Background and Objectives

- Patients with cancer are at higher risk of dying of COVID-19
- Known risk factors for 30-days all-cause mortality (ACM-30) in patients with cancer are older age, gender, smoking status, performance status, obesity, and co-morbidities.

Objectives: 1) Identify informative common clinical and laboratory parameters predictive of a higher risk of 30-days ACM; 2) Build and validate a machine learning model based on clinical and laboratory values to estimate individual survival probability of ACM at 30 days.

COVID-19 and Cancer Consortium (CCC19) international registry study
12,668 patients were enrolled between March 17, 2020 and December 31, 2021.
891 patients were excluded due to quality score<4 (poor quality data).
Data collected & managed using REDCap at Vanderbilt University Medical Center.

Shapley plot showing the contributions of the twenty selected variables on the prediction of the final model.

Methods

Data Analysis

- ACM-30 days defined death from any cause within 30 days of COVID19 diagnosis.
- Pre-specified variables: age, sex, race, smoking status, ECOG performance status, timing of cancer treatment relative to COVID19 diagnosis, severity of COVID19, type of cancer, and other laboratory measurements.
- Imputed missing variables using random forest proximity.
- Randomly split data into training and testing sets allocation ratio 2:1 stratifying on US census division.
- Two-step random forest utilized to model ACM-30: top twenty variables first selected based on feature importance given by random forest; model was refitted using the selected variables on the training set.

Area under the curve (AUC) computed measure of predictive accuracy with out-of-bag (OOB) and testing prediction.
Optimal predicted probability cutoff ACM-30 selected maximum Youden’s Index.
Sensitivity and specificity computed using the optimal probability cutoff.
Monte Carlo cross-validation: repeated random splitting and model fitting 100 times and obtained variable selection frequency and distribution of the AUCs.

Variable Contribution

- Display plot showing the contributions of the twenty selected variables on the prediction of the final model.

Baseline Characteristics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>Yes (30+)</td>
<td>No (70+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer status</td>
<td></td>
<td>Yes (1-3m)</td>
<td>No (4-12m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity of COVID19</td>
<td></td>
<td>Yes (2-3)</td>
<td>No (0-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td>Yes (&gt;30)</td>
<td>No (≤30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>Male</td>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td>Non-Hispanic White</td>
<td>Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td>Never</td>
<td>Current (0 or 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td>North East Central</td>
<td>West North Central</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
<td>Mexican</td>
<td>Non Mexican</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Median (Q1, Q3)</td>
<td></td>
<td>64.0 (54.0, 73.0)</td>
<td>73.0 (64.0, 81.0)</td>
<td>65.0 (55.0, 74.0)</td>
<td></td>
</tr>
<tr>
<td>Median AUC (Testing)</td>
<td></td>
<td>0.89 (0.87-0.90)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median AUC (OOB)</td>
<td></td>
<td>0.88 (0.87-0.89)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results

- Model can be used to estimate individual survival probability within 30-days for COVID-19.
- ACM-30 days model includes readily available clinical and laboratory variables.
- Validation cut points, for treatment selection, prophylaxis prioritization, and/or enrollment in clinical trials.

Variable Risk Group by Selected Variables

- Top predictors: Severity of COVID19, cancer status and age
- ACM-30 days model includes readily available clinical and laboratory variables
- Model can be used to estimate individual survival probability within 30-days for COVID-19.
- Model can be used to select or classify patients with cancer and COVID-19 into risk groups based on validated cut points, for treatment selection, prophylaxis prioritization, and/or enrollment in clinical trials.
- Future work includes validation from external dataset of patients with COVID-19 and cancer.

Acknowledgement

We thank the CCC19 Steering Committee and investigators for their contribution to this registry. Funding: CCC19 supported by NCI P30 CA166724 to Vanderbilt University.