Risk Stratification Models for Stroke in Patients Hospitalized with COVID-19 Infection

Alexander E. Merkler, MD, MS,[a] Cenai Zhang, MS,[a] Ivan Diaz, PhD,[b] Carolyn Stewart, BA,[a] Natalie M. LeMoss, BS,[a] Saad Mir, MD,[a] Neal Parikh, MD, MS,[a] Santosh Murthy, MD, MPH,[a] Ning Lin, MD,[c] Ajay Gupta, MD, MS,[d] Costantino Iadecola, MD,[a] Mitchell S.V. Elkind, MD, MS, MPhil,[e] Hooman Kamel, MD, MS,[a] and Babak B. Navi, MD, MS[a]

Objectives: To derive models that identify patients with COVID-19 at high risk for stroke. *Materials and Methods:* We used data from the AHA's Get With The Guidelines® COVID-19 Cardiovascular Disease Registry to generate models for predicting stroke risk among adults hospitalized with COVID-19 at 122 centers from March 2020-March 2021. To build our models, we used data on demographics, comorbidities, medications, and vital sign and laboratory values at admission. The outcome was a cerebrovascular event (stroke, TIA, or cerebral vein thrombosis). First, we used Cox regression with cross validation techniques to identify factors associated with the outcome in both univariable and multivariable analyses. Then, we assigned points for each variable based on corresponding coefficients to create a prediction score. Second, we used machine learning techniques to create risk estimators using all available covariates. *Results:* Among 21,420 patients hospitalized with COVID-19, 312 (1.5%) had a cerebrovascular event. Using traditional Cox regression, we created/validated a COVID-19 stroke risk score with a C-statistic of 0.66 (95% CI, 0.60−0.72). The CANDLE score assigns 1 point each for prior cerebrovascular disease, afebrile temperature, no prior pulmonary disease, history of hypertension, leukocytosis, and elevated systolic blood pressure. CANDLE stratified risk of an acute cerebrovascular event according to low- (0−1: 0.2% risk), medium- (2−3: 1.1% risk), and high-risk (4−6: 2.1−3.0% risk) groups. Machine learning estimators had similar discriminatory performance as CANDLE: C-statistics, 0.63−0.69. *Conclusions:* We developed a practical clinical score, with similar performance to machine learning estimators, to help stratify stroke risk among patients hospitalized with COVID-19.

Introduction

Coronavirus Disease 2019 (COVID-19) is the most impactful pandemic of our lifetime. As of March 2022, there have been more than 470 million confirmed cases of COVID-19, leading to more than 6.0 million deaths worldwide.[1] Although COVID-19 is primarily a respiratory illness, multiple studies have found that the SARS-CoV-2 infection promotes immune dysregulation, a hypercoagulable state, and thrombotic complications.[2-8] Additionally, multiple studies have reported that COVID-19 is associated with an increased risk of stroke.[9-14] Furthermore, strokes appear to be more severe and associated with worse outcomes in patients with COVID-19 infection.[15,16]

Identifying patients with COVID-19 who have an elevated risk for stroke may aid in management and therapeutic decisions. We therefore used multicenter data from the American Heart Association's (AHA) Get With The Guidelines® (GWTG) COVID-19 Cardiovascular Disease Registry to create risk stratification models to help identify incident stroke among patients hospitalized with COVID-19. We derived and validated an easy-to-use clinical score encompassing clinical factors at hospital presentation associated with incident stroke, as well as constructed three different machine learning algorithms using all available covariates. Our prespecified study hypothesis was that older age, vascular risk factors, and laboratory markers of inflammation and thrombosis would be associated with an increased risk of stroke in patients hospitalized with COVID-19.

Methods

Design, setting, and participants

We conducted a retrospective cohort study using prospectively collected data from patients enrolled in the AHA COVID-19 Cardiovascular Disease Registry. The details of this registry have been previously described, but in brief it is powered by the AHA GWTG quality improvement program and aims to elucidate the characteristics and cardiovascular outcomes of patients hospitalized with COVID-19 infection in the United States.[17] The registry includes over 200 data elements on consecutive patients hospitalized with COVID-19 at participating sites, which comprised urban and rural hospitals of all sizes across all geographic regions of the continental United States. Each site that participated in the registry obtained institutional review board approval or exemption and was granted a waiver of informed consent under the common rule. The Weill Cornell Medicine institutional review board confirmed exemption status for this study.

The data for this analysis are maintained by the AHA and can be made available through written application. Our analysis followed guidelines from the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) statement.[18]

We included all patients aged 18 years or older who were hospitalized at 122 centers from March 1, 2020 to March 31, 2021 with active COVID-19 infection confirmed by laboratory testing (either a positive PCR or IgM antibody test for SARS-CoV-2). All analyzed patients had been discharged from the hospital or died and had complete data on age, sex, medical history, clinical presentation, and in-hospital events.

Measurements

To build risk stratification models, we included data on sociodemographics (age, sex, race, ethnicity, and payment source), medical history, home medications, and initial vital signs and laboratory values at admission. Medical history encompassed pertinent vascular, neoplastic, and immunological diseases. Home medications included medications commonly used to prevent or treat cardiovascular disease and its risk factors as well as chemotherapy and immunosuppressive medicines. Laboratory values included standard blood and chemistry parameters, cardiac biomarkers, and measures of inflammation and thrombosis. The complete online data collection form, powered by IQVIA (Parsippany, New Jersey), is available at https://www.heart.org/-/media/files/professional/quality-improvement/covid-19-cvd-registry/ahacovidcvdcrf428-fillable-pdf.pdf?la=en. As we aimed to build a stroke risk stratification score that could be implemented at the time of hospital presentation, we did not include data on events, biomarkers, or treatments that occurred during the hospitalization (e.g., mechanical ventilation, rising plasma D-dimer, prophylactic anticoagulation, etc.).

The primary outcome was an acute cerebrovascular event, defined as any acute ischemic stroke, intracerebral hemorrhage, subarachnoid hemorrhage, subdural hemorrhage, epidural hemorrhage, transient ischemic attack (TIA), or cerebral vein thrombosis diagnosed during the index hospitalization. The secondary outcome was an ischemic cerebrovascular event, defined as any acute ischemic stroke or TIA diagnosed during the index hospitalization.

Statistical analysis

We used descriptive statistics with exact binomial confidence intervals to characterize the patient cohort. We

divided the analytical dataset into a derivation cohort (70% of the cohort) and a validation cohort (30% of the cohort). We created separate risk stratification models using different analytical techniques. This included a simple-to-use standard clinical score and more complex machine learning estimators using different statistical approaches. For both models, two neurovascular specialists (A.M. and B.N.) reviewed all available clinical data at admission and selected exposure variables they believed could be associated with incident cerebrovascular events based on biological plausibility and prior data.[9,15,19-21] Variables that were missing in more than 25% of patients were excluded. After applying these criteria, 32 variables were analyzed as potential exposures for incident cerebrovascular events (Table 1).

For the standard clinical score, we used data from the derivation cohort and performed univariate Cox regression with 5-fold cross validation techniques to identify clinical factors associated with the primary outcome at a *p*-value of <0.10. Factors significantly associated at the univariate level were then entered into a multivariable Cox regression model and any factor independently associated with the primary outcome at a *p*-value <0.05 was selected for the final model. We then applied the final model to the validation cohort and measured Harrel's C-statistic to internally validate the results. To facilitate the score's clinical applicability, we dichotomized continuous variables according to normal and abnormal values per standard criteria. For instance, white blood cell count was dichotomized as ≤ 11 K/uL (normal) or >11 K/uL (abnormal) and temperature was dichotomized as <38.3°Celsius (afebrile) or ≥ 38.3 °C (febrile). As the six variables selected for the final model had overlapping hazard ratios and beta-coefficients for their association with the primary outcome when dichotomized, we assigned one point for each variable.

For the machine learning models, we used data from the derivation cohort and performed regularized Cox regression, XGBoost, and Random Forest machine learning techniques to create separate risk stratification estimators using all available covariates.[22] Continuous variables were analyzed as continuous unlike for the standard clinical score. We used nonparametric bootstrap methods to calculate 95% confidence intervals (CI). We applied the final estimator models to the validation cohort and measured Harrel's C-statistic to internally validate the results.

Regularized Cox regression uses regularization techniques to achieve variable selection and provide accurate inference on an outcome's predicted survival when a moderate number of potential predictors are available.[22] In our analysis, an elastic net penalty was used to regularize the model. Gradient tree boosting is an ensemble method that seeks to create a strong classifier (model) based on "weak" classifiers.[23] It fits a new model to the residuals of the previous prediction and then corrects the errors of the previous model by adding the new model.

XGBoost implements gradient tree boosting with an additional custom regularization term in the objective function to achieve better model performance and faster execution speed. Random Survival Forest is an extension of the Random Forest method for analyzing survival data. The Random Survival Forest technique first draws bootstrap samples from the original data, then grows a survival tree for each bootstrap sample, and finally calculates a cumulative hazard function for each tree.[22,23] The ensemble cumulative hazard function is obtained by averaging individual cumulative hazard functions. Prediction error is calculated by using the ensemble cumulative hazard function.

In secondary analysis, we measured the discriminatory performance of both the standard clinical score and the machine learning estimators for the secondary outcome of ischemic cerebrovascular events.

Plasma D-dimer was not analyzed as a potential exposure in our risk stratification models because 60% of patients had missing values at admission. However, as prior studies have reported that D-dimer may be associated with an elevated risk of ischemic stroke among patients with COVID-19 infection, we performed an exploratory analysis restricted to the 40% of patients with D-dimer values at admission.[9,19] For this analysis, we followed the same methodology used to derive the risk stratification models above except all eligible patients were used for both derivation and validation.

Deidentified data from this registry were analyzed by C.Z. through the AHA's online Precision Medicine Platform (https://precision.heartorg/) using RStudio, version 3.6.0 (R Foundation). We used multiple imputation to account for missing data (mean imputation for continuous variables and median imputation for binary variables).

Results

Patient characteristics and outcomes

We evaluated 21,420 patients hospitalized with COVID-19. Their median age was 62 years (interquartile range [IQR], 49−75) and 54% were men. During a median hospitalization duration of 11 days (IQR, 6−18), there were 312 (1.5%) patients diagnosed with an acute cerebrovascular event, including 168 with acute ischemic stroke, 48 with intracerebral hemorrhage, 33 with subarachnoid hemorrhage, 22 with subdural/epidural hemorrhage, 9 with TIA, 2 with cerebral venous thrombosis, and 48 with stroke not otherwise specified (some patients had multiple events, so events sum to more than 312). Patients with a cerebrovascular event were on average older, more often men, had more vascular risk factors, higher systolic blood pressures, lower temperatures, and higher white blood cell and platelet counts than patients without a cerebrovascular event. The median duration from hospital admission to cerebrovascular event diagnosis was 2 days

Table 1. *Baseline Characteristics of Hospitalized COVID-19 Patients with and without an Acute Cerebrovascular Event.*

Characteristic[a]	No Cerebrovascular Event (n=21,108)	Cerebrovascular Event (n=312)
Demographics		
Age, y, median (IQR)	62 (49−75)	66 (59−74)
Male sex	11,389 (54)	196 (63)
Race/Ethnicity		
Non-Hispanic White	8041 (38)	127 (41)
Non-Hispanic Black	5419 (26)	92 (29)
Hispanic	5387 (26)	54 (17)
Asian	836 (4)	18 (6)
Other/Unknown	1425 (7)	21 (7)
Insurance		
Medicare/Medicaid	5450 (26)	71 (23)
Private	13,484 (64)	203 (65)
Other	2174 (10)	38 (12)
Medical History		
Hypertension	13,592 (64)	242 (78)
Hyperlipidemia	9878 (47)	167 (54)
Diabetes	7945 (38)	134 (43)
Prior stroke/TIA	2530 (12)	74 (24)
Heart failure	7277 (34)	126 (40)
Coronary artery disease	2078 (10)	42 (13)
Peripheral vascular disease	573 (3)	9 (3)
Chronic pulmonary disease	3945 (19)	36 (12)
Chronic kidney disease	2714 (13)	45 (14)
Atrial fibrillation/flutter	2025 (10)	49 (16)
Cancer	2626 (12)	43 (14)
HIV	217 (1)	2 (1)
Autoimmune disorder	947 (5)	9 (3)
Tobacco use	1382 (7)	23 (7)
Alcohol use	5586 (12.1)	41,875 (2.5)
Home Medications		
Antiplatelet	5689 (27)	113 (36)
Anticoagulant	2900 (14)	56 (18)
Initial Vital Signs (IQR)		
Temperature	37.2 (36.7-37.8)	37.0 (36.6−37.4)
Oxygen saturation	95 (93−97)	96 (93−98)
Systolic BP	131 (116−144)	134 (117−152)
Heart rate	94 (80−106)	87 (75−102)
Respiratory rate	20 (18−24)	20 (18−24)
BMI, kg/m^2	30 (26−34)	30 (25−33)
Admission Labs (IQR)		
WBC count, K/uL	7.1 (5.2−9.5)	8.1 (6.5−11.4)
Hemoglobin, g/dL	12.9 (11.5−14.3)	12.9 (10.9−14.1)
Platelet count, K/uL	208 (159−263)	223 (166−286)
Hemoglobin A1c, %	6.9 (6.1−8.9)	6.5 (5.9−7.9)
Creatinine, mg/dL	1.04 (0.80−1.62)	1.10 (0.87−1.68)

Abbreviations: IQR, interquartile range; TIA, transient ischemic attack; HIV, human immunodeficiency virus; WBC, white blood cell
[a]Data are presented as number (%) unless otherwise specified.

(IQR, 1-8). The median NIH Stroke Scale was 10 (IQR, 3-20). In-hospital mortality was 35% among patients diagnosed with a cerebrovascular event and 14% among patients not diagnosed with a cerebrovascular event (p<0.001).

Standard clinical score

Among the 32 analyzed variables, we identified 6 variables at admission that were independently associated with increased risk of an acute cerebrovascular event

during hospitalization with COVID-19 infection (Table 2). These variables were prior stroke or TIA, lower body temperature, no previous pulmonary disease (COPD, asthma, other pulmonary disease), history of hypertension, elevated serum white blood cell count, and elevated systolic blood pressure. In the validation cohort, a clinical score comprising these six variables had a Harrel's C-statistic of 0.66 (95% CI, 0.60−0.72) for predicting an acute cerebrovascular event. For the secondary outcome of ischemic stroke or TIA, the score's C-statistic was 0.67 (95% CI, 0.59−0.76).

After dichotomizing continuous variables to enable bedside calculation and assigning 1 point for each variable, the CANDLE risk stratification score was derived as follows: Cerebrovascular disease history, Afebrile temperature (<38.3 °C), No pulmonary disease history, Disorder of hypertension, Leukocytosis (white blood cell count >11 K/uL), and an Elevated systolic blood pressure (>140 mm Hg). The magnitude and precision of the independent associations between CANDLE's dichotomized variables and an acute cerebrovascular event are provided in Table 3. We stratified the risk of an acute cerebrovascular event according to low (0-1), medium (2-3), and high (4-6) risk groups. In the derivation cohort, the low-risk group had a 0.2% risk (95% CI, 0.1−0.7%) of incident cerebrovascular events, while the medium-risk group had a 1.1% risk (95% CI, 0.9−1.3%), and the high-risk group had a 3.0% risk (95% CI, 2.5−3.6%) (Fig. 1). In the validation cohort, the low-risk group had a 0.2% risk (95% CI, 0.0−1.2%) of incident cerebrovascular events, while the medium-risk group had a 1.1% risk (95% CI, 0.8−1.4%), and the high-risk group had a 2.1% risk (95% CI, 1.5−2.9%)

Machine learning models

The machine learning risk stratification models had similar discriminatory performance to CANDLE for the primary outcome. Random Forest performed best with a validation cohort C-statistic of 0.69 (95% CI, 0.65−0.72), regularized Cox regression had a validation cohort C-statistic of 0.67 (95% CI, 0.60−0.73), and XGBoost had a validation cohort C-statistic of 0.63 (95% CI, 0.56−0.70). For the secondary outcome of ischemic cerebrovascular events, the C-statistics were 0.69 (95% CI, 0.64−0.75) for the Random Forest model, 0.63 (95% CI, 0.54−0.73) for the regularized Cox regression model, and 0.64 (95% CI, 0.58−0.71) for the XGBoost model.

Exploratory D-dimer analysis

Plasma D-dimer at admission was associated with an incident cerebrovascular event in univariate (hazard ratio per unit in ug/mL, 1.06; 95% CI, 1.03−1.09) but not multivariable (hazard ratio per unit in ug/mL, 1.03; 95% CI, 1.00-1.06; p=0.15) Cox regression analyses. As D-dimer was not independently associated with an increased risk of incident cerebrovascular events, we did not evaluate whether its addition would improve the discriminatory ability of the CANDLE score.

When D-dimer was added as a potential exposure variable to the three machine learning estimators, discriminatory performance of these models were similar to their original versions. The Random Forest model had a C-statistic of 0.67 (95% CI, 0.65−0.68), the regularized Cox regression model had a C-statistic of 0.72 (95% CI, 0.67−0.77), and the XGBoost model had a C-statistic of 0.65 (95% CI, 0.60−0.69). While discriminatory performance did not improve, D-dimer was a selected exposure in the regularized Cox regression estimator and an important feature of the Random Forest and XGBoost models.

Discussion

Using data from over 21,000 patients enrolled into the AHA's multicenter COVID-19 Cardiovascular Disease registry, we created an easy-to-use clinical score, entitled CANDLE, to help stratify stroke risk among patients hospitalized with COVID-19 infection. CANDLE includes 6

Table 2. *Baseline variables independently associated with an acute cerebrovascular event in the derivation cohort.*

Variable[a]	Univariable HR (95% CI)	Multivariable HR (95% CI)
Prior stroke or TIA	2.6 (1.9−3.5)	2.4 (1.8−3.3)
Temperature[b]	0.7 (0.6−0.9)	0.8 (0.7−0.9)
History of pulmonary disease	0.6 (0.4−0.9)	0.6 (0.4−0.8)
History of hypertension	2.1 (1.5−3.0)	2.0 (1.4−2.7)
White blood cell count[b]	1.05 (1.02−1.08)	1.05 (1.02−1.07)
Systolic blood pressure[c]	1.09 (1.03−1.15)	1.09 (1.03−1.15)

Abbreviations: HR, hazard ratio; CI, confidence interval; TIA, transient ischemic attack.

[a]Cox regression with 5-fold cross validation was used to identify baseline clinical factors associated with the primary outcome of an acute cerebrovascular event. Factors associated at the univariate level (p<0.10) were then entered into a multivariable Cox regression model. Factors independently associated with the primary outcome at p<0.05 are described herein. Temperature, white blood cell count, and systolic blood pressure were analyzed as continuous variables, while prior stroke or TIA, history of pulmonary disease, and history of hypertension were analyzed as dichotomous variables.

[b]Temperature and white blood cell count were analyzed per unit of each.

[c]Systolic blood pressure was analyzed per 10 units mmHg.

Table 3. *CANDLE: a clinical score for stratifying stroke risk among patients hospitalized with COVID-19 Infection.*

CANDLE Parameters	Multivariable HR (95% CI)
Cerebrovascular disease history	2.5 (1.9−3.4)
Afebrile temperature (<38.3 ˚C)	2.1 (1.3−3.4)
No pulmonary disease history	1.8 (1.2−2.6)
Disorder of hypertension	1.9 (1.4−2.7)
Leukocytosis (WBC count >11 K/uL)	1.7 (1.2−2.2)
Elevated systolic blood pressure (>140 mm Hg)	1.7 (1.3−2.2)

Abbreviations: HR, hazard ratio; CI, confidence interval; WBC, white blood cell.

[a]Multivariable Cox regression was used to examine the independent association between baseline clinical parameters and the risk of an acute cerebrovascular event.

variables—history of cerebrovascular disease, lack of fever, no history of pulmonary disease, history of hypertension, serum leukocytosis, and an elevated systolic blood pressure—that can be easily ascertained at hospital presentation though medical history, vital signs, and basic laboratory evaluation. The score's C-statistic was 0.66 for any cerebrovascular event and 0.67 for an ischemic cerebrovascular event, indicating moderate discriminatory performance for both outcomes. When grouped into different risk categories, a low-risk score of 0-1 estimated a 0.2% risk of an incident cerebrovascular event, while medium- and high-risk scores of 2-6 indicated considerably higher cerebrovascular event risks, ranging from 1.1% to 3.0%. Therefore, the score may be most useful in identifying patients who are highly unlikely (i.e., one-in-five hundred) to develop stroke during COVID-19 hospitalization and consequently may not warrant the same degree of monitoring or prophylactic treatments as other patients.

We also derived three machine learning estimators utilizing different statistical approaches to predict stroke risk during COVID-19 hospitalization. Despite more sophisticated models incorporating more quantitative and qualitative data from hospital admission, the discriminatory performance of these machine learning estimators was similar to that of the easy-to-use CANDLE score. In exploratory analyses restricted to the 40% of patients with available plasma D-dimer values at admission (n=8523), including D-dimer in the three machine learning estimators did not significantly improve their discriminatory ability, indicating that D-dimer is not a reliable risk marker for stroke among patient's hospitalized with COVID-19.

Some of our findings align with existing literature while others are novel and require further discussion. Prior studies have reported that history of cerebrovascular disease and hypertension may be risk factors for stroke in patients with COVID-19 infection.[13,15,24] This includes a descriptive study that we published using a smaller sample from the AHA's COVID-19 Cardiovascular Disease Registry.[25] Further, acute hypertension is a known short-term risk factor or trigger for cerebrovascular events, particularly hemorrhagic stroke, and it sometimes reflects a physiological response to stroke that has already manifested.[26, 27] Serum leukocytosis is a nonspecific marker of systemic inflammation, and inflammation is a probable risk factor for stroke in people with and without COVID-19 infection.[28-30]

Our finding that patients without prior pulmonary disease face a higher risk of stroke with COVID-19 was less expected. As history of pulmonary disease is associated with an increased risk of hospitalization and respiratory failure from COVID-19 infection,[31,32] it is possible that patients without prior pulmonary disease who became hospitalized with COVID-19 infection had a more severe systemic syndrome resulting in higher degrees of inflammation, endotheliopathy, and hypercoagulability. Conversely, it is also possible that patients with pulmonary disease were more likely to succumb earlier to COVID-19 infection, reducing the time available to develop stroke (i.e., severe COVID-19 infection served as a competing risk) or they were more likely to be sedated in an ICU making stroke detection more difficult.

We also found that lack of fever at presentation was associated with heightened stroke risk in patients hospitalized with COVID-19. A potential explanation for this finding is that patients who can mount an appropriate early febrile reaction to SARS2-CoV-2 may be less likely to develop a maladaptive delayed inflammatory response which is linked to more severe forms of COVID-19 infection and corresponding increased risks of thromboembolism.[33,34] In support of this hypothesis, high fever has been associated with a lower risk of death among patients hospitalized with COVID-19 pneumonia.[35] Alternatively, the observed association between a normal presenting temperature and increased stroke risk among patients hospitalized with COVID-19 could be because in some patients the viral infection was asymptomatic or mild and stroke or conditions predisposing to stroke were the primary reason for hospital admission—the registry does not distinguish patients hospitalized for COVID-19 versus those hospitalized with COVID-19.

Our study had several notable limitations. First, it was limited to patients enrolled into the AHA's COVID-19 Cardiovascular Disease registry. While this registry includes data from urban and rural hospitals from all
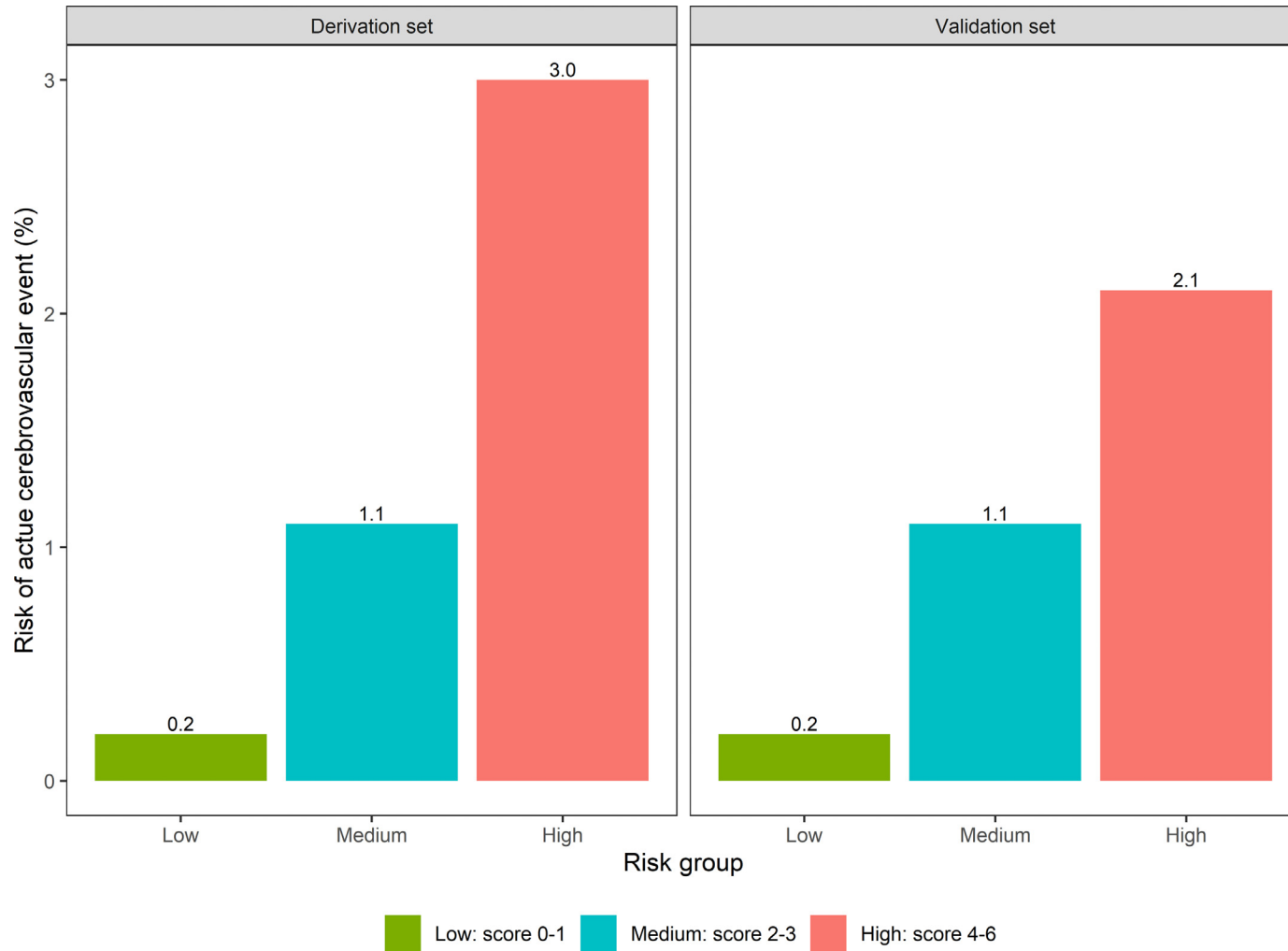
Fig. 1. *CANDLE: a stroke risk stratification score for patients hospitalized with COVID-19 infection. The CANDLE score ranges from 0-6 and assigns 1 point each for Cerebrovascular disease history, Afebrile temperature (<38.3 °C), No pulmonary disease history, Disorder of hypertension, Leukocytosis (white blood cell count >11 K/uL), and an Elevated systolic blood pressure (>140 mm Hg). Bar graphs display patients' risk of an acute cerebrovascular event during hospitalization with COVID-19 infection according to low-risk (score 0−1), medium-risk (score 2−3), and high-risk (score 4−6) groups. Separate graphs are shown for the derivation and validation cohorts.*

geographic regions and settings in the U.S. during the first 13 months of the pandemic, this study's findings may not generalize to non-participating U.S. hospitals, other countries, or infection with novel COVID-19 variants, such as the Delta or Omicron variants. Similarly, we lacked data on vaccination status and most patients were enrolled in 2020 before vaccines became widely available; therefore, the validity of our results amongst vaccinated patients is uncertain. Second, we had few data on advanced laboratory tests that reflect heightened coagulability and inflammation (e.g., interleukin-6, anti-phospholipid antibodies) and therefore did not include these biomarkers in our risk stratification models. Prior studies have found that elevated markers of inflammation and coagulation, particularly D-dimer, are associated with an increased risk for stroke in patients with COVID-19 infection.[19,36] Because of a high rate of missingness predisposing to selection bias, we restricted our analysis of D-dimer to an exploratory analysis, the results of which are hypothesis-generating. Third, as we aimed to build a stroke risk stratification score that could be implemented upon hospital presentation, we did not account for in-hospital events such as acute respiratory distress syndrome, mechanical ventilation, venous thromboembolism, or administered anti-viral and anti-thrombotic medications, which could have affected stroke risk among patients hospitalized with COVID-19.[3,13] Fourth, we lacked data on stroke mechanisms and therefore could not evaluate our risk stratification models for differential discriminatory performance according to individual stroke subtypes. Fifth, we could not account for potentially important patient- and hospital-level factors that varied over time, such as patient willingness to visit the hospital, evidence-based treatments for COVID-19 treatment, and hospital resources and thresholds for brain imaging.

In conclusion, we created and internally validated an easy-to-use clinical score, entitled CANDLE, and several complex machine learning estimators, to help clinicians in stratifying patients' stroke risk at the time of hospitalization with COVID-19 infection. Before clinical use, these risk stratification models should be validated in external cohorts with more contemporaneous SARS-CoV-2 variants. Further, they should be evaluated for their ability to predict all thromboembolic events, not just stroke, as that may be more helpful for frontline providers when deciding which patients to treat with prophylactic antithrombotic therapy. In the meantime, these risk stratification models, which were derived from one of the largest and most diverse COVID-19 cohorts in the world, may provide clinicians with useful estimates for stroke risk among patients presenting to their hospital with COVID-19 infection.

Declaration of Competing Interest

Alexander Merkler and Babak Navi report an institutional research grant from Weill Cornell Medicine to examine stroke risk in patients with COVID-19. Costantino Iadecola serves on the scientific advisory board for Broadview Ventures. Mitchell Elkind receives study drug in kind from Bristol Myers Squibb-Pfizer alliance for Eliquis and ancillary research funding from Roche for an NIH-funded stroke prevention trial; he also reports royalties from UpToDate for chapters on COVID-19 and neurological disease and is an uncompensated officer of the American Heart Association. Hooman Kamel serves as co-PI for the NIH-funded ARCADIA trial which receives in-kind study drug from the BMS-Pfizer Alliance and in-kind study assays from Roche Diagnostics, serves as Deputy Editor for *JAMA Neurology*, serves as a steering committee member of Medtronic's Stroke AF trial (uncompensated), serves on an endpoint adjudication committee for a trial of empagliflozin for Boehringer-Ingelheim, and has served on an advisory board for Roivant Sciences related to Factor XI inhibition. The other authors have no disclosures.

References

1. WHO. WHO coronavirus (COVID-19) dashboard. https://covid19.who.int/. Accessed March 22, 2022.
2. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical characteristics of Covid-19 in New York city. N Engl J Med 2020;382:2372-2374.
3. Helms J, Tacquard C, Severac F, et al. High risk of thrombosis in patients with severe SARS-CoV-2 infection: a multicenter prospective cohort study. Intensive Care Med 2020;46:1089-1098.
4. Tang N, Li D, Wang X, et al. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020;18:844-847.
5. Zhang Y, Xiao M, Zhang S, et al. Coagulopathy and anti-phospholipid antibodies in patients with Covid-19. N Engl J Med 2020;382:e38.
6. Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. Clin Infect Dis 2020;71:762-768.
7. Bilaloglu S, Aphinyanaphongs Y, Jones S, et al. Thrombosis in hospitalized patients with COVID-19 in a New York city health system. JAMA 2020;324:799-801.
8. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex immune dysregulation in COVID-19 patients

with severe respiratory failure. Cell Host Microbe 2020;27:992-1000. e1003.

9. Yaghi S, Ishida K, Torres J, et al. SARS2-CoV-2 and stroke in a New York healthcare system. Stroke 2020;51:2002-2011.

10. Oxley TJ, Mocco J, Majidi S, et al. Large-vessel stroke as a presenting feature of Covid-19 in the young. N Engl J Med 2020;382:e60.

11. Beyrouti R, Adams ME, Benjamin L, et al. Characteristics of ischaemic stroke associated with COVID-19. J Neurol Neurosurg Psychiatry 2020;91:889-891.

12. Mao L, Jin H, Wang M, et al. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. JAMA Neurol 2020;77:683-690.

13. Merkler AE, Parikh NS, Mir S, et al. Risk of ischemic stroke in patients with coronavirus disease 2019 (COVID-19) vs patients with influenza. JAMA Neurol 2020;77:1-7.

14. Katsanos AH, Palaiodimou L, Zand R, et al. The impact of SARS-CoV-2 on stroke epidemiology and care: a meta-analysis. Ann Neurol 2021;89:380-388.

15. Qureshi AI, Baskett WI, Huang W, et al. Acute ischemic stroke and COVID-19: an analysis of 27 676 patients. Stroke 2021;52:905-912.

16. Srivastava PK, Zhang S, Xian Y, et al. Acute ischemic stroke in patients with COVID-19: An analysis from get with the guidelines-stroke. Stroke 2021;52:1826-1829.

17. Alger HM, Rutan C, JHt Williams, et al. American heart association COVID-19 CVD registry powered by get with the guidelines. Circ Cardiovasc Qual Outcomes 2020;13: e006967.

18. Benchimol EI, Smeeth L, Guttmann A, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLOS Medicine 2015;12:e1001885.

19. Esenwa C, Cheng NT, Luna J, et al. Biomarkers of coagulation and inflammation in COVID-19-associated ischemic stroke. Stroke 2021;52:e706-e709.

20. Katz JM, Libman RB, Wang JJ, et al. Cerebrovascular complications of COVID-19. Stroke 2020;51:e227-e231.

21. Rothstein A, Oldridge O, Schwennesen H, et al. Acute cerebrovascular events in hospitalized COVID-19 Patients. Stroke 2020;51:e219-e222.

22. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions, 2. SN Computer Science; 2021. p. 160.

23. Moncada-Torres A, van Maaren MC, Hendriks MP, et al. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep 2021;11:6968.

24. Nannoni S, De Groot R, Bell S, et al. Stroke in COVID-19: A systematic review and meta-analysis. Int J Stroke 2021;16:137-149.

25. Shakil SS, Emmons-Bell S, Rutan C, et al. Stroke among patients hospitalized with COVID-19: results from the American heart association COVID-19 cardiovascular disease registry. Stroke 2022;53:800-807.

26. Fischer U, Cooney MT, Bull LM, et al. Acute post-stroke blood pressure relative to premorbid levels in intracerebral haemorrhage versus major ischaemic stroke: a population-based study. Lancet Neurol 2014;13:374-384.

27. Qureshi AI. Acute hypertensive response in patients with stroke. Circulation 2008;118:176-187.

28. Esenwa CC, Elkind MS. Inflammatory risk factors, biomarkers and associated therapy in ischaemic stroke. Nat Rev Neurol 2016;12:594-604.

29. McAlpine LS, Zubair AS, Maran I, et al. Ischemic stroke, inflammation, and endotheliopathy in COVID-19 patients. Stroke 2021;52:e233-e238.

30. Guiraud V, Amor MB, Mas J-L, et al. Triggers of ischemic stroke. Stroke 2010;41:2669-2677.

31. Zhao Q, Meng M, Kumar R, et al. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. J Med Virol 2020;92:1915-1921.

32. Gerayeli FV, Milne S, Cheung C, et al. COPD and the risk of poor outcomes in COVID-19: a systematic review and meta-analysis. EClinicalMedicine 2021;33:100789.

33. Fajgenbaum DC, June CH. Cytokine storm. N Engl J Med 2020;383:2255-2273.

34. Klok FA, Kruip MJHA, Van Der Meer NJM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res 2020;191:145-147.

35. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med 2020;180:934-943.

36. Li Y, Li M, Wang M, et al. Acute cerebrovascular disease following COVID-19: a single center, retrospective, observational study. Stroke Vasc Neurol 2020;5:279-284.