original report

Machine-Learning and Stochastic Tumor Growth Models for Predicting Outcomes in Patients With Advanced Non–Small-Cell Lung Cancer

Kien Wei Siah, SM[1]; Sean Khozin, MD, MPH[2]; Chi Heem Wong, SM[1]; and Andrew W. Lo, PhD[1,3]

abstract

PURPOSE The prediction of clinical outcomes for patients with cancer is central to precision medicine and the design of clinical trials. We developed and validated machine-learning models for three important clinical end points in patients with advanced non–small-cell lung cancer (NSCLC)—objective response (OR), progression-free survival (PFS), and overall survival (OS)—using routinely collected patient and disease variables.

METHODS We aggregated patient-level data from 17 randomized clinical trials recently submitted to the US Food and Drug Administration evaluating molecularly targeted therapy and immunotherapy in patients with advanced NSCLC. To our knowledge, this is one of the largest studies of NSCLC to consider biomarker and inhibitor therapy as candidate predictive variables. We developed a stochastic tumor growth model to predict tumor response and explored the performance of a range of machine-learning algorithms and survival models. Models were evaluated on out-of-sample data using the standard area under the receiver operating characteristic curve and concordance index (C-index) performance metrics.

RESULTS Our models achieved promising out-of-sample predictive performances of 0.79 area under the receiver operating characteristic curve (95% CI, 0.77 to 0.81), 0.67 C-index (95% CI, 0.66 to 0.69), and 0.73 C-index (95% CI, 0.72 to 0.74) for OR, PFS, and OS, respectively. The calibration plots for PFS and OS suggested good agreement between actual and predicted survival probabilities. In addition, the Kaplan-Meier survival curves showed that the difference in survival between the low- and high-risk groups was significant (log-rank test $P <$ .001) for both PFS and OS.

CONCLUSION Biomarker status was the strongest predictor of OR, PFS, and OS in patients with advanced NSCLC treated with immune checkpoint inhibitors and targeted therapies. However, single biomarkers have limited predictive value, especially for programmed death-ligand 1 immunotherapy. To advance beyond the results achieved in this study, more comprehensive data on composite multiomic signatures is required.

INTRODUCTION

Non–small-cell lung cancer (NSCLC) accounts for approximately 85% of patients with lung cancer. The majority of patients with NSCLC are diagnosed at advanced stages.[2] Because NSCLC is a heterogeneous group of diseases, there is wide variation in the effectiveness of different therapies. Predictive models play an important role in therapeutic planning by allowing patients and physicians to make informed decisions on the basis of specific characteristics of the individual rather than on general population statistics. Despite their clinical relevance, however, no predictive model for NSCLC is widely implemented.

In this article, we describe a pooled analysis of randomized clinical trials in NSCLC. These trials evaluated chemotherapy, targeted therapy, and immunotherapy treatments in patients with advanced NSCLC. We characterize the tumor dynamics, response, and patient survival with these different modalities to develop predictive models that reflect recent advances in the treatment of NSCLC.

We explore the utility of machine-learning algorithms and survival models using data routinely collected in clinical trials and propose a stochastic tumor growth model based on clinical data to predict tumor response. We describe our training and validity testing methodology in Methods. Using the trained models, we identify baseline variables that are strongly associated with response and survival, and compare our findings with relevant related studies in the literature.

METHODS

Data

We specified 17 randomized clinical trials submitted to the US Food and Drug Administration between

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on August 16, 2019 and published at ascopubs.org/journal/cci on September 20, 2019: DOI https://doi.org/10.1200/CCI.19.00046

CONTEXT

Key Objective

What are the drivers behind objective response, progression-free survival, and overall survival while receiving new treatment modalities for patients with advanced non–small-cell lung cancer?

Knowledge Generated

Our models achieved promising predictive performances for objective response, progression-free survival, and overall survival. We found that variables beyond single biomarker positivity that are not routinely collected and analyzed in clinical trials can help develop better predictive models.

Relevance

Predictive models allow the personalization of treatment decisions using specific patient and disease characteristics. However, single biomarkers have limited predictive value.[1] Complex signatures based on data from multiomics pipelines are needed to achieve optimal precision in selecting patients for clinical trial enrollment and administration of anticancer therapies in advanced non–small-cell lung cancer.

January 2007 and February 2017 to support New Drug Applications as our initial data set. These trials evaluated treatment with nine approved drugs for NSCLC, consisting of three programmed death-ligand 1 (PD-L1) immune checkpoint inhibitors (ICIs), three epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs), and three anaplastic lymphoma kinase (ALK) TKIs. The characteristics of the clinical trials are listed in Table 1. The data set included 8,925 patients in the intention-to-treat population.

We extracted survival data, tumor measurements, response outcomes, baseline demographics, medical history, and laboratory test results from the patient-level Study Data Tabulation Model and Analysis Data Model databases. After standardizing the common features across all trials, there were 46 categorical variables and five continuous variables in the data set (excluding end points; Table 2).

For our analysis, we excluded patients who had no tumor measurements in the database or had ambiguous records, patients who were given a placebo or were untreated by therapy in the trial before their discontinuation, and patients who had missing features in their records necessary for subsequent analyses. The final sample was composed of 7,805 patients (Fig 1). In the Data Supplement, we provide summary statistics on their key baseline variables, tumor measurements, and response and survival end points.

Stochastic Model for Tumor Growth

We first developed a stochastic model for tumor growth (Data Supplement). We assumed that net tumor size followed an exponential model, commonly used to describe macroscopic tumor growth.[3-6] To identify the factors driving tumor regression, we modeled the rate constant as a linear function of the treatment group (as defined in the Data Supplement), demographic background, medical history, and laboratory test features in the data set (Table 2). We incorporated additive Gaussian white noise to account for any randomness in the tumor growth process, uncertainty in measurement, or errors as a result of model misspecification. We estimated parameters of the model using the longitudinal patient-level tumor sum of longest diameter (SLD) measurements in our data set.

When patients trigger the discontinuation criteria of a clinical study, for example, through progressive disease, they are typically withdrawn from the trial. This discontinuation affects the observation of SLD measurements. To obtain accurate estimates, the tumor growth model must therefore incorporate this effect. We propose to model the discontinuation process as a sequence of Bernoulli trials in which patients have a probability of being discontinued at each visit. This probability is conditioned on each individual's target lesion response as derived from observed SLD measurements. We combined the tumor growth and the discontinuation models, and estimated parameters of both jointly through maximum likelihood estimation. Because the set of features under consideration was large (Table 2; Data Supplement), we used the stepwise forward selection algorithm with the Akaike information criterion to identify a parsimonious set of factors for our model.

This model can be used to simulate tumor growth in patients with different intrinsic characteristics under different types of therapy. By aggregating outcomes from multiple bootstrap simulations, we can predict the probability of objective response (OR) in patients. This acts as a baseline predictive model to compare against the machine-learning models discussed in the following section.

Machine-Learning Models for Tumor Response

The prediction of tumor response can be formulated as a supervised binary classification problem to predict the probability of OR in individuals given a set of input features. In this article, we consider a total of 45 categorical variables and five continuous variables as input features based on known clinical and prognostic importance. We converted the categorical features to binary variables and standardized the continuous variables (Data Supplement).

TABLE 1. Characteristics of Clinical Trials in the Data Set

Therapy Type	Experimental Arm	Therapy Type	Control Arm	Design	Phase	Line	ITT	Initiation	Cutoff
EGFR	Gefitinib	Chemo	Docetaxel	R, OL	3	2	1,466	Mar 2004	Mar 2007
EGFR	Gefitinib	Chemo	Carboplatin with paclitaxel	R, OL	3	1	1,217	Mar 2006	Apr 2008
EGFR	Erlotinib	Chemo	Pemetrexed or docetaxel	R, OL	3	2	424	Apr 2006	Aug 2010
EGFR	Erlotinib	Chemo	Docetaxel or gemcitabine with cisplatin or carboplatin	R, OL	3	1	173	Feb 2007	Apr 2012
EGFR	Afatinib	Placebo	Best supportive care	R, DB	2/3	3	585	Apr 2008	Jun 2010
EGFR	Afatinib	Chemo	Pemetrexed with cisplatin	R, OL	3	1	345	Aug 2009	Nov 2013
ALK	Crizotinib	Chemo	Pemetrexed or docetaxel	R, OL	3	2	347	Sep 2009	Aug 2015
ALK	Crizotinib	Chemo	Pemetrexed with cisplatin or carboplatin	R, OL	3	1	343	Jan 2011	Nov 2013
EGFR	Erlotinib	Chemo	Gemcitabine with cisplatin	R, OL	3	1	217	Mar 2011	Apr 2014
EGFR	Afatinib	EGFR	Erlotinib	R, OL	3	2	795	Mar 2012	Feb 2015
PD-L1	Nivolumab	Chemo	Docetaxel	R, OL	3	2	272	Oct 2012	Dec 2014
PD-L1	Nivolumab	Chemo	Docetaxel	R, OL	3	2	582	Nov 2012	Feb 2015
ALK	Ceritinib	Chemo	Pemetrexed or docetaxel	R, OL	3	2	231	Jun 2013	Jan 2016
PD-L1	Atezolizumab	Chemo	Docetaxel	R, OL	2	2	287	Aug 2013	Dec 2015
PD-L1	Pembrolizumab	Chemo	Docetaxel	R, OL	2/3	2	1,033	Aug 2013	Oct 2015
ALK	Alectinib	ALK	Crizotinib	R, OL	3	1	303	Aug 2014	Feb 2017
PD-L1	Pembrolizumab	Chemo	Platinum based	R, OL	3	1	305	Sep 2014	May 2016
Total							8,925		

Abbreviations: ALK, anaplastic lymphoma kinase; Chemo, chemotherapy; DB, double blind; EGFR, epidermal growth factor receptor; ITT, intention-to-treat population; OL, open label; PD-L1, programmed death-ligand 1; R, randomized.

We randomly split the data set into two disjoint sets, a training set (70%; n = 5,463) and a testing set (30%; n = 2,342). We used the training set to develop predictive models and kept the testing set as an out-of-sample data set for model validation.

We explored several machine-learning algorithms: penalized logistic regression, decision trees, random forests, and multilayer perceptrons. We tuned their hyperparameters using five-fold cross-validation with the area under the receiver operating characteristic curve (AUC) as the standard metric for performance (the receiver operating characteristic curve plots the true-positive rate and false-positive rate of a classifier as its decision threshold is varied). AUC is the estimated probability that a classifier will rank a responder higher than a nonresponder.[7] An AUC of 0.5 corresponds to a random classifier, whereas 1.0 corresponds to a perfect classifier. We repeated the experiment 100 times for each model to obtain CIs for the expected performance. To investigate the drivers behind tumor response, we examined the top 20 most important variables of the best-performing predictive model.

Because patients with different biomarkers may have distinct drivers for tumor response, it is possible that classifiers trained on group-specific data for treatment will outperform the general models trained on the entire data set. To test this, we built and analyzed specialized models by filtering the data set by treatment group before training and testing. We then computed the AUC for the entire data set by aggregating predictions from all six specialized models. As a comparison, we also broke down the performance of the general models by treatment group.

Survival Models

For progression-free survival (PFS) and overall survival (OS), we implemented two standard methods used in survival modeling: the Cox proportional hazards model[8] and the accelerated failure time model with log-normal distribution. Again, we performed stepwise forward selection with Akaike information criterion as criteria to identify a parsimonious set of features for each model. We also explored two nonlinear and nonparametric survival models, the random survival forest model[9] and the neural network survival model.[10] For the PFS models, we considered the same features as we did for tumor response. However, because of the confounding effects of crossovers and post-trial therapies, we replaced the treatment group feature for the OS models with biomarker positivity in PD-L1, EGFR, or ALK (Data Supplement).

We used the same training and validation methodology described earlier. Instead of the AUC, however, we assessed performance using the concordance index (C-index), a measure of the concordance between

TABLE 2. List of Variables Extracted From Study Data Tabulation Model and Analysis Data Model Databases

Variable	Value
Demographics	
Age	Years
Weight	Kilograms
Sex	Male, female
Race group	Asian,[a] white, others
Region	APAC, NAM, WEUR, others
Medical history	
Time since diagnosis	Days
Performance status[b]	0, 1, 2, or higher
Smoking status	Ever, never
Stage at screening	IIIB or lower, IV
Prior chemotherapy	Yes, no
Histology	Adeno, SCC, others[c]
Metastases in brain, bone, liver, and others	Yes, no
No. of metastasis sites	Count
Biomarker status in PD-L1, EGFR, and ALK[d]	Positive, negative, not tested
No. of baseline target lesions	1, 2, 3, 4, 5, or more
Baseline SLD[e]	Millimeters
Comorbidities in 23 system organ class levels[f]	Yes, no
Laboratory measurements[g]	
Alkaline phosphate	High, normal, low
ALT	High, normal, low
AST	High, normal, low
Bilirubin	High, normal, low
Creatine	High, normal, low
Hemoglobin	High, normal, low
Platelets count	High, normal, low
WBC count	High, normal, low
Therapy type	
Therapy received	Chemotherapy, PD-L1 ICI, EGFR TKI, ALK TKI
End points	
Overall survival[h]	Days
Overall survival censor	Yes, no
Progression-free survival[i]	Days
Progression-free survival censor	Yes, no

(Continued on following page)

orderings of observed survival times and predicted times.[11] This represents the probability that, for a pair of randomly chosen patients, the individual with the higher risk prediction will experience an event before the other. Like the AUC, C-indices range between 0.5 (random) and 1.0 (perfect).

In addition, we assessed model calibration by comparing the actual and the predicted survival probabilities at different times: at 6, 12, and 24 months for PFS, and at 12, 24, and 36 months for OS. For each cutoff, we divided the test set into quintiles on the basis of the predicted risk scores. We then computed and plotted the true survival probability for each quintile against the corresponding average predicted score. Finally, we stratified the test set into different risk groups on the basis of their predicted risk scores and examined the differences in their Kaplan-Meier survival curves using the log-rank test.

RESULTS

Models for Tumor Response

The test set AUC results are listed in Table 3 (baseline stochastic model and best-performing model) and presented in Fig 2 (receiver operating characteristic curve) and the Data Supplement (all other models). We found that the machine-learning models substantially outperformed the baseline stochastic model, with improvements up to 0.07 AUC. Logistic regression achieved the best performance, with 0.79 AUC (95% CI, 0.77 to 0.81). We did not observe any appreciable difference in predictive power between the general models and the specialized models.

We extracted the top 20 largest coefficients of the general stochastic model (Data Supplement) and the best-performing general logistic regression model (Table 4). The biomarker status was among the strongest drivers of regression and response in patients treated with immunotherapies and targeted therapies. (We noted that EGFR TKIs seemed to have weaker effects compared with ALK TKIs and PD-L1 ICIs, and discuss possible reasons in Discussion.)

Survival Models

We list the test set C-index results in Table 3 (best-performing models) and the Data Supplement (all other models). The performance was similar across linear and nonlinear models for both PFS and OS, suggesting that the use of more complex models did not improve prediction. Thus, we decided to focus on the Cox model because of its ease of implementation and interpretability. The model achieved C-indices of 0.67 (95% CI, 0.66 to 0.69) and 0.73 (95% CI, 0.72 to 0.74) on out-of-sample data for PFS and OS, respectively. Our experiments indicated that

TABLE 2. List of Variables Extracted From Study Data Tabulation Model and Analysis Data Model Databases (Continued)

Variable	Value
RECIST best overall response	CR, PR, SD,[j] PD, NE
Objective response[k]	Yes, no
Time point SLD	Millimeters
Depth of response	Percent

Abbreviations: Adeno, adenocarcinoma; ALK, anaplastic lymphoma kinase; APAC, Asia-Pacific; CR, complete response; EGFR, epidermal growth factor receptor; ICI, immune checkpoint inhibitor; NAM, North America; NE, not evaluable; PD, progressive disease; PD-L1, programmed death-ligand 1; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SCC, squamous cell carcinoma; SD, stable disease; SLD, sum of longest diameter; TKI, tyrosine kinase inhibitor; WEUR, western Europe.

[a]Includes Pacific Islanders.

[b]Eastern Cooperative Oncology Group or WHO score.

[c]Includes large cell carcinoma and not otherwise specified.

[d]Patients are tested for, at most, one biomarker, depending on the experimental arm of the clinical trial they are from: patients from PD-L1 ICI trials are tested for PD-L1 expression, from EGFR TKI trials for EGFR mutation, and from ALK TKI trials for ALK-translocation.

[e]Measurements under RECIST version 1.0 are scaled to reconcile with version 1.1 (Data Supplement).

[f]As defined in Medical Dictionary for Regulatory Activities.

[g]High, normal, low, as determined by investigators on site.

[h]Defined as the time from randomization in a clinical trial until death from any cause.

[i]Defined as the time from randomization to tumor progression or death.

[j]Includes non-CR/non-PD.

[k]Defined as having either CR or PR as best overall response.

specialized PFS models performed more poorly than their general counterparts.

We show calibration plots of the Cox models at multiple time points in the Data Supplement. The curves lay close to the diagonal, indicating that the models were well calibrated. We further stratified patients into two risk groups using a median split of the risk scores. The difference in survival between the low- and high-risk groups was significant (log-rank test $P < .001$) for both PFS and OS (Fig 2).

As listed in Table 4, we extracted the top 20 coefficients in the Cox models to identify the specific factors that predict PFS and OS. We found that biomarker positivity was associated with better PFS for patients treated with inhibitor therapies. For OS, the presence of a proven biomarker led to improved survival. (Because the end points for OR, PFS, and OS are related to one another—PFS is dependent on both response and mortality data, from which OR and OS, respectively, are derived—many of the important features were common across the models.)

DISCUSSION

We developed predictive models for OR, PFS, and OS in patients with advanced NSCLC receiving chemotherapy, targeted therapy, and immunotherapy. By examining the top coefficients in our linear predictive models, we identified the specific factors that have the greatest predictive values for tumor response and survival.

We found the treatment modality to be among the strongest predictors of response and survival end points (Table 4). Biomarker status had a substantial effect on the efficacy of inhibitor therapies. Among patients treated with ICIs and TKIs, those who tested positive for biomarkers had a higher probability of OR and a more favorable PFS than those who tested negative or were untested. Patients in the former treatment group also had a more favorable prognosis than patients treated with chemotherapy. Our findings were consistent with observations of the OR rate (ORR; Data Supplement) and the survival data (Data Supplement). For

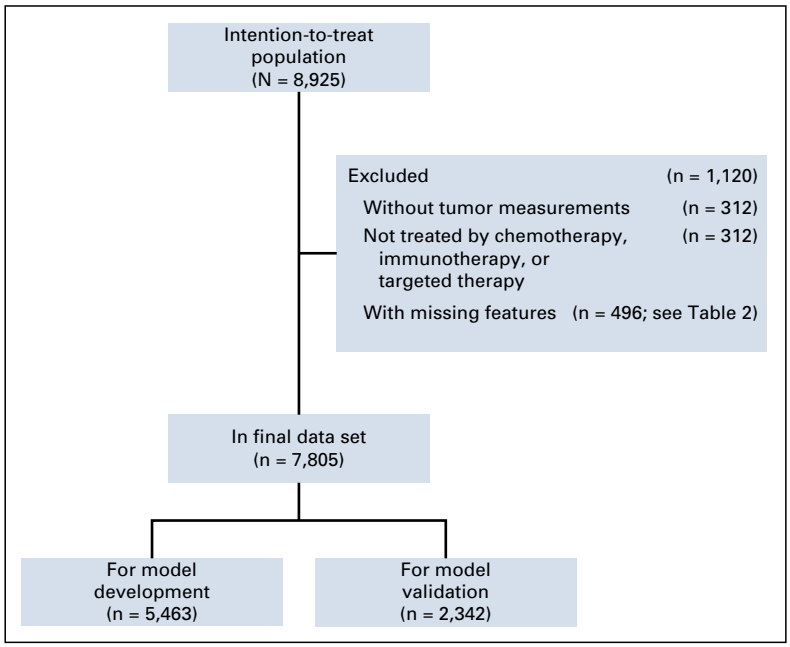

FIG 1. Sample size of the data set after filtering. We excluded patients who either (1) did not have tumor measurements in the database or had ambiguous records, such as nonmeasurable disease, no target lesions, or 0-mm baseline sum of longest diameter; (2) were receiving placebo (eg, the placebo comparator arm in the third-line afatinib trial) or were not treated by chemotherapy, immunotherapy, or targeted therapy in the clinical trial before discontinuation; or (3) had missing features that were necessary for subsequent analyses (Table 2). We randomly selected 70% of the data set as the development cohort and used the other 30% as the validation cohort (see Methods).

TABLE 3. Test Set Performance of Selected Predictive Models for Objective Tumor Response, PFS, and OS, Broken Down by Treatment Group

	Test Set Average AUC/C-Index (95% CI)			
	Objective Tumor Response		PFS	OS
Treatment Modality	Statistical Model	Logistic Regression	Cox Proportional Hazards	Cox Proportional Hazards
General model				
All	0.716 (0.695 to 0.740)	0.787 (0.770 to 0.805)	0.673 (0.664 to 0.685)	0.729 (0.721 to 0.739)
Chemotherapy	0.628 (0.594 to 0.670)	0.707 (0.675 to 0.740)	0.660 (0.646 to 0.675)	–
PD-L1 positive and receiving PD-L1 ICI	0.522 (0.459 to 0.589)	0.561 (0.507 to 0.620)	0.602 (0.568 to 0.639)	–
EGFR positive and receiving EGFR TKI	0.583 (0.517 to 0.645)	0.719 (0.669 to 0.764)	0.693 (0.667 to 0.722)	–
ALK positive and receiving ALK TKI	0.550 (0.482 to 0.614)	0.625 (0.575 to 0.685)	0.654 (0.619 to 0.690)	–
Negative biomarker but receiving inhibitor therapy	0.504 (0.436 to 0.581)	0.530 (0.458 to 0.595)	0.603 (0.577 to 0.630)	–
Not tested but receiving inhibitor therapy	0.739 (0.695 to 0.784)	0.813 (0.777 to 0.853)	0.642 (0.626 to 0.658)	–
Specialized models				
All*	0.724 (0.707 to 0.738)	0.775 (0.749 to 0.789)	0.629 (0.620 to 0.639)	–
Chemotherapy	0.627 (0.581 to 0.661)	0.706 (0.676 to 0.736)	0.651 (0.636 to 0.667)	–
PD-L1 positive and receiving PD-L1 ICI	0.516 (0.461 to 0.579)	0.528 (0.467 to 0.581)	0.561 (0.516 to 0.603)	–
EGFR positive and receiving EGFR TKI	0.515 (0.419 to 0.591)	0.690 (0.643 to 0.732)	0.677 (0.646 to 0.708)	–
ALK positive and receiving ALK TKI	0.478 (0.409 to 0.541)	0.605 (0.544 to 0.672)	0.631 (0.591 to 0.664)	–
Negative biomarker but receiving inhibitor therapy	0.563 (0.485 to 0.629)	0.561 (0.491 to 0.632)	0.597 (0.571 to 0.621)	–
Not tested but receiving inhibitor therapy	0.721 (0.660 to 0.776)	0.797 (0.766 to 0.838)	0.636 (0.616 to 0.655)	–

NOTE. General models were trained on the entire data set. Specialized models were trained on treatment-group–specific data. We implemented the logistic regression models in Python using the scikit-learn[33] package and the Cox proportional hazards models in R using the survival[34] package.

Abbreviations: ALK, anaplastic lymphoma kinase; AUC, area under the curve; C-index, concordance index; EGFR, epidermal growth factor receptor; ICI, immune checkpoint inhibitor; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; TKI, tyrosine kinase inhibitor.

*C-index derived from predictions aggregated from specialized models.

OS, the presence of an actionable biomarker in PD-L1, EGFR, or ALK was associated with improved survival. This reflects the clinical benefit associated with the personalization of anticancer therapies on the basis of these biomarkers.

Consistent with previous findings, our models predicted favorable tumor response in nonsmokers, women, and those with good performance status.[12-15] We also found that patients treated with ICI and TKI therapies who had undergone prior chemotherapy treatments had poorer survival rates than chemotherapy-naive patients, highlighting the prognostic implications of line of therapy in this patient population. Our models did not identify age as a significant feature, supporting previous findings that performance status is a better prognostic variable than age in patients with advanced NSCLC.[16,17]

In contrast, we found that patients with organ dysfunction as classified by abnormal baseline laboratory measurements tended to have a poorer prognosis. Baseline organ dysfunction may interfere with cancer treatment delivery and adversely affect survival,[18-20] underscoring the importance of model-informed risk stratification for management of these patients. Consistent with other studies, we identified liver metastasis[21] and squamous cell carcinoma (SCC) histology as important risk factors. Because driver mutations EGFR and ALK are rarely found in SCC,[22] this group of patients may have worse clinical outcomes than patients without SCC because they have fewer effective options.

Unusually, the coefficient for patients with PD-L1–positive disease receiving PD-L1 ICI treatment was almost twice that of the analogous coefficient for patients with EGFR-positive disease receiving EGFR TKI treatment (Table 4), especially given that patients in the former group had a lower ORR than those in the latter group (Data Supplement). A closer look at the testing set AUCs listed in Table 3 reveals that the

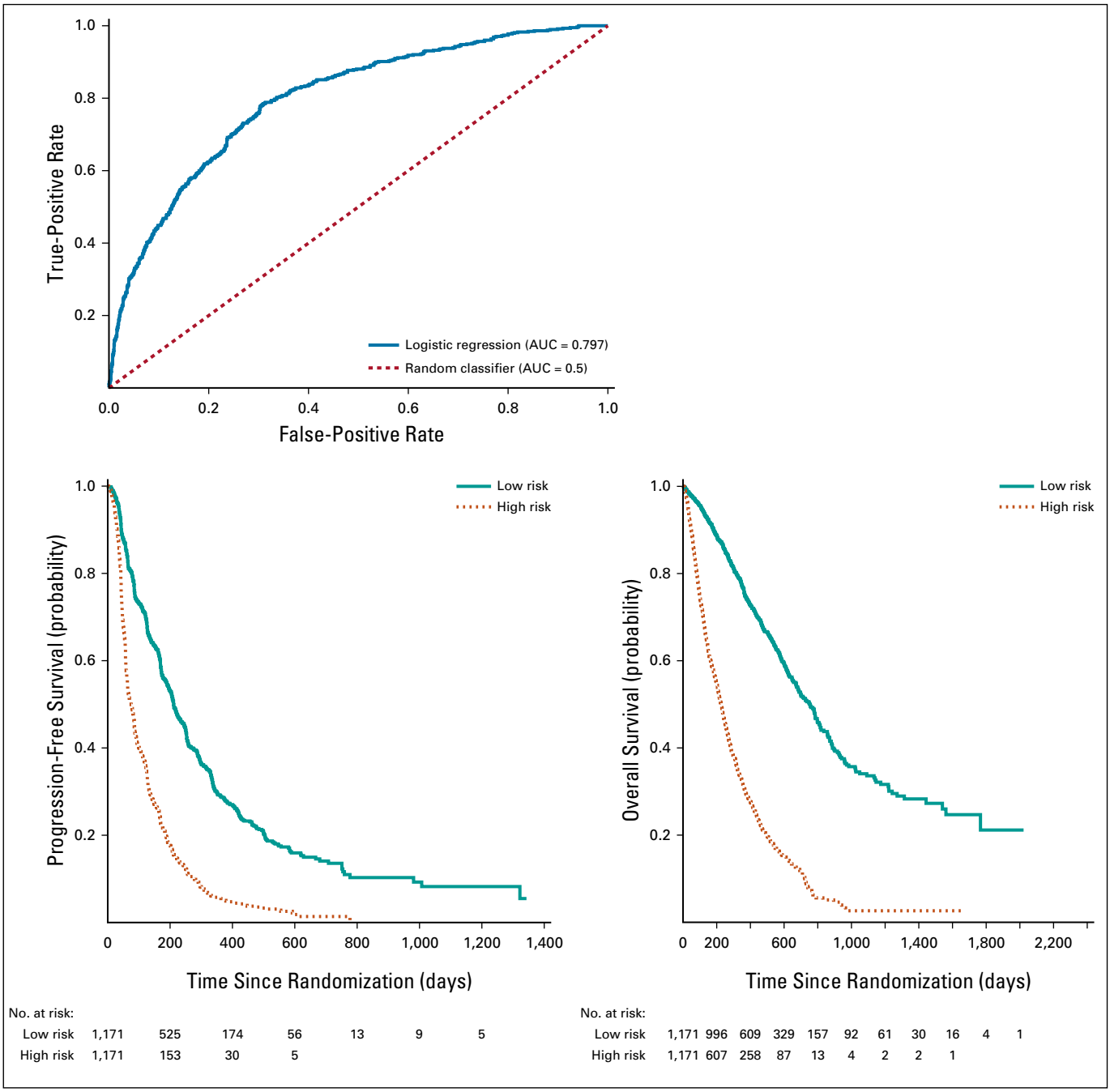

FIG 2. Test set performance of predictive models for a random experiment iteration. The top figure shows the receiver operating characteristic curve of the logistic regression model for objective tumor response: area under the curve (AUC), 0.80. The bottom-left figure shows the risk stratification by the Cox proportional hazards model for progression-free survival: hazard ratio, 2.58 (95% CI, 2.35 to 2.85; log-rank test $P < .001$). The bottom-right figure shows the risk stratification by the Cox proportional hazards model for overall survival: hazard ratio, 3.94 (95% CI, 3.52 to 4.42; log-rank test $P < .001$).

logistic regression model performed much more poorly in the PD-L1–positive treatment group than in the EGFR-positive subgroup. The same trend was also found in the specialized and nonlinear models.

We believe that this poor performance was the result of a lack of strong predictors within the PD-L1–positive treatment subgroup. In the absence of other useful factors, the algorithm will attribute the higher-than-average ORR to the effectiveness of ICIs on PD-L1–positive patients and assign greater weight to the corresponding treatment group indicator so that it dominates all other factors in the data set. (This also explains the large coefficient for the ALK positive with ALK TKI treatment group.)

Our results suggest that factors beyond PD-L1 positivity are at work in response to PD-L1 ICI, but absent from the current feature set. Ruling out factors already present in the data set,

TABLE 4. Top 20 Coefficients of the Objective Tumor Response Logistic Regression Model, the PFS Cox Proportional Hazards Model, and the OS Cox Proportional Hazards Model

Response Logistic Regression				PFS Cox Proportional Hazards				OS Cox Proportional Hazards			
Type	**Variable**	**Avg**	**SD**	**Type**	**Variable**	**Avg**	**SD**	**Type**	**Variable**	**Avg**	**SD**
Trt	ALK+ with ALK TKI	1.65	0.32	Trt	ALK+ with ALK TKI	–0.93	0.04	Hist	PS – 2 or higher	0.60	0.05
Trt	PD-L1+ with PD-L1 ICI	1.12	0.28	Trt	PD-L1+ with PD-L1 ICI	–0.79	0.05	Hist	Prior chemo – no	–0.51	0.03
Hist	Prior chemo – no	0.84	0.11	Hist	Prior chemo – no	–0.49	0.03	Bio+	PD-L1+, EGFR+, or ALK+	–0.48	0.03
Trt	EGFR+ with EGFR TKI	0.66	0.16	Trt	EGFR+ with EGFR TKI	–0.39	0.04	Demo	Race – Asian	–0.37	0.09
Hist	Histol – SCC	–0.44	0.07	Hist	PS – 2 or higher	0.37	0.04	Hist	PS – 0	–0.37	0.02
Lab	Hgb – low	–0.20	0.05	Meta	Liver – yes	0.24	0.03	Lab	WBC – high	0.34	0.03
Hist	PS – 2 or higher	–0.20	0.08	Lab	WBC – high	0.22	0.02	Lab	BILI – high	0.30	0.13
Hist	Smoking – never	0.20	0.05	Lab	BILI – high	0.18	0.13	Hist	Histol – SCC	0.26	0.03
Meta	Liver – yes	–0.18	0.05	Lab	ALP – low	0.17	0.15	Lab	ALP – high	0.25	0.03
Demo	Sex – female	0.17	0.06	Meta	Others – no	0.17	0.07	Lab	Hgb – low	0.24	0.02
Lab	ALT – high	–0.15	0.07	Lab	AST – low	–0.15	0.17	Meta	Liver – yes	0.22	0.03
Demo	Region – WEUR	–0.14	0.08	Lab	ALT – high	0.13	0.06	Hist	Smoking – never	–0.20	0.03
Hist	PS – 0	0.14	0.05	Demo	Race – others	–0.13	0.10	Como	Eye – yes	–0.18	0.06
Trt	Not tested with ICI or TKI	–0.13	0.09	Demo	Sex – female	–0.13	0.03	Demo	Sex – female	–0.18	0.03
Lab	WBC – high	–0.12	0.05	Hist	Histol – SCC	0.13	0.04	Hist	BSLD	0.17	0.01
Como	GI – yes	–0.12	0.04	Lab	Hgb – low	0.13	0.02	Hist	Histol – others	0.15	0.04
Lab	ALP – high	–0.12	0.04	Como	Blood and lymphatic system – yes	0.12	0.05	Como	Nervous system – yes	–0.13	0.03
Demo	Race – Asian	0.10	0.08	Hist	PS – 0	–0.12	0.02	Como	Renal and urinary – yes	–0.13	0.08
Demo	Region – others	–0.09	0.08	Como	GI – yes	0.11	0.02	Demo	Weight	–0.13	0.01
Meta	Brain – yes	–0.08	0.06	Hist	Smoking – never	–0.11	0.04	Como	Skin and subcutaneous tissue – yes	–0.12	0.06

NOTE. The parameters were estimated with good precision. Plus sign (+) indicates positivity. Factors with positive (negative) coefficients in the response model increase (decrease) the odds of objective tumor response. Factors with negative (positive) coefficients in the survival models increase (decrease) the probability of survival. In addition, the impact of each factor is proportional to the absolute value of the corresponding coefficient; that is, the greater the absolute value, the stronger the impact.

Abbreviations: ALK, anaplastic lymphoma kinase; ALP, alkaline phosphatase; Avg, average; BILI, bilirubin; Bio+, biomarker positivity; BSLD, baseline sum of longest diameter; Como, comorbidities; Demo, demographics; EGFR, epidermal growth factor receptor; Hgb, hemoglobin; Hist, medical history; Histol, histology; ICI, immune checkpoint inhibitor; Lab, laboratory measurements; Meta, metastasis; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PS, performance status; SCC, squamous cell carcinoma; SD, standard deviation; TKI, tyrosine kinase inhibitor; Trt, treatment; WEUR, western Europe.

we suspect that patients with PD-L1–positive disease may have additional factors such as complex somatic mutations, germline single-nucleotide polymorphisms, and exposomal influences that may affect treatment efficacy and predispose patients to specific responses. In fact, there is emerging evidence that variables such as tumor-mutational burden are predictive of response to immunotherapy.[23] We believe the same also applies, albeit to a lesser extent, to EGFR and ALK TKIs because not all patients with EGFR- and ALK-positive disease responded to treatment. Unfortunately, although genomic profiling is now almost routinely performed in clinical trials, such data are typically not examined holistically and are rarely submitted to the US Food and Drug Administration.

Our study has two important limitations. First, the data set was based on clinical trial data. Clinical trials have strict inclusion and exclusion eligibility criteria. As a result, the patients enrolled may not be representative of the heterogeneous, real-world patient population.[24] In this study, we increased the heterogeneity of our patient cohort by pooling patients from multiple trials. However, the population in our pooled data set can still be prone to external validity deficits arising from the strict exclusion criteria in traditional clinical trials independent of our study sample. It would be desirable to validate our models with patients outside the clinical trial setting.

To our knowledge, our work is one of the largest studies of NSCLC to consider biomarker and inhibitor therapy as candidate predictive variables. Putila et al[13] developed a prognostic model for OS on the basis of almost two decades of data (1998 to 2006) from the SEER database. However, although the sample size is impressive (more than 230,000 patients), the data set does not capture recent advances in treatment. Alexander et al[12] proposed the Lung Cancer Prognostic Index (LCPI) to predict OS. The model includes actionable mutations in EGFR, ALK, and KRAS as features, but lacks the PD-L1 biomarker. Its derivational cohort is also much smaller than our training set (700 *v* 5,400 patients).

Unfortunately, a direct comparison of our performance with the SEER and LCPI models is limited by differences in data and features. Both models focus on prognosis at the time of diagnosis, whereas our models focus on prediction before therapy for OR and PFS, and any time after diagnosis for OS. In addition, the SEER model requires TNM cancer staging information, which was not available in our data set. Nor did we have information on weight loss at diagnosis used in the LCPI model.

In this article, we aggregated data from 17 clinical trials to estimate OR, PFS, and OS models applicable to patients receiving different treatment modalities, including chemotherapy, targeted therapy, and immunotherapy. The models included established and novel predictive factors in lung cancer. We offered an interpretation of the effects of the variables and found them to be largely consistent with other NSCLC prognostic tools in the literature. The models illuminated the drivers behind response and survival, which may be useful for patient selection in future clinical trials. Survival risk scores may also be used in randomized trials to stratify patients into groups with homogeneous risk for analysis.

Our current results are promising for chemotherapy and targeted therapies, but less so for immunotherapy. We hypothesize that the lackluster performance in the PD-L1–positive subgroup receiving PD-L1 ICI treatment was specifically the result of a lack of relevant predictors. We believe that more powerful predictive models can be developed with a collection of diverse data types that more fully describe patient outcomes and phenotypic manifestations of the oncogenic process.

To advance beyond the results achieved in this article, current clinical trial data collection pipelines need to include composite multiomic signatures. For example, radiomic features from deep learning models have shown immense potential in NSCLC prognostication.[25-31] Such variables will inevitably become more important as we gradually reach the limits of biomedical reductionism—that is, the target-based treatment paradigm—and shift toward systems biology approaches for drug discovery.[32] This will allow us to develop better predictive models to truly tailor treatments and clinical trial enrollment strategies.

AFFILIATIONS

[1]Massachusetts Institute of Technology, Cambridge, MA
[2]US Food and Drug Administration, Silver Spring, MD
[3]Santa Fe Institute, Santa Fe, NM

CORRESPONDING AUTHOR

Andrew W. Lo, PhD, MIT Laboratory for Financial Engineering, Sloan School of Management, 100 Main St, E62-618, Cambridge, MA 02142; e-mail: alo-admin@mit.edu.

AUTHOR CONTRIBUTIONS

Conception and design: Kien Wei Siah, Sean Khozin, Andrew W. Lo
Financial support: Andrew W. Lo
Administrative support: Andrew W. Lo
Collection and assembly of data: Kien Wei Siah, Sean Khozin, Chi Heem Wong,
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated.

Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Chi Heem Wong

Stock and Other Ownership Interests: Editas Medicine, Sarepta Therapeutics, Intellia Therapeutics, Immunogen, Bluebird Bio

Andrew W. Lo

Leadership: Roivant Sciences, BridgeBio Pharma

Stock and Other Ownership Interests: Bridgebio Pharma, Nautilus Biotechnology, LS Polaris Innovation Fund, MPM Capital, Royalty Pharma, BillionToOne, QLS Advisors

Travel, Accommodations, Expenses: QLS Advisors

Other Relationship: BridgeBio Pharma, MIT Whitehead Institute, Beth Israel Deaconess Medical Center, BrightEdge Impact Fund, National Institutes of Health

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank three anonymous referees for specific comments on this manuscript and Jayna Cummings for editorial assistance. Research support from the Massachusetts Institute of Technology Laboratory for Financial Engineering is gratefully acknowledged. The views and opinions expressed in this article are those of the authors only and do not necessarily represent the views and opinions of any institution or agency, any of their affiliates or employees, or any of the individuals acknowledged here. This research effort was in part facilitated by the US Food and Drug Administration's Information Exchange and Data Transformation Program.

REFERENCES

1. Beaver JA, Tzou A, Blumenthal GM, et al: An FDA perspective on the regulatory implications of complex signatures to predict response to targeted therapies. Clin Cancer Res 23:1368-1372, 2017
2. Morgensztern D, Ng SH, Gao F, et al: Trends in stage distribution for patients with non-small cell lung cancer: A National Cancer Database survey. J Thorac Oncol 5:29-33, 2010
3. Yorke ED, Fuks Z, Norton L, et al: Modeling the development of metastases from primary and locally recurrent tumors: Comparison with a clinical data base for prostatic cancer. Cancer Res 53:2987-2993, 1993
4. Kimmel M, Gorlova O: Stochastic models of progression of cancer and their use in controlling cancer-related mortality. Proc 2002 Am Control Conf 5:3443-3448, 2002
5. Chia YL, Salzman P, Plevritis SK, et al: Simulation-based parameter estimation for complex models: A breast cancer natural history modelling illustration. Stat Methods Med Res 13:507-524, 2004
6. Talkington A, Durrett R: Estimating tumor growth rates in vivo. Bull Math Biol 77:1934-1954, 2015
7. Fawcett T: An introduction to ROC analysis. Pattern Recognit Lett 27:861-874, 2006
8. Cox DR: Regression models and life-tables. J R Stat Soc B 34:187-220, 1972
9. Ishwaran H, Kogalur UB, Blackstone EH, et al: Random survival forests. Ann Appl Stat 2:841-860, 2008
10. Faraggi D, Simon R: A neural network model for survival data. Stat Med 14:73-82, 1995
11. Harrell FE Jr, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. JAMA 247:2543-2546, 1982
12. Alexander M, Wolfe R, Ball D, et al: Lung cancer prognostic index: A risk score to predict overall survival after the diagnosis of non-small-cell lung cancer. Br J Cancer 117:744-751, 2017
13. Putila J, Remick SC, Guo NL: Combining clinical, pathological, and demographic factors refines prognosis of lung cancer: A population-based study. PLoS One 6:e17493, 2011 [Erratum: PLoS One 6(3), 2011]
14. Lin J, Carter CA, McGlynn KA, et al: A prognostic model to predict mortality among non-small-cell lung cancer patients in the U.S. military health system. J Thorac Oncol 10:1694-1702, 2015
15. Blanchon F, Grivaux M, Asselain B, et al: 4-year mortality in patients with non-small-cell lung cancer: Development and validation of a prognostic index. Lancet Oncol 7:829-836, 2006
16. Gridelli C, Ardizzoni A, Le Chevalier T, et al: Treatment of advanced non-small-cell lung cancer patients with ECOG performance status 2: Results of an European Experts Panel. Ann Oncol 15:419-426, 2004
17. Langer CJ: Older age itself rarely a restriction to treatment of lung cancer. http://www.cancernetwork.com/asco-lung-cancer/older-age-itself-rarely-restriction-treatment-lung-cancer
18. Mandrekar SJ, Schild SE, Hillman SL, et al: A prognostic model for advanced stage nonsmall cell lung cancer. Pooled analysis of North Central Cancer Treatment Group trials. Cancer 107:781-792, 2006
19. Kanz BA, Pollack MH, Johnpulle R, et al: Safety and efficacy of anti-PD-1 in patients with baseline cardiac, renal, or hepatic dysfunction. J Immunother Cancer 4:60, 2016
20. Kanz BA, Pollack MLH, Eroglu Z, et al: Anti-PD-1 in patients with advanced malignancies and baseline organ dysfunction. J Clin Oncol 34, 2017 (abstr e14539) doi:10.1200/JCO.2016.34.15_suppl.e14539
21. Hoang T, Dahlberg SE, Sandler AB, et al: Prognostic models to predict survival in non-small-cell lung cancer patients treated with first-line paclitaxel and carboplatin with or without bevacizumab. J Thorac Oncol 7:1361-1368, 2012
22. Derman BA, Mileham KF, Bonomi PD, et al: Treatment of advanced squamous cell carcinoma of the lung: A review. Transl Lung Cancer Res 4:524-532, 2015
23. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. N Engl J Med 377:2500-2501, 2017
24. Khozin S, Blumenthal GM, Pazdur R: Real-world data for clinical evidence generation in oncology. J Natl Cancer Inst 109:djx187, 2017
25. Sun R, Limkin EJ, Vakalopoulou M, et al: A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. Lancet Oncol 19:1180-1191, 2018
26. Rizzo S, Raimondi S, de Jong EEC, et al: Genomics of non-small cell lung cancer (NSCLC): Association between CT-based imaging features and EGFR and K-RAS mutations in 122 patients-An external validation. Eur J Radiol 110:148-155, 2019

27. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5: 4006, 2014 [Erratum: Nat Commun 5:4644, 2014]
28. Hosny A, Parmar C, Coroller TP, et al: Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS Med 15:e1002711, 2018
29. Khorrami M, Khunger M, Zagouras A, et al: Combination of peri- and intratumoral radiomic features on baseline CT scans predicts response to chemotherapy in lung adenocarcinoma. Radiology: Artificial Intelligence 1:180012, 2019
30. Xu Y, Hosny A, Zeleznik R, et al: Deep learning predicts lung cancer treatment response from serial medical imaging. Clin Cancer Res 25:3266-3275, 2019
31. Lambin P, Leijenaar RTH, Deist TM, et al: Radiomics: The bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749-762, 2017
32. Earm K, Earm YE: Integrative approach in the era of failing drug discovery and development. Integr Med Res 3:211-216, 2014
33. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825-2830, 2011
34. Therneau TM: A Package for Survival Analysis in S. https://CRAN.R-project.org/package=survival