Managing COVID-19 with a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation

Michael P. McRae, Isaac P. Dapkins, Iman Sharif, Judd Anderson, David Fenyo, Odai Sinokrot, Stella K. Kang, Nicolaos J. Christodoulides, Deniz Vurmaz, Glennon W. Simmons, Timothy M. Alcorn, Marco J. Daoura, Stu Gisburne, David Zar, John T. McDevitt

Table of Contents

Managing COVID-19 with a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation

Michael P. McRaePhD, ; Isaac P. DapkinsMD, ; Iman SharifMD, MPH, ; Judd Anderson; David FenyoPhD, ; Odai SinokrotMD, ; Stella K. KangMD, MSc, ; Nicolaos J. ChristodoulidesPhD, ; Deniz Vurmaz; Glennon W. Simmons; Timothy M. AlcornPhD, ; Marco J. DaouraMS, ; Stu Gisburne; David ZarMS, ; John T. McDevittPhD,

Corresponding Author:
John T. McDevittPhD,
Phone: +12129989204
Email: mcdevitt@nyu.edu

Abstract

Background: The COVID-19 pandemic has resulted in significant morbidity and mortality, with large numbers of patients requiring intensive care threatening to overwhelm healthcare systems globally. There is an urgent need for a COVID-19 disease severity assessment that can assist in patient triage and resource allocation for patients at risk for severe disease.

Objective: The goal of this study was to develop, validate, and scale a clinical decision support system and mobile app to assist in COVID-19 severity assessment, management, and care.

Methods: Model training data from 701 patients with COVID-19 were collected across practices within the Family Health Centers network at New York University Langone Health. A two-tiered model was developed. Tier 1 uses easily available, non-laboratory data to help determine whether biomarker-based testing and/or hospitalization is necessary. Tier 2 predicts probability of mortality using biomarker measurements (CRP, PCT, D-dimer) and age. Both Tier 1 and Tier 2 models were validated using two external datasets from hospitals in Wuhan, China comprising 160 and 375 patients, respectively.

Results: All biomarkers were measured at significantly higher levels in patients that died vs. those that were not hospitalized or discharged (P < .001). The Tier 1 and Tier 2 internal validation had AUC (95% confidence interval) of 0.79 (0.74–0.84) and 0.95 (0.92–0.98), respectively. The Tier 1 and Tier 2 external validation had AUCs of 0.79 (0.74–0.84) and 0.97 (0.95–0.99), respectively.

Conclusions: Our results demonstrate validity of the clinical decision support system and mobile app, which are now ready to assist healthcare providers in making evidence-based decisions in managing COVID-19 patient care. The deployment of these new capabilities has potential for immediate impact in community clinics, sites whereby application of such tools could lead to improvements in patient outcomes and cost containment.

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

   Only make the preprint title and abstract visible.

   No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Managing COVID-19 with a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation

Authors

1. Michael P McRae, PhD
    Department of Biomaterials, Bioengineering Institute, New York University, New York, NY, USA
    michael.mcrae@nyu.edu
2. Isaac P Dapkins, MD
    Family Health Centers at NYU Langone, Department of Population Health and Internal Medicine, New York University School of Medicine, New York, NY, USA
    Isaac.Dapkins@nyulangone.org
3. Iman Sharif, MD, MPH
    Family Health Centers at NYU Langone, Departments of Pediatrics and Population Health, New York University School of Medicine, New York, NY, USA
    Iman.Sharif@nyulangone.org
4. Judd Anderman
    Family Health Centers at NYU Langone
    Judd.Anderman@nyulangone.org
5. David Fenyo, PhD
    Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA
    david.fenyo@nyulangone.org
6. Odai Sinokrot, MD
    Department of Medicine, New York University School of Medicine, New York, NY, USA
    odai.sinokrot@nyulangone.org
7. Stella K Kang, MD, MSc
    Departments of Radiology, Population Health New York University School of Medicine, New York, NY, USA
    stella.kang@nyulangone.org
8. Nicolaos J Christodoulides, PhD
    Department of Biomaterials, Bioengineering Institute, New York University, New York, NY, USA
    nicolaoschristo19@gmail.com
9. Deniz Vurmaz
    Department of Chemical and Biomolecular Engineering, NYU Tandon School of Engineering, New York University, New York, NY, USA
    dv832@nyu.edu
10. Glennon W Simmons
    Department of Biomaterials, Bioengineering Institute, New York University, New York, NY, USA
    glennon.simmons@nyu.edu
11. Timothy M Alcorn, PhD
    Latham BioPharm Group, Cambridge, MA, USA
    talcorn@lathambiopharm.com

12. Marco J Daoura, MS
    OraLiva, Inc., Naples, FL, USA
    marco@marcowerx.com
13. Stu Gisburne
    OraLiva, Inc., Naples, FL, USA
    stuville@me.com
14. David Zar, MS
    OraLiva, Inc., Naples, FL, USA
    dzar@wustl.edu
15. John T McDevitt, PhD
    Department of Biomaterials, Bioengineering Institute, New York University, New York, NY,
    USA
    mcdevitt@nyu.edu

Corresponding Author: John T McDevitt, PhD, Department of Biomaterials, Bioengineering
Institute, New York University, 433 First Avenue, Room 820, New York, NY 10010-4086, USA.
Email: mcdevitt@nyu.edu. Phone: 212-998-9204.

Abstract

Background: The COVID-19 pandemic has resulted in significant morbidity and mortality, with
large numbers of patients requiring intensive care threatening to overwhelm healthcare systems
globally. There is an urgent need for a COVID-19 disease severity assessment that can assist in
patient triage and resource allocation for patients at risk for severe disease.

Objective: The goal of this study was to develop, validate, and scale a clinical decision support
system and mobile app to assist in COVID-19 severity assessment, management, and care.

Methods: Model training data from 701 patients with COVID-19 were collected across practices
within the Family Health Centers network at New York University Langone Health. A two-tiered
model was developed. Tier 1 uses easily available, non-laboratory data to help determine whether
biomarker-based testing and/or hospitalization is necessary. Tier 2 predicts probability of mortality
using biomarker measurements (CRP, PCT, D-dimer) and age. Both Tier 1 and Tier 2 models were
validated using two external datasets from hospitals in Wuhan, China comprising 160 and 375
patients, respectively.

Results: All biomarkers were measured at significantly higher levels in patients that died vs. those
that were not hospitalized or discharged ($P < .001$). The Tier 1 and Tier 2 internal validation had
AUC (95% confidence interval) of 0.79 (0.74–0.84) and 0.95 (0.92–0.98), respectively. The Tier 1
and Tier 2 external validation had AUCs of 0.79 (0.74–0.84) and 0.97 (0.95–0.99), respectively.

Conclusions: Our results demonstrate validity of the clinical decision support system and mobile
app, which are now ready to assist healthcare providers in making evidence-based decisions in
managing COVID-19 patient care. The deployment of these new capabilities has potential for
immediate impact in community clinics, sites whereby application of such tools could lead to
improvements in patient outcomes and cost containment.

Keywords: COVID-19; coronavirus; clinical decision support system; point of care; mobile app;
disease severity; biomarkers; artificial intelligence; app; family health center

Introduction

The 2019-20 coronavirus disease (COVID-19) was first reported in Wuhan, Hubei, China, in
December 2019 [1], and it was declared a pandemic by the World Health Organization (WHO) [2]
soon thereafter. As of June 15, 2020, about 8 million cases have been confirmed with approximately

435,000 deaths from the disease globally [3]. The COVID-19 crisis has exposed critical gaps in diagnostic testing and population-level surveillance [4]. With hospitalization rates of 20-31% and intensive care unit (ICU) admission rates of 5-12% [5], surges of patients requiring care have overwhelmed local healthcare systems and depleted reserves of medical resources.

Physicians are tasked with evaluating large amounts of rapidly changing patient data and making critical decisions in a short amount of time. Well-designed clinical decision support systems (CDSS) deliver pertinent knowledge and individualized patient information to healthcare providers to enhance medical decisions [6]. Such systems may rely on surveys of similar cases, while others may use a "black box" approach [7]. Traditional scores like SOFA [8-10] and APACHE-2 [11, 12] are commonly used in hospitals for determining disease severity and mortality, whereas clinical decision management systems like electronic ICU (eICU) allow for systematic collection of comprehensive data [13]. However, CDSS that use conventional variables, such as demographics, symptoms, and medical history, often do not reach full diagnostic potential [14]. There is a compelling need for a COVID-19 disease severity assessment to help prioritize care for patients at elevated risk of mortality and manage low risk patients in outpatient settings or at home through self-quarantine.

Several scoring systems for COVID-19 severity have been developed or adapted from existing tools, such as the Brescia-COVID Respiratory Severity Scale [15], African Federation for Emergency Medicine COVID-19 Severity Scoring Tool [16], Berlin Criteria for Acute Respiratory Distress Syndrome [17, 18], and Epic Deterioration Index [19]. However, these tools have either (a) not yet been externally validated in peer-reviewed publications or (b) developed specifically for COVID-19 patient populations. Recently, we developed an integrated point-of-care COVID-19 Severity Score and CDSS which combines multiplex biomarker measurements and risk factors in a statistical learning algorithm to predict mortality with excellent diagnostic accuracy [20]. The COVID-19 Severity Score was trained and evaluated using data from 160 hospitalized COVID-19 patients from Wuhan, China. The COVID-19 Severity Scores were significantly higher for patients who died as compared with patients who were discharged with median (interquartile range [IQR]) scores of 59 (40–83) and 9 (6–17), respectively, and area under the curve (AUC) of 0.94 (95% confidence interval [CI] 0.89–0.99).

The COVID-19 condition has caused and continues to cause significant morbidity and mortality globally. A validated tool to assess and quantify viral sepsis severity and patient mortality risk would address the urgent need for disease severity categorization. Towards the goal of improving prognostic judgement and outcomes, we have assembled a multidisciplinary team representing stakeholders from technology, machine learning, engineering, primary care, and *in vitro* diagnostic testing to develop a COVID-19 disease severity test. The unfolding novel COVID-19 pandemic has greatly illuminated the important role of community health centers in providing safe and effective patient care. The Family Health Centers (FHC) at New York University (NYU) Langone is a large Federally Qualified Health Center providing comprehensive primary and preventive healthcare to a diverse population of patients across the New York City metropolitan area and is well-positioned to improve survival by fast-tracking hospitalization of patients at high risk of severe disease. This current study describes a clinical decision support tool for COVID-19 disease severity developed using recent data from the FHC and externally validated using data from two recent studies from hospitals in Wuhan, China. We describe a practical and efficient tiered approach which involves a model with non-laboratory inputs (Tier 1), a model with biomarkers commonly measured in ambulatory settings (Tier 2), and a mobile app to deliver and scale these tools. The deployment of these new capabilities has potential for immediate clinical impact in community clinics whereby such tools could lead to improvements in patient outcomes and prognostic judgment.

Methods

Patient Data

Data from 701 patients with COVID-19 were collected across 9 clinics and hospitals within the FHC network at NYU Langone, one of the largest Federally Qualified Health Center networks in the US. All patients had detectable SARS-CoV-2 infection by polymerase chain reaction (PCR) testing. The following outcomes were recorded: not hospitalized, discharged, ventilated, and deceased. Validation data for the Tier 1 Outpatient Model were derived from a study of 160 hospitalized COVID-19 patients from Zhongnan Hospital of Wuhan University. Validation data for the Tier 2 Biomarker Model were derived from a study of 375 hospitalized COVID-19 patients from Tongji Hospital in Wuhan, China [21].

Clinical Decision Support Tool

This current study describes the development of a 2-tiered CDSS for the assessment of COVID-19 disease severity, using similar methods as described previously [20, 22]. The Tier 1 Outpatient Model uses non-laboratory data that are readily available prior to laboratory measurements and is intended to help determine whether Tier 2 biomarker-based testing and/or hospitalization is necessary. Here, a lasso logistic regression model was trained to distinguish between patients that were not hospitalized or were hospitalized and discharged home without need for ventilation versus patients that were ventilated or died. Patients who were still hospitalized when the data were compiled were excluded. The following predictors were considered in model training: age, gender, body mass index, systolic blood pressure, temperature, symptoms (cough, fever, or shortness of breath), known cardiovascular comorbidities (patient problem list includes one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial infarction, peripheral vascular disease, and hypertension), pulmonary comorbidities (asthma and chronic obstructive pulmonary disease), and diabetes.

The Tier 2 Biomarker Model predicts disease severity using biomarker measurements and patient characteristics. A lasso logistic regression model was trained to distinguish patients that died versus patients that were either never hospitalized or discharged home. Patients who were ventilated and/or still hospitalized when the data were compiled were excluded. The following predictors were considered in model training: age, gender, comorbidities, C-reactive protein (CRP), cardiac troponin I (cTnI), D-dimer, procalcitonin (PCT), and N-terminal pro–B-type natriuretic peptide (NT-proBNP). Predictors that were not relevant to the model (i.e., coefficients equal to zero) were removed. Laboratory measurements across all time points were log-transformed. Patients with no measurements for the aforementioned biomarkers were excluded. Biomarker values below the limits of detection were set to the minimum measured value divided by the square root of 2.

Model Development and Statistical Analysis

Both Tier 1 and 2 models were developed using the same procedure. All continuous predictors were standardized with mean of zero and variance of one. Missing data were imputed using the multivariate imputation by chained equations algorithm in statistical software R [23]. Predictive mean matching and logistic regression imputation models were used to generate 10 imputations for continuous and categorical predictors, respectively. Samples in the training and test sets were partitioned using stratified 5-fold cross-validation to preserve the relative proportions of outcomes in each fold. Model training and selection were performed on each of the 10 imputation datasets for 10 Monte Carlo repetitions and optimized for the penalty parameter corresponding to one standard error above the minimum deviance for additional shrinkage. After initial training, only predictors with nonzero regression coefficients were retained, and the model was retrained with a reduced number of predictors. The training process was repeated until all predictors yielded nonzero coefficients. Model

performance was documented in terms of mean (95% CI) of AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Median (IQR) cross-validated COVID-19 Scores were compared across disease outcomes. The COVID-19 Scores for both models and biomarker measurements were compared using Wilcoxon rank sum test. Normally distributed predictors were compared using an independent *t*-test. Proportions were compared using the Chi-squared test [24, 25]. Two-sided tests were considered statistically significant for *P* < .05.

External Validation

We externally validated the Tier 1 Outpatient Model using data from a study of 160 hospitalized COVID-19 patients from Zhongnan Hospital of Wuhan University. Only patients with complete information (age, systolic blood pressure, gender, diabetes, and cardiovascular comorbidities) were included. Model performance was documented in terms of AUC, sensitivity, specificity, PPV, and NPV. Results were presented in a scatter/box plot of COVID-19 Outpatient Scores on patients that were discharged and those that died.

Similarly, we externally validated the Tier 2 Biomarker Model using data from a study of 375 hospitalized COVID-19 patients from Tongji Hospital in Wuhan, China collected between January 10 and February 18, 2020 [21]. While most patients had multiple lab measurements over time, the first available lab value for each biomarker was used to validate the model to maximize lead time. Patients with one or more missing predictor values were excluded. Model performance was documented in terms of AUC, sensitivity, specificity, PPV, and NPV. Results were presented in a scatter/box plot of COVID-19 Biomarker Scores on patients that were discharged and those that died.

To demonstrate how the COVID-19 Biomarker Score could be used to track changes in disease severity over time, the model was evaluated on time series biomarker data. Since lab measurements were reported asynchronously, the model was reevaluated every time a new biomarker measurement became available. Time series plots of the COVID-19 Biomarker Score were generated for each patient.

Results

This current study describes the development of a 2-tiered CDSS to assess COVID-19 disease severity using similar methods as described previously [20, 22]. The Tier 1 Outpatient Model uses non-laboratory data that are readily available prior to laboratory measurements and is intended to help determine whether Tier 2 biomarker-based testing and/or hospitalization are warranted. The Tier 2 Biomarker Model predicts disease severity using biomarker measurements and patient characteristics.

The CDSS and mobile app are designed to support decisions made in multiple settings, including (1) home care, (2) primary care or urgent care clinics, (3) emergency departments, and (4) hospital and intensive care (Figure 1). The process starts with symptomatic patients who are positive or presumably positive for COVID-19 and seeking care at a family health center or emergency room. In the family health center, decisions are made in two key stages, or tiers. The Tier 1 algorithm is intended for individuals in an outpatient setting where laboratory data are not yet readily available, (e.g., age, gender, blood pressure, and comorbidities). Patients with a low COVID-19 Outpatient Score may be managed through a home or telemedicine setting, while patients with a high COVID-19 Outpatient Score are referred for a blood draw and Tier 2 biomarker-based test. The Tier 2 algorithm, which is directly related to mortality risk, predicts disease severity using biomarker measurements and age. Patients with a low COVID-19 Biomarker Score are expected to be managed in a low-to-moderate risk group (e.g., 5 day Telehealth follow-up), while patients with a high COVID-19 Biomarker Score are expected to be hospitalized in most cases or managed in a high risk

group (e.g., 24-48 hour follow-up). Providers encountering clinically evident severe cases, as in urgent care or emergency departments, may choose to bypass the Tier 1 Outpatient Score and perform biomarker testing and Tier 2 triage on all COVID-19 patients. Lastly, in the hospital setting, patients are serially monitored for their COVID-19 Biomarker Score. Such personalized time series information directly related to mortality risk has strong potential to optimize therapy, improve patient care, and ultimately save lives. For both algorithms, we selected cutoffs that balanced sensitivity and specificity; however, these algorithms can be easily tuned for high sensitivity or high specificity by adjusting the weighting or relative importance of sensitivity and specificity in clinical practice.
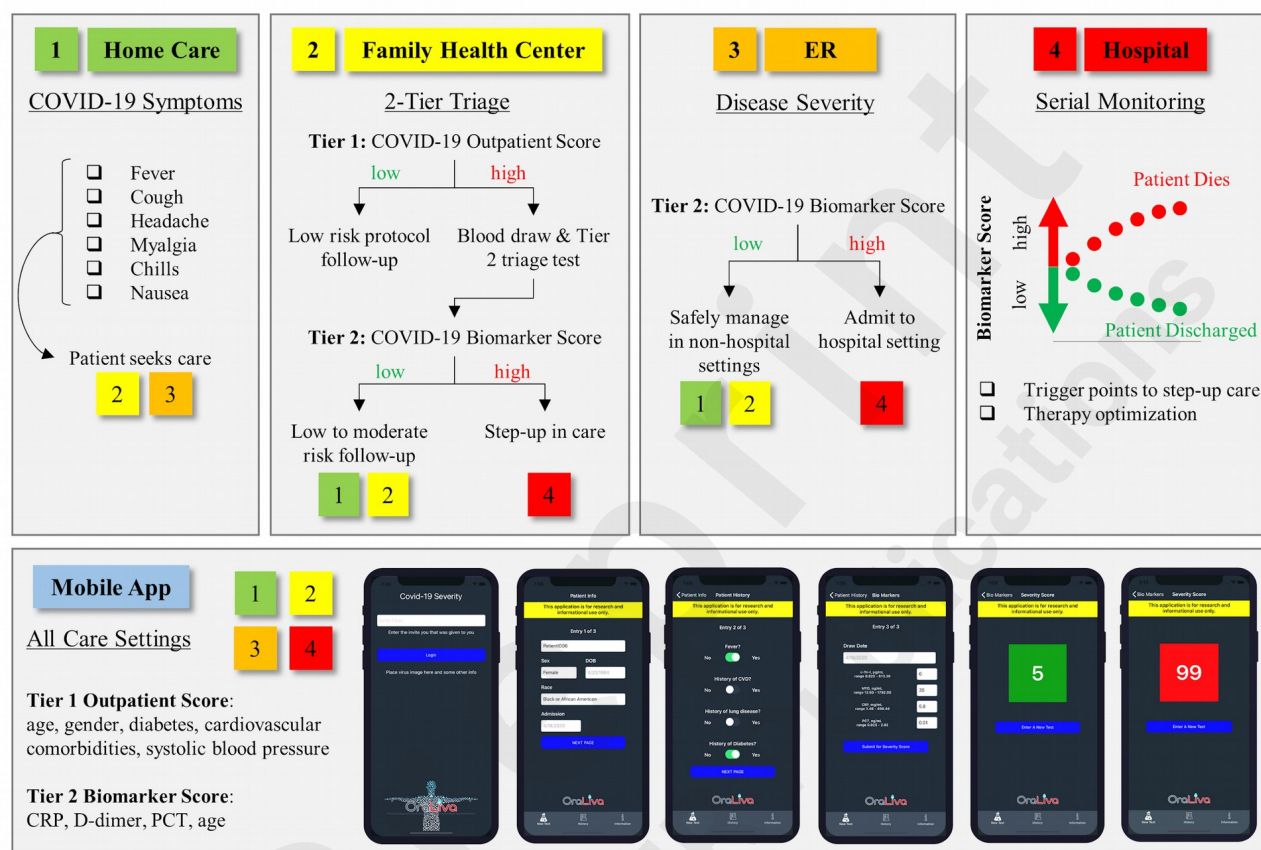

Figure 1. Clinical decision support system and mobile app for managing COVID-19 care.

Out of the 701 patients with detectable COVID-19 infection cared for by one of the 9 clinics within the FHC network, 402 were not hospitalized, 185 were hospitalized and discharged, 19 were ventilated, and 95 died (Table 1). Ventilated and deceased patients were older than those that were not hospitalized or discharged (P = .03 and <.001, respectively). Males accounted for 74% and 63% of patients who were ventilated and deceased vs. 46% for patients with less severe disease (i.e., not hospitalized or discharged) (P = .02 and .002, respectively). Diabetes was also a statistically significant factor with 47% and 55% in the ventilated and deceased groups vs. 25% in the non-hospitalized and discharged groups (P = .03 and <.001, respectively). Likewise, 53% of ventilated patients (P = .04) and 68% of deceased patients (P < .001) had one or more cardiovascular comorbidities vs. 31% for the less severe disease categories, with hypertension being the most common. Interestingly, systolic blood pressure was significantly higher for patients who were not hospitalized vs. those that were discharged (P = .004), and patients who died had abnormally low blood pressure relative to less severe disease (P <.001). All biomarkers (cTnI, CRP, PCT, D-dimer, and NT-proBNP) were measured at significantly higher levels in patients that died vs. those that were not hospitalized or discharged (P < .001).

Table 1. Characteristics of patients included in model training. Data are represented as n (%), mean ± standard deviation, or median (IQR).

	Not hospitalized	Discharged	Ventilated	Deceased
	n = 402	n = 185	n = 19	n = 95
Age, years	48 ± 17	50 ± 17	58 ± 20	67 ± 14
Gender	182 (45)	89 (48)	14 (74)	60 (63)
BMI[a], kg/m²	25 ± 4	28 ± 6	29 ± 5	25 ± 6
Systolic BP[b], mmHg	132 ± 14	123 ± 19	126 ± 20	94 ± 40
Diastolic BP[b], mmHg	82 ± 8	71 ± 11	70 ± 12	54 ± 26
Temperature	99 ± 1	98 ± 5	99 ± 1	100 ± 2
Pulse, beats per min.	90 ± 18	84 ± 14	93 ± 14	74 ± 54
Asthma	44 (11)	12 (6)	3 (16)	6 (6)
COPD[c]	60 (15)	17 (9)	3 (16)	15 (16)
Cancer	13 (3)	5 (3)	2 (11)	14 (15)
Cardiovascular comorbidities[d]	120 (30)	61 (33)	10 (53)	65 (68)
Diabetes	96 (24)	53 (29)	9 (47)	52 (55)
HIV/AIDS	3 (1)	2 (1)	0 (0)	3 (3)
Liver disease	11 (3)	10 (5)	2 (11)	4 (4)
Renal disease	20 (5)	17 (9)	3 (16)	21 (22)
cTnI, pg/mL	7.07 (7.07–7.07)	7.07 (7.07–7.07)	20.00 (7.07–63.75)	73.50 (7.07–712.00)
CRP, mg/L	51.40 (16.55–101.35)	67.90 (17.95–121.50)	37.30 (27.30–139.72)	176.00 (115.00–287.00)
PCT, ng/mL	0.12 (0.06–0.36)	0.10 (0.05–0.31)	0.69 (0.07–1.91)	1.61 (0.35–8.31)
D-Dimer, μg/mL	0.39 (0.20–0.71)	0.27 (0.18–0.56)	0.86 (0.50–3.02)	1.58 (0.72–5.35)
NT-proBNP, pg/mL	93.00 (36.50–375.25)	88.00 (28.50–298.00)	217.00 (78.00–394.25)	937.00 (160.25–5728.50)

[a]BMI: body mass index
[b]BP: blood pressure
[c]COPD: chronic obstructive pulmonary disease
[d]Cardiovascular comorbidities: one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial infarction, peripheral vascular disease, and hypertension

Tier 1 Outpatient Model

The Tier 1 Outpatient Model for COVID-19 disease severity was developed and internally validated using data from the FHCs at NYU Langone (Figure 2). The model retained the following predictors: age, gender, systolic blood pressure, cardiovascular comorbidities (one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial infarction, peripheral vascular disease, and hypertension), and diabetes. Median COVID-19 Outpatient Scores were 11, 13, 20, and 27 for not hospitalized, discharged, ventilated, and deceased patients, respectively. The model's AUC (95% CI) was 0.79 (0.74–0.84) at the optimal cutoff COVID-19 Outpatient Score of 18 (Table 2). Median scores (Figure 2) had statistically significant differences for comparisons between all patient groups except not hospitalized vs. discharged (P = .18).
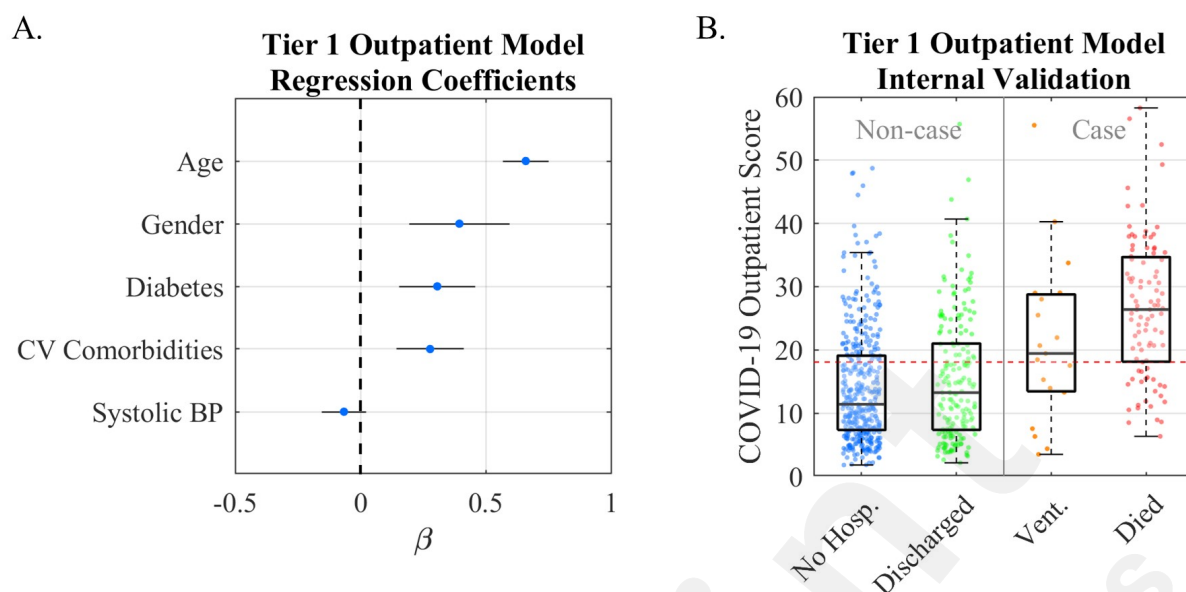
Figure 2. Tier 1 Outpatient Model results. Lasso logistic regression coefficients (A) reveals the relative predictor importance in generating the score. The box/scatter plot (B) from internal validation shows Tier 1 Outpatient Scores for the four outcomes. A cutoff score of 18 (red dotted line) balances sensitivity and specificity for "Non-case" vs. "Case" patients (gray line) (No Hosp. = patients who were not hospitalized, Vent. = patients who were ventilated, CV comorbidities = cardiovascular comorbid conditions).

Tier 2 Biomarker Model

The Tier 2 Biomarker Model for COVID-19 disease severity was developed and internally validated using data from the FHCs at NYU Langone (Figure 3). Patients who were ventilated (n=19) and still hospitalized (n=19) were excluded. Patients with fewer than one biomarker measurement were excluded (n=190 not hospitalized, n=64 discharged, n=1 deceased). The remaining 427 patients with one or more biomarker measurement were included in the analysis (n=212 not hospitalized, n=121 discharged, n=94 deceased). The model retained the following predictors after shrinkage and selection: age, D-dimer, PCT, and CRP. Median COVID-19 Outpatient Scores were 5, 5, and 64 for not hospitalized, discharged, and deceased patients, respectively. The model's AUC (95% CI) was 0.95 (0.92–0.98) at the optimal cutoff COVID-19 Outpatient Score of 27 (Table 2). Median COVID-19 Outpatient Scores (Figure 3) had statistically significant differences for comparisons between not hospitalized vs. died (P <.001) and discharged vs. died (P <.001).
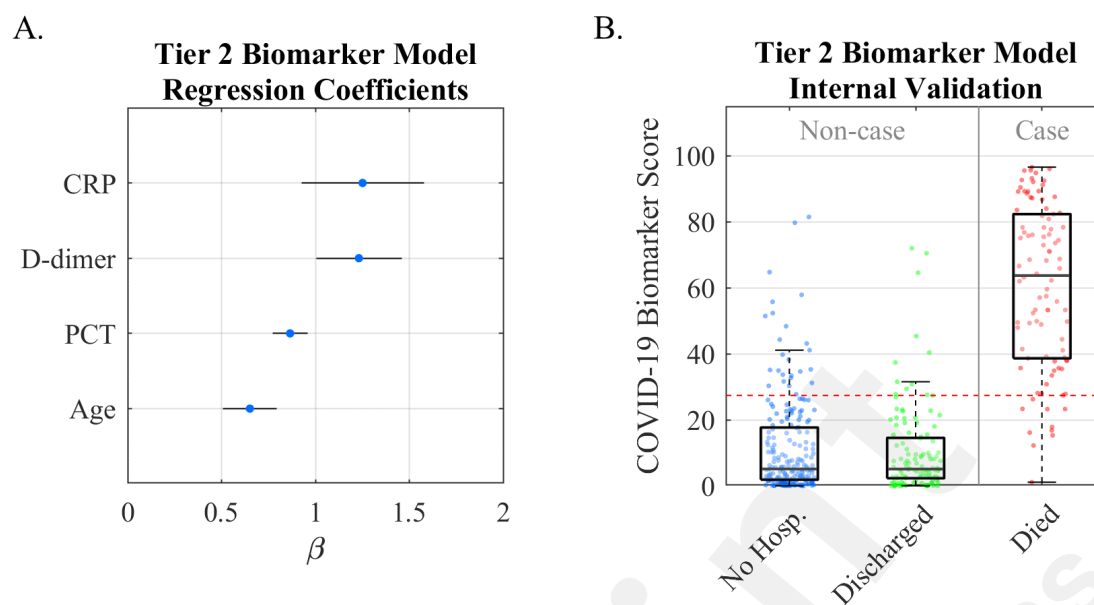
Figure 3. Tier 2 Biomarker Model results. Lasso logistic regression coefficients (A) reveals relative predictor importance in generating the score. The box/scatter plot (B) from internal validation shows Tier 2 Biomarker Scores for the three patient outcomes. A cutoff score of 27 (horizontal red dotted line) balances sensitivity and specificity for "Non-case" vs. "Case" patients (vertical gray line) (No Hosp. = patients who were not hospitalized).

Table 2. Internal validation performance in terms of AUC, sensitivity, specificity, PPV, and NPV (95% CI) from 5-fold cross-validation. Tier 1 and 2 models were trained and tested using data from FHCs at NYU.

	Tier 1 Outpatient Model	Tier 2 Biomarker Model
AUC	0.79 (0.74–0.84)	0.95 (0.92–0.98)
Sensitivity	0.73 (0.69–0.76)	0.89 (0.86–0.92)
Specificity	0.73 (0.69–0.76)	0.89 (0.86–0.92)
PPV	0.34 (0.30–0.38)	0.70 (0.65–0.74)
NPV	0.93 (0.91–0.95)	0.97 (0.94–0.98)

External Validation

We externally validated the Tier 1 Outpatient Model using data from a study of 160 hospitalized COVID-19 patients with hypertension from Zhongnan Hospital of Wuhan University, Wuhan, China [26]. Out of the 160 patients in the study, 4 were missing one or more predictors and were excluded from the analysis. The COVID-19 Biomarker Scores were evaluated for 115 patients who were discharged and 41 patients who died (Figure 4A). The median (IQR) COVID-19 Biomarker Scores were 27.9 (22.0–36.4) for patients that were discharged and 39.7 (34.2–47.4) for patients that died. The external validation diagnostic performance was determined using a cutoff score of 34 (Table 3).

We externally validated the Tier 2 Biomarker Model using data from a study of 375 hospitalized COVID-19 patients from Tongji Hospital in Wuhan, China collected between January 10 and February 18, 2020 [21]. In order to maximize potential lead time, the first available laboratory measurements during hospitalization were used to generate cross-sectional COVID-19 Biomarker Scores, representing the first in a series of measurements collected for hospital stays lasting a median

(IQR) of 12.5 (8–17.5) days prior to the outcomes (discharged or deceased). Out of the 375 patients in the study, 133 were missing one or more lab value and excluded from the analysis. The COVID-19 Biomarker Scores were evaluated for 112 patients who were discharged and 130 patients who died (Figure 4B). The median (IQR) COVID-19 Biomarker Scores were 1.6 (0.5–6.2) for patients that were discharged and 59.1 (36.6–78.9) for patients that died. The external validation diagnostic performance was determined using a cutoff score of 19 (Table 3).
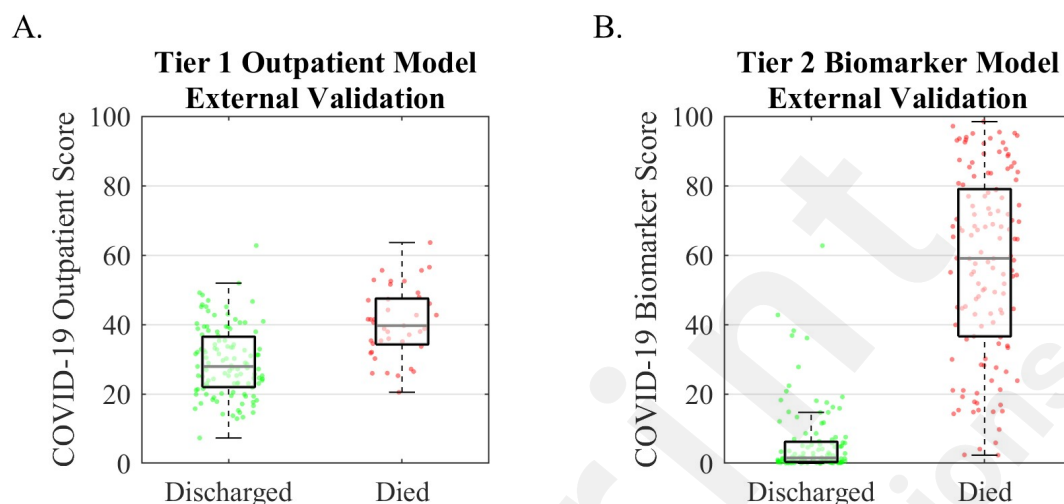

Figure 4. External validation results. The Tier 1 Outpatient Model was evaluated on data from COVID-19 patients at Zhongnan Hospital of Wuhan University [26]. The Tier 2 Biomarker Model was evaluated on data from COVID-19 patients at Tongji Hospital [21].

Table 3. External validation performance in terms of AUC, sensitivity, specificity, PPV, and NPV (95% CI). The Tier 1 Outpatient Model was evaluated on Zhongnan Hospital dataset [26]. The Tier 2 model evaluated on Tongji Hospital dataset [21].

	Tier 1 Outpatient Model	Tier 2 Biomarker Model
AUC	0.79 (0.70–0.88)	0.97 (0.95–0.99)
Sensitivity	0.76 (0.68–0.82)	0.89 (0.84–0.93)
Specificity	0.73 (0.65–0.80)	0.93 (0.89–0.96)
PPV	0.50 (0.42–0.58)	0.94 (0.90–0.96)
NPV	0.89 (0.83–0.94)	0.88 (0.83–0.92)

We also evaluated the COVID-19 Biomarker Scores for patients over time using longitudinal biomarker measurement data from individual patients in the external validation set (Figure 5). When comparing the first scores after admission vs. the final measurements prior to discharge/death, patients who recovered and were discharged had an average decrease in score of 4.7 while patients who died had an average increase in score of 11.2.
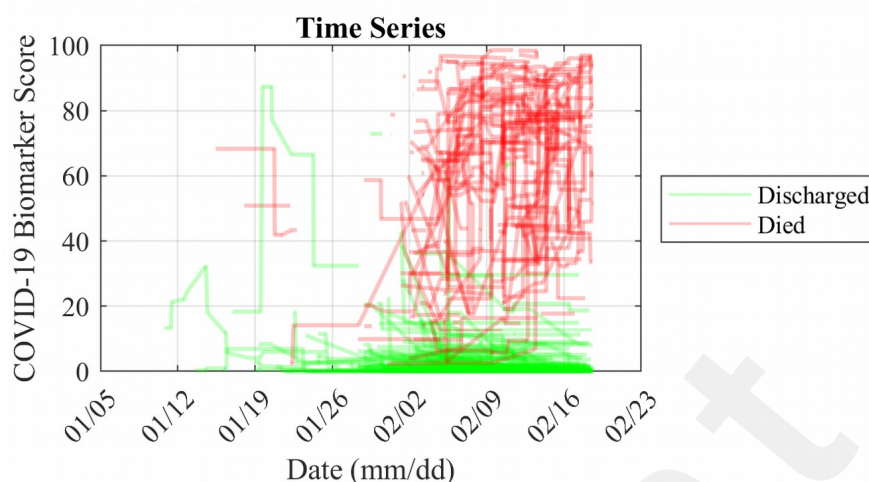
Figure 5. Spaghetti plot of longitudinal COVID-19 Biomarker scores for patients in the external validation set from Tongji Hospital [21] between January 10 and February 18, 2020. These data represent individual patients' scores over a median (IQR) of 12.5 (8–17.5) days between admission and outcomes of discharged or deceased. The first scores available after admission were significantly higher in those that died vs. those that were discharged (AUC 0.97, cutoff score of 19), and over time patients who were discharged had an average decrease in score (-4.7) while those that died had an average increase in score (+11.2).

Discussion

As the COVID-19 pandemic continues to create surges and resurgences without an effective vaccine, the goal of this multidisciplinary team was to develop a triage and prognostication tool that strengthens community-level testing and disease severity monitoring. A CDSS and mobile app for COVID-19 disease severity have been designed, developed, and validated using data from 1236 patients with COVID-19 across numerous clinics and hospitals in the coronavirus disease epicenters of Wuhan, China and New York, USA. These clinically validated tools have potential to assist healthcare providers in making evidence-based decisions in managing COVID-19 patient care. The significance of this work is realized by the algorithms developed and validated here, which are accurate, interpretable, generalizable.

With respect to accuracy, both Tier 1 and Tier 2 models were effective at discriminating disease outcomes with statistically significant differences between the most relevant patient groups (AUCs of 0.79 and 0.97 for Tier 1 and Tier 2 external validation, respectively). As expected, the Tier 1 Outpatient Model diagnostic accuracy in terms of AUC was lower than Tier 2 Biomarker Model, which demonstrates the importance of biomarker data in determining disease severity. Accurately identifying patients with elevated risk for developing severe COVID-19 complications can empower healthcare providers to save lives by prioritizing critical care, medical resources, and therapies.

Another strength of this approach is the interpretability of the models. While many predictive tools rely on 'black box' methods in which algorithmic decisions and the logic supporting those decisions are uninterpretable, the lasso logistic regression method is transparent through its coefficients (i.e., log odds) and probabilistic output. The Tier 1 Outpatient Score is the probability of severe disease (ventilation or death) based on the predictors (age, gender, diabetes, cardiovascular comorbidities, and systolic blood pressure). Likewise, the Tier 2 Biomarker Score is the probability of mortality based on CRP, D-dimer, PCT, age. Predictive models such as these are more likely to be adopted for clinical applications which value transparency and interpretability.

One of the most clinically relevant features of this new CDSS is the capacity to monitor individual patients over time. The use of this *precision diagnostic* approach allows for the amplification of early signs of disease that can be achieved by focusing on time-course changes of biomarker signatures that are referenced not to population metrics, but rather back to the individual patient. As an example, the use of time course changes in individual biomarker fingerprints has been explored previously in the study of early detection in ovarian cancer [27]. Studies demonstrated that CA-125 by itself for a single time point was a poor diagnostic marker due to overlapping reference range problems across the population. However, when each patient was treated as their own point of reference and biomarker slopes for individual patients were considered, the diagnostic accuracy for this same biomarker increased significantly. Similarly, the COVID-19 Biomarker Score time series (Figure 5) reveals a strong capacity to separate patients who die of COVID-19 complications from those who are discharged from the hospital. Note that the app includes capabilities to use the proximal biomarker measurements allowing for biomarker measurements to be collected over time without the rigid restriction of having all biomarker measurements be completed at the same time for all time points. This flexibility is anticipated to afford more convenience for longitudinal monitoring of patients.

Lastly, the models developed here demonstrated generalizability through external model validation. External validation is essential before implementing prediction models in clinical practice [28]. We found that the AUCs for both Tier 1 and Tier 2 models were similar for internal vs. external validation, demonstrating that the models are generalizable to making predictions for these disease indications despite different care settings and patient demographics. Usually, prediction models perform better on the training data than on new data; however, in this study, we found that the external validation results were approximately the same or better (Tier 1: AUC of 0.79 vs. 0.79; Tier 2: 0.95 and 0.97 for internal and external validation, respectively), suggesting that patients in the external validation sets may have suffered from more severe disease.

Despite the potential for CDSSs to transform health care, major challenges remain for translating and scaling such tools. Future data and, thus, model performance may have large heterogeneity, which is exacerbated by missing data (potentially not missing at random), non-standard definitions of outcomes, and incomplete laboratory measurements and follow-up times [29]. The mobile app developed here is intended to reduce heterogeneity by encouraging the harmonization of data collection across multiple care settings. Further, models may be tuned through optimization of cutoffs for certain patient subpopulations. Another challenge in deploying a CDSS that relies on biomarker measurements is accounting for differences in laboratory testing across hospitals and clinics. The variability of such measurements across institutions may have a large impact on the distribution of COVID-19 Biomarker Scores. This challenge creates a unique opportunity for standardized, well-calibrated, and highly scalable point-of-care tests for COVID-19 disease severity [20, 30, 31].

A commercial app has been developed in collaboration with OraLiva, Inc. for deployment of these tools to frontline healthcare workers managing COVID-19 patients. Plans are now in place to assess the usability, user satisfaction, and confidence in results of this CDSS and mobile app in the FHCs at NYU. Future efforts will focus on point-of-care testing capabilities to more rapidly assess the Tier 2 Biomarkers described in this study using a previously developed and published platform [20, 30, 31]. The deployment of these new capabilities has potential for immediate clinical impact in community clinics, where the application of such tools could significantly improve the quality of care.

Acknowledgements

Conflicts of Interest

MPM has served as a paid consultant for SensoDx, LLC and has a provisional patent pending. NJC has a provisional patent pending. JTM has a provisional patent pending. In addition, he has an ownership position and an equity interest in both SensoDx II, LLC and OraLiva, Inc. and serves on their advisory boards. All other authors declare no competing interests.

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020 Jan 30;395(10223):507-13. doi: 10.1016/S0140-6736(20)30211-7.

2. World Health Organization. [April 2020]; Available from: https://www.who.int/.

3. Coronavirus Disease 2019 (COVID-19) Situation Report - 133. World Health Organization, 1 Jun 2020. Report No.: 133.

4. Sharfstein JM, Becker SJ, Mello MM. Diagnostic testing for the novel coronavirus. JAMA. 2020 Mar 9;323(15):1437-8. doi: 10.1001/jama.2020.3864.

5. Morbidity and Mortality Weekly Report (MMWR), Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. [April 2020]; Available from: http://dx.doi.org/10.15585/mmwr.mm6912e2.

6. The Office of the National Coordinator for Health Information Technology. Clinical Decision Support. [cited 2020 June 16]; Available from: https://www.healthit.gov/topic/safety/clinical-decision-support.

7. Wasylewicz ATM, Scheepers-Hoeks A. Clinical decision support systems. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham (CH): Springer; 2019. p. 153-69.

8. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054-62. doi: 10.1016/S0140-6736(20)30566-3.

9. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762-74. doi: 10.1001/jama.2016.0288.

10. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intens Care Med. 1996 Jul 01;22(7):707-10. doi: 10.1007/BF01709751.

11. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J, et al. Acute Physiology and Chronic Health Evaluation II Score as a predictor of hospital mortality in patients of coronavirus disease 2019. Crit Care Med. 2020 May 01. doi: 10.1097/ccm.0000000000004411.

12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985 Oct;13(10):818-29. PMID:3928249.

13. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS. A multicenter study of ICU telemedicine reengineering of adult critical care. CHEST. 2014 Mar;145(3):500-7. doi: 10.1378/chest.13-1973.

14. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018 Sep;5(1):180178. doi: 10.1038/sdata.2018.178.

15.   Duca A, Piva S, Focà E, Latronico N, Rizzi M. Calculated Decisions: Brescia-COVID Respiratory Severity Scale (BCRSS)/Algorithm. Emerg Med Pract. 2020 Apr 16;22(5 Suppl):Cd1-cd2. PMID: 32297727.

16.   Wallis LA. African Federation for Emergency Medicine resources for managing COVID-19 in low resourced settings. Afr J Emerg Med. 2020 Jun;10(2):49. doi: 10.1016/j.afjem.2020.06.001.

17.   The ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin Definition. JAMA. 2012 Jun;307(23):2526-33. doi: 10.1001/jama.2012.5669.

18.   Fan E, Brodie D, Slutsky AS. Acute respiratory distress syndrome: advances in diagnosis and treatment. JAMA. 2018 Feb 20;319(7):698-710. PMID: 29466596. doi: 10.1001/jama.2017.21907.

19.   Singh K, Valley TS, Tang S, Li BY, Kamran F, Sjoding MW, et al. Validating a widely implemented deterioration index model among hospitalized COVID-19 patients. medRxiv. 2020 Apr 29:1-22. doi: 10.1101/2020.04.24.20079012.

20.   McRae MP, Simmons GW, Christodoulides NJ, Lu Z, Kang SK, Fenyo D, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. Lab Chip. 2020 Jun 03;20(12):2075-85. doi: 10.1039/D0LC00373E.

21.   Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020 May 01;2(5):283-8. doi: 10.1038/s42256-020-0180-7.

22.   McRae MP, Bozkurt B, Ballantyne CM, Sanchez X, Christodoulides N, Simmons G, et al. Cardiac ScoreCard: A diagnostic multivariate index assay system for predicting a spectrum of cardiovascular disease. Expert Sys Appl. 2016 Jul 15;54:136-47. doi: 10.1016/j.eswa.2016.01.029.

23.   Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. J Stat Software. 2011 Dec 12;45(3). doi: 10.18637/jss.v045.i03.

24.   Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med. 2007 Aug 30;26(19):3661-75. PMID: 17315184. doi: 10.1002/sim.2832.

25.   Richardson JTE. The analysis of 2 × 2 contingency tables—Yet again. Stat in Med. 2011 Apr 15;30(8):890. doi: 10.1002/sim.4116.

26.   Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19). JAMA Cardiol. 2020 Mar 27. doi: 10.1001/jamacardio.2020.1017.

27.   Skates SJ, Pauler DK, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. J Am Stat Assoc. 2001 Dec;96(454):429-39. doi: 10.1198/016214501753168145

28.   Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research:: A clinical example. J Clin Epidemiol. 2003 Sep 01;56(9):826-32. doi: 10.1016/S0895-4356(03)00207-5.

29.   Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016 Jun 22;353:1-11. doi: 10.1136/bmj.i3140.

30.   McRae MP, Simmons G, Wong J, McDevitt JT. Programmable bio-nanochip platform: a point-of-care biosensor system with the capacity to learn. Acc Chem Res. 2016 Jul 19;49(7):1359-68. doi: 10.1021/acs.accounts.6b00112.

31.   McRae MP, Simmons GW, Wong J, Shadfan B, Gopalkrishnan S, Christodoulides N, et al. Programmable bio-nano-chip system: a flexible point-of-care platform for bioscience and clinical measurements. Lab Chip. 2015 Aug 18;15(20):4020-31. doi: 10.1039/C5LC00636H.

Abbreviations

AUC: area under the curve
CDSS: Clinical decision support system
CI: confidence interval
COVID-19: coronavirus disease of 2019
CRP: C-reactive protein
cTnI: cardiac troponin I

eICU: electronic intensive care unit
FHC: Family Health Centers
ICU: intensive care unit
IQR: interquartile range
NPV: negative predictive value
NT-proBNP: N-terminal pro–B-type natriuretic peptide
NYU: New York University
PCR: polymerase chain reaction
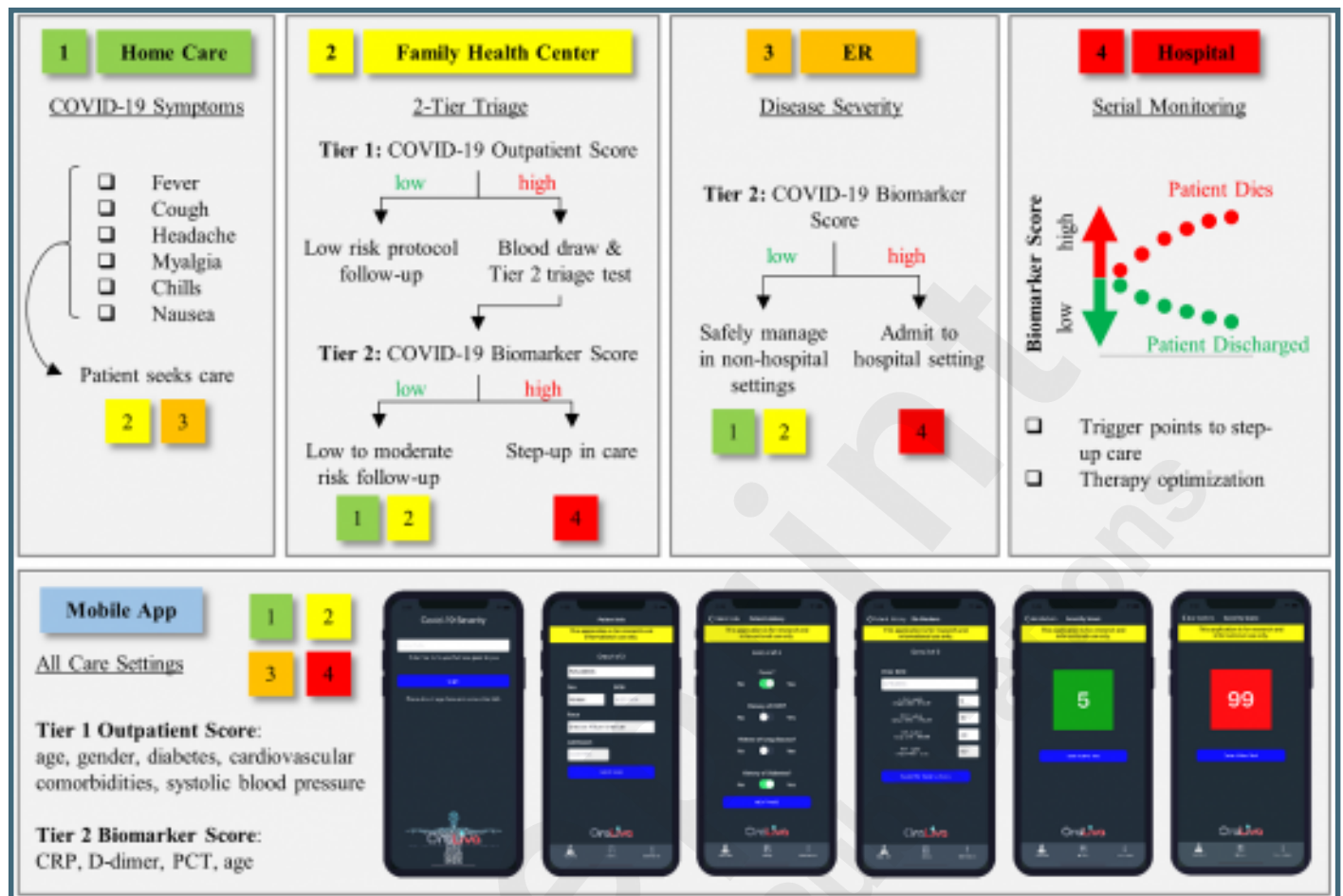PCT: procalcitonin
PPV: positive predictive value
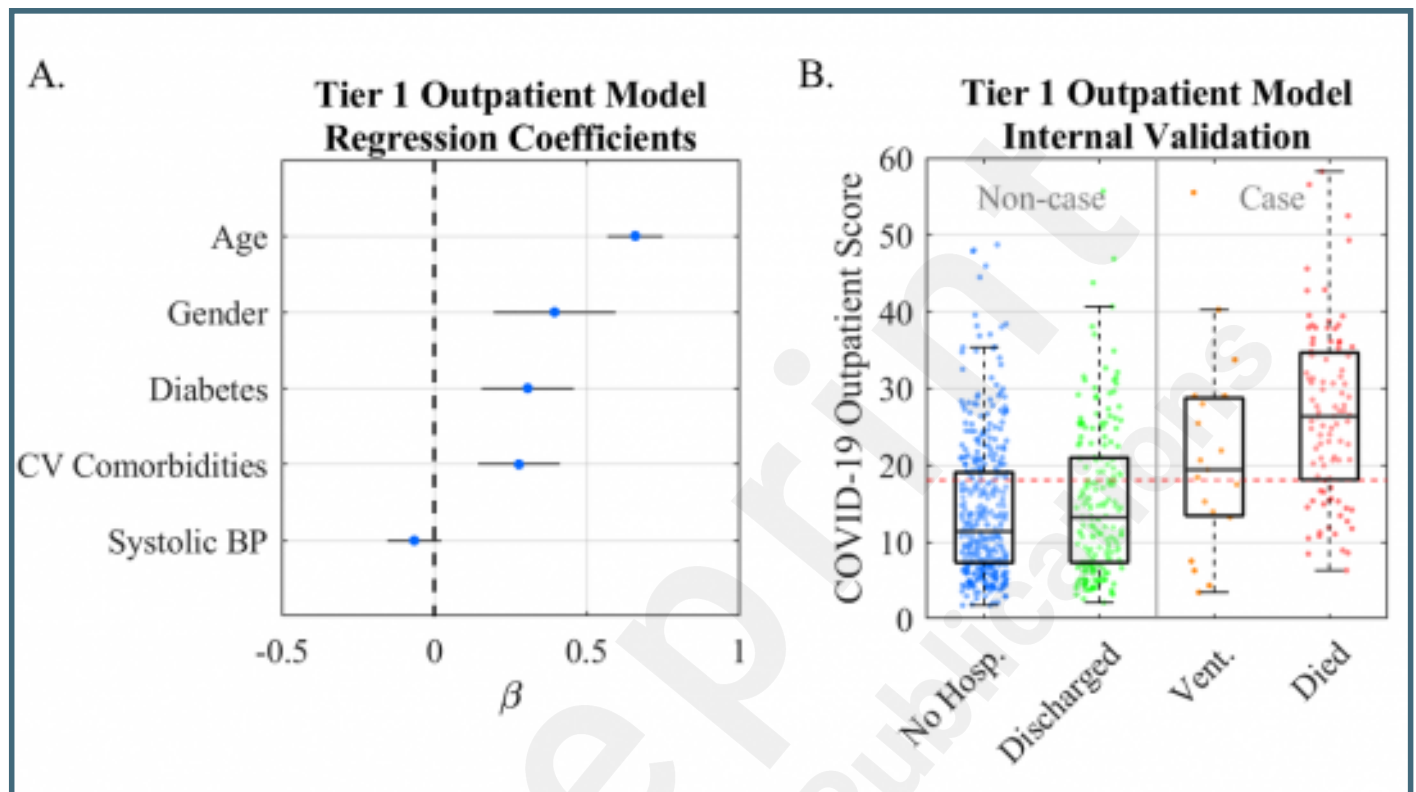SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
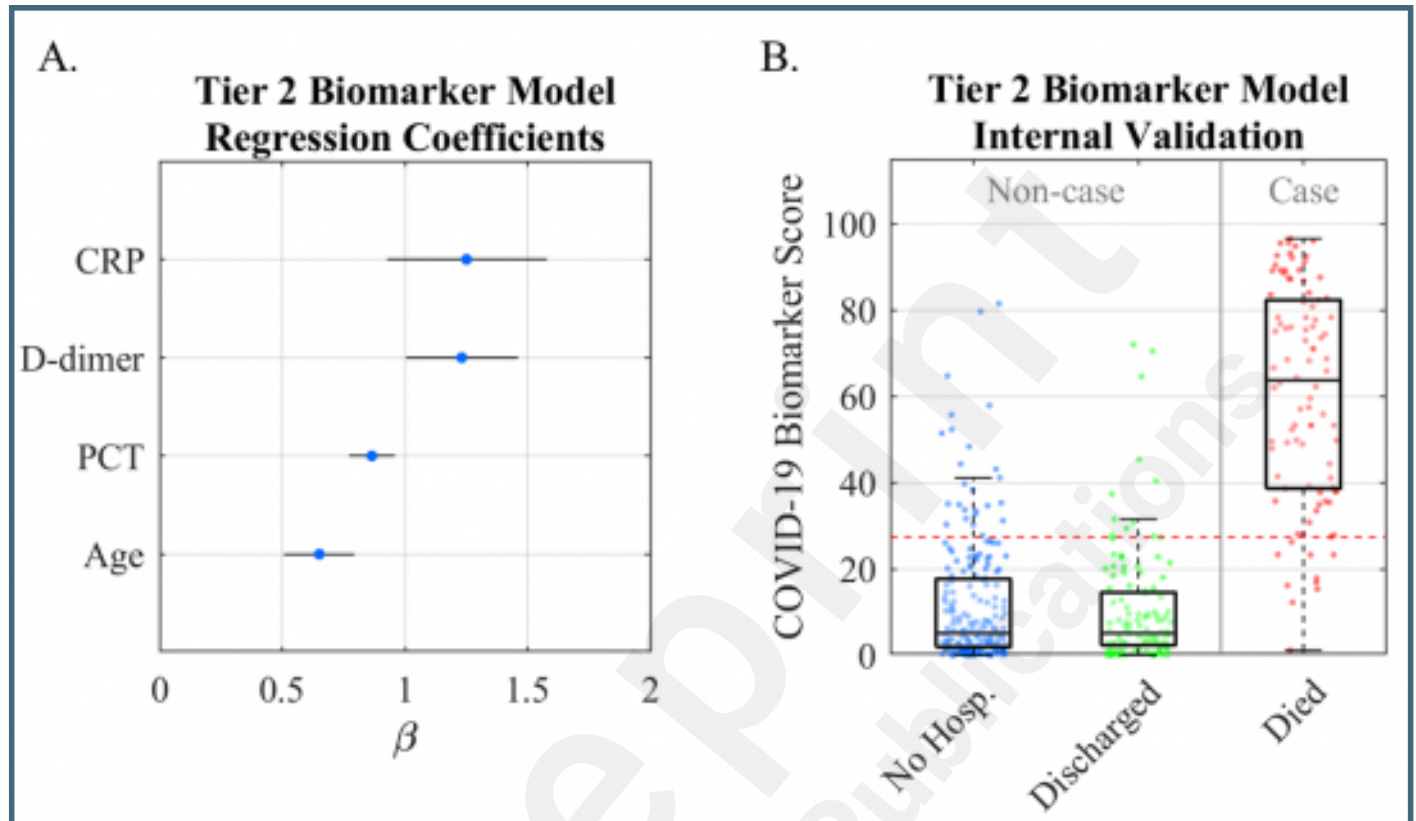
Supplementary Files

Figures

Clinical decision support system and mobile app for managing COVID-19 care.
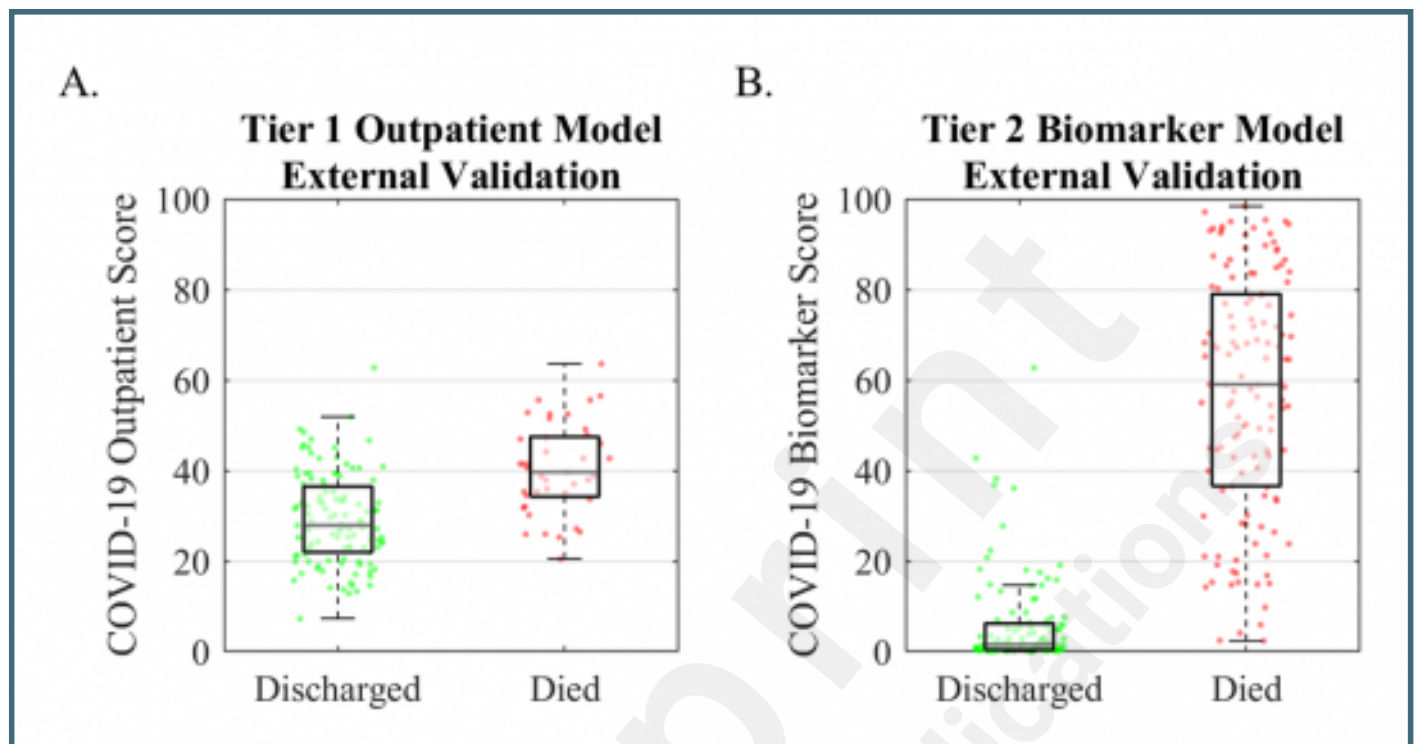
Tier 1 Outpatient Model results. Lasso logistic regression coefficients (A) reveals the relative predictor importance in generating the score. The box/scatter plot (B) from internal validation shows Tier 1 Outpatient Scores for the four outcomes. A cutoff score of 18 (red dotted line) balances sensitivity and specificity for "Non-case" vs. "Case" patients (gray line) (No Hosp. = patients who were not hospitalized, Vent. = patients who were ventilated, CV comorbidities = cardiovascular comorbid conditions).

Tier 2 Biomarker Model results. Lasso logistic regression coefficients (A) reveals relative predictor importance in generating the score. The box/scatter plot (B) from internal validation shows Tier 2 Biomarker Scores for the three patient outcomes. A cutoff score of 27 (horizontal red dotted line) balances sensitivity and specificity for "Non-case" vs. "Case" patients (vertical gray line) (No Hosp. = patients who were not hospitalized).

External validation results. The Tier 1 Outpatient Model was evaluated on data from COVID-19 patients at Zhongnan Hospital of Wuhan University [26]. The Tier 2 Biomarker Model was evaluated on data from COVID-19 patients at Tongji Hospital [21].

Spaghetti plot of longitudinal COVID-19 Biomarker scores for patients in the external validation set from Tongji Hospital [21] between January 10 and February 18, 2020. These data represent individual patients' scores over a median (IQR) of 12.5 (8–17.5) days between admission and outcomes of discharged or deceased. The first scores available after admission were significantly higher in those that died vs. those that were discharged (AUC 0.97, cutoff score of 19), and over time patients who were discharged had an average decrease in score (-4.7) while those that died had an average increase in score (+11.2).