45-14.33], $P<.001$) was found to be significant on multivariate modeling. Thus, the final clinical nomogram developed (Fig. 3) was based on age, CCI, WHO PS, smoking history, and tumor diameter. The C-statistic for the multivariate training dataset model was 0.69, demonstrating good discrimination.

Nomogram validation

On univariate analysis of the internal validation dataset, WHO PS (HR: 1.69 for 2 vs 0, $P=.013$), CCI (HR: 1.19 per 1 unit increase, $P<.001$), and tumor diameter (HR: 1.16 per 10-mm increase, $P=.028$) remained significantly prognostic. Age was of borderline significance (HR: 1.09 per 5-year increase, $P=.082$), and smoking was no longer prognostic ($P=.591$). Although only WHO PS ($P=.004$) and CCI remained prognostic ($P<.001$) for multivariate modeling of the internal validation dataset, the model continued to demonstrate good discrimination with a C-index of 0.66.

Calibration plots confirmed a high correlation between observed and predicted probability of 5-year OS for the internal, surgical, and SABR validation sets, where $r^2=0.97$, 0.91, and 0.79, respectively (Fig. 4). As a sensitivity analysis, 4-year OS was also evaluated and demonstrated r^2 values of 0.98, 0.94, and 0.84 for the same datasets, respectively. An electronic version of the clinical nomogram is available for download (Supplemental File E2; available online at www.redjournal.com).

Discussion

The increasing use of SABR for ES-NSCLC is due to its low morbidity, uncommon treatment-related mortality, and convenience relative to longer radiation fractionation schemes. Despite the high rates of local control obtained, the OS in many lung SABR series are low due to other competing risks of death (1). In a review of 44 reports consisting of 3641 ES-NSCLC SABR patients with varying levels of comorbidities, 1-year OS ranged from 35% to 96%, with a weighted average of 70%. The challenges in interpreting this heterogeneous data, as well as historical fears of toxicity, may contribute to the nihilistic undertreatment of ES-NSCLC at the population level (2, 11). To assist clinicians in determining the appropriateness of radical treatment for ES-NSCLC, we report development of a novel OS prognostication tool for lung SABR patients, the Amsterdam prognostic model.

Our study is unique in that it is 1 of a few studies that uses separate training and internal validation sets for model creation, followed by external validation in a SABR and surgical cohort. The key finding was that competing risks, as measured by patient factors (age, CCI, smoking history, and WHO PS) and tumor factors (diameter), were found to be strong predictors of survival. RPA-predicted classes demonstrated modest discrimination in SABR patients but performed less favorably in surgical patients. This finding

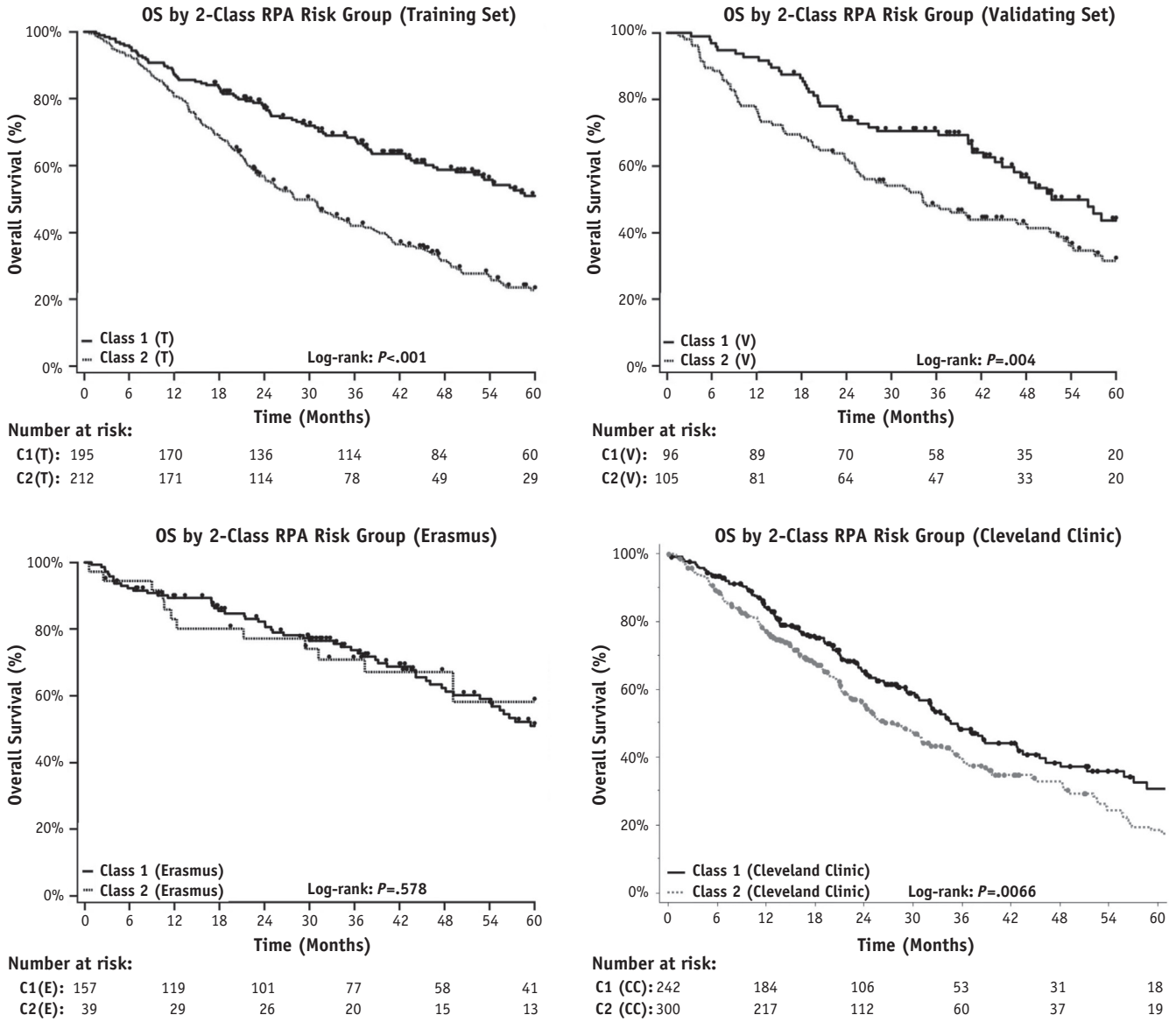

Fig. 2. Kaplan-Meier curves show OS stratified by RPA classes in (top left panel) VU University Medical Center (VUMC) training set; (top right panel) Erasmus Medical Center internal validation set; Erasmus Medical Center surgical validation set (bottom left panel); and Cleveland Clinic SABR validation set (bottom right panel). OS = overall survival; RPA = recursive partitioning analysis; SABR = stereotactic ablative radiation therapy; T = training dataset; V = internal validating dataset.

may be due to overfitting of data, which is a well-described issue in prediction modeling as a tradeoff for practicality (18). Regarding the poorer performance of the RPA in the surgical dataset, it is important to note that EMC patients tended to be younger and fitter and have earlier stage disease and thus the low proportion of RPA class 2 patients resulted in limited statistical power. Conversely, the nomogram showed good discrimination and calibration in both surgery and SABR cohorts, suggesting that this tool warrants use in clinical practice.

The findings of this study are consistent with those of existing reports of prognostic tools for ES-NSCLC and build on those findings in several important ways. A nomogram for OS using multi-institutional Chinese registry data for patients with resected NSCLC was validated with a separate cohort from the International Association for the Study of Lung Cancer database (7). That model, based on age, sex, histology, number of lymph nodes obtained, blood loss volume, and T and N stage resulted in more precise prognostication of OS in both Chinese and IASLC datasets than the traditional TNM staging system. Although similar to our APM nomogram, that tool did not include comorbidity (coded in the model as yes/no rather than a more comprehensive metric like CCI) or smoking history (data were unavailable on greater than 10% of patients). Inclusion of these covariates may have allowed for increased discrimination and precision, as they have also been shown to be important factors in tools used to prognosticate and guide the use of adjuvant therapy for resected NSCLC (eg, Adjvuant! Online, www.adjuvantonline.com).

Table 2 Univariate and multivariate Cox regression models predictive of OS for all eligible factors entered into RPA for training and validating sets

Factor	Training dataset				Internal validation dataset			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
Independent variables	HR (95% CI)	*P* value	HR (95% CI)	*P* value	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Age per 5-y increase	1.13 (1.05-1.20)	<.001	1.13 (1.06-1.22)	<.001	1.09 (0.99-1.21)	.082	1.07 (0.96-1.20)	.241
WHO PS								
1 vs 0	1.47 (0.96-2.25)	.073	1.41 (0.91-2.18)	.126	0.99 (0.57-1.73)	.977	0.88 (0.50-1.54)	.643
2 vs 0	1.95 (1.26-3.01)	.003	1.90 (1.22-2.98)	.005	2.01 (1.16-3.50)	.013	1.69 (0.96-2.98)	.068
3 vs 0	3.36 (1.89-5.97)	<.001	3.81 (2.11-6.89)	<.001	2.16 (0.95-4.89)	.066	1.64 (0.67-4.02)	.278
Overall effect analysis		<.001		<.001		<.001		.004
Smoker	3.78 (1.21-11.80)	.022	4.56 (1.45-14.33)	.009	1.37 (0.44-4.31)	.591	1.32 (0.41-4.28)	.648
Charlson comorbidity index per 1 unit increase	1.09 (1.03-1.16)	.004	1.09 (1.02-1.16)	.011	1.19 (1.10-1.29)	<.001	1.17 (1.07-1.27)	<.001
Tumor diameter per 10 mm increase	1.21 (1.11-1.32)	<.001	1.17 (1.06-1.29)	.001	1.16 (1.02-1.33)	.028	1.07 (0.92-1.25)	.364
T stage								
T2 vs T1	1.38 (1.11-1.73)	.004	-	-	1.48 (1.07-2.06)	.020	-	-
T stage								
T1b vs T1a	1.48 (1.08-2.02)	.014	-	-	1.12 (0.71-1.75)	.632	-	-
T2a vs T1a	1.59 (1.18-2.13)	.002	-	-	1.60 (1.05-2.42)	.027	-	-
T2b vs T1a	2.50 (1.59-3.93)	<.001	-	-	1.34 (0.60-3.03)	.478	-	-
Overall effect analysis		<.001	-	-		.118	-	-
Fractions								
5 vs 3	1.45 (1.12-1.86)	.004	-	-	1.48 (1.02-2.15)	.042	-	-
8 vs 3	1.19 (0.87-1.64)	.283	-	-	0.99 (0.61-1.60)	.961	-	-
Overall effect analysis		.017	-	-		.062	-	-

Abbreviations: CI = confidence interval; HR = hazard ratio; OS = overall survival; PS = performance status; RPA = recursive partitioning analysis; WHO = World Health Organization.

Other lung cancer surgical risk models include the Thoracoscore (5), the European Society Objective Score (ESOS) (19), and Society of Thoracic Surgeons (STS) models (6), created from French national, European multi-institutional, and American volunteer registry data, respectively. In addition to prognostic factors described in the APM, these models comprehensively included a number of covariates relevant to surgery, such as the extent of surgery, urgency of surgery, and American Society of Anesthesiologists score. Thoracoscore, ESOS, and STS are used to predict in-hospital mortality (STS is also used to predict major morbidity) and can be seen as complementary to the APM for lung SABR, for patients weighing the relative merits of surgery versus SABR.

Although most lung cancer patients treated with SABR worldwide are medically inoperable (20), there is an increasing trend toward treating patients who are younger and have fewer comorbidities and a better performance status (21). Indeed, there is growing equipoise for the role of SABR for these patients despite the fact that 3 randomized controlled trials comparing SABR to surgery have failed to accrue subjects (22, 23). Recently, a meta-analysis of 40 SABR and 23 surgery studies for ES-NSCLC was conducted (3). Of the 27 SABR studies reporting on proportion of patients who were potentially operable, the mean operability rate was 20.1%. The meta-analysis found that, when adjusting for age and potential operability, there were no significant differences in OS between the 2 treatment options ($P=.36$), a finding that must be considered in the context of heterogeneous data and the potential bias of unmeasured confounders. Ultimately, both surgery and radiation therapy will be crucial to address the unmet therapeutic needs of ES-NSCLC.

Prognosis as determined through the APM following SABR was also found to be highly dependent on tumor size, a finding consistent with previous SABR and surgical studies (24-27). In the APM, the RPA determined a tumor diameter cutoff point of 20 mm. Although this conforms to the difference between a T1a and T1b in the most recent TNM staging system, dichotomizing tumor size may overfit

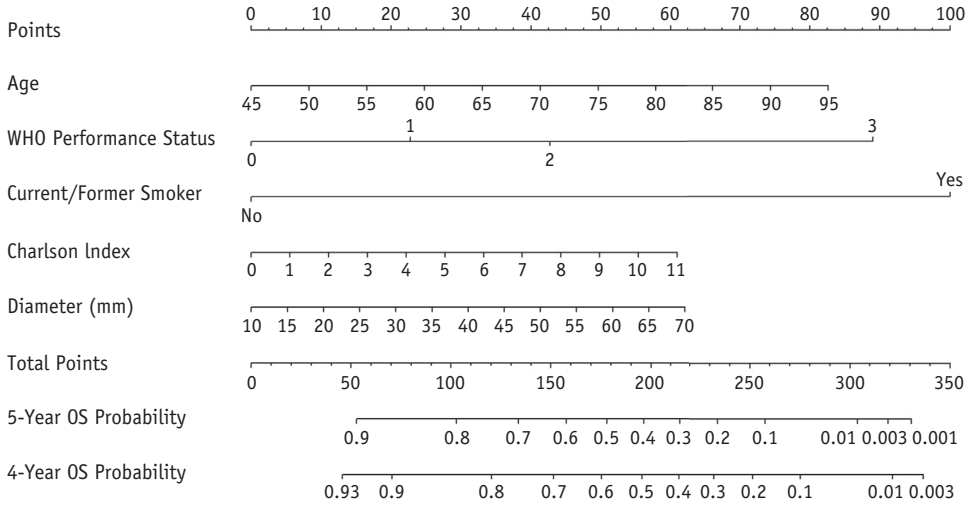

Fig. 3. Nomograms predicting 5- and 4-year overall survival (OS) based on the training set from the primary dataset. WHO = World Health Organization.

the data. Tumor diameter was modeled as a continuous variable in our nomogram, similar to a recent SEER-based model, which found that incorporating data in such a manner resulted in OS predictions that were superior to those in AJCC (28).

Limitations of our model may be improved upon in the future to further guide clinical practice. While SABR populations from 2 continents performed well on validation of the models herein, the RPA did not apply to the surgical series. As the EMC dataset was mostly comprised of RPA class 1 patients (fitter, younger, with smaller tumors), a larger more heterogeneous group of patients including a larger case mix of borderline operable patients would have been more useful to evaluate the proposed stratification. In addition, the cohort of patients in this study was comprised of patients treated at academic centers, where findings may not be generalizable to community practice. Nonetheless, although a variety of methods are used worldwide for lung SABR, early multi-institutional data suggests that outcomes appear to be generalizable across various delivery platforms, image-guidance, and dose fractionations in different geographic regions, provided that a BED_{10} greater than 100 Gy is delivered (2, 29).

Conclusions

In conclusion, we developed the APM, consisting of a novel 2-class RPA system and 5-year OS nomogram as a prognostic tool for patients with ES-NSCLC treated with SABR. Our findings indicate that the nomogram may be used to guide individual patient decision making, and the

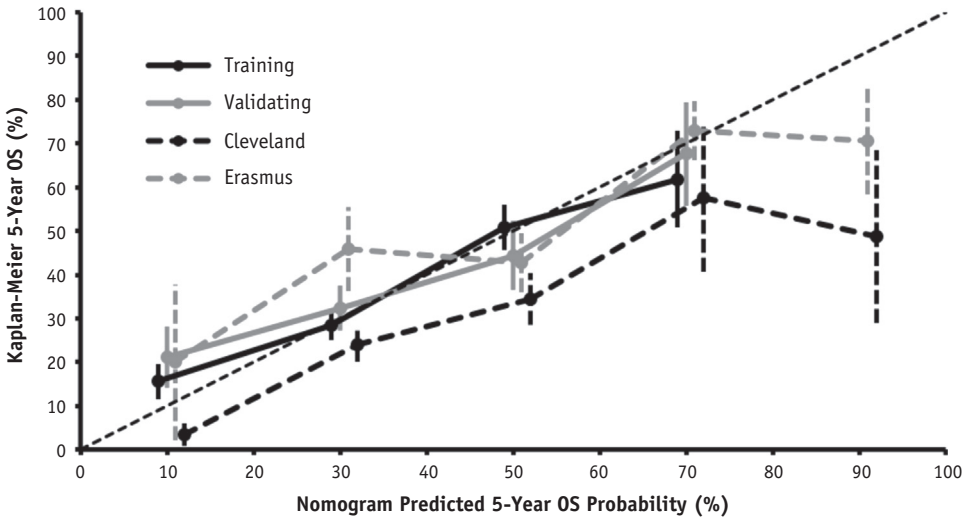

Fig. 4. Calibration plots for nomograms predicting 5-year OS compared to Kaplan-Meier 5-year overall survival (OS) estimates for internal training, internal validation and Erasmus Medical Center and Cleveland Clinic datasets. Global r^2 value of 0.79 reveals a high correlation between observed and predicted probabilities.

RPA may be helpful in stratification of patients for clinical trials. The proposed models can be refined based on future work, which may include neural network analyses to evaluate the effect of potential unmeasured confounders.

References

1. Louie AV, Palma DA, Dahele M, et al. Management of early-stage non-small cell lung cancer using stereotactic ablative radiotherapy: Controversies, insights, and changing horizons. *Radiother Oncol* 2015;114:138-147.
2. Haasbeek CJ, Palma D, Visser O, et al. Early-stage lung cancer in elderly patients: A population-based study of changes in treatment patterns and survival in the Netherlands. *Ann Oncol* 2012;23:2743-2747.
3. Zheng X, Schipper M, Kidwell K, et al. Survival outcome after stereotactic body radiation therapy and surgery for stage I non-small cell lung cancer: A meta-analysis. *Int J Radiat Oncol Biol Phys* 2014;90:603-611.
4. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: A decade of progress. *Chest* 2002;122:1037-1057.
5. Falcoz PE, Conti M, Brouchet L, et al. The thoracic surgery scoring system (thoracoscore): Risk model for in-hospital death in 15,183 patients requiring thoracic surgery. *J Thorac Cardiovasc Surg* 2007;133:325-332.
6. Kozower BD, Sheng S, O'Brien SM, et al. STS database risk models: Predictors of mortality and major morbidity for lung cancer resection. *Ann Thorac Surg* 2010;90:875-881. discussion 881-883.
7. Liang W, Zhang L, Jiang G, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* 2015;33:861-869.
8. Rogers SO Jr., Gray SW, Landrum MB, et al. Variations in surgeon treatment recommendations for lobectomy in early-stage non-small-cell lung cancer by patient age and comorbidity. *Ann Surg Oncol* 2010;17:1581-1588.
9. Brada M, Pope A, Baumann M. SABR in NSCLC—The beginning of the end or the end of the beginning? *Radiother Oncol* 2015;114:135-137.
10. Mellemgaard A, Luchtenborg M, Iachina M, et al. Role of comorbidity on survival after radiotherapy and chemotherapy for nonsurgically treated lung cancer. *J Thorac Oncol* 2015;10:272-279.
11. Koshy M, Malik R, Spiotto M, et al. Disparities in treatment of patients with inoperable stage I non-small cell lung cancer: A population-based analysis. *J Thorac Oncol* 2015;10:264-271.
12. Vansteenkiste J, De Ruysscher D, Eberhardt WE, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2013;24(Suppl 6):vi89-vi98.
13. Senthi S, Lagerwaard FJ, Haasbeek CJ, et al. Patterns of disease recurrence after stereotactic ablative radiotherapy for early stage non-small-cell lung cancer: A retrospective analysis. *Lancet Oncol* 2012;13:802-809.
14. Mokhles S, Nuyttens JJ, Maat AP, et al. Survival and treatment of non-small cell lung cancer stage I-II treated surgically or with stereotactic body radiotherapy: Patient and tumor-specific factors affect the prognosis. *Ann Surg Oncol* 2015;22:316-323.
15. Marwaha G, Stephans KL, Woody NM, et al. Lung stereotactic body radiation therapy: Regional nodal failure is not predicted by tumor size. *J Thorac Oncol* 2014;9:1693-1697.
16. Fowler JF, Tome WA, Fenwick JD, et al. A challenge to traditional radiation oncology. *Int J Radiat Oncol Biol Phys* 2004;60:1241-1256.
17. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008;26:1364-1370.
18. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 2009;14:323-348.
19. Brunelli A, Varela G, Van Schil P, et al. Multicentric analysis of performance after major lung resections by using the European Society Objective Score (ESOS). *Eur J Cardiothorac Surg* 2008;33:284-288.
20. Chi A, Liao Z, Nguyen NP, et al. Systemic review of the patterns of failure following stereotactic body radiation therapy in early-stage non-small-cell lung cancer: Clinical implications. *Radiother Oncol* 2010;94:1-11.
21. Peguret N, Dahele M, Lagerwaard F, et al. A brief report of 10-year trends in the use of stereotactic lung radiotherapy at a Dutch academic medical center. *J Thorac Oncol* 2014;9:114-117.
22. Louie AV, Senthi S, Palma DA. Surgery versus SABR for NSCLC. *Lancet Oncol* 2013;14:e491.
23. Moghanaki D, Karas T. Surgery versus SABR for NSCLC. *Lancet Oncol* 2013;14:e490-e491.
24. Ricardi U, Frezza G, Filippi AR, et al. Stereotactic ablative radiotherapy for stage I histologically proven non-small cell lung cancer: An Italian multicenter observational study. *Lung Cancer* 2014;84:248-253.
25. Allibhai Z, Taremi M, Bezjak A, et al. The impact of tumor size on outcomes after stereotactic body radiation therapy for medically inoperable early-stage non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 2013;87:1064-1070.
26. Ou SH, Zell JA, Ziogas A, et al. Prognostic factors for survival of stage I nonsmall cell lung cancer patients: A population-based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003. *Cancer* 2007;110:1532-1541.
27. Dunlap NE, Larner JM, Read PW, et al. Size matters: A comparison of T1 and T2 peripheral non-small-cell lung cancers treated with stereotactic body radiation therapy (SBRT). *J Thorac Cardiovasc Surg* 2010;140:583-589.
28. Zhang J, Gold KA, Lin HY, et al. Relationship between tumor size and survival in non-small-cell lung cancer (NSCLC): An analysis of the Surveillance, Epidemiology, and End Results (SEER) registry. *J Thorac Oncol* 2015;10:682-690.
29. Guckenberger M, Allgauer M, Appold S, et al. Safety and efficacy of stereotactic body radiation therapy for stage I non-small-cell lung cancer in routine clinical practice: A patterns-of-care and outcome analysis. *J Thorac Oncol* 2013;8:1050-1058.