Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: a machine learning approach

Samuel Peter Heilbroner [ID] ,[1] Reed Few,[1] Tomas G Neilan,[2] Judith Mueller,[1] Jitesh Chalwa,[1] Francois Charest,[1] Somasekhar Suryadevara,[1] Christine Kratt,[3] Andres Gomez-Caminero,[3] Brian Dreyfus[3]

¹Data Science, ConcertAI, New York, New York, USA
²Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA
³Bristol Myers Squibb, New York, New York, USA

Correspondence to
Dr Samuel Peter Heilbroner; sheilbroner@concertai.com

ABSTRACT

Background Treatment with immune checkpoint inhibitors (ICIs) has been associated with an increased rate of cardiac events. There are limited data on the risk factors that predict cardiac events in patients treated with ICIs. Therefore, we created a machine learning (ML) model to predict cardiac events in this at-risk population.

Methods We leveraged the CancerLinQ database curated by the American Society of Clinical Oncology and applied an XGBoosted decision tree to predict cardiac events in patients taking programmed death receptor-1 (PD-1) or programmed death ligand-1 (PD-L1) therapy. All curated data from patients with non-small cell lung cancer, melanoma, and renal cell carcinoma, and who were prescribed PD-1/PD-L1 therapy between 2013 and 2019, were used for training, feature interpretation, and model performance evaluation. A total of 356 potential risk factors were included in the model, including elements of patient medical history, social history, vital signs, common laboratory tests, oncological history, medication history and PD-1/PD-L1-specific factors like PD-L1 tumor expression.

Results Our study population consisted of 4960 patients treated with PD-1/PD-L1 therapy, of whom 418 had a cardiac event. The following were key predictors of cardiac events: increased age, corticosteroids, laboratory abnormalities and medications suggestive of a history of heart disease, the extremes of weight, a lower baseline or on-treatment percentage of lymphocytes, and a higher percentage of neutrophils. The final model predicted cardiac events with an area under the curve–receiver operating characteristic of 0.65 (95% CI 0.58 to 0.75). Using our model, we divided patients into low-risk and high-risk subgroups. At 100 days, the cumulative incidence of cardiac events was 3.3% in the low-risk group and 6.1% in the high-risk group (p<0.001).

Conclusions ML can be used to predict cardiac events in patients taking PD-1/PD-L1 therapy. Cardiac risk was driven by immunological factors (eg, percentage of lymphocytes), oncological factors (eg, low weight), and a cardiac history.

BACKGROUND

Immune checkpoint inhibitor (ICI) therapy with programmed death receptor-1 (PD-1) or programmed death ligand-1 (PD-L1) inhibitors has dramatically improved outcomes in patients with cancer.[1–8] However, ICI therapy has also been associated with several immune-related adverse events (irAEs).[9] Cardiac irAEs with ICI therapy are not common but have a mortality rate of up to 30%,[9] and the occurrence of cardiac events with ICI therapy is likely under-reported.[10–14] Myocarditis is the most well-described cardiac irAE with ICI therapy[15–19]; however, cardiac irAEs such as pericarditis, pericardial effusions, and acute vascular events are increasingly described.[11 12 15 20–22]

The risk factors for the development of adverse cardiac events with ICI therapy are poorly understood.[14] Beyond combination immune therapy,[15] there are no established risk factors for cardiac events with ICI therapy. Identification of patients at increased risk of ICI-related cardiac side effects may support testing of rational surveillance strategies in such high-risk populations. However, efforts to identify a high-risk population have been limited by the sample size of most of the studies to date and identification of the optimal approach for such risk stratification. Machine learning (ML) has improved our ability to predict outcomes in oncology.[23] Furthermore, ML techniques better account for the complex interaction between risk factors which have gone relatively unevaluated so far. Using Shapley additive explanations (SHAP), ML can be used to discover novel insights into the disease process.[24–27] Therefore, we created an ML decision tree model for predicting cardiac events among patients being treated with a PD-1 or PD-L1 inhibitor in the CancerLinQ database. We also used SHAP to identify novel risk factors for future cardiac events.[28]

METHODS

To predict cardiac events, we trained an XGBoosted decision tree model on all

patients with non-small cell lung cancer (NSCLC), melanoma, and renal cell carcinoma (RCC) who had PD-1/PD-L1 therapy.

Data source

Our model was trained on ConcertAI data. ConcertAI derives its data from several sources, most notably the CancerLinQ database (American Society of Clinical Oncology), which in turn aggregates data from oncological practices across the USA. ConcertAI receives copies of each oncological practice's complete electronic medical record (EMR), including unstructured notes and scanned reports. Structured data from the EMR such as ICD codes, medications, and laboratory values are extracted programmatically, combined with important hand curated variables from provider notes, and entered into a standardised database which can be easily analysed and queried. Data are provided as far back as the practices have history.

This real-world dataset provides information on patient demographics, tumor characteristics, comorbidities, treatment information, treatment toxicities, and outcomes. A subset of patients with at least 10 documents in their chart were selected for deep curation, a process in which a team of oncology nurses read a patient's notes and abstracted additional information, including treatment toxicities and tumor response/progression information. Our final analysis was conducted only on patients who underwent curation.

Cohort selection and labeling

Our cohort included all patients with NSCLC, melanoma, and RCC who (1) underwent nurse curation of treatment toxicities, (2) did not have evidence of a second malignancy, (3) had PD-1/PD-L1 therapy regardless of stage, and (4) were at least 18 years old at index. Index was defined as the date of first PD-1/PD-L1 administration. If a patient received multiple distinct PD-1/PD-L1 drugs, the last PD-1/PD-L1 drug taken was used for indexing. This was done to maximize the number of patients with combination immunotherapy available for analysis.

A cardiac event was defined as the first documentation of arrhythmia (eg, complete heart block and ventricular fibrillation), heart failure, myocarditis, or pericardial disease as defined by the Medical Dictionary for Regulatory Activities (MedDRA) and International Classification of Diseases (ICD) codes (table 1). Codes from both the native EMRs and nurse curation were included to define endpoints. Mild or non-specific cardiac conditions, such as chest pain, were not included because they are difficult to attribute to a new onset cardiac event and may represent documentation of alternate etiologies. Some providers routinely document chronic conditions in a patient's chart. To ensure that all our events represented new cardiac disease, we did not include an event if the patient had a history of similar events prior to the index date. For example, if a patient had a history of heart failure prior to starting immunotherapy, heart failure ICD codes were ignored when defining cardiac events for this patient. ICD and MedDRA codes were defined as similar if they fell into the same category (table 1). If a patient had any ICD or MedDRA code for one of the conditions listed in table 1 in the 6 months prior to the index date, then they became ineligible for a cardiac event in the same category.

To avoid missing cardiac events, patients were censored after their last date of continuous follow-up. Continuous

Table 1				
Disease category	Count	Percent	ICD codes	MedDRA codes*
Atrial fibrillation	143	34	I48.x	10003658
Heart failure	95	23	I50.x Exclude I50.2(2-4)x	10019279
Pericardial disease	84	20	I3(0–1).x Exclude I31.3x	10034474, 10034484, 10053565
Unspecified arrhythmia	39	9	I49, I49.(89)x	
Cardiac arrest	20	5	I46.x	10007515
Premature depolarization	20	5	I49.(1–4)x	
Atrioventricular block	13	3	I44.x	10003673
Sick sinus syndrome	3	1	I149.5x	
Myocarditis	1	0	I4(0–1).x Exclude I40.0x	
Ventricular arrhythmia	0	0	I49.0x	
Total	418	100		

Table 1. Distribution of cardiac events among patients receiving PD-1/PD-L1 therapy. PD-1, Programmed cell death protein 1; PD-L1, Programmed death-ligand 1.

*Only MedDRA codes present in the ConcertAI database are listed. This is not an exhaustive list.

ICD, International Classification of Diseases; MeDRA, Medical Dictionary for Regulatory Activities.

follow-up was defined as at least two EMR entries within a 6-month period. To better establish a causal relationship between PD-1/PD-L1 therapy and cardiac events, patients were also censored 100 days after their last PD-1/PD-L1 administration.

Time windowing

To avoid the effects of immortal time bias[29] and reverse causality, most patient variables were calculated using data from 6 months before index to index. For vitals, laboratory results, and Eastern Cooperative Oncology Group (ECOG), some patients had missing values if we did not include data from immediately after the index date. To minimize bias secondary to missing data, we allowed these values to be calculated using data from up to 30 days after index. If a patient suffered a cardiac event within 30 days of index, the data were truncated 15 days before the cardiac event. To better distinguish between baseline and on-treatment laboratory values, we conducted a sensitivity analysis using only data that was strictly before the index date. The results are reported in the appendix (online supplemental figure 1) and did not change significantly from our primary analysis.

Model variables

A total of 356 variables were abstracted from the data for modeling. Demographics, such as age at diagnosis, sex, race, and ethnicity, were extracted directly from the electronic medical record. Cancer type was abstracted through nurse curation and documented using ICD codes. A patient's pseudostage was defined as stage 4 if the patient had progressed to metastatic disease by index date and their stage at diagnosis otherwise. Because patients with NSCLC represented such a large percentage of our cohort, we also included NSCLC histology as a covariate by mapping ICD for Oncology, Third Edition codes in the data to categories defined in the Surveillance, Epidemiology, and End Results Database guidelines.[30] When a patient had multiple records which disagreed (eg, multiple different histology codes), the record closest to the index was used for analysis.

Comorbidities were calculated using ICD codes documented in the EMR and by nurse curators. The Charlson Comorbidity Index was calculated using the method described by Deyo.[31] Common comorbidities, such as congestive heart failure, were calculated using ICD codes enumerated in the Centers for Medicare and Medicaid Services' Chronic Conditions Data Warehouse algorithm.[32] Autoimmune conditions were calculated using codes listed in online supplemental table 1. ECOG performance status was taken directly from EMR records and curated data. Smoking status was binned into current-smoker, ex-smoker, and never-smoker categories. When a single patient had multiple discordant smoking statuses, the worst was used for analysis (ie, current smoker >ex-smoker>never smoker).

Vitals, including heart rate, respiratory rate, blood pressure, O_2 saturation, temperature, weight, and body mass index (BMI) were collected using Logical Observation Identifiers Names and Codes (LOINC) codes (online supplemental table 2). Change in weight per day (lb/day) was calculated as the slope of the trend line fitting the patient's weight in the 6 months prior to diagnosis. A negative slope indicated that the patient was losing weight; a positive slope indicated that they were gaining weight. Common laboratory and panel values, including complete blood count (CBC), basic metabolic panel (BMP), magnesium, phosphorous, liver function tests, lactate dehydrogenase, B-type natriuretic peptide (BNP), troponin, erythrocyte sedimentation rate (ESR), C-reactive protein, prothrombin time, partial thromboplastin time, international normalised ratio, D-dimer, and fibrinogen, were also collected using LOINC codes (online supplemental table 2). These laboratory parameters were either recorded prior to the start of an ICI or within 30 days after start of an ICI. The absolute neutrophil count and absolute lymphocyte count (ALC) were calculated using the corresponding percent values from the CBC and the white blood count. PD-1/PD-L1 expression was available directly in the data.

Using RxNorm codes, all of a patient's oncological and non-oncological medications were collected. To simplify the model, medications of the same class (eg, platinum-based chemotherapies, angiotensin-converting enzyme inhibitors) were grouped together using the Food and Drug Administration (FDA) Established Pharmacology Class (EPC).[33] The EPC classification system already includes categories for PD-1, PD-L1, and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) therapies. However, we also included the number of distinct immunotherapies a patient has undergone to explicitly model the interaction between these drugs.

Missing data

Because our data were derived from oncology clinics, many of the non-oncological laboratory tests (eg, troponin, BNP) were missing. XGBoost naturally handles missing data. However, because patients with cardiac events after index date but before 30 days have passed have their data truncated, missing data were erroneously associated with early adverse events. As a sensitivity analysis, we masked missing values with multivariate imputation by chained equations using the 10 nearest features.[34] We also analysed how the model changed when we excluded features with missing rates above 70%, 80%, and 90%.

Modeling

We used an XGBoosted decision tree to predict time to cardiac events. To properly account for censoring, we modified the model's loss function to the same one used in the Cox proportional hazard (CoxPH) model.[35 36] For each patient, the model predicted an HR. As in a CoxPH model, a higher HR indicated greater risk. The tree depth and learning rate were tuned to 2.0 and 0.15, respectively, based on a stochastic search with cross validation (n=10).

Validation and performance

The data were randomly divided into training (80%) and testing (20%) sets. The model's ability to predict cardiac events within 20, 40, 60, 80, 100, 120, and 140 days of index was assessed using area under the curve–receiver operating characteristics (AUC-ROCs).[37] We also calculated the concordance index, which represents the percentage of time a model, given a random pair of patients, correctly predicts which patient will have a cardiac event first. Finally, we divided the PD-1/PD-L1 test set into two equal-sized high-risk and low-risk subcohorts. The high-risk group had predicted HRs above the median, and patients in the low-risk group had HRs below the median. The time elapsed prior to an adverse cardiac event for these two subcohorts was compared using the cumulative incidence function and the Aalen-Johansen estimator.

Model interpretation and statistical analysis

We used SHAP to determine the variables most predictive of cardiac risk.[38] Bootstrapping (n=500) was used to calculate p values for each feature under the null hypothesis that the mean absolute SHAP score for that feature was zero.

RESULTS

There were 4960 patients in our cohort. Of these patients, 4103 were treated for NSCLC; 534 were treated for melanoma; and 323 were treated for RCC. The median follow-up time was 154 (IQR 57–342) days. A total of 418 patients experienced a cardiac event before they were censored. The most common cardiac events were atrial fibrillation (34%), heart failure (23%), and pericardial disease (20%). The rates of all cardiac events are listed in table 1. Model training was performed in 80% of patients (3937 PD-1/PD-L1 patients). The remaining patients were used for model validation. We also conducted several sensitivity analyses where we included ACS related events or removed atrial fibrillation from the list of events. The results were unchanged.

Demographics are shown in table 2, stratified by whether the patient had a cardiac event within 100 days or not. Patients censored before 100 days are not included in table 2 because one cannot determine if they would have had a cardiac event after they were censored. However, censored patients were not excluded from the modeling because the CoxPH loss function can adjust for this source of bias. Compared with those without an event, patients with a cardiac event were more likely to be white (75% vs 68%) or African–American (15% vs 14%), male (61% vs 53%), have lung cancer (88% vs 80%), be at a higher stage, be a current smoker or ex-smoker, have chronic obstructive pulmonary disease (41% vs 31%), have a higher ECOG status, or have a higher Charlson score. These results were all significant at a p<0.05 level.

The rates of missing data varied by feature (full results in online supplemental table 3). In general, missing data

in demographic variables (such as age) were negligible. Common vital signs and laboratory values (like BMP and CBC) had missing rates between 10% and 60%. Laboratory tests that are less commonly ordered in the oncology clinic (such as ESR, BNP, and troponin) had missing rates of as high as 99%. Despite these high missing rates, we opted to include these variables in the model because missingness can be an informative feature. We conducted several sensitivity analyses where we excluded features with missing rates above 70%, 80%, and 90%. We also tried imputing missing values with multiple imputation with chained equations. The results of these sensitivity analyses are shown in the appendix (online supplemental table 4 and figures 2–5) and did not substantially differ from our primary analysis.

Using SHAP, we identified several variables that helped predict time to adverse cardiac event. The 40 variables with the highest mean absolute SHAP score are shown in figure 1. Some of the variables associated with an increased risk of cardiac events are as follows: increased age (p<0.002); immunological labs including a lower baseline or peri-index percentage of lymphocytes (p=0.002), a higher percentage of neutrophils (p=0.02), a lower absolute lymphocyte count (p<0.002), a higher absolute neutrophil count (p=0.006), and a higher WBC (p=0.02); administration of immunomodulators including absence of corticosteroids (p=0.008) and receipt of a PD-L1 antibody instead of a PD-1 antibody (p=0.028); a higher platelet count (p<0.002); vital signs such an abnormal (high or low) heart rate (p<0.002), abnormal temperature (p<0.002), abnormal weight (p=0.008), increase in weight over time (p=0.002), or abnormal BMI (p=0.01); various laboratory abnormalities and medications associated with heart failure including not receiving an ACE inhibitor (p=0.026), not receiving a loop diuretic (p=0.01), lower hemoglobin (p<0.002), lower sodium (p<0.002), and lower chloride (p<0.002); liver abnormalities including a lower aspartate aminotransferate (AST) (p=0.01), a lower alanine aminotransferase (ALT) (p=0.002), and lower alkaline phosphatase (p=0.02); renal abnormalities including a lower creatinine (p<0.002) or higher blood urea nitrogen (BUN) (p=0.002); a missing prothrombin time (p=0.004); and a higher lactate dehydrogenase (LDH) (p=0.004). Note that although therapy with PD-1 therapy significantly impacted cardiac risk, it was not a top 40 predictor and cannot be seen in figure 1. A complete list of all significant risk factors is listed in online supplemental table 5. To determine the clinical significance of vitals sign changes detected by the model, we examined the model's trees by hand: heart rate had three major cut-offs: ~60, ~100, and ~150 beats/min; temperature had cut-offs at around 36.1°C and 37.8°C.

To maximize model performance, we included laboratory, vitals, and ECOG values up to 30 days after index. However, this choice also limits our ability to interpret which variables contribute most to cardiac risk. In general, we found that ~75% of these values were baseline and ~25% of values were in the immediate postindex period.

Table 2 Demographics of patients in our cohort, stratified by cardiac event

		No cardiac event in 100 days	Cardiac event in 100 days	Total*	P value
n		2928	301	3229	
Age		64.2 (10.7)	65.6 (11.3)	64.4 (10.8)	0.066
Race	White	1996 (68.2)	227 (75.4)	2223 (68.8)	0.004
	Black or African–American	410 (14.0)	45 (15.0)	455 (14.1)	
	Other race	446 (15.2)	23 (7.6)	469 (14.5)	
	Missing	76 (2.6)	6 (2.0)	82 (2.5)	
Ethnicity	Hispanic or Latino	76 (2.6)	6 (2.0)	82 (2.5)	0.244
	Not Hispanic or latino	2275 (77.7)	224 (74.4)	2499 (77.4)	
	Unknown	577 (19.7)	71 (23.6)	648 (20.1)	
Sex	Female	1379 (47.1)	118 (39.2)	1497 (46.4)	0.031
	Male	1548 (52.9)	183 (60.8)	1731 (53.6)	
	Unknown	1 (0.0)		1 (0.0)	
Cancer type	Lung	2355 (80.4)	266 (88.4)	2621 (81.2)	0.004
	Melanoma	365 (12.5)	23 (7.6)	388 (12.0)	
	RCC	208 (7.1)	12 (4.0)	220 (6.8)	
Stage at index	Missing	143 (4.9)	21 (7.0)	164 (5.1)	0.028
	Stage 0	2 (0.1)		2 (0.1)	
	Stage 1	92 (3.1)	5 (1.7)	97 (3.0)	
	Stage 2	118 (4.0)	7 (2.3)	125 (3.9)	
	Stage 3	480 (16.4)	34 (11.3)	514 (15.9)	
	Stage 4	2093 (71.5)	234 (77.7)	2327 (72.1)	
Stage at Diagnosis	Missing	154 (5.3)	25 (8.3)	179 (5.5)	0.050
	Stage 1	207 (7.1)	13 (4.3)	220 (6.8)	
	Stage 2	214 (7.3)	20 (6.6)	234 (7.2)	
	Stage 3	675 (23.1)	60 (19.9)	735 (22.8)	
	Stage 4	1678 (57.3)	183 (60.8)	1861 (57.6)	
Smoking status	Ex-smoker	1111 (37.9)	125 (41.5)	1236 (38.3)	0.024
	Missing/unknown	447 (15.3)	31 (10.3)	478 (14.8)	
	Never smoker	397 (13.6)	31 (10.3)	428 (13.3)	
	Smoker	973 (33.2)	114 (37.9)	1087 (33.7)	
ECOG score	0	574 (19.6)	49 (16.3)	623 (19.3)	0.006
	1	1040 (35.5)	110 (36.5)	1150 (35.6)	
	2	311 (10.6)	41 (13.6)	352 (10.9)	
	3	51 (1.7)	13 (4.3)	64 (2.0)	
	4	2 (0.1)	1 (0.3)	3 (0.1)	
	Missing	950 (32.4)	87 (28.9)	1037 (32.1)	
Charlson score	0	4 (0.1)	2 (0.7)	6 (0.2)	0.004
	1–2	436 (14.9)	28 (9.3)	464 (14.4)	
	3+	2488 (85.0)	271 (90.0)	2759 (85.4)	
AIDS–HIV		9 (0.3)	1 (0.3)	10 (0.3)	1.000
Cerebrovascular disease		84 (2.9)	13 (4.3)	97 (3.0)	0.220
Chronic pulmonary disease		907 (31.0)	122 (40.5)	1029 (31.9)	0.001
Congestive heart failure		265 (9.1)	35 (11.6)	300 (9.3)	0.173

Continued

Table 2 Continued

		No cardiac event in 100 days	Cardiac event in 100 days	Total*	P value
Dementia		9 (0.3)	2 (0.7)	11 (0.3)	0.274
Diabetes mellitus with chronic complication		44 (1.5)	2 (0.7)	46 (1.4)	0.314
Diabetes mellitus without chronic complication		434 (14.8)	51 (16.9)	485 (15.0)	0.370
Hemiplegia or paraplegia		8 (0.3)		8 (0.2)	1.000
Malignancy		2924 (99.9)	299 (99.3)	3223 (99.8)	0.101
Metastatic solid tumor		2128 (72.7)	228 (75.7)	2356 (73.0)	0.283
Mild liver disease		133 (4.5)	17 (5.6)	150 (4.6)	0.469
Moderate or severe liver disease		7 (0.2)		7 (0.2)	1.000
Myocardial infarction		132 (4.5)	17 (5.6)	149 (4.6)	0.451
Peptic ulcer disease		27 (0.9)	3 (1.0)	30 (0.9)	0.755
Peripheral vascular disease		118 (4.0)	17 (5.6)	135 (4.2)	0.236
Renal disease		179 (6.1)	24 (8.0)	203 (6.3)	0.254
Rheumatic disease		38 (1.3)	6 (2.0)	44 (1.4)	0.296
PD-1/PD-L1 therapy	Atezolizumab	171 (5.8)	12 (4.0)	183 (5.7)	0.233
	Avelumab	3 (0.1)		3 (0.1)	
	Durvalumab	65 (2.2)	2 (0.7)	67 (2.1)	
	Nivolumab	1921 (65.6)	207 (68.8)	2128 (65.9)	
	Pembrolizumab	768 (26.2)	80 (26.6)	848 (26.3)	
PD-1 testing	Missing	2331 (79.6)	235 (78.1)	2566 (79.5)	0.497
	Negative	143 (4.9)	11 (3.7)	154 (4.8)	
	Not interpretable	133 (4.5)	15 (5.0)	148 (4.6)	
	Positive	321 (11.0)	40 (13.3)	361 (11.2)	

*Excludes patients censored in first 100 days.
†Mean (SD).
‡
ECOG, Eastern Cooperative Oncology Group; PD-1, programmed death receptor-1.

As a sensitivity analysis, we repeated the model using only values that were before index. The resulting SHAP plots are shown in online supplemental figure 1. Results did not vary substantially from our primary analysis.

The model's ability to predict cardiac events within 20, 40, 60, 80, 100, 120, and 140 days of index is shown in figure 2. At 100 days, the AUC-ROC was 0.65 (95% CI 0.58 to 0.75). The model had a concordance index of 0.66 (95% CI 0.57 to 0.71). Using our model, we divided patients into low-risk and high-risk subgroups. The cumulative incidence of cardiac events in these two groups is shown in figure 3. At 100 days, the cumulative incidence of cardiac events was 3.3% in the low-risk group and 6.1% in the high-risk group (p<0.001), 80% higher.

In the real clinical setting, not all the 356 variables in our model would always be available. Therefore, we tested how our model performed if only the top 20 features from our SHAP plots were used. The AUC-ROC of this simpler model was 0.65 and the

concordance index was also 0.65. Complete results are shown in online supplemental table 1.

Our model evaluated several immunotherapy-related factors, including the number of distinct immunotherapies undergone, the type of immunotherapy, and PD-1/PD-L1 tumor expression values. Only receipt of PD-1 versus PD-L1 therapy was significant.

DISCUSSION
We leveraged a large cross-sectional database of patients with cancer and created an ML model that predicts cardiac events in patients taking PD-1/PD-L1 therapy. The model had solid performance (AUC-ROC=65% at 100 days from index). We also showed that a simpler model using only the top 20 features had comparable performance. In practice, a provider could use this simpler model to enter values by hand or integrate the more complex model into their
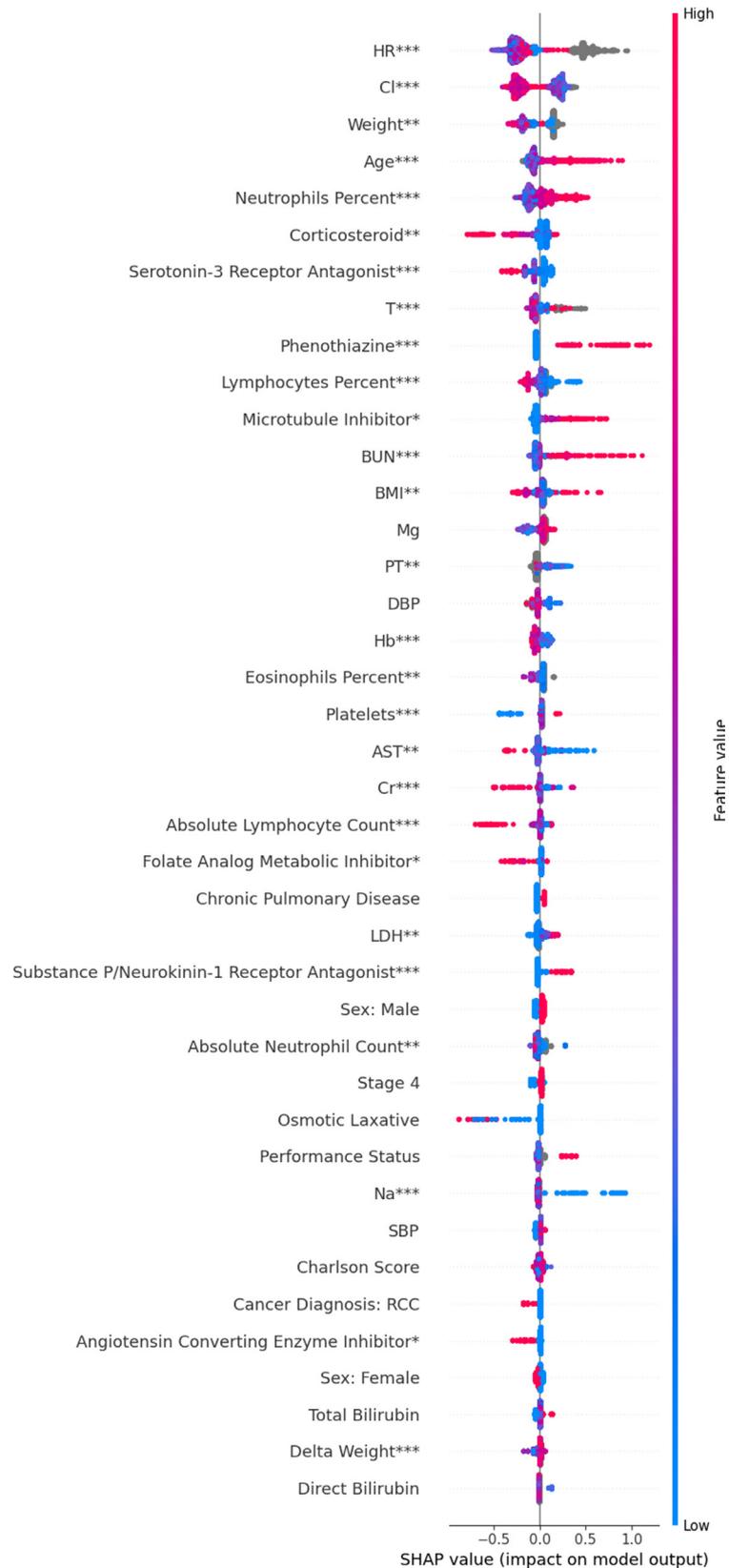
Figure 1 SHAP summary plot for interpreting the impact of features on our model. Each row shows the impact of a single feature on the model's predictions. Within each row, each dot represents a patient. Red means patients had a high feature value; blue means patients had a low value; gray means patients had a missing value. The position of the dot along the x-axis indicates whether that feature increased or decreased a patient's predicted risk. When all the red dots are on the right, a high feature value was associated with increased risk. When all the blue dots are on the right, a low feature value increased risk. Statistical significance is indicated as follows: *p<0.05, **p<0.01, ***p<0.002. BMI, body mass index; Cr, Creatinine; DBP, Diastolic Blood Pressure; SBP, Systolic Blood Pressure; Hb, hemoglobin; SHAP, Shaply additive explanations.
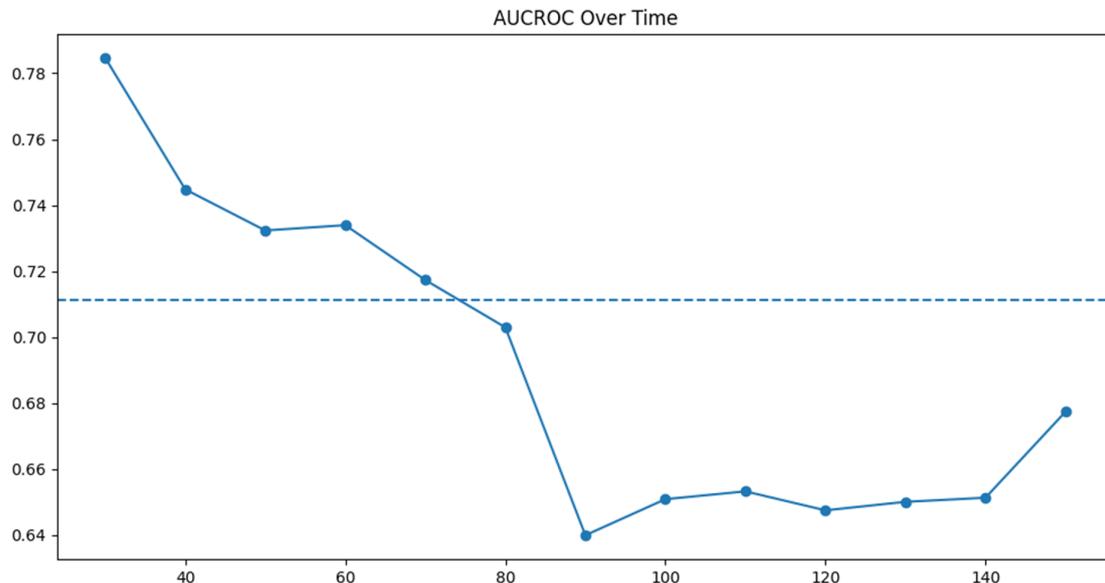
Figure 2 Plot of the cumulative dynamic AUC-ROC of our model on PD-1/PD-L1 patients.[61] Model's ability to predict cardiac adverse events within 20, 40, 60, 80, 100, 120, and 140 days of index is shown. The model's performance varied between 63% and 72% as the time window for predictions changed. AUC-ROC, area under the curve–receiver operating characteristic; PD-1, programmed death receptor-1; PD-L1, programmed death ligand-1.

EMR, thus avoiding manual data entry. Because our model can handle missing values and was trained on a real world dataset that replicates the distribution of missing values seen in clinical practice, a provider need not order any additional laboratory or imaging tests to use our model. These values may simply be entered as missing.
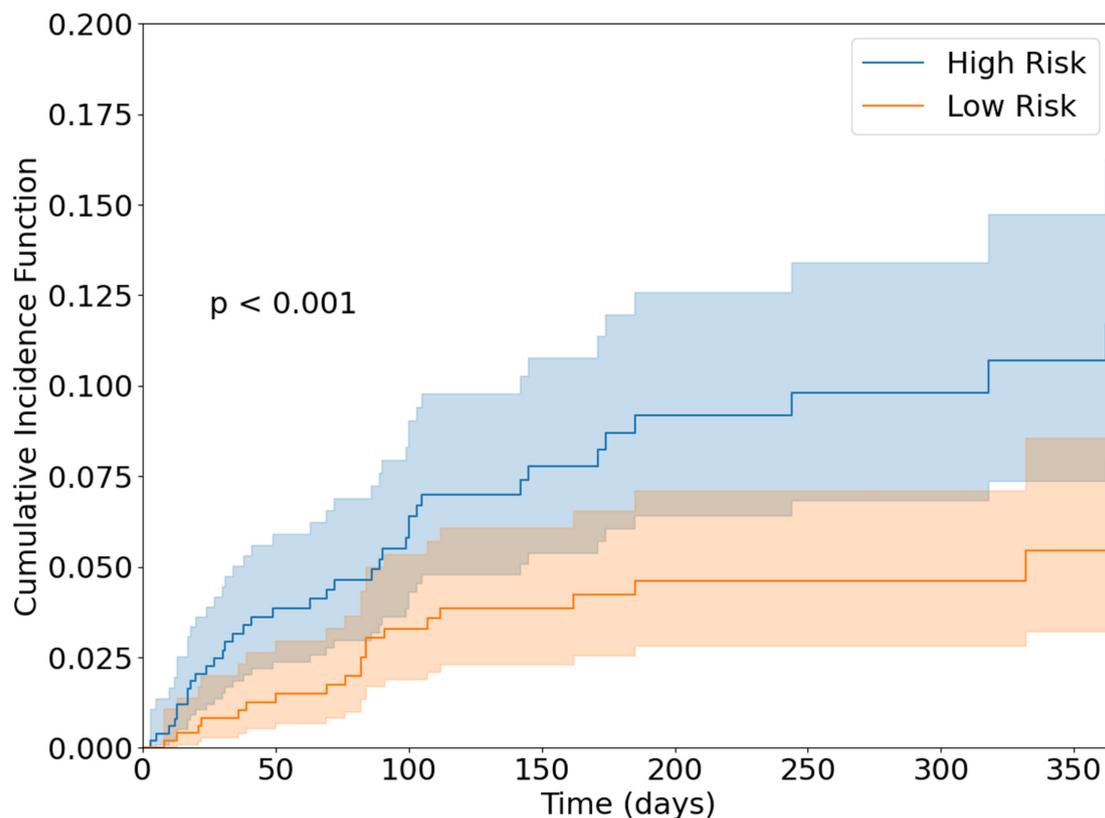
Figure 3 Cumulative incidence of cardiac events in low-risk and high-risk PD-1/PD-L1 patients from the test set. Groups were stratified by our model's median predicted HR. The cumulative incidence function was calculated using the method described by Aalen and Johansen, taking into account the competing risk of death.[62] High-risk patients had a significantly higher incidence of cardiac events. PD-1, programmed death receptor-1; PD-L1, programmed death ligand-1.

The results are complementary and additive to what has previously been reported in the literature. For example, Kartolo *et al* were able to predict any irAE with an AUC-ROC of 82.4% using clinically driven factors, similar to our model.[39] Our performance was probably lower than that found by Kartolo *et al* because our model was predicting very specific and rare cardiac events. Using SHAP, we also elucidated several novel risk factors for cardiac adverse events. In general, a combination of immunological factors, oncological factors, and cardiac history was associated with subsequent cardiac events.

We found that patients with a higher percent lymphocyte count, lower percent neutrophil count, or lower platelet count were at lower risk of cardiac disease. The neutrophil to lymphocyte ratio (NLR) is thought to be a marker of the inflammatory tumor microenvironment and therefore could represent a patient's potential risk of ICI-mediated cardiac toxicity. Traditionally, a low NLR has been associated with both a higher risk of toxicity and a higher response rate.[40] We found that patients with a low NLR were at decreased risk of cardiac events. This suggests that in patients with a low NLR, ICI therapies' ability to treat the tumor and reduce cardiopulmonary burden is protective against cardiac events. Platelets are known to express PD-1 receptors and interact with immunotherapy[41] and have traditionally been associated with a better response and higher rates of toxicity.[40] This is consistent with our findings.

In our model, corticosteroids were associated with a lower rate of cardiac events. The use of corticosteroids in patients on immunotherapy is well studied.[42] PD-1 therapy (vs PD-L1) was also associated with a lower risk of cardiac disease. Traditionally, treatment with PD-1/CTLA-4 combined therapy has been associated with the highest risk of myocarditis.[15] However, treatment with PD-1 therapy has also been associated with lower rates of myocarditis.[15] Given the borderline significance (p=0.028) of this result, low number of PD-L1 patients, and potential for a type I error in the setting of multiple comparisons, it is important to be cautious when interpreting this result.

We identified abnormal heart rate, both high and low, and abnormal temperature as predictors of cardiac events. Baseline vital signs help identify baseline disease in the patient, even if the patient does not carry a formal diagnosis with an ICD code for that condition. For example, our final model identified patients with a heart rate of <60 or >100 beats/min as being at high risk of a future cardiac event. Having a heart rate outside of the normal range could suggest subclinical arrhythmias with an abnormal rate (<60 and >100 beats/min) or more severe heart failure (>100 beats/min). It is interesting that blood pressure was not a significant risk factor, while heart rate was. Although blood pressure can be abnormal in heart failure, we believe that heart rate is a more indicative finding. This is supported by the Boston criteria, which use tachycardia to diagnose heart failure but do not use blood pressure.[43] Our results also identified

low temperature as a risk for cardiac events. A low body temperature has previously been associated with a low cardiac output and poor prognosis in patients with heart failure.[44]

We identified several parallel and aligned factors related to heart failure and severity of heart failure that were associated with cardiac events after PD-1/PD-L1 therapy. Specifically, an elevated heart rate, lower hemoglobin, lower sodium, and lower chloride are all variables associated with the severity of heart failure and were also associated with cardiac events in our study. We found that a high BUN and low creatinine also increased cardiac risk. The BUN to creatinine ratio has long been associated with heart failure severity. ACE inhibitors and loop diuretics were associated with a decreased rate of cardiac events. The novel finding of an association between features of heart failure and future events after PD-1 and PD-L1 inhibitory therapies is logical as a high incidence of the events post-PD-1 and PD-L1 therapies is related to heart failure. PD-1 and PD-L1 therapies may associate with heart failure through a variety of mechanisms. The most common mechanism is through the development of an acute and fulminant myocarditis.[14] The second, and less common and less well-understood mechanism, is via a progressive decline in the left ventricular ejection fraction without a clear acute fulminant process. Independent of the mechanism, the likelihood is that an impaired baseline cardiac reserve increases the probability that cardiac injury with an ICI will become clinically manifest as a heart failure event. What is unclear, and needs to be the focus of future studies, is whether the baseline elevated risk with heart failure is related to heart failure with a reduced ejection fraction, heart failure with a preserved ejection fraction, or both, and how baseline cardiac medications such as inhibitors of the angiotensin system or beta-blockers for heart failure may attenuate this risk.

The most common cardiac event on an ICI was the occurrence of atrial fibrillation. Atrial fibrillation and heart failure frequently coexist. We also found that the presence of a prothrombin time lab measurement was predictive and this may relate to the use of anticoagulation in the setting of atrial fibrillation.

Additionally, advanced age was associated with a higher rate of cardiac events in our model. Age-related immunosenescence is generally thought to increase irAE risk through a paradoxically higher concentration of inflammatory cytokines and autoantibodies.[45] Abnormal weight/BMI (high or low) and higher LDH levels also predicted cardiac disease (p<0.002). Weight loss can be driven by a variety of factors in patients with cancer, but immunological derangements mediated by cytokines are thought to be a primary driver.[46–54] Weight loss is also a general indicator of cancer severity[55] as are higher LDH levels.[56] Conversely, the association between an elevated weight/BMI and cardiac disease is well known. By checking the trees of our XGBoost model by hand, we identified a BMI of >40 as a cut-off for increased risk of cardiac disease.

Our study does have several limitations. To define cardiac events, we used both curated adverse events that were documented by a team of nurse abstractors and ICD codes documented in the patient chart. The frequency of different cardiac events is described in table 1. In our study, 8% of patients had a cardiac event. In general, we found a high incidence of atrial fibrillation, heart failure, and pericardial disease in our cohort. While the rate of myocarditis in ICI patients is thought to be less than 1%,[57] the rate of other cardiac events secondary to ICI-therapy is still unknown. A recent meta-analysis found that 0.5% of patients with cancer treated with ICIs developed myocarditis, 0.3% developed heart failure, and 4.6% developed atrial fibrillation. In addition, pericardial effusion occurred in 0.5% of patients, cardiomyopathy in 0.3% of patients, myocardial infarction in 0.4% of patients, and cardiac arrest in 0.4% of patients.[58] Our findings are in line with other studies using EHR data. For example, Waheed *et al* reported 15% of ICI patient developed new cardiac disease: 0.4% developed cardiomyopathy; 5% developed heart failure; 6% developed an arrhythmia; 2% developed pericardiac disease; 2% developed heart block; and 0.2% developed myocarditis.[59]

It is possible that pre-existing atrial fibrillation/heart failure is being routinely documented by a patient's oncologist, even if the patient is not experiencing new disease. To address this concern, we did not count ICD/MedDRA codes as cardiac events if similar codes were already part of the patient's medical history. However, whether these cardiac events truly represent new heart failure, an exacerbation of existing heart failure, or routine documentation of existing disease is impossible to tell with real-world data and is a major limitation of our study.

Another limitation is that the ConcertAI database does not have robust documentation of Common Terminology Criteria for Adverse Events, and it is impossible for us to distinguish between severe and non-severe events using the traditional definition.

Finally, while demographics and oncological fields like stage had a very low rate of missingness, laboratory values and vital signs suffered high missingness rates of between 20% and 99%. Traditionally, high missingness rates can introduce bias because (1) patients with missing values have to be excluded from the analysis or (2) missing values are replaced with imputed values. In our case, neither of these mechanisms are possible. Because XGBoost treats missing values as a separate entity, no patients are excluded in the modeling process and no values need be imputed. Missing values are simply treated as another value in SHAP plots and can be interpreted separately from the non-missing values. Indeed, XGBoost has been previously sited for its ability to robustly manage missing data in large EMR datasets.[60] In light of these findings, we did not remove features with high missingness rates from our final model. However, in several sensitivity analyses, we did show that removing variables with high missingness or imputing missing values has a minimal impact on our results.

To minimize missing data, we included data from up to 30 days after start of ICI. With this analysis alone, it is unclear whether the laboratory, vital, or other patient measurements were baseline or on-treatment. Therefore, we conducted a sensitivity analysis using only baseline values. The performance of our model decreased minimally, but interpretations with SHAP remained largely the same, suggesting that interpreting these values as baseline and not on-treatment is most appropriate. This is also supported by our general observation that ~75% of these values were collected before index date.

To summarize, ML was able to predict cardiac adverse events with a high performance. Using SHAP, we identified multiple risk factors for cardiac events including immunological labs and medications, oncological factors, and elements of the patient cardiac history. Further research is needed to triage PD-1/PD-L1 patients' risk of cardiac disease.

ORCID iD
Samuel Peter Heilbroner http://orcid.org/0000-0003-0036-5440

REFERENCES

1 Robert C, Ribas A, Wolchok JD, *et al*. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. *Lancet* 2014;384:1109–17.

2 Hamid O, Robert C, Daud A, *et al*. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N Engl J Med* 2013;369:134–44.

3 Reck M, Rodríguez-Abreu D, Robinson AG, *et al*. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016;375:1823–33.

4 Robert C, Long GV, Brady B, *et al*. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* 2015;372:320–30.

5 Brahmer JR, Tykodi SS, Chow LQM, *et al*. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* 2012;366:2455–65.

6 Lejeune FJ. The conquest of melanoma by immunotherapy. *Melanoma Res* 2015;25:373–5.

7 Eggermont AMM, Maio M, Robert C. Immune checkpoint inhibitors in melanoma provide the cornerstones for curative therapies. *Semin Oncol* 2015;42:429–35.

8 Ansell SM, Lesokhin AM, Borrello I, *et al*. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med* 2015;372:311–9.

9 Brumbaugh AD, Narurkar R, Parikh K, *et al*. Cardiac immune-related adverse events in immune checkpoint inhibition therapy. *Cardiol Rev* 2019;27:97–107.

10 Tay RY, Blackley E, McLean C, *et al*. Successful use of equine anti-thymocyte globulin (ATGAM) for fulminant myocarditis secondary to nivolumab therapy. *Br J Cancer* 2017;117:921–4.

11 Norwood TG, Westbrook BC, Johnson DB, *et al*. Smoldering myocarditis following immune checkpoint blockade. *J Immunother Cancer* 2017;5:91.

12 Behling J, Kaes J, Münzel T, *et al*. New-onset third-degree atrioventricular block because of autoimmune-induced myositis under treatment with anti-programmed cell death-1 (nivolumab) for metastatic melanoma. *Melanoma Res* 2017;27:155–8.

13 Gujral DM, Mouhayar EN, Bhattacharyya S. Cardiac adverse events related to immune checkpoint inhibitors. *JACC Case Rep* 2020;2:200–2.

14 Zhang L, Reynolds KL, Lyon AR, *et al*. The evolving immunotherapy landscape and the epidemiology, diagnosis, and management of cardiotoxicity*JACC: CardioOncology* Primer. *JACC CardioOncol* 2021;3:35–47.

15 Mahmood SS, Fradley MG, Cohen JV, *et al*. Myocarditis in patients treated with immune checkpoint inhibitors. *J Am Coll Cardiol* 2018;71:1755–64.

16 Zhang L, Awadalla M, Mahmood SS, *et al*. Cardiovascular magnetic resonance in immune checkpoint inhibitor-associated myocarditis. *Eur Heart J* 2020;41:1733–43.

17 Bersanelli M, Giannarelli D, Castrignanò P, *et al*. Influenza vaccine indication during therapy with immune checkpoint inhibitors: a transversal challenge. The INVIDIa study. *Immunotherapy* 2018;10:1229–39.

18 Moslehi JJ. Cardiovascular toxic effects of targeted cancer therapies. *N Engl J Med* 2016;375:1457–67.

19 Thavendiranathan P, Zhang L, Zafar A, *et al*. Myocardial T1 and T2 Mapping by Magnetic Resonance in Patients With Immune Checkpoint Inhibitor-Associated Myocarditis. *J Am Coll Cardiol* 2021;77:1503–16.

20 Varricchi G, Galdiero MR, Marone G, *et al*. Cardiotoxicity of immune checkpoint inhibitors. *ESMO Open* 2017;2:e000247.

21 Lichtman AH. The heart of the matter: protection of the myocardium from T cells. *J Autoimmun* 2013;45:90–6.

22 Drobni ZD, Alvi RM, Taron J, *et al*. Association between immune checkpoint inhibitors with cardiovascular events and atherosclerotic plaque. *Circulation* 2020;142:2299–311.

23 Kourou K, Exarchos TP, Exarchos KP, *et al*. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015;13:8–17.

24 Janizek JD, Celik S, Lee S-I. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. *bioRxiv* 2018:331769.

25 Li R, Shinde A, Liu A, *et al*. Identification of nonlinear interaction effects in prostate cancer survival using machine learning-based modeling. *Int J Radiat Oncol Biol Phys* 2019;105:S121.

26 Siddhartha M, Maity P, Nath R. Explanatory artificial intelligence (XAI) in the prediction of post-operative life expectancy in lung cancer patients. *Int J Sci Res* 2020;8.

27 Sarkar RR, Bryant AK, Guss ZD, *et al*. Prognostic impact of neutrophil-to-lymphocyte ratio in solid cancers. *Int J Radiat Oncol Biol Phys* 2019;105:S121–2.

28 Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowl Inf Syst* 2014;41:647–65.

29 Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008;167:492–9.

30 NIH National Cancer Institute. SEER training modules, morphology & grade: ICD-0-3 morphology codes. Available: https://training.seer.cancer.gov/lung/abstract-code-stage/morphology.html [Accessed 29 Dec 2020].

31 Quan H, Sundararajan V, Halfon P, *et al*. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.

32 Gorina Y, Kramarow EA. Identifying chronic conditions in Medicare claims data: evaluating the chronic condition data Warehouse algorithm. *Health Serv Res* 2011;46:1610–27.

33 RxNorm API. Lister Hill national center for biomedical communications. Available: https://rxnav.nlm.nih.gov/RxNormAPIs.html# [Accessed 29 Dec 2020].

34 Buck SF. A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *J R Stat Soc Series B Stat Methodol* 1960;22:302–6.

35 Chen T, He T, Benesty M, *et al*. Xgboost: extreme gradient boosting. R package version 04-2 2015:1–4.

36 Hothorn T, Bühlmann P, Dudoit S, *et al*. Survival ensembles. *Biostatistics* 2006;7:355–73.

37 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.

38 Lundberg SM, Lee S-I. *A unified approach to interpreting model predictions. paper presented at: 31st International Conference on neural information processing systems; December 4-9, 2017*. Long Beach, CA, 2020.

39 Kartolo A, Sattar J, Sahai V, *et al*. Predictors of immunotherapy-induced immune-related adverse events. *Curr Oncol* 2018;25:e403–10.

40 Liu J, Li S, Zhang S, *et al*. Systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio can predict clinical outcomes in patients with metastatic non-small-cell lung cancer treated with nivolumab. *J Clin Lab Anal* 2019;33:e22964.

41 Kanikarla-Marie P, Lam M, Sorokin AV, *et al*. Platelet metabolism and other targeted drugs; potential impact on immunotherapy. *Front Oncol* 2018;8:107.

42 Arbour KC, Mezquita L, Long N, *et al*. Impact of baseline steroids on efficacy of programmed cell death-1 and programmed Death-Ligand 1 blockade in patients with non-small-cell lung cancer. *J Clin Oncol* 2018;36:2872–8.

43 Marantz PR, Tobin JN, Wassertheil-Smoller S, *et al*. The relationship between left ventricular systolic function and congestive heart failure diagnosed by clinical criteria. *Circulation* 1988;77:607–12.

44 Ahmed A, Aboshady I, Munir SM, *et al*. Decreasing body temperature predicts early rehospitalization in congestive heart failure. *J Card Fail* 2008;14:489–96.

45 Franceschi C, Bonafè M, Valensin S, *et al*. Inflamm-aging. an evolutionary perspective on immunosenescence. *Ann N Y Acad Sci* 2000;908:244–54.

46 Falconer JS, Fearon KC, Plester CE, *et al*. Cytokines, the acute-phase response, and resting energy expenditure in cachectic patients with pancreatic cancer. *Ann Surg* 1994;219:325–31.

47 Hellerstein MK, Meydani SN, Meydani M, *et al*. Interleukin-1-induced anorexia in the rat. influence of prostaglandins. *J Clin Invest* 1989;84:228–35.

48 Mantovani G, Macciò A, Mura L, *et al*. Serum levels of leptin and proinflammatory cytokines in patients with advanced-stage cancer at different sites. *J Mol Med* 2000;78:554–61.

49 Sonti G, Ilyin SE, Plata-Salamán CR. Anorexia induced by cytokine interactions at pathophysiological concentrations. *Am J Physiol* 1996;270:R1394–402.

50 Staal-van den Brekel AJ, Dentener MA, Schols AM, *et al*. Increased resting energy expenditure and weight loss are related to a systemic inflammatory response in lung cancer patients. *J Clin Oncol* 1995;13:2600–5.

51 Tocco-Bradley R, Georgieff M, Jones CT, *et al*. Changes in energy expenditure and fat metabolism in rats infused with interleukin-1. *Eur J Clin Invest* 1987;17:504–10.

52 Van der Poll T, Romijn JA, Endert E, *et al*. Tumor necrosis factor mimics the metabolic response to acute infection in healthy humans. *Am J Physiol* 1991;261:E457–65.

53 Walsh D, Mahmoud F, Barna B. Assessment of nutritional status and prognosis in advanced cancer: interleukin-6, C-reactive protein, and the prognostic and inflammatory nutritional index. *Support Care Cancer* 2003;11:60–2.

54 Zeisler H, Tempfer C, Joura EA, *et al*. Serum interleukin 1 in ovarian cancer patients. *Eur J Cancer* 1998;34:931–3.

55 Martin L, Senesse P, Gioulbasanis I, *et al*. Diagnostic criteria for the classification of cancer-associated weight loss. *J Clin Oncol* 2015;33:90–9.

56 Forkasiewicz A, Dorociak M, Stach K, *et al*. The usefulness of lactate dehydrogenase measurements in current oncological practice. *Cellular & Molecular Biology Letters* 2020;25:1–14.

57 Palaskas N, Lopez-Mattei J, Durand JB, *et al*. Immune checkpoint inhibitor myocarditis: pathophysiological characteristics, diagnosis, and treatment. *J Am Heart Assoc* 2020;9:e013757.

58 Nso N, Antwi-Amoabeng D, Beutler BD, *et al*. Cardiac adverse events of immune checkpoint inhibitors in oncology patients: a systematic review and meta-analysis. *World J Cardiol* 2020;12:584.

59 Waheed N, Fradley MG, DeRemer DL, *et al*. Newly diagnosed cardiovascular disease in patients treated with immune checkpoint inhibitors: a retrospective analysis of patients at an academic tertiary care center. *Cardiooncology* 2021;7:1–7.

60 Vaid A, Somani S, Russak AJ, *et al*. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation. *J Med Internet Res* 2020;22:e24018.

61 Uno H, Cai T, Tian L, *et al*. Evaluating prediction rules for *t*-year survivors with censored regression models. *J Am Stat Assoc* 2007;102:527–37.

62 Andersen PK, Borgan O, Gill RD, *et al*. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

Correction: *Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: a machine learning approach*

Heilbroner SP, Few R, Mueller J, *et al.* Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: a machine learning approach. *J ImmunoTher Cancer* 2021;9:e002545. doi: 10.1136/jitc-2021-002545

A correction has been made to the order of authors in this article since it was first published. The author order should be as follows: Samuel P Heilbroner, Reed Few, Tomas G Neilan, Judith Mueller, Jitesh Chalwa, Francois Charest, Somasekhar Suryadevara, Christine Kratt, Andres Gomez-Caminero, Brian Dreyfus.

Check for updates