

ORIGINAL RESEARCH

Development and Validation of the COVID-NoLab and COVID-SimpleLab Risk Scores for Prognosis in 6 US Health Systems

Mark H. Ebell, MD, MS, Xinyan Cai, MPH, Robert Lennon, MD, Derjung M. Tarn, MD, PhD, Arch G. Mainous III, PhD, Aleksandra E. Zgierska, MD, PhD, Bruce Barrett, MD, PhD, Wen-Jan Tuan, DHA, MS, MPH, Kevin Maloy, MD, Munish Goyal, MD, and Alex Krist, MD, MPH

Purpose: Develop and validate simple risk scores based on initial clinical data and no or minimal laboratory testing to predict mortality in hospitalized adults with COVID-19.

Methods: We gathered clinical and initial laboratory variables on consecutive inpatients with COVID-19 who had either died or been discharged alive at 6 US health centers. Logistic regression was used to develop a predictive model using no laboratory values (COVID-NoLab) and one adding tests available in many outpatient settings (COVID-SimpleLab). The models were converted to point scores and their accuracy evaluated in an internal validation group.

Results: We identified 1340 adult inpatients with complete data for nonlaboratory parameters and 741 with complete data for white blood cell (WBC) count, differential, c-reactive protein (CRP), and serum creatinine. The COVID-NoLab risk score includes age, respiratory rate, and oxygen saturation and identified risk groups with 0.8%, 11.4%, and 40.4% mortality in the validation group (AUROC = 0.803). The COVID-SimpleLab score includes age, respiratory rate, oxygen saturation, WBC, CRP, serum creatinine, and comorbid asthma and identified risk groups with 1.0%, 9.1%, and 29.3% mortality in the validation group (AUROC = 0.833).

Conclusions: Because they use simple, readily available predictors, developed risk scores have potential applicability in the outpatient setting but require prospective validation before use. (J Am Board Fam Med 2021;34:S127–S135.)

Keywords: Clinical Decision Support, Clinical Prediction Rule, COVID-19, Logistic Models

Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 virus, named COVID-19, has created an

unprecedented health crisis. There have been more than 10 million confirmed cases and more than 500,000 deaths worldwide,¹ with an estimated 10 undetected cases per confirmed case.² The case fatality rate is estimated to be approximately 0.5 to 1.0%, approximately 5 to 10 times higher than seasonal influenza, with older patients having much higher case fatality rates.³ The spectrum of illness is broad, ranging from completely asymptomatic carriers to those with critical illness and death. This

This article was externally peer reviewed.
Submitted 4 September 2020; revised 19 October 2020; accepted 3 December 2020.

From the Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens (MHE, XC); Department of Family and Community Medicine, Penn State College of Medicine, Hershey (RL); Department of Family Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles (DMT); Department of Health Services Research, Management and Policy, University of Florida, Gainesville (AGM); Departments of Public Health Sciences, and Anesthesiology and Perioperative Medicine, Penn State College of Medicine, Hershey (AEZ); Department of Family Medicine and Community Health, University of Wisconsin, Madison (BB, WJT); Department of Emergency Medicine, MedStar Washington Hospital Center, Washington, DC

(KM, MG); Department of Family Medicine, Virginia Commonwealth University, Richmond (AK).

Funding: None.

Conflict of interest: None.

Corresponding author: Mark H. Ebell, MD, MS, 125 BS Miller Hall, UGA Health Sciences Campus, Athens, GA 30602 (E-mail: ebelle@uga.edu).

breadth of presentation makes optimal disposition difficult at the time of initial presentation, because the clinical presentation may not correlate with the patient's actual risk of a bad outcome.

A major concern is that hospital beds and in particular intensive care unit (ICU) beds and mechanical ventilators may be overwhelmed when cases rise in an area. This makes it critical that physicians have the tools needed to identify patients both at lower and elevated mortality risk at the time of initial presentation. An accurate risk assessment tool using simple parameters available on presentation to the emergency department and other settings could aid clinicians in rapidly making optimal patient disposition decisions. For patients who are hospitalized, it could guide the intensity of monitoring and the initial admission location (hospital ward, telemetry, or ICU). If validated in the outpatient setting, it could also guide hospitalization decisions. Key risk factors for mortality have been identified and include increasing age, male sex, comorbidities, and certain laboratory parameters.³⁻⁵ Systematic review of laboratory parameters found that lymphopenia and elevated levels of c-reactive protein (CRP), neutrophil count, interleukin-6, d-dimer, lactate dehydrogenase, and troponin I were all associated with a poor outcome in hospitalized COVID-19 patients.^{6,7}

Researchers have attempted to develop prediction models for poor prognosis in COVID-19 patients, combining demographic, comorbidity, physical examination, laboratory, and imaging predictors into multivariate models. In some cases, these have been simplified into clinical prediction rules (CPRs) or online calculators.⁸⁻¹¹ However, many have not been externally validated, and none have been externally validated in a US population. In addition, many of these CPRs or models use laboratory tests and imaging that would not readily allow their extension to primary care or urgent care settings.¹⁰⁻¹² As more COVID-19 patients are managed via telehealth, having a CPR that can be applied early in the disease course and that does not rely on any laboratory testing would be desirable to avoid having to bring low-risk patients to a laboratory or outpatient office for an in-person visit.

Therefore, the primary goal of the current study is to develop and validate 2 simple CPRs to predict COVID-19 mortality risk, 1 that relies only on nonlaboratory parameters (COVID-NoLab) and another that adds simple laboratory tests commonly

available in primary or urgent care settings (COVID-SimpleLab). As the goal is to decide decision making on initial presentation, only data from the first 24 hours will be used to develop the CPRs. To accomplish this, we used data from a diverse multicenter US population of adults hospitalized with COVID-19. Secondly, we will use this population's data to evaluate several previously developed risk scores for COVID-19 prognosis.

Methods and Materials

Study Organization

The lead investigator (MHE) identified colleagues at 6 major US universities (University of Wisconsin-Madison, Penn State University, University of Florida, Virginia Commonwealth University, University of California at Los Angeles, and Georgetown University) with inpatient health centers to participate in a study of COVID-19 prognosis. Each site obtained Institutional Review Board (IRB) approval for this project, which was deemed to be exempt research due to using deidentified, previously collected patient data extracted retrospectively from each health system's electronic health record. Data use agreements were established between each university and the University of Georgia. The overall project was approved by the University of Georgia IRB.

Data Collection

A standardized data set of demographic, clinical, and laboratory parameters was assembled using extant literature and with input from the group (Appendix 1). Comorbidities were defined using Clinical Classifications Software categories for the following disease clusters: cardiovascular disease (CCS 101), chronic obstructive pulmonary disease (CCS 127), asthma (CCS 128), and diabetes mellitus (CCS 49).¹³ Inclusion criteria included any adult inpatient with a positive polymerase chain reaction test for COVID-19 hospitalized at one of the participating institutions whose disposition was already determined (discharged or deceased) at the time of data extraction. The primary outcome was in-hospital mortality. We also conducted exploratory analyses for prediction of the combined outcome of death or need for mechanical ventilation.

Each site was responsible for its own data extraction from its electronic health record, following the standardized approach to each variable definition. Gender, age, and predictor variable were collected.

Because the goal is to be able to predict prognosis at admission, only predictor variables available within 24 hours of admission date/time were included. Each patient's extracted data were deidentified at the collection site. As age over 90 could be considered identifying, patients aged 90 years or over had their age listed as 90. Each center had a different range of dates for data collection, beginning as early as March 1, 2020 and extending as far as June 12, 2020. Deidentified data were securely transferred from each institution to a central repository at the University of Georgia, where they were combined for analysis.

Validation of Existing Clinical Risk Scores

The lead investigator's systematic review of individual risk factors, risk scores, and prognostic models to predict critical illness or death in patients with COVID-19 (manuscript in review) was used to identify 2 simple risk scores^{9,11} and a simple multivariate model¹⁰ for COVID-19 mortality in the literature (all in inpatients). For each patient with the predictor variables in the risk score, the score was calculated. The proportion of patients with the outcome of interest (eg, death) in each risk group and where possible the area under the receiver operating characteristic curve (AUROC) were calculated for each score or model.

Development and Internal Validation of Novel Risk Scores Using Our Data Set

Continuous variables were presented as the median and interquartile range, and categorical variables were presented as frequencies and percentages of occurrence. For the univariate analysis, the bivariate associations between predictor variables and mortality were assessed using the chi squared test for categorical variables and Wilcoxon rank-sum test for continuous variables.

We then randomly divided the data into derivation and validation groups with a ratio of 60:40 and built logistic regression models in the derivation set with in-hospital mortality as the outcome or dependent variable. In the first model, we only considered the patient's age, comorbidities, and vital signs (including oxygen saturation) as independent predictors. In the second model, we added the white blood cell (WBC) count, white cell differential, serum creatinine, and CRP to the models. Imputation of laboratory data were considered, but given the large number of missing cases, we performed complete

case analyses. Continuous variables were converted to categorical variables to simplify calculations in the final risk score based on inspection of histograms. We used stepwise backward selection with $P < .1$ for retention in the model.¹⁴ Once the predictors were selected, β coefficients were determined from the final multivariable logistic regression model. We then created a simple point score by dividing each β coefficient by the smallest β value and rounded it to the nearest integer. The low-risk, moderate-risk, and high-risk groups were created based on visual inspection of the point score distribution to create groups that would be most useful for clinical decision making, with a particular goal of having the low-risk group be at or near 1% mortality.

The performance of the point scores was internally validated using the validation data set. This included evaluation of how accurately the score classified patients into low-, moderate-, and high-risk groups. We used the Hosmer–Lemeshow test and a calibration curve to evaluate calibration, which indicates how well predicted mortality matched observed mortality. The AUROC was used as a measure of overall discrimination.

Results

Characteristics of the Study Population

The characteristics of the study population are summarized in Table 1, stratified by health system. The number of patients available for analysis at each center ranged from 69 to 582, and the mortality rate ranged from 1.4% to 16.7%, with an overall mortality rate of 13.1%. The median age of participants at the 6 sites ranged from 52 to 62 years; there was a slight male preponderance.

The bivariate analysis of the association between clinical variables and mortality is shown in Table 2. Nonlaboratory parameters positively associated with mortality ($P < .05$) included increasing age, several comorbidities (cardiovascular disease, diabetes mellitus, and chronic obstructive pulmonary disease), increased body mass index, decreased oxygen saturation, and increased respiratory rate. Laboratory parameters positively associated with mortality included increased CRP, WBC count, neutrophil count, serum creatinine, and decreased lymphocyte count.

Development and Validation of Simple Risk Scores

Table 3 summarizes the 2 multivariate models to predict COVID-19 mortality using basic data

Table 1. Characteristics of Included Patients From Each Institution

	Georgetown University (n = 582)	Virginia Commonwealth University (n = 223)	University of Wisconsin (n = 102)	Penn State University (n = 69)	University of Florida (n = 133)	University of California, LA (n = 333)	Total (n = 1442)
Disposition							
Discharge home	485 (83.3%)	190 (85.2%)	88 (86.3%)	68 (98.6%)	121 (91%)	301 (90.4%)	1253 (86.9%)
Deceased	97 (16.7%)	33 (14.8%)	14 (13.7%)	1 (1.4%)	12 (9%)	32 (9.6%)	189 (13.1%)
Demographics and vitals							
Age: median (interquartile range)	61 (22)	61 (22)	62 (25)	63 (29)	57 (39)	52 (30)	61 (25)
Male: female	287:295	119:104	55:47	34:35	53:80	197:143	745:704

available at initial presentation. Complete case data were available for 1342 patients for the COVID-NoLab model and 741 for the COVID-SimpleLab model. The COVID-NoLab model had an AUROC of 0.771 in the derivation group and 0.803 in the validation group. The COVID-SimpleLab model had an AUROC of 0.835 in the derivation group and 0.833 in the validation group.

Calibration in the validation groups was good based on visual inspection of calibration plots, with nonstatistically significant values for the Hosmer–Lemeshow goodness of fit test ($P=.759$ for the COVID-NoLab model and $P=.400$ for the COVID-SimpleLab model). The receiver operating characteristic (ROC) curves and calibration plots for each model are shown in Appendix 2.

The COVID-NoLab and COVID-SimpleLab risk scores were created based on the derivation set data, using β -coefficients as described above. The COVID-NoLab and COVID-SimpleLab risk scores and their classification accuracy are summarized in Table 4 for the derivation and validation groups for each risk score. Both simple risk scores had similar classification accuracy in the derivation and validation groups. However, the score that adds simple laboratory tests classifies a higher percentage of patients as low risk (29% vs 21% in derivation and 33% vs 24% in validation) who could potentially be managed as outpatients. It also classifies more patients as high risk who will require closer monitoring or intensive care (29% vs 12% in derivation and 34% vs 11% in validation).

Models were also developed and internally validated for settings where only the WBC count might be available, or only the CRP test. These models' risk scores are summarized in Appendix 3.

Although both models were able to identify high-risk patients, in each case the low-risk group in the validation data sets had an appreciably higher mortality rate than in the derivation data (4.4% vs 0.0% for both models). Their calibration was good, based on visual inspection of the calibration plots and the Hosmer–Lemeshow test.

Evaluation of Previous Risk Scores

We evaluated 3 existing simple models for predicting COVID-19 mortality. Five clinical variables were included in the 3 tools: age, CRP, lactate dehydrogenase, lymphopenia, and oxygen saturation. Two tools used classification trees and had not been externally validated,^{9,11} and 1 was a simple multivariate model that had been validated at a single Chinese hospital.¹⁵ We were unable to evaluate the accuracy of other risk scores due to either the unavailability of some of the predictors in our data set or because they predicted outcomes other than mortality.⁸ The performance of each of the 3 prediction models in the US study population is summarized in Table 5.

Discussion

We have developed and internally validated 2 simple CPRs, 1 of which requires no laboratory testing (COVID-NoLab) and another that only requires clinical variables plus simple laboratory tests that are commonly and rapidly available in many outpatient settings (COVID-SimpleLab). The score that includes simple lab tests classifies more patients as low or high risk and is therefore potentially more clinically useful. Previous risk scores have either not been internally validated, have not been validated in

Table 2. Association Between Individual Clinical Variables and the Outcome of Mortality

Clinical Predictors	Discharged Alive	Deceased	P-value
Categorical variable (n [%])	1253	189	
Female sex	615 (49.1)	84 (44.4)	0.267
Chronic obstructive pulmonary disease	180 (14.4)	44 (23.3)	0.002
Asthma	139 (11.1)	23 (12.2)	0.754
Diabetes	421 (33.6)	91 (48.1)	<0.001
Cardiovascular disease	217 (17.3)	56 (29.6)	<0.001
Temperature $\geq 39^{\circ}\text{C}$	69 (5.8)	11 (6.0)	1
Temperature $\geq 38^{\circ}\text{C}$	203 (17.0)	34 (18.6)	0.684
Age (years)			<0.01
<50	396 (32.3)	9 (5)	
50 to 65	415 (33.8)	42 (23.2)	
66 to 79	270 (22)	79 (43.7)	
>80	146 (11.9)	51 (28.1)	
Body mass index $\geq 40\text{ kg/m}^2$	147 (14.0)	27 (15.9)	0.604
Body mass index $\geq 30\text{ kg/m}^2$	513 (49.0)	69 (40.6)	0.051
Oxygen saturation < 93%	82 (7.0)	36 (19.3)	<0.001
Respiratory rate ≥ 25 breaths/minute	215 (17.9)	67 (35.6)	<0.001
Respiratory rate ≥ 30 breaths/minute	96 (8.0)	40 (21.3)	<0.001
C-reactive protein			
> 10 mg/dL	228 (33.0)	77 (67.5)	<0.001
> 15 mg/dL	119 (17.2)	46 (40.4)	<0.001
> 20 mg/dL	42 (6.1)	15 (13.2)	0.011
> 40 mg/dL	2 (0.3)	2 (1.8)	0.18
White blood cell count $> 10 \times 10^9/\text{mL}$	172 (14.9)	60 (33.0)	<0.001
Lymphocytes $< 0.8 \times 10^9/\text{mL}$	271 (26.0)	68 (40.5)	<0.001
Neutrophils $> 8 \times 10^9/\text{mL}$	145 (13.9)	49 (29.2)	<0.001
Serum creatinine $\geq 2.0\text{ mg/dL}$	114 (9.9)	66 (36.1)	<0.001
Continuous variables (median [interquartile range])			
Age in years	59 [44.5, 71]	72.5 [64.8, 80]	<0.001
Temperature $^{\circ}\text{C}$	37.2 [36.8, 37.8]	37.2 [36.8, 37.8]	0.831
Oxygen saturation	97% [95%, 98%]	95% [93%, 97%]	<0.001
Respiratory rate (breaths/minute)	20 [18, 23]	20 [18, 27]	<0.001
White blood cell count ($\times 10^9/\text{mL}$)	6.1 [4.6, 8.1]	7.5 [5.5, 11.8]	<0.001
Lymphocytes ($\times 10^9/\text{mL}$)	1.1 [0.8, 1.5]	0.8 [0.6, 1.3]	<0.001
Neutrophils ($\times 10^9/\text{mL}$)	4.4 [3, 6.2]	5.9 [4, 8.7]	<0.001
Body mass index (kg/m^2)	29.8 [25.3, 35.8]	28.5 [24.6, 35.8]	0.254
Serum creatinine (mg/dL)	0.9 [0.8, 1.3]	1.5 [1.1, 2.7]	<0.001
C-reactive protein (mg/dL)	6.2 [2.4, 12.3]	13 [9, 17.3]	<0.001
Lactate dehydrogenase (U/L)	313 [241, 409]	441 [349, 590]	<0.001

the United States, or have required tests not commonly available in outpatient settings such as procalcitonin, lactate dehydrogenase, or chest radiography. Our risk scores performed well in an internal validation, although external, prospective validation in other populations would be desirable. The risk scores are simple enough for clinicians to memorize or keep on a pocket card. In the future, they could be made available as a mobile app for

point-of-care use or integrated into electronic health records.

The COVID-NoLab score has important potential utility in the telehealth setting, which has become a common venue for assessing and monitoring COVID-positive patients while minimizing the risk of viral transmission to clinical staff. Although the score does require an oxygen saturation level, patients with COVID-19 are increasingly

Table 3. Multivariate Models Using Limited Clinical Data to Predict Mortality in Hospitalized Patients With COVID-19

Model using no laboratory values (COVID-NoLab)				
Predictors	β -Coefficient	Std. Error	Z Value	Pr(> z)
Constant	-4.010	0.465	-8.620	0.001
Age				
50 to 65 years	1.631	0.498	3.270	0.001
> 65 years	2.602	0.478	5.450	< 0.001
Respiratory rate \geq 30/min	1.352	0.293	4.620	< 0.001
Oxygen saturation < 93%	1.036	0.288	3.600	< 0.001
AUROCC = 0.771 in the derivation group and 0.803 in the validation group.				
Model adding simple blood tests (COVID-SimpleLab)				
Predictors	β -Coefficient	Std. Error	Z Value	Pr(> z)
Constant	-4.899	0.669	-7.330	0.000
C-reactive protein				
11–20 mg/dL	1.316	0.329	4.010	0.000
> 20 mg/dL	1.057	0.510	2.080	0.038
Respiratory rate \geq 30/min	1.522	0.403	3.770	0.000
Oxygen saturation < 93%	0.911	0.418	2.180	0.029
Age				
50 to 65 years	1.310	0.661	1.980	0.047
> 65 years	2.233	0.640	3.490	0.000
Asthma	0.953	0.428	2.230	0.026
White blood cell count > 10×10^9 /mL	0.579	0.348	1.660	0.097
Serum creatinine > 2.0 mg/dL	1.152	0.348	3.310	0.001
AUROCC = 0.835 in the derivation group and 0.833 in the validation group.				

AUROCC, area under the receiver operating characteristic curve.

being given devices for home assessment of oxygen saturation as a way to remotely monitor their symptoms. Our study reinforces the value of knowing this parameter as a way to predict mortality risk and, potentially, health decline. Our findings, although not yet conclusive, may encourage innovative health systems to consider home oxygen saturation as a means to safely manage COVID-infected patients at home. For example, one could have patients measure oxygen saturation twice daily, have a daily telehealth visit with a health care professional who could evaluate respiratory rate, and recalculate the risk score daily. It is also something that could be used by emergency response personnel when evaluating patients in the field, where blood tests are not available but oxygen saturation monitors are readily available.

The COVID-SimpleLab risk score was somewhat more accurate than the COVID-NoLab risk

score and is appropriate for outpatient settings where the WBC count, CRP, and serum creatinine are available. We also developed risk scores that included only clinical variables and either WBC or CRP, because outpatient settings around the world often have different tests available. For example, although the WBC is often available in the US primary care setting at the point of care, CRP is rarely available. On the other hand, the opposite is true in many European countries.^{16,17} Although the “Clinical + WBC” and “Clinical + CRP” risk scores did not perform as well in validation, particularly at identifying a very low-risk group, they should still be prospectively validated in lower-risk outpatients with COVID-19 to see if they perform better in that population.

The previously reported risk scores originally developed in Chinese populations^{9–11} were less accurate in our US population. This may be

Table 4. Calculation of the COVID-NoLab and COVID-SimpleLab Risk Scores and Their Classification Accuracy in Derivation and Validation Groups

COVID-NoLab Score			Derivation Group		Validation Group	
			Mortality	SSLR	Mortality	SSLR
Clinical Predictor	Points	Risk Group	n/Total (%)		n/Total (%)	
Age		Low (0 to 1)	3/167 (1.8%)	0.10	1/129 (0.8%)	0.06
50 to 65 years	3	Moderate (2 to 5)	74/543 (13.6%)	0.89	40/350 (11.4%)	0.95
> 65 years	5	High (6+)	44/96 (45.8%)	4.79	23/57 (40.4%)	4.99
Respiratory rate \geq 30	3					
O ₂ saturation < 93%	2					

COVID-SimpleLab Score			Derivation Group		Validation Group	
			Mortality	SSLR	Mortality	SSLR
Clinical Predictor	Points	Risk Group	n/Total (%)		n/Total (%)	
C-reactive protein > 10 mg/dL	5	Low (0 to 7)	0/129 (0.0%)	0.0	1/97 (1.0%)	0.07
Respiratory rate \geq 30	5	Moderate (8 to 11)	10/131 (8.3%)	0.46	9/99 (9.1%)	0.66
O ₂ saturation < 93%	4	High (12+)	58/185 (45.7%)	2.53	29/99 (29.3%)	2.73
Age						
50 to 65 years	6					
> 65 years	8					
Asthma	4					
White blood cell count > 10×10^9 /mL	3					
Serum creatinine > 2.0 mg/dL	4					

SSLR, stratum specific likelihood ratio.

because of overfitting of the early models, differences in the spectrum of illness, or differences between the health care systems in China and the United States. In addition, these models were developed early in the pandemic when mortality rates were higher.

We hope to work with investigators at other institutions to evaluate the COVID-NoLab and COVID-SimpleLab models in their populations. We

only gathered data on 4 comorbidities and in the future would want to explore adding other clinical variables such as hypertension, chronic liver disease, and tobacco use. It would be preferable to use prospective data collection and add patient symptoms such as dyspnea, although respiratory rate and oxygen saturation measurements may covary with dyspnea, making it less important. Including patients identified in a range of settings and managed as

Table 5. Selected Clinical Prediction Rules for COVID-19 Prognosis

Study	Performance in Original Study	Performance in 5 US Sites
Lu et al., 2020 ⁹	Low risk: 0% Moderate risk: 6% High risk: 33%	Low risk: 2/108 (1.8%) Moderate risk: 24/359 (6.7%) High risk: 87/324 (26.9%)
Xie et al., 2020 ¹⁰	AUROC (derivation) = 0.893 AUROC (test) = 0.980	AUROC = 0.7981 Hosmer–Lemeshow $P = .23$
Yan et al., 2020 ¹¹	Low risk: 3/189 (1.6%) High risk: 157/162 (96.9%)	Low risk: 10/243 (4.1%) High risk: 85/374 (22.7%)

AUROC, area under the receiver operating characteristic curve.

See appendix for details regarding calculation of risk scores.

outpatients will be important. Finally, this work should be ongoing, because as treatments will hopefully improve, the prognosis will change and predictive models will require updating.

Strengths and Limitations

An important strength of this study is that our model was developed using data from 6 geographically diverse sites in the United States, sites that serve racially and ethnically diverse populations. Further, by generating risk scores that use either no laboratory variables or limited laboratory testing, if appropriately validated our results could potentially be useful in outpatient settings or in telehealth to guide decisions regarding the need for admission or the intensity of outpatient follow-up that is needed. The risk scores are also quite simple and have good face validity, making them practical for busy clinical settings.

Our study has several limitations. This is a convenience sample, and we only included data for patients who had been discharged alive or who died. Thus, patients still in the hospital were not included; this may bias the sample. Importantly, the data collected is restricted to COVID-19 patients in an inpatient setting who have a narrower and more severe spectrum of illness than patients managed at home without hospitalization. Thus, our work requires validation in other populations, including primary care and urgent care settings, before clinical application in outpatients. Changes in the virus itself and changes in treatment may also affect prognosis over time, so any risk score may eventually require updating or recalibration. Finally, we used a split-sample internal validation, which may inflate calibration, and the model should be prospectively validated before adoption by clinicians.

Conclusion

The COVID-NoLab and COVID-SimpleLab scores derived in a large, diverse population of hospitalized COVID-19 patients in the United States had good discrimination, calibration, and classification accuracy using an internal validation (split-sample) approach. If validated in a new population of hospitalized patients, they provide a rapid, simple way to determine prognosis for hospitalized patients and identify a low-risk group that could be considered for outpatient management in a bed shortage, for example. Because they were designed to use no or minimal laboratory tests, these risk scores may also be generalizable to outpatient

settings. This could potentially provide clinicians a useful aid for decision making regarding hospital admission and the intensity of outpatient follow-up. However, it is important that the risk scores be prospectively validated in the outpatient setting before its use there.

To see this article online, please go to: <http://jabfm.org/content/34/Supplement/S127.full>.

References

1. Johns Hopkins COVID-19 Dashboard. Available from: <https://coronavirus.jhu.edu/map.html>. Accessed June 27, 2020.
2. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* 2020;396(10247):P313–9.
3. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20(6):669–77.
4. 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.
5. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395(10229):1054–62.
6. Fu L, Wang B, Yuan T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. *J Infect* 2020;80(6):656–65.
7. Moutchia J, Pokharel P, Kerri A, et al. Clinical laboratory parameters associated with severe or critical novel coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *PLoS One* 2020;15(10):e239802.
8. Liang W, Liang H, Ou L, for the China Medical Treatment Expert Group for COVID-19, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med* 2020;180:1081.
9. Lu J, Hu S, Fan R, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan. *medRxiv*. Preprint posted online February 23, 2020.
10. Xie J, Hungerford D, Chen H, Abrams ST. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *medrxiv.org*. Preprint posted online April 7, 2020.
11. Yan L, Zhang H-T, Xiao Y, Wang M. Prediction of criticality in patients with severe Covid-19

- infection using three clinical features. medrxiv.org. Preprint posted online March 3, 2020.
12. Galloway JB, Norton S, Barker RD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: an observational cohort study. *J Infect* 2020;81(2):282–8.
13. Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project website. 2020. Available from: <https://www.hcup-us.ahrq.gov/>.
14. Hocking RR. The analysis and selection of variables in linear regression. *Biometrics* 1976;32:1.
15. Xie J, Hungerford D, Chen H, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *SSRN Electron J*. Preprint posted online April 6, 2020.
16. Hardy V, Thompson M, Keppel GA, et al. Qualitative study of primary care clinicians' views on point-of-care testing for C-reactive protein for acute respiratory tract infections in family medicine. *BMJ Open* 2017;7:e012503.
17. Howick J, Cals JWL, Jones C, et al. Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open* 2014;4:e005611–e005611.

Appendix 1. Full List of Requested Clinical Variables; Predictor Variables only Included if Ordered Within 24 Hours of Admission

Clinical Variable	Normal Range	Units
Demographics		
Health system		
Hospital		
Age in years		years
Sex		
Race		
Comorbidities		
COPD		
Asthma		
Cardiovascular disease		
Hypertension		
Diabetes mellitus		
Vitals		
Respiratory rate	12 to 20	breaths/minute
Temperature	36.5 to 37.5	degrees Celsius
Heart rate	60 to 100	beats/minute
Systolic blood pressure	90 to 139	mm Hg
Diastolic blood pressure	50 to 89	mm Hg
BMI	20 to 24.9	kg/m ²
O ₂ saturation room air	95% to 100%	%
Laboratory tests		
White blood cell count	4.5 to 10	1000 cells/microliter
Lymphocyte count	1000 to 4800	cells/microliter
Neutrophil count	2500 to 7500	cells/microliter
Platelets	150,000 to 450,000	platelets/microliter
Serum creatinine	0.5 to 1.2	mg/dL
Blood urea nitrogen	7 to 20	mg/dL
Lactate dehydrogenase (LDH)	140 to 280	units/L
Aspartate aminotransferase (AST)	10 to 40	units/L
Alanine aminotransferase (ALT)	7 to 56	units/L
Ferritin	12 to 300	ng/mL
Troponin T or I	0 to 0.4	ng/mL
C-reactive protein (CRP)	< 10	mg/dL
D-dimer	< 0.5	mg/L
Interleukin-6 (IL6)	0 to 16	pg/mL
Outcome variables		
Vasopressor needed in first 24 hours		
Discharge disposition (discharged home, still hospitalized, deceased)		
ICU admit during hospitalization (Y/N)		
Mechanical ventilation (Y/N)		
Number of days hospitalized (including observation status)		
Number of days in the ICU		
Number of days on ventilator		

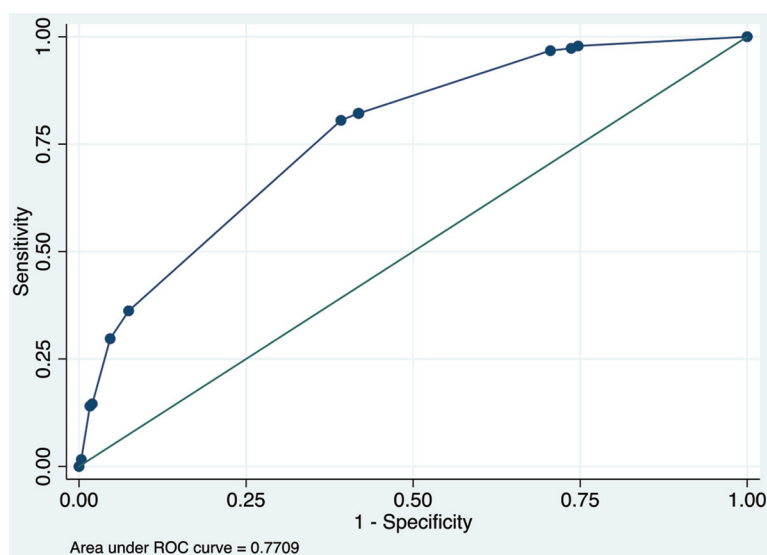
COPD, Chronic obstructive pulmonary disease; BMI, body mass index.

Appendix 2. This Summarizes the Receiver Operating Characteristic (ROC) Curves and Calibration Plots for Each Model.

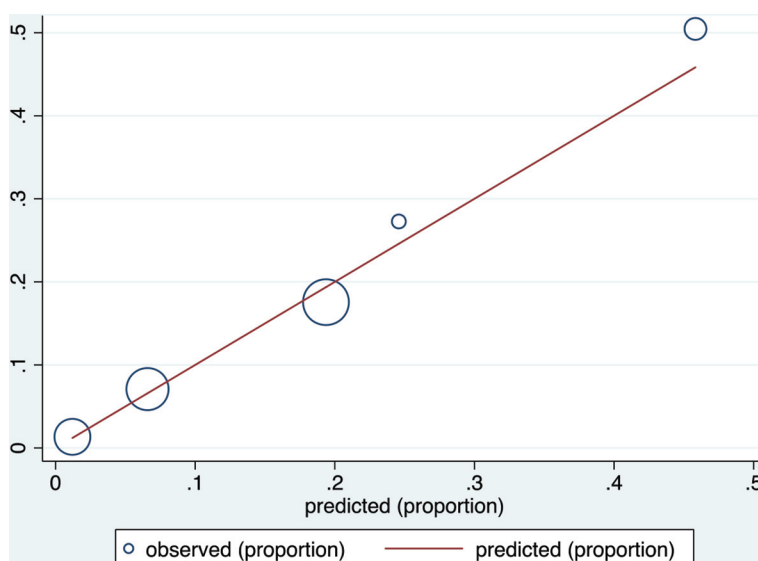
Model using clinical predictors only (COVID-NoLab)

Derivation data set

Receiver operating characteristic (ROC) curve

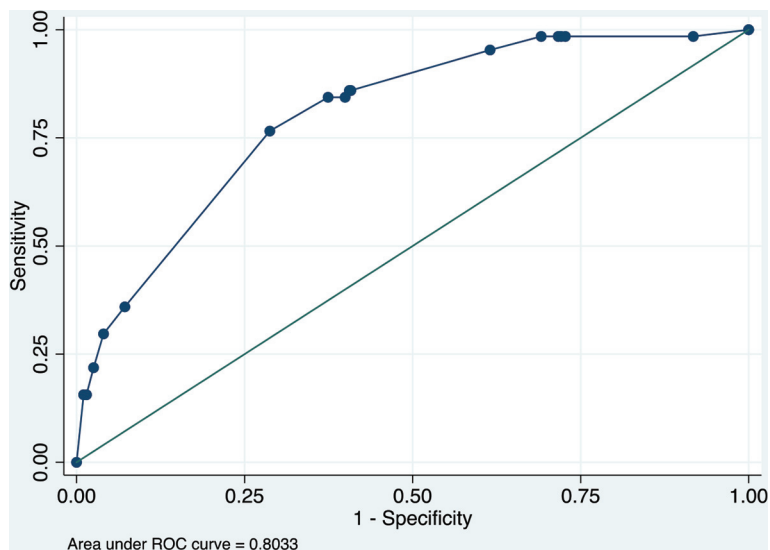


Calibration plot

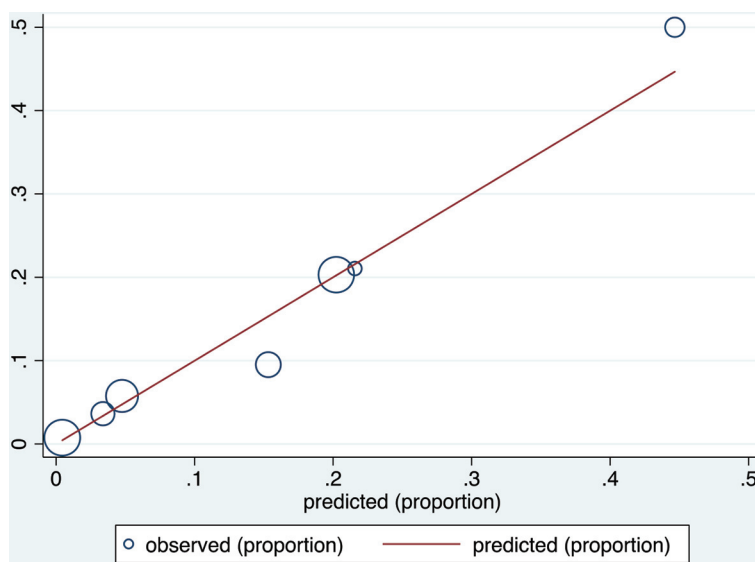


Number of observations = 1343
 Number of groups = 5
 Hosmer–Lemeshow $\chi^2(3) = 2.34$
 Prob > $\chi^2 = 0.5051$

Validation data set ROC curve



Calibration plot



Number of observations = 537

Number of groups = 7

Hosmer–Lemeshow $\chi^2(5) = 2.62$

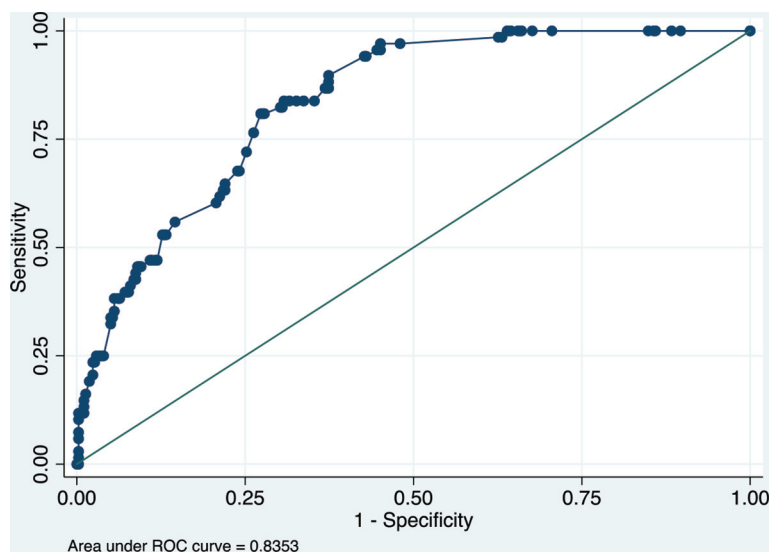
Prob > $\chi^2 = 0.7590$

Model using clinical + complete blood count

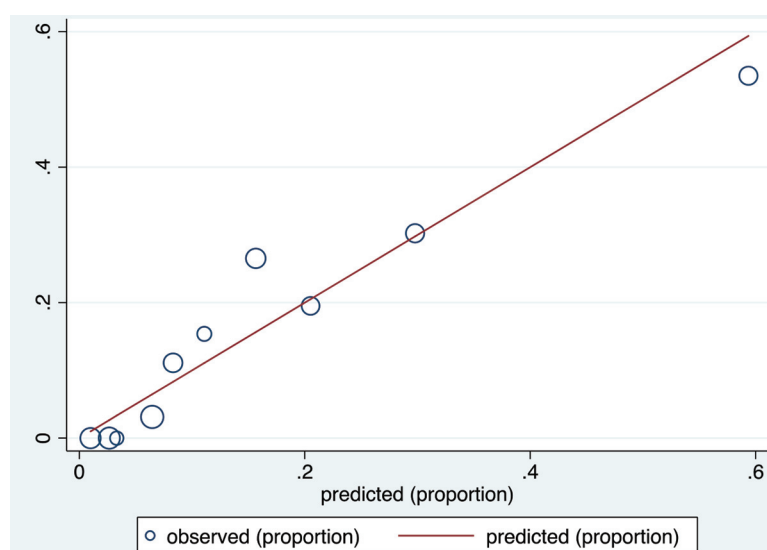
+ c-reactive protein + creatinine (COVID-SimpleLab)

Derivation data set

ROC curve

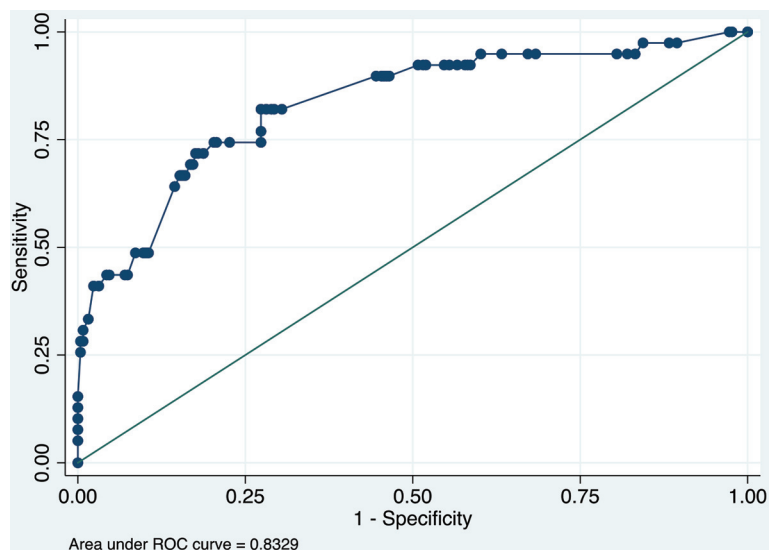


Calibration plot

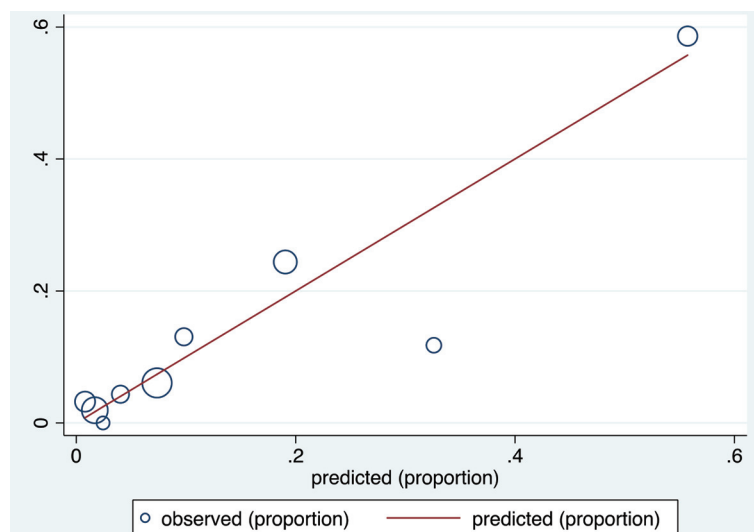


Number of observations = 445
 Number of groups = 10
 Hosmer–Lemeshow $\chi^2(8) = 10.07$
 Prob > $\chi^2 = 0.2601$

Validation data set ROC curve



Calibration plot



Number of observations = 295
 Number of groups = 9
 Hosmer–Lemeshow $\chi^2(7) = 7.29$
 Prob > $\chi^2 = 0.3998$

Appendix 3. Additional Model Using Clinical Variables and Complete Blood Count Only.

Model using clinical variables + complete blood count only Logistic regression model

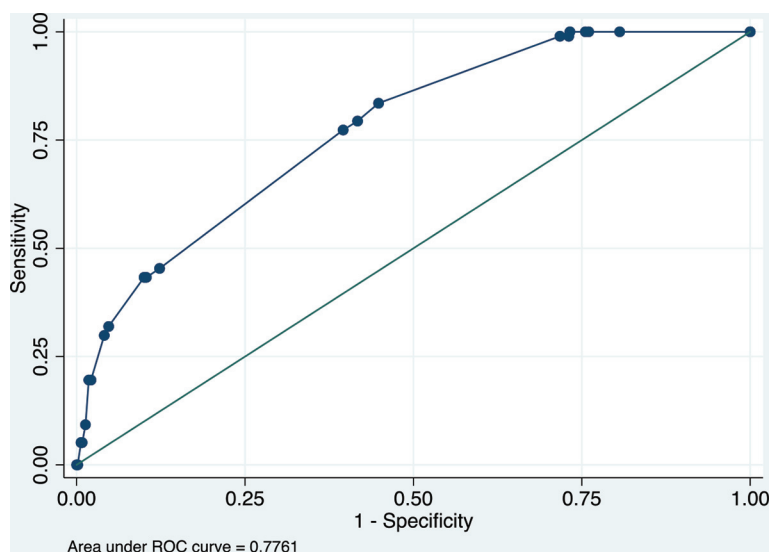
	Coef.	Std. Err.	z	P > z	Points
White blood cell count (WBC) > 10	0.586	0.279	2.100	0.036	1
Resp rate ≥ 30	1.404	0.344	4.080	0.000	2
O ₂ sat < 93 %	0.986	0.337	2.930	0.003	2
Age					
50 to 65	3.101	1.031	3.010	0.003	5
> 65	4.179	1.020	4.100	0.000	7
_cons	−5.699	1.018	−5.600	0.000	

Proposed point score and its accuracy for prediction of mortality in development and validation data sets

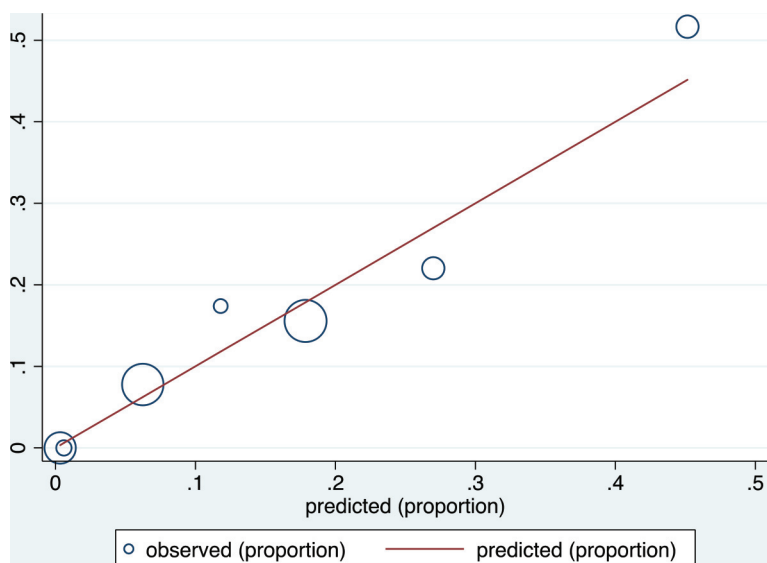
Development Risk Group (Points)	Deaths	Survivors	Total	Mortality	LR
Low: 0 to 2	0	163	163	0.0%	0.00
Mod: 3 to 6	20	192	212	10.4%	0.65
High: 7+	77	254	331	30.3%	1.90
	97	609	706		
Validation					
Risk Group (Points)	Deaths	Survivors	Total	Mortality	LR
Low: 0 to 2	5	108	113	4.4%	0.27
Mod: 3 to 6	6	113	119	5.0%	0.31
High: 7+	58	180	238	24.4%	1.87
	69	401	470		

Derivation data set

Receiver operating characteristic (ROC) curve

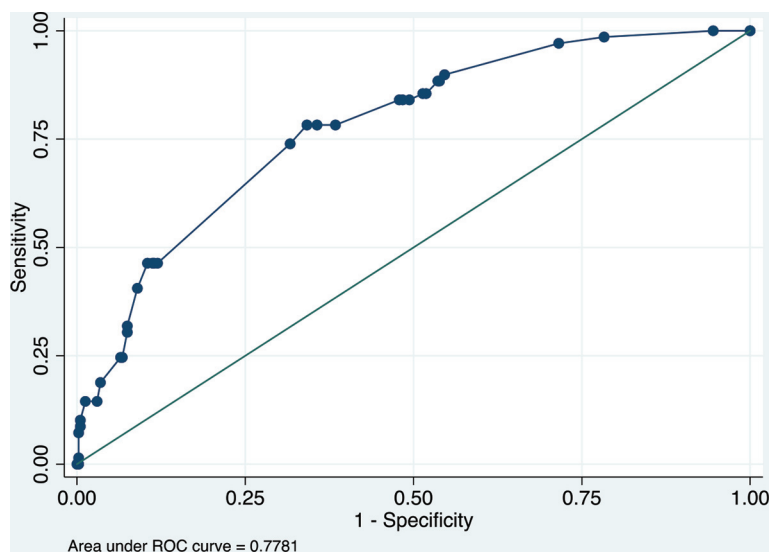


Calibration plot

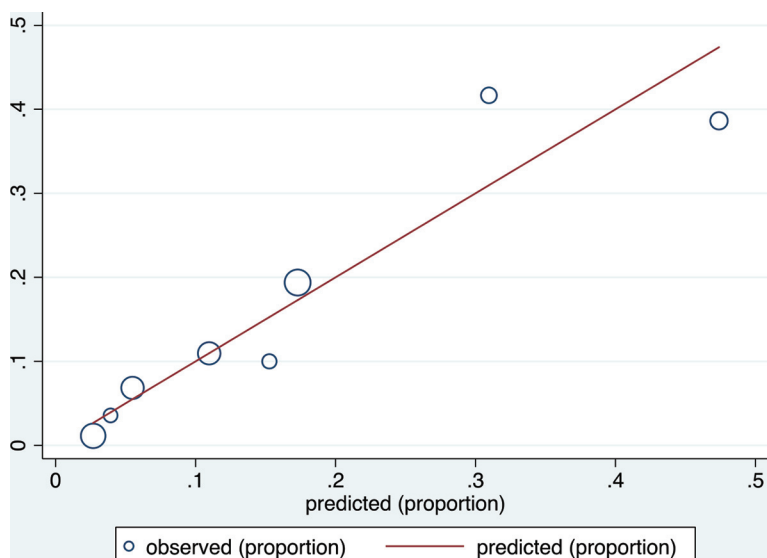


Number of observations = 706
 Number of groups = 7
 Hosmer–Lemeshow $\chi^2(5) = 4.61$
 Prob > $\chi^2 = 0.4649$

Validation data set ROC curve



Calibration plot



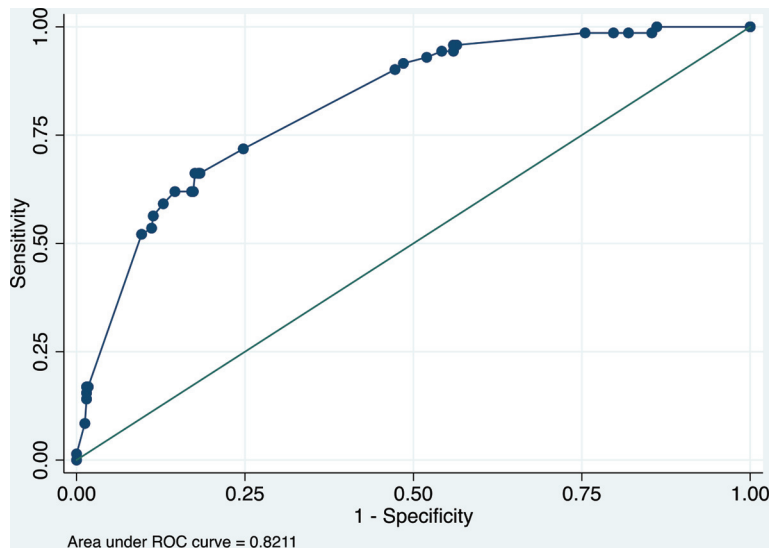
Number of observations = 470
 Number of groups = 8
 Hosmer–Lemeshow $\chi^2(6) = 5.30$
 Prob > $\chi^2 = 0.5055$

Model using clinical predictors + c-reactive protein (CRP) only

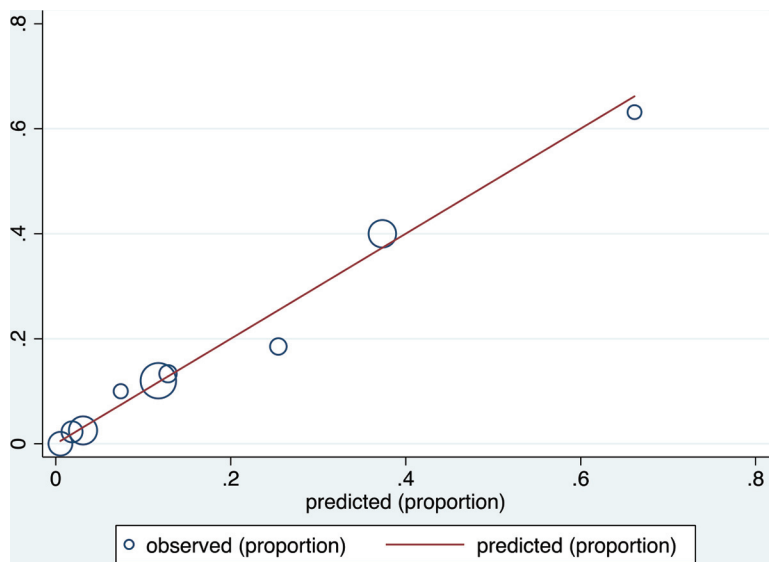
	Coef.	Std. Err.	z	P>z	Points
CRP					
> 10 to 20 mg/dL	1.522	0.309	4.920	0.000	3
> 20 mg/dL	0.974	0.539	1.810	0.071	2
Respiratory rate ≥ 30	1.246	0.420	2.970	0.003	3
Age					
50 to 65	1.798	0.786	2.290	0.022	4
> 65	3.279	0.765	4.290	0.000	7
Asthma	0.977	0.417	2.340	0.019	2
_cons	-5.236	0.782	-6.690	0.000	

Training Risk Group	Deaths	Survivors	Total	Prev	LR
0 to 4	3	176	179	1.7%	0.10
5 to 8	21	157	178	11.8%	0.76
9+	47	71	118	39.8%	3.77
	71	404	475		
Testing					
Risk Group	Deaths	Survivors	Total	Prev	LR
0 to 4	5	108	113	4.4%	0.30
5 to 8	11	115	126	8.7%	0.62
9+	26	51	77	33.8%	3.33
	42	274	316		

Derivation data set
ROC curve

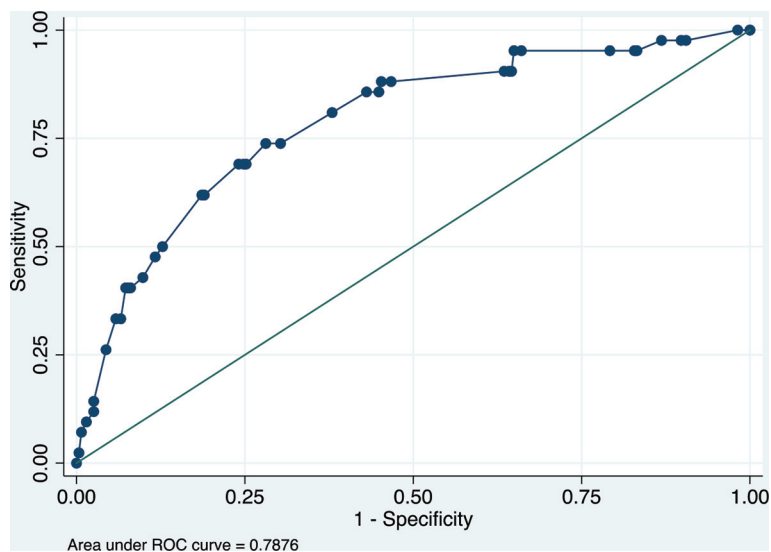


Calibration plot

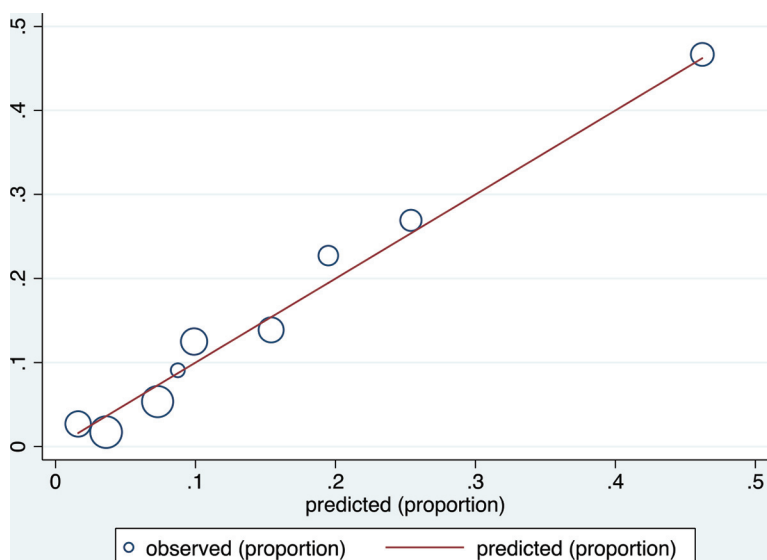


Number of observations = 475
 Number of groups = 9
 Hosmer–Lemeshow $\chi^2(7) = 1.62$
 Prob > $\chi^2 = 0.9777$

Validation data set ROC curve



Calibration plot



Number of observations = 316
 Number of groups = 9
 Hosmer–Lemeshow $\chi^2(7) = 1.73$
 Prob > $\chi^2 = 0.9733$