Performance of Predictive Models for 30-Day Hospitalization and Mortality after COVID-19 Infection

Identifying which patients have the highest risk of hospitalization for coronavirus disease (COVID-19) can inform personalized recommendations about exposure risk, proactive screening and treatments at the health system level, and vaccine distribution across a population. Previously developed COVID-19 risk tools estimate mortality of those admitted to the hospital (1). We developed and implemented a real-time COVID-19 risk score (CRS) predicting 30-day nonelective hospitalization to inform elective surgery workflows (2), which we subsequently updated to incorporate newer data. Similar health system scores, including the Veterans Health Administration COVID-19 (VACO) index, have been developed to predict 30-day mortality across a population (3). Here, we report performance of CRS for predicting both 30-day nonelective hospitalization and mortality and compare the performance of CRS with the VACO index for these outcomes. These analyses were uniquely feasible given our position in an integrated health system (Kaiser Permanente Northern California [KPNC]) with population-level data.

Methods

We identified patients’ first positive severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) polymerase chain reaction result between February 1, 2020 and September 30, 2020 in adult members (aged ≥18 yr) at KPNC, which serves 4.5 million members across 21 medical centers.

We developed and internally validated CRS using a split 75% train/25% test dataset. We performed multivariable logistic regression using age, sex, and comorbidity point score version 2 (COPS2) (4). COPS2 is a weighted score of 75 hierarchical condition categories (5) over the 1 year preceding the first positive test. COPS2 is easily generated as a sum and has been externally validated in Canada with better discrimination than Elixhauser or Charlson (6). Because of low turnover in KPNC, more than 90% of patients had data for 12 months prior to COVID-19 diagnosis, indicating that there were few missing data. We used a truncated power function to spline at 3 knots for age and COPS2 (7). The final model included age, age (2), age (3), cubic spline of age, sex, COPS2, COP2 (2), COP3 (3), and cubic spline of COPS2. Our outcome was 30-day nonelective hospitalization, defined as hospitalizations originating in the emergency department. By regional policy, patients seen in clinic and sent to the hospital must pass through the emergency room for stabilization and triage, so there are no direct admits from clinic. We chose this outcome because it is a more proximal outcome than 30-day mortality and potentially more intervenable. Each patient could contribute one hospitalization in the dataset, which was the hospitalization closest to the first positive test. We also examined CRS performance for 30-day mortality outcome.

We calculated the VACO index on the same cohort using the published variable specifications and coefficients for age, sex, history of myocardial infarction/peripheral vascular disease, and interaction terms for age and Charlson comorbidity (3).

We report concordance statistics (c-statistics) of CRS and the VACO index for predicting 30-day nonelective hospitalization and 30-day mortality with bootstrapped 95% confidence intervals (CIs) and sensitivity and specificity at a positive predictive value threshold near 20%. We also calculated the net reclassification index for CRS relative to the VACO index for both outcomes to assess comparative performance (8). We show calibration plots for the outcome for which the model was originally derived (hospitalization for CRS and mortality for VACO) along with calibration statistics (calibration-in-the-large and calibration slope) and 95% CIs. We performed calibration on the validation cohort for CRS and the full cohort for VACO.

The KPNC Institutional Review Board approved the project with a waiver of informed consent. We used SAS 9.4.

Results

From more than 3.2 million patients, we examined 36,137 patients with positive tests, of whom 3,397 (9.4%) had nonelective hospitalization and 361 (1.4%) died within 30 days of the test. Most tests (34,435, 95.3%) were performed in outpatient (nonemergency department) settings. Mean age was 43.4 ± 16.3 years (Table 1).

The c-statistics for CRS were 0.82 (95% CI, 0.81–0.83) for nonelective hospitalization and 0.93 (95% CI, 0.92–0.94) for 30-day mortality in the test set. The c-statistics for the VACO index were 0.78 (95% CI, 0.77–0.79) for nonelective hospitalization and 0.93 (95% CI, 0.92–0.94) for 30-day mortality.

At a positive predictive value threshold of 20%, CRS had higher sensitivity but lower specificity for hospitalization compared with the VACO index, resulting in net reclassification of 8.0% more hospitalizations in the higher risk category and 3.5% fewer nonhospitalizations in the lower risk category compared with the results from the VACO index (Table 2). For 30-day mortality, CRS had lower sensitivity but slightly higher specificity compared with the VACO index, resulting in net reclassification of 11.0% fewer deaths in the higher risk category and 0.7% more nondeaths in the lower risk category.

The calibration plots reveal that CRS is well calibrated for patients at low risk of hospitalization (0–15% risk), which is the vast majority of the cohort (88%) (Figure 1A). In contrast, VACO overestimates risk across the full range of risk (Figure 1B). The calibration-in-the-large was 0.17 (95% CI, 0.03–0.31) for nonelective hospitalization for CRS on the validation cohort and −0.39 (95% CI, −0.56 to −0.22) for mortality for VACO on the full cohort. The calibration slope was 1.04 (95% CI, 0.97–1.10) and 1.24 (1.16–1.32) for CRS and VACO, respectively.
Table 1. Baseline characteristics of patients included in COVID-19 risk score from Kaiser Permanente Northern California

| Variable                                | Total (N = 36,137) | No Hospitalization (n = 32,734) | Hospitalization (n = 3,403) |
|-----------------------------------------|--------------------|---------------------------------|-----------------------------|
| Age                                     | 43.4 (16.3)        | 41.8 (15.3)                     | 59.5 (17.1)                 |
| Male                                    | 17,165 (47.5%)     | 15,331 (46.8%)                  | 1,837 (54.0%)               |
| Neighborhood deprivation index          | 0.2 (0.9)          | 0.2 (0.9)                       | 0.2 (1.0)                   |
| Race                                    |                    |                                 |                             |
| Hispanic                                | 17,792 (49.2%)     | 16,329 (49.9%)                  | 1,463 (43.0%)               |
| White                                   | 8,257 (22.8%)      | 7,409 (22.6%)                   | 0,848 (24.9%)               |
| Asian                                   | 4,339 (12.0%)      | 3,834 (11.7%)                   | 505 (14.8%)                 |
| Other                                   | 3,235 (9.0%)       | 2,996 (9.2%)                    | 0,239 (7.0%)                |
| Black                                   | 2,514 (7.0%)       | 2,166 (6.8%)                    | 346 (10.2%)                 |
| Comorbidity burden                      |                    |                                 |                             |
| COPS2 score                             | 14.2 (17.8)        | 12.6 (13.1)                     | 30.4 (37.7)                 |
| Hypertension                            | 5,768 (16.0%)      | 4,358 (13.3%)                   | 1,410 (41.4%)               |
| Diabetes without complications          | 3,619 (10.0%)      | 2,737 (8.4%)                    | 882 (25.9%)                 |
| Obesity                                 | 3,432 (9.5%)       | 2,848 (8.7%)                    | 584 (17.2%)                 |
| Diabetes with complications             | 2,778 (7.7%)       | 1,929 (5.9%)                    | 849 (24.9%)                 |
| Chronic pulmonary disease               | 2,751 (7.6%)       | 2,196 (6.7%)                    | 556 (16.3%)                 |
| Peripheral vascular disease             | 2,232 (6.1%)       | 1,410 (4.3)                     | 822 (24.1%)                 |
| Depression                              | 1,362 (3.8%)       | 1,136 (3.5%)                    | 226 (6.6%)                  |
| Hypothyroidism                          | 1,268 (3.5%)       | 1,007 (3.1%)                    | 261 (7.7%)                  |
| Deficiency anemia                       | 1,197 (3.3%)       | 878 (2.7%)                      | 319 (9.4%)                  |
| Liver disease                           | 1,073 (3.0%)       | 866 (2.6%)                      | 207 (6.1%)                  |
| Renal failure                           | 1,029 (2.8%)       | 540 (1.6%)                      | 489 (14.4%)                 |
| Psychoses                               | 1,021 (2.8%)       | 778 (2.4%)                      | 243 (7.1%)                  |
| Other neurological disorders            | 943 (2.6%)         | 618 (1.9%)                      | 325 (9.6%)                  |
| Fluid and electrolyte disorders         | 592 (1.6%)         | 369 (1.1%)                      | 223 (6.6%)                  |
| Congestive heart failure                | 484 (1.3%)         | 237 (0.7%)                      | 247 (7.3%)                  |
| Rheumatoid arthritis                    | 418 (1.2%)         | 322 (1.0%)                      | 306 (9.0%)                  |
| Solid tumor without metastasis         | 382 (1.1%)         | 288 (0.9%)                      | 194 (5.6%)                  |
| Renal transplantation                   | 259 (0.7%)         | 166 (0.5%)                      | 93 (2.7%)                   |
| Coagulation deficiency                  | 209 (0.6%)         | 137 (0.4%)                      | 72 (2.1%)                   |
| Weight loss                             | 217 (0.6%)         | 140 (0.4%)                      | 077 (2.3%)                  |
| Blood loss anemia                       | 229 (0.6%)         | 210 (0.6%)                      | 019 (0.6%)                  |
| Alcohol abuse                           | 227 (0.6%)         | 181 (0.6%)                      | 046 (1.4%)                  |
| Drug abuse                              | 208 (0.6%)         | 165 (0.5%)                      | 043 (1.3%)                  |
| Valvular disease                        | 180 (0.5%)         | 99 (0.3%)                       | 81 (2.4%)                   |
| Metastatic cancer                       | 102 (0.3%)         | 74 (0.2%)                       | 28 (0.8%)                   |
| Acquired immunodeficiency disease       | 61 (0.2%)          | 43 (0.1%)                       | 18 (0.5%)                   |
| Lymphoma                                | 83 (0.2%)          | 49 (0.1%)                       | 34 (1.0%)                   |
| Pulmonary circulatory disorders         | 55 (0.2%)          | 34 (0.1%)                       | 21 (0.6%)                   |
| Chronic peptic ulcer disease           | 39 (0.1%)          | 25 (0.1%)                       | 14 (0.4%)                   |

Definition of abbreviations: COPS2 = comorbidity point score version 2; COVID-19 = coronavirus disease.
Mean and standard deviation are reported for continuous variables. Number and percent are reported for categorical variables. Scale for Neighborhood Deprivation Index is 2 to 3.72 with higher values indicating lower socioeconomic status. Comorbidities from the Elixhauser comorbidity index are listed in decreasing frequency in the overall cohort column.

Table 2. Model performance for predicting 30-day nonelective hospitalization and 30-day mortality where scores are set at a threshold near 20%

| Outcome                  | Score  | Positive Predictive Value | Number Needed to Evaluate | Risk Score Threshold | Sensitivity | Specificity |
|--------------------------|--------|---------------------------|---------------------------|---------------------|-------------|-------------|
| 30-d nonelective hospitalization | CRS    | 20%                       | 5                         | 0.07                | 80%         | 66%         |
|                          | VACO Index | 20%                       | 5                         | 0.02                | 71%         | 70%         |
| 30-d mortality           | CRS    | 19%                       | 5                         | 0.49                | 49%         | 97%         |
|                          | VACO Index | 19%                       | 5                         | 0.20                | 60%         | 96%         |

Definition of abbreviations: COVID-19 = coronavirus disease; CRS = COVID-19 risk score; VACO index = Veterans Health Administration COVID-19 index. The COVID-19 risk score was developed at Kaiser Permanente Northern California to predict 30-day nonelective hospitalization. The Veterans Administration COVID-19 index was developed using nationwide veterans’ data originally to predict 30-day mortality.
Figure 1. (A and B) Calibration plots for COVID-19 risk score (A) and Veterans Administration COVID-19 index (B). Calibration plots with observed risk shown as a line and predicted risk shown as a straight dotted line. The outcome tested is the one for which the score was originally developed (hospitalization for COVID-19 risk score and mortality for Veterans' Administration COVID-19 index). The number of events in each risk stratum is listed, either nonelective hospitalization or death. The sample for COVID-19 risk score is the validation cohort (n = 9,165), whereas the sample for VACO is the broader COVID-19 positive adult population (n = 36,137) because it had been derived outside of Kaiser Permanente Northern California. The majority (>88%) has low risk (0–15% risk of the outcome), which is to the left of the dotted line. COVID-19 = coronavirus disease.
Discussion
We evaluated the performance of models for two patient-centered outcomes following COVID-19 infection using data from a large, integrated health system. We found that CRS model discrimination was very good for nonelective hospitalization, the outcome for which it was developed (c-statistic, 0.82). For 30-day mortality, CRS was comparable to the VACO index (both 0.93).

CRS showed good calibration for patients with low risk of hospitalization, which was the majority of the cohort (88%). Given this result, the model has been used in KPNC to target patients for outpatient procedures who were at low risk of hospitalization due to COVID-19 during the initial reopening of operating rooms. External validation of CRS is necessary. VACO overestimated risk across the range of risk in this external validation, which is consistent with its previous validation at the Veterans Administration (3).

Because the veteran population is older, predominantly male, with higher chronic disease burden (9), external validation of VACO is important for generalizability. KPNC members are diverse and tend to reflect the general population (10). In our external validation, VACO index discrimination was better (0.93) than what was reported in the two internal validations (0.81–0.84) (3). This finding could be due to a more complete dataset at KPNC or a more rare outcome rate (1.4% mortality in the KPNC cohort vs. 6.8% across the two veterans’ validation cohorts). Regardless, the VACO index showed strong potential for generalizability in an independent cohort despite differences in the composition of the veteran and KPNC cohorts.

Both models focus on a selected set of variables known to account for key drivers of adverse outcomes in COVID-19 that can be easily identified using electronic health record data. Both models have already been implemented in real-time systems or online calculators to facilitate rapid dissemination and everyday use. For example, CRS has been deployed at KPNC to assist with diverse health system workflows for patients with COVID-19; the VACO index is a publicly available tool on MDCalc.

We did not include race in CRS because the model could be used for resource allocation. Our group has previously shown that severity of illness scores used during crisis standard of care could prioritize white patients over Black patients because they overestimate risk for Black patients (11). Therefore, we did not want to create a model that prioritized one race to receive resources, which might propagate health disparities. This is a complex, important, and emerging area. Further work is necessary to demonstrate equity of this model across subgroups of races.

This study fills important gaps. A previous model by Jehi and colleagues predicting 30-day hospitalization in COVID-19 infection required 10 input variables, including smoking status, symptoms, and chronic medications (12). Our model is much simpler and had similar performance (0.82 in this study compared with 0.81) (12). While numerous models have been developed to predict COVID-19 mortality, models calibrated for hospitalization offer higher potential for health system decision-making, which can include proactive screening and outreach, recommendations for specific treatments, and patient education and counseling. With both scores, the use of data through late 2020 also increases confidence that these models show good discrimination over multiple waves of infection.

There are several limitations. Patients may have had positive tests outside our health system that were not counted. Similarly, deaths that occurred outside the hospital may not have been reported to KPNC at the time of analysis. The strengths of this study include its large, population-based sample with minimal missing data spanning all waves of the pandemic. Few data sources contain the longitudinal data necessary to evaluate these models on the population level.

Author disclosures are available with the text of this article at www.atsjournals.org.

Laura C. Myers, M.D., M.P.H.*
Patricia Kipnis, Ph.D.
Liam O’Sullivanbhan, M.S.
Gabriel Escobar, M.D.
Vincent X. Liu, M.D., M.Sc.
Kaiser Permanente Northern California
Oakland, California

ORCID ID: 0000-0002-2872-3388 (L.C.M.).

*Corresponding author (e-mail: laura.c.myers@kp.org).

References
1. Berenguer J, Borobia AM, Ryan P, Rodriguez-Baño J, Bellón JM, Jarrín I, et al.; COVID-19 @ Spain and COVID@HULP Study Groups. Development and validation of a prediction model for 30-day mortality in hospitalised patients with COVID-19: the COVID-19 SEIMC score. Thorax 2021;76:920–929.
2. Rouillard S, Liu VX, Corley DA. COVID-19 and long-term planning for procedure-based specialties during extended mitigation and suppression strategies. Gastroenterology 2021;160:4–9.
3. King JT Jr, Yoon JS, Rentsch CT, Tate JP, Park LS, Kidwai-Khan F, et al. Development and validation of a 30-day mortality index based on pre-existing medical administrative data from 13,323 COVID-19 patients: The Veterans Health Administration COVID-19 (VACO) Index. PLoS One 2020;15:e0241825.
4. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. Med Care 2013;51:446–453.
5. Healthcare Cost and Utilization Project. Clinical classification software (CCS) 2015, Rockville, MD: Agency for Healthcare Quality and Research; 2015 [updated 2016 Mar; accessed 2021 Jan 19]. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf.
6. van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. J Clin Epidemiol 2010;63:798–803.
7. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning, 3rd ed. New York, NY: Springer; 2008.
8. Pencina MJ, Steyerberg EW, D’Agostino RB Sr. Net reclassification index at event rate: properties and relationships. Stat Med 2017;36:4455–4467.
9. Hoerster KD, Lehavot K, Simpson T, McFall M, Reiber G, Nelson KM. Health and health behavior differences: U.S. military, veteran, and civilian men. Am J Prev Med 2012;43:483–489.
10. Gordon N. Similarity of adult Kaiser Permanente members to the adult population in Kaiser Permanente’s Northern California service area: comparisons based on the 2017/2018 cycle of the California health interview survey; 2020 [updated 8 Nov 2020; accessed 2021 Jan 21]. Available from: https://divisionofresearch.kaiserpermanente.org/projects/memberhealthsurvey/SiteCollectionDocuments/compare_kp_ncal_chis2017–18.pdf.
11. Ashana DC, Anesi GL, Liu VX, Escobar GJ, Chesley C, Eneanya ND, et al. Equitably allocating resources during crises: racial differences in mortality prediction models. Am J Respir Crit Care Med 2021;204:178–186.
12. Jehi L, Ji X, Milinovich A, Erzurrum S, Merlino A, Gordon S, et al. Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. PLoS One 2020;15:e0237419.