



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Short communication

## A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results



Rohan P. Joshi<sup>a</sup>, Vikas Pejaver<sup>b</sup>, Noah E. Hammarlund<sup>b</sup>, Heungsup Sung<sup>c</sup>, Seong Kyu Lee<sup>d</sup>, Al'ona Furmanchuk<sup>e</sup>, Hye-Young Lee<sup>c,f</sup>, Gregory Scott<sup>a</sup>, Saurabh Gombar<sup>a</sup>, Nigam Shah<sup>g</sup>, Sam Shen<sup>h</sup>, Anna Nassiri<sup>i</sup>, Daniel Schneider<sup>j</sup>, Faraz S. Ahmad<sup>k</sup>, David Liebovitz<sup>l</sup>, Abel Kho<sup>l</sup>, Sean Mooney<sup>b</sup>, Benjamin A. Pinsky<sup>a,m,n</sup>, Niaz Banaei<sup>a,m,o,\*</sup>

<sup>a</sup> Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

<sup>b</sup> Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine, Seattle, WA, USA

<sup>c</sup> Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Republic of Korea

<sup>d</sup> Department of Laboratory Medicine, Bundang Jesaeng General Hospital, Seongnam, Republic of Korea

<sup>e</sup> Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>f</sup> Department of Laboratory Medicine, U2Bio Laboratories, Seoul, Republic of Korea

<sup>g</sup> Department of Medicine, Division of Stanford University School of Medicine, Stanford, CA, USA

<sup>h</sup> Department of Emergency Medicine, Stanford University School of Medicine, Stanford, CA, USA

<sup>i</sup> Northwestern Medicine, Chicago, IL, USA

<sup>j</sup> Research Analytics at Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

<sup>k</sup> Department of Medicine, Division of Cardiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>l</sup> Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>m</sup> Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, CA, USA

<sup>n</sup> Clinical Virology Laboratory, Stanford University Medical Center, Palo Alto, CA, USA

<sup>o</sup> Clinical Microbiology Laboratory, Stanford University Medical Center, Palo Alto, CA, USA

## ARTICLE INFO

## Keywords:

COVID-19  
SARS-CoV-2  
Rapid testing  
Machine learning  
Prediction tool

## ABSTRACT

**Background:** Testing for COVID-19 remains limited in the United States and across the world. Poor allocation of limited testing resources leads to misutilization of health system resources, which complementary rapid testing tools could ameliorate.

**Objective:** To predict SARS-CoV-2 PCR positivity based on complete blood count components and patient sex.

**Study design:** A retrospective case-control design for collection of data and a logistic regression prediction model was used. Participants were emergency department patients > 18 years old who had concurrent complete blood counts and SARS-CoV-2 PCR testing. 33 confirmed SARS-CoV-2 PCR positive and 357 negative patients at Stanford Health Care were used for model training. Validation cohorts consisted of emergency department patients > 18 years old who had concurrent complete blood counts and SARS-CoV-2 PCR testing in Northern California (41 PCR positive, 495 PCR negative), Seattle, Washington (40 PCR positive, 306 PCR negative), Chicago, Illinois (245 PCR positive, 1015 PCR negative), and South Korea (9 PCR positive, 236 PCR negative). **Results:** A decision support tool that utilizes components of complete blood count and patient sex for prediction of SARS-CoV-2 PCR positivity demonstrated a C-statistic of 78 %, an optimized sensitivity of 93 %, and generalizability to other emergency department populations. By restricting PCR testing to predicted positive patients in a hypothetical scenario of 1000 patients requiring testing but testing resources limited to 60 % of patients, this tool would allow a 33 % increase in properly allocated resources.

**Conclusions:** A prediction tool based on complete blood count results can better allocate SARS-CoV-2 testing and other health care resources such as personal protective equipment during a pandemic surge.

Health care systems worldwide are struggling to meet the demands of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

pandemic [1]. Weeks after local transmission was first recognized in the US, laboratory testing for COVID-19, the disease caused by SARS-CoV-

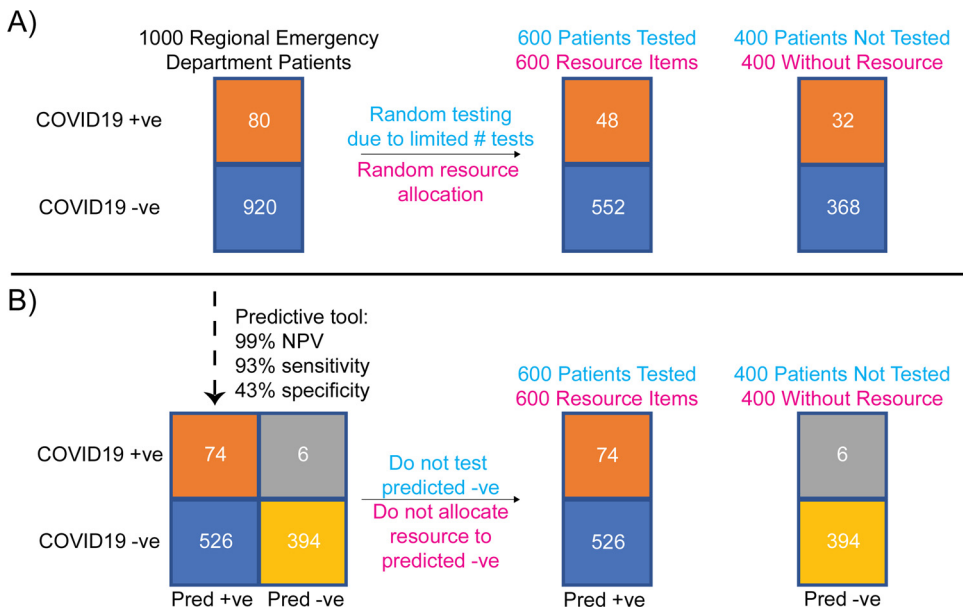
\* Corresponding author at: Stanford University School of Medicine, 3375 Hillview Ave, Rm 1602, Palo Alto, CA, 94304, USA.

E-mail address: [nbanaei@stanford.edu](mailto:nbanaei@stanford.edu) (N. Banaei).

<https://doi.org/10.1016/j.jcv.2020.104502>

Received 3 June 2020; Accepted 7 June 2020

1386-6532/ © 2020 Elsevier B.V. All rights reserved.



**Fig. 1.** Value of a predictive COVID-19 rule-out tool in improving utilization of health care resources during a pandemic. Example contains a cohort comprising 1000 hypothetical patients with respiratory symptoms presenting to emergency departments across a region and assumes 8% COVID-19 prevalence, a highly accurate SARS-CoV-2 test, and 600 of a limited hospital resource (e.g. SARS-CoV-2 tests, personal protective equipment). (Panel A) If patients are randomly tested or randomly allocated a hospital resource during the wait for results, many patients with COVID-19 patients may not get tested or allocated the resource. (Panel B) With availability of a predictive tool of high sensitivity and negative predictive value based on readily available routine test results, utilization of limited confirmatory SARS-CoV-2 testing or other resources is reserved for those patients more likely to have COVID-19, with a 33% improvement (48/80 to 74/80) in resource allocation. **COVID19 +ve:** COVID-19 positive patients; **COVID19 -ve:** COVID-19 negative patients; **Pred + ve:** predicted positive; **Pred -ve:** predicted negative.

2, was limited by assay complexity and reagent shortages. Limited testing and long turnaround times led to misutilization of resources resulting in strained health systems in high burden regions. While testing capacity is improving, alternative testing approaches and specimen sources are needed to handle COVID-19 surges, particularly in low-resource health systems. The aim of this study was to develop a decision support tool that integrates readily available routine lab values to predict negative SARS-CoV-2 results in patients presenting to the emergency department (ED). Use of this tool could reserve confirmatory SARS-CoV-2 testing and health system resources such as personal protective equipment and isolation rooms for those patients more likely to have COVID-19 (Fig. 1).

A number of studies reported associations between certain non-SARS-CoV-2 test results and COVID-19 disease [2,3]. We therefore collected complete blood count (CBC) data from 3/1/2020 to 3/20/2020 ordered within 24 h of a SARS-CoV-2 PCR [4] (based off of the WHO assay) order, which resulted in a dataset of 33 confirmed SARS-CoV-2 PCR positive and 357 PCR negative ED patients at Stanford Health Care. CBC were generally concurrently ordered on ED patients with an Emergency Severity Index [5] (ESI) 1–3 and 50% of ED patients with SARS-CoV-2 PCR tests had an accompanying CBC. ED patients with both CBC and SARS-CoV-2 PCR tests were therefore of older age (median 59 years old) compared to ED patients with SARS-CoV-2 PCR tests alone (median 37 years old), with a similar sex ratio and rate of SARS-CoV-2 PCR positivity. CBC orders were placed before positive SARS-CoV-2 PCR tests resulted and 83% of patients had CBC orders placed within 4 h of the PCR order.

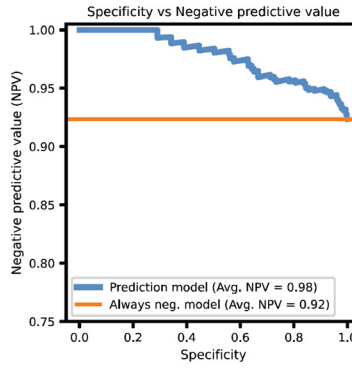
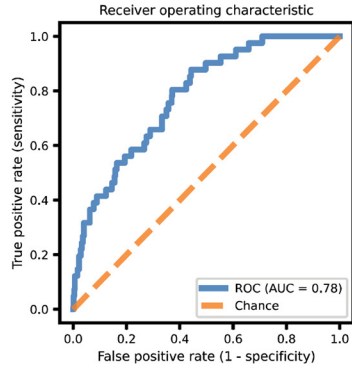
For training, we selected 3 CBC components, absolute neutrophil count (ANC), absolute lymphocyte count (ALC), and hematocrit (HCT), based on a univariate analysis suggesting association with positive SARS-CoV-2 PCR, evidence of ANC and ALC association with disease severity in the literature [2,3], and a low Pearson correlation of these features to each other. We included male sex as a feature in our model since it was associated with positive SARS-CoV-2 PCR status independent of hematocrit. In cross-validation within the training set, this manual variable selection method performed comparably to other model-based variable selection methods (e.g. recursive feature elimination, L1-penalization, and L2-penalization). Using ANC, ALC, hematocrit, and patient sex, we trained an L2-regularized logistic regression model [6]. ANC and ALC were negative predictors while male sex and HCT were positive predictors of SARS-CoV-2 PCR positivity. Using a receiver operating characteristic curve, we chose a test

probability threshold to optimize for > 50% specificity and > 80% sensitivity. Validation sets consisted of data that were not seen by the model during training from emergency department patients who received SARS-CoV-2 PCR testing and concurrent CBC.

In testing on a validation set from patients presenting to Stanford Health Care from 3/21/2020 to 4/7/2020 (confirmed SARS-CoV-2 PCR positive, n = 41; negative, n = 495), the decision support tool showed a diagnostic C-statistic of 0.78 (Fig. 2A). We examined the tradeoff between negative predictive value (NPV) and specificity using our model and found high NPV was maintained across a range of specificities, with a specificity-weighted average NPV of 98%. Using the operating threshold defined using the training set, the model accurately ruled out SARS-CoV-2 in 40% of total test patients with an NPV of 99% and sensitivity of 93%. Results of seasonal virus rapid respiratory panel (RRP) PCR, when positive, have been used by health systems to avoid the need for testing for COVID-19. We trained an L2-regularized logistic regression model that included the CBC features as well as results of seasonal rapid respiratory panel (RRP) PCR as predictors. The predictive performance of this RRP model was indistinguishable from the CBC-only model (data not shown). Consistent with previous reports [7], these results suggest that RRP does not have additional predictive value compared to CBC.

We note that these models were trained on our local population and could be sensitive to the choice of patients for whom our ED providers chose to order CBCs, in addition to other institutional factors affecting CBC and SARS-CoV-2 testing. We sought to test the generalizability of the predictive value of these CBC and sex features in our greater geographic area. Our Northern California population trained model performed similarly in a Seattle, Washington based patient cohort, with a diagnostic C-statistic of 0.75 and a specificity-weighted average NPV of 95%. Applying our Northern California-defined operating threshold, the model accurately ruled out SARS-CoV-2 in 31% of total patients, with an NPV of 96% and sensitivity of 88% (Fig. 2B). In a Chicago, Illinois-based patient cohort, the model demonstrated a C-statistic of 0.75 and a specificity-weighted average NPV of 90%. Applying our Northern California-defined operating threshold, the model accurately ruled out SARS-CoV-2 in 39% of total patients, with an NPV of 92% and sensitivity of 85% (Fig. 2C). The model also performed similarly in a South Korean based patient cohort, with a diagnostic C-statistic of 0.81 and a specificity-weighted average NPV of 99%. Applying our Northern California-defined operating threshold, the model accurately ruled out SARS-CoV-2 in 53% of patients in this South Korean

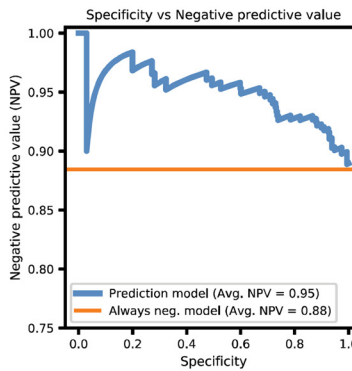
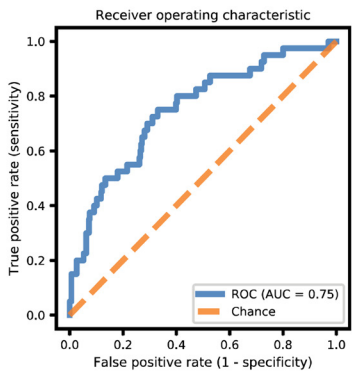
### A) Northern California



	Pred +ve	Pred -ve
COVID19 +ve	38	3
COVID19 -ve	282	213

Sensitivity = 93%  
 Specificity = 43%  
 PPV = 12%  
 NPV = 99%

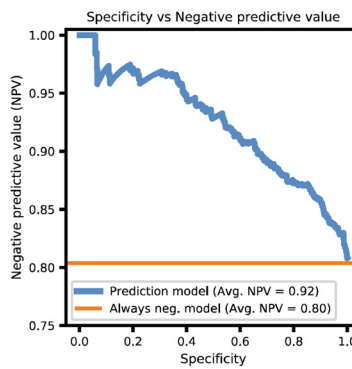
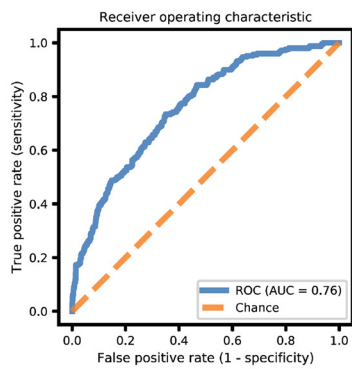
### B) Seattle, Washington



	Pred +ve	Pred -ve
COVID19 +ve	35	5
COVID19 -ve	199	107

Sensitivity = 88%  
 Specificity = 35%  
 PPV = 15%  
 NPV = 96%

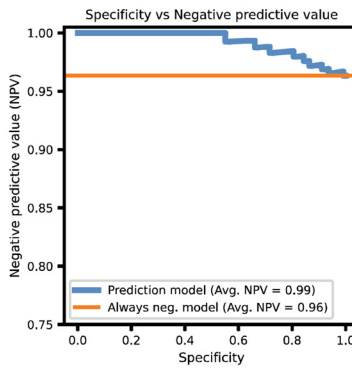
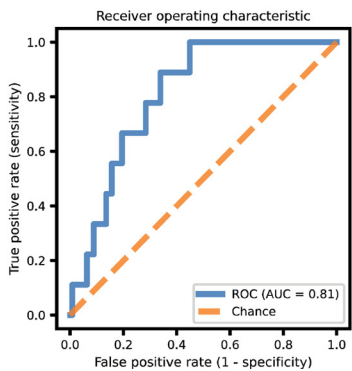
### C) Chicago, Illinois



	Pred +ve	Pred -ve
COVID19 +ve	213	35
COVID19 -ve	516	499

Sensitivity = 86%  
 Specificity = 49%  
 PPV = 29%  
 NPV = 93%

### D) South Korea



	Pred +ve	Pred -ve
COVID19 +ve	8	1
COVID19 -ve	106	130

Sensitivity = 89%  
 Specificity = 55%  
 PPV = 7%  
 NPV = 99%

(caption on next page)

**Fig. 2.** Performance of complete blood count (CBC)-based predictive COVID-19 rule-out tool. **A)** Results of predictive tool on Stanford Health Care emergency department patient cohort. **B)** Results of predictive tool on a Seattle, Washington emergency department patient cohort. **C)** Results of predictive tool on a Chicago, Illinois emergency department patient cohort. **D)** South Korean cohort of emergency department patients. All four cohorts represented validation sets not previously seen by the decision support tool. SARS-CoV-2 PCR performed in local laboratories was used as the reference method. **Left:** Receiver operating characteristic curves. **Middle:** Specificity versus negative predictive value across all operating thresholds. **Right:** Confusion matrix calculated using operating point defined using Stanford Health Care training cohort. **AUC:** Receiver operating characteristic area under curve. **PPV:** positive predictive value. **NPV:** negative predictive value. **Avg. NPV:** Weighted average of negative predictive values with specificity as weights across all probability thresholds. **Always neg. model:** baseline negative predictive value expected by a classifier that always predicts SARS-CoV-2 negative.

population, with an NPV of 99 % and sensitivity of 89 % (Fig. 2D).

Our results demonstrate that a decision support tool based on CBC component results has high negative predictive value for SARS-CoV-2 PCR results in diverse patient populations from the Northwest US and South Korea. In the setting of limited testing where only 60 % of 1000 patients entering emergency departments could be tested, our tool would result in a 33 % better allocation of testing and other hospital resources (Fig. 1B). As a proof-of-concept, we made the CBC model available as a simple web-based tool, where individuals can enter ANC, ALC, HCT, sex, and receive a result of whether SARS-CoV-2 testing is predicted negative (<http://web.stanford.edu/~gscott2/cgi-bin/Covid-Tool/>). Given that this predictive model trades high sensitivity for tolerable specificity, specific cost-benefit analyses that include resource scarcity, model operating point, and model performance as factors should be performed prior to implementation in a local setting. Our predictive model and others like it should be useful for COVID-19 and future pandemics as an interim stopgap early in and during surges of the pandemic when testing is limited. Integration of this tool can aid in optimizing utilization of health care resources, reducing turnaround-time of viral testing, and improving patient care.

#### Declaration of Competing Interest

The authors declare no competing interests.

#### Acknowledgements

We thank Ethan Steinberg for review of data curation and the

prediction model. Our study was supported by Stanford University Department of Pathology.

#### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jcv.2020.104502>.

#### References

- [1] E.J. Emanuel, et al., Fair allocation of scarce medical resources in the time of Covid-19, *N. Engl. J. Med.* 0 (2020).
- [2] W. Guan, et al., Clinical characteristics of coronavirus disease 2019 in China, *N. Engl. J. Med.* (2020) 0, null.
- [3] N. Chen, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (2020) 507–513.
- [4] C.A. Hogan, M.K. Sahoo, B.A. Pinsky, Sample pooling as a strategy to detect community transmission of SARS-CoV-2, *JAMA* (2020), <https://doi.org/10.1001/jama.2020.5445>.
- [5] D.R. Eitel, D.A. Travers, A.M. Rosenau, N. Gilboy, R.C. Wuerz, The emergency severity index triage algorithm version 2 is reliable and valid, *Acad. Emerg. Med. Off. J. Soc. Acad. Emerg. Med.* 10 (2003) 1070–1080.
- [6] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition, Springer, 2016.
- [7] D. Kim, J. Quinn, B. Pinsky, N.H. Shah, I. Brown, Rates of co-infection between SARS-CoV-2 and other respiratory pathogens, *JAMA* (2020), <https://doi.org/10.1001/jama.2020.6266>.