ORIGINAL ARTICLE

International Journal of Laboratory Hematology — ISLH — WILEY

A meta-analysis of SARS-CoV-2 patients identifies the combinatorial significance of D-dimer, C-reactive protein, lymphocyte, and neutrophil values as a predictor of disease severity

Kunwar Singh[1] 🔘 | Sasha Mittal[2] | Sumanth Gollapudi[2] | Alexandra Butzmann[2] |
Jyoti Kumar[1] | Robert S. Ohgami[2] 🔘

[1]Department of Pathology, Stanford University, Stanford, CA, USA

[2]Department of Pathology, University of California, San Francisco, San Francisco, CA, USA

Correspondence
Robert S. Ohgami, Department of Pathology, University of California, San Francisco, San Francisco, CA 94143, USA.
Email: robert.ohgami@ucsf.edu

Abstract

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), known to be the causative agent of COVID-19, has led to a worldwide pandemic. At presentation, individual clinical laboratory blood values, such as lymphocyte counts or C-reactive protein (CRP) levels, may be abnormal and associated with disease severity. However, combinatorial interpretation of these laboratory blood values, in the context of COVID-19, remains a challenge.

Methods: To assess the significance of multiple laboratory blood values in patients with SARS-CoV-2 and develop a COVID-19 predictive equation, we conducted a literature search using PubMed to seek articles that included defined laboratory data points along with clinical disease progression. We identified 9846 papers, selecting primary studies with at least 20 patients for univariate analysis to identify clinical variables predicting nonsevere and severe COVID-19 cases. Multiple regression analysis was performed on a training set of patient studies to generate severity predictor equations, and subsequently tested on a validation cohort of 151 patients who had a median duration of observation of 14 days.

Results: Two COVID-19 predictive equations were generated: one using four variables (CRP, D-dimer levels, lymphocyte count, and neutrophil count), and another using three variables (CRP, lymphocyte count, and neutrophil count). In adult and pediatric populations, the predictive equations exhibited high specificity, sensitivity, positive predictive values, and negative predictive values.

Conclusion: Using the generated equations, the outcomes of COVID-19 patients can be predicted using commonly obtained clinical laboratory data. These predictive equations may inform future studies evaluating the long-term follow-up of COVID-19 patients.

KEYWORDS

blood, CBC, COVID-19, CRP, D-dimer, lymphocyte, neutrophil, SARS-CoV-2

Singh and Mittal contributed equally.

1 | INTRODUCTION

In December 2019, a cluster of atypical pneumonia cases of unknown etiology emerged. This phenomenon was epidemiologically linked to a seafood and wet animal wholesale market in Wuhan, Hubei Province, China.[1] The causative agent was isolated and sequenced from human airway epithelial cells from infected patients and identified as a novel beta coronavirus, named 2019-nCoV.[2] To date, this is the seventh member of the *Coronaviridae* family that is known to infect humans. The 229E, OC43, NL63, and HKU1 strains usually cause mild illness and are associated with the common cold. SARS-CoV and MERS-CoV caused epidemic outbreaks of severe respiratory distress in 2002-2003 and 2012, respectively.[3,4] Full-length sequencing of the 2019-nCoV[2] virus showed that 88% of the genome was shared with the bat SARS coronavirus[5] and 80% of the nucleotides were identified in the SARS-CoV that caused the 2002-2003 epidemic.[6] Based on published genomic information, the nomenclature for the causative strain was changed to SARS-CoV-2. After identification of this novel virus, the World Health Organization (WHO) named the disease COVID-19. As of July 2020, over 14 million cases have been identified attributing to approximately 600 000 deaths.[7] Due to the overwhelming burden this pandemic has caused on healthcare systems worldwide, there has been a concerted effort to understand the etiology and epidemiology of this disease. To date, there have been multiple articles published to establish clinical features, severity, and mortality. However, it has been difficult to identify diagnostic data that could reliably predict patient outcomes.

In both adult and pediatric patients, several studies have suggested that hematological data may be key to determining severity and outcomes in COVID-19. This is particularly attractive in the clinical setting due to rapid availability of these tests in both inpatient settings.[8,9,10] Here, we perform a comprehensive meta-analysis of globally reported laboratory data of hospitalized COVID-19 patients to generate COVID-19 severity predictive equations. Clinical severity predictor equations may potentially help physicians more accurately prognosticate outcomes of COVID-19 patients.

2 | MATERIALS AND METHODS

2.1 | Literature review, data collection, and statistical analysis

Articles with information regarding COVID-19 were identified in PubMed in May 2020 using COVID-19, COVID, and SARS-CoV-2 as search keywords yielding 9846 articles. We identified 160 articles that were printed or translated into English, published after December 1, 2019, and containing specific clinical blood laboratory data values (Figure 1). Papers that did not have full CBC data, had <20 patients, and were not original research papers were removed through manual curation. We reviewed these papers for reported
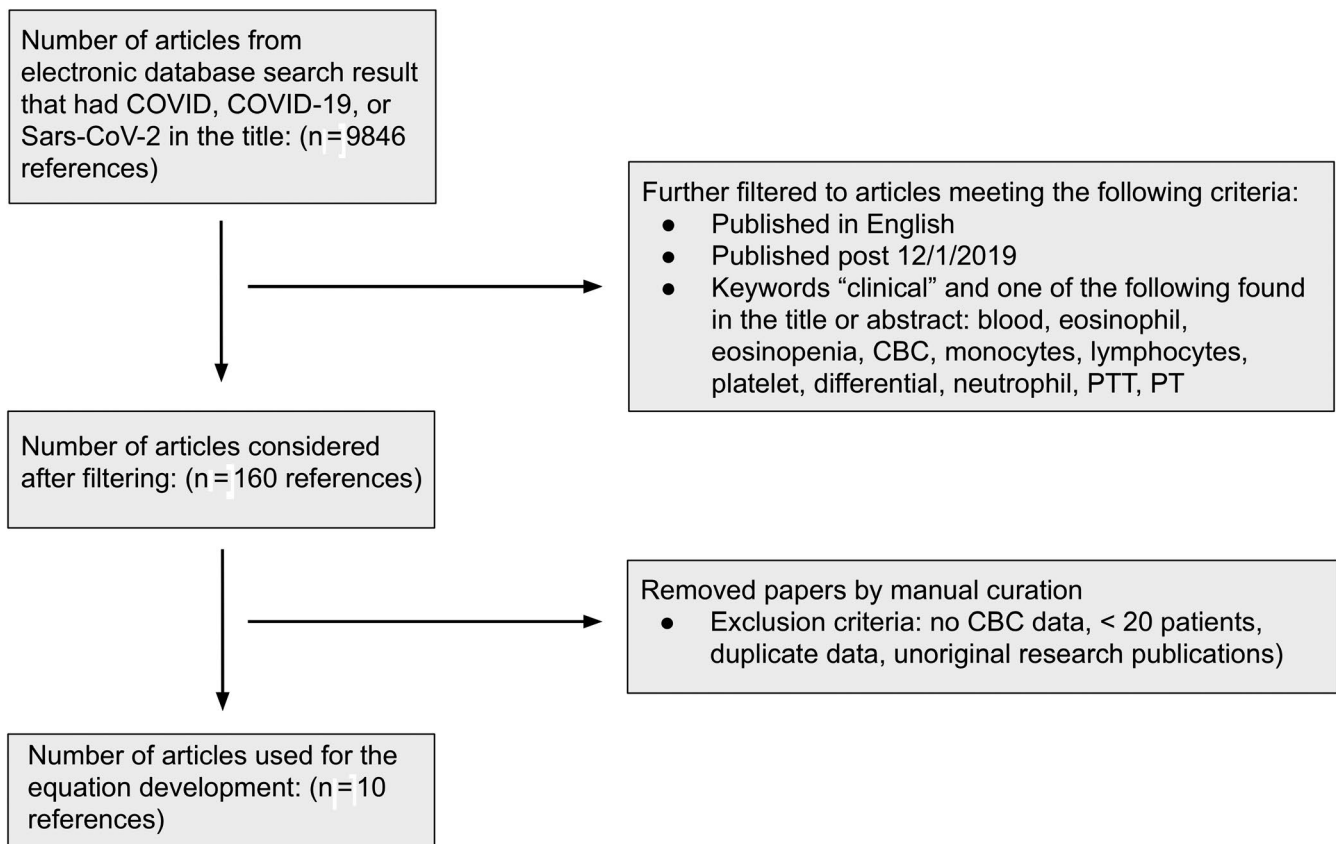

FIGURE 1 Summary of literature search, and articles reviewed and used for analysis

TABLE 1 COVID-19 severity prediction equations

Equation		Multiple R	R^2
1	$y = 0.97 - 0.92 \times$ (LYM K/μl) $+ 0.070 \times$ (NEU K/μl) $+ 0.0038 \times$ (CRP mg/L) $+ 0.033 \times$ (DD mg/L)	0.86	.75
2	$y = 0.79 - 0.82 \times$ (LYM K/μl) $+ 0.090 \times$ (NEU K/μl) $+ 0.0045 \times$ (CRP mg/L)	0.82	.68

significant findings with regard to laboratory data and their univariate significance in outcomes in SARS-CoV-2 patients. This yielded the following possible parameters that were reported significant in outcomes: lymphocyte count, neutrophil count, eosinophil count, D-dimer values, procalcitonin, lactate dehydrogenase, creatinine, albumin, and CRP. We narrowed the relevant variables and papers to 51 remaining papers that had such data and were then evaluated to see whether they contained four variables (D-dimer, neutrophil count, lymphocyte count, and CRP) as these were identified as most significant in univariate analyses in two or more papers in the literature. This resulted in a total of 10 papers, which were used to develop SARS-CoV-2 severity prediction equations. Quantitative analysis, t test, linear regression analysis, and multiple regression analysis were performed using XLSTAT (Addinsoft) and Microsoft Excel™ (Microsoft Corporation. [2016]. *Microsoft Excel*.). *P*-values <.05 were considered statistically significant. This study was approved by the University of San Francisco's institutional review board (IRB-28225).

3 | RESULTS

3.1 | Development of COVID-19 severity prediction equations

Laboratory data variables that were identified as significant in predicting outcomes of COVID-19 patients in two or more papers in literature, which had 20 or more patients, and were from curated studies as described in Materials and Methods were identified and used to generate COVID-19 severity predictive equations (Table S1). Ten studies from three different countries (China, Italy, and the United States of America), with annotated patient data, were used to develop COVID-19 severity predictive equations. In these cohorts, severe patients were those that were managed in the intensive care unit (ICU) with symptoms of organ failure or hypoxemic respiratory failure leading to intubation and ventilation, while those who were nonsevere were either patients that did not require hospital admissions or those that only placed on low supplemental oxygen requirements. Using multiple regression analysis, two equations were generated (Table 1). Equation 1 consisted of the four variables:

absolute lymphocyte count (LYM; K/μL), absolute neutrophil count (NEU; K/μL), CRP (mg/L), and D-dimer levels (DD; mg/L). Equation 1 was defined as $y = 0.97 - 0.92 \times$ (LYM) $+ 0.070 \times$ (NEU) $+ 0.0038 \times$ (CRP) $+ 0.033 \times$ (DD). Equation 2 included the variables LYM, NEU, and CRP and was defined as $y = 0.79 - 0.82 \times$ (LYM) $+ 0.090 \times$ (NEU) $+ 0.0045 \times$ (DD). Numerical ranges for predicted severity outcomes were evaluated across all values at iterations of 0.1. The optimal y value range for predicting outcomes was identified as: $y < 0.5$ predicted to be nonsevere, $y > 0.8$ predicted to be severe, and y between 0.5 and 0.8 considered inconclusive.

3.2 | Validation of the COVID-19 severity prediction equations in adult patients

Both COVID-19 severity predictor equations were tested against validation data sets in the adult population (Tables S2 and S3). The validation set for Equation 1 included 62 adult patients with all four values available for analysis. The median age was 62.5 years (range 24-91 years), with 28 females and 34 males. Evaluation of Equation 1 resulted in the following data: 0.76 sensitivity, 0.79 specificity, 0.73 positive predictive value (PPV), and 0.82 negative predictive value (NPV) with a test yield of 79% (Table 2). Equation 2 was evaluated against a data set that included 138 adult patients with all three variables available for analysis. The median age was 60 years (range 24-94 years), with 68 females and 70 males. The test yield was 84% (percentage of cases that could be classified and resulted in the following outcome data: 0.68 sensitivity, 0.83 specificity, 0.68 PPV, and 0.83 NPV.

3.3 | Assessment of the COVID-19 severity prediction equations in pediatric patients

As pediatric populations have been shown to have altered responses to SARS-CoV-2,[11] we tested both equations that were generated using adult data sets in pediatric cases (Tables S4 and S5). The validation set for Equation 1 included 18 patients where all four values were available. The median age was 42 months (range 2-156 months),

TABLE 2 Evaluation of the performance of COVID-19 predictor equations in adult patients

Equation	Test yield (%)	Positive predictive value	Negative predictive value	Sensitivity	Specificity
1	79	0.73	0.82	0.76	0.79
2	84	0.68	0.83	0.68	0.83

Note: Test yield = percentage of cases that can be classified.

TABLE 3 Evaluation of the performance of COVID-19 predictor equations in pediatric patients

Equation	Test yield (%)	Positive predictive value	Negative predictive value	Sensitivity	Specificity
1	89	1.00	0.64	0.29	1.00
2	92	1.00	0.68	0.13	1.00

Note: Test yield = percentage of cases that can be classified.

with nine females and nine males. Evaluation of Equation 1 resulted in the following data: 0.29 sensitivity, 1.00 specificity, 1.00 PPV, and 0.64 NPV with a test yield of 89% (Table 3). Equation 2 was evaluated against a data set that included 25 patients where all three variables were available. The median age was 48 months (range 2-168 months), with 13 females and 12 males. The test yield was 92% (percentage of cases that could be classified and resulted in the following outcome data: 0.13 sensitivity, 1.00 specificity, 1.00 PPV, and 0.68 NPV.

4 | DISCUSSION

Over the past several months, clinicians and researchers have been working to understand the 2019-nCoV that has afflicted many countries and continues to spread throughout the community. Our study aimed to analyze presently published data and compare clinical outcomes along with determining laboratory blood data that may be predictors of disease severity. There have been many working theories on how to manage COVID-19–positive patients and what biologic markers can be used to assess these patients.[12-14] To date, there has not been a potential multiple regression routine diagnostic equation purposed that could triage patients and allow clinicians to determine management at day of presentation to a healthcare facility with symptomatic COVID-19 infections and positive PCR results.

Here, we report a meta-analysis of all English-language published articles including laboratory-confirmed COVID-19 patients and their clinical data during hospitalization. In this cohort, we extrapolated common hematological markers used to determine clinical severity and disease progression based on current literature.[15-17] We observed that common markers that influenced patient outcomes were white blood cell values and acute phase reactants. Interestingly, we noted that neutrophil count, lymphocyte count, CRP, and D-dimer levels showed trends on whether a patient would have a mild-to-moderate course of disease vs a severe course or eventual death. Specifically, neutropenia, lymphopenia along with elevated CRP, and D-dimer levels were associated with progression, while moderate changes to these values were seen in patients that had mild-to-moderate disease. Several studies have looked at these values individually but have not compared them in totality.

Based on these findings, we created two COVID-19 severity predictor regression equations demonstrating relationships between the predictor variables (gathered laboratory data) and the outcome variable (disease severity). The equations were tested against an adult validation set. Power analysis of multiple regression equations showed significant power of >0.9 at an anticipated effect size of 0.3, adequate for achieving statistical significance <0.05. Testing showed that Equation 1 had a higher positive predictive value (73%) as compared to Equation 2 (68%). This finding suggested that Equation 1 with four predicator values was more accurate. However, Equation 2 had a higher test yield (84% vs 79%) in comparison with Equation 1, indicating that Equation 2 is more precise despite having only three predicator values. Of note, given that there are no published data on comparing these variables, we set arbitrary diagnostic cutoffs, which explains the cases with inconclusive outcomes.

COVID-19 initially was thought to only affect the adult population; however, as cases continued to rise throughout the world it was quickly seen that more children were testing positive with contrasting symptoms to that seen in adults.[18] Given that this disease was not recognized in children or adolescents until recently, there is a lack a published data regarding this patient population. We were able to identify 43 reported cases and test our regression equation against these data to see whether age was not a factor in our predictive formula. Both equations had a 100% positive predictive value along with specificity and a high-test yield, but sensitivity was suboptimal. However, the data set we used here was small, and as such, these results require evaluation in a larger data set of pediatric patients.

Our results of predicative outcomes show that there is a relationship between these commonly acquired laboratory values. As compared to current outcome critical care predictors such as APACHE II and SOFA scores, our approach augments these models by combining only blood laboratory values. Our results of predicative outcomes from children and adults show that there is a relationship between these commonly acquired laboratory values. These findings could serve as a clinical diagnostic tool to triage patients and determine whether patients would be eligible for specific treatments given their severity score. To date, studies have shown that early intervention in severe cases with newly approved drugs for treating COVID-19 such as remdesivir[19] and dexamethasone[20] has shown to statistically improve mortality rates.

Due to this study being a retrospective meta-analysis, there are limitations that might lead to potential biases, such as standardized average laboratory values over hospitalization course and intrahospital laboratory variances. We also recognize that another limitation may be on how our regression equation may be applied in an outpatient basis given that most infections may be asymptomatic or do not

require hospitalizations. However, we aim to define potentially hospitalized patients as they have shown to cause a burden on healthcare systems along with filtering out patients that may have a nonsevere clinical course. Our results do show that the regression equations purposed can be performed in a controlled prospective environment that would provide validation and eliminate these biases. Our severity outcome score could also potentially link patients to ongoing treatment trials. Furthermore, hospital systems could use this as a screening tool to determine potential demand on the healthcare infrastructure (ie, ICU-bed usage, ventilator capacity). In summary, our study defines a novel diagnostic algorithm that when clinically implemented at day 0 of presenting with COVID-19–like symptoms or COVID-19 PCR–confirmed disease, these easily acquirable laboratory data points along with our regression algorithm would impact patient care and potentially reduce mortality in COVID-19 patients.

CONFLICT OF INTEREST

The authors have no significant conflicts of interests to note.

AUTHOR CONTRIBUTIONS

K Singh performed data collection, data analysis, and wrote the paper. S Mittal performed data collection, data analysis, and wrote the paper. S Gollapudi and A Butzmann performed data collection and edited the manuscript. J Kumar performed data analysis and wrote the manuscript. RS Ohgami conceived of the idea for this manuscript, performed data analysis, and wrote the paper.

ORCID

Kunwar Singh https://orcid.org/0000-0002-9024-5751
Robert S. Ohgami https://orcid.org/0000-0003-1881-3440

REFERENCES

1. Pneumonia of unknown cause – China; 2020. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/. Accessed July 15, 2020.
2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733.
3. Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*. 2013;348(20):1967-1976. https://pubmed.ncbi.nlm.nih.gov/12690091/
4. Zaki AM, Boheemen SV, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19):1814-1820. https://doi.org/10.1056/nejmoa1211721
5. Tan WJ, Zhao X, Ma XJ, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020. *China CDC Weekly*. 2020;2:61-62.
6. Rambaut A.Preliminary Phylogenetic Analysis of 11 nCoV2019 Genomes; 2020. http://virological.org/t/preliminary-phylogenetic-analysis-of-11-ncov2019-genomes-2020-01-19/329. Accessed February 12, 2020

7. Coronavirus Disease (COVID-19) Situation Reports. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ Accessed July 15, 2020.
8. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe*. 2020;27(6):992-1000.
9. Tan L, Wang Q, Zhang D, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Sig Transduct Target Ther*. 2020;5:33.
10. Lindsley AW, Schwartz JT, Rothenberg ME. Eosinophil responses during COVID-19 infections and coronavirus vaccination. *J Allergy Clin Immunol*. 2020;146(1):1-7.
11. Molloy EJ, Bearer CF. COVID-19 in children and altered inflammatory responses. *Pediatr Res*. 2020;88(3):340-341.
12. Tabata S, Imai K, Kawano S, et al. Clinical characteristics of COVID-19 in 104 people with SARS-CoV-2 infection on the Diamond Princess cruise ship: a retrospective analysis. *Lancet Infect Dis*. 2020;20(9):1043-1050.
13. Bonetti G, Manelli F, Patroni A, et al. Laboratory predictors of death from coronavirus disease 2019 (COVID-19) in the area of Valcamonica, Italy. *Clin Chem Lab Med*. 2020;58(7):1100-1105.
14. Cummings MJ, Baldwin MR, Abrams D, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet*. 2020;395(10239):1763-1770.
15. Wang L. C-reactive protein levels in the early stage of COVID-19. *Méd Mal Infect*. 2020;50(4):332-334.
16. Li Q, Ding X, Xia G, et al. Eosinopenia and elevated C-reactive protein facilitate triage of COVID-19 patients in fever clinic: a retrospective case-control study. *EClinicalMedicine*. 2020;23:100375.
17. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J Transl Med*. 2020;18(1): https://doi.org/10.1186/s12967-020-02374-0.
18. Götzinger F, Santiago-García B, Noguera-Julián A, et al. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. *Lancet Child AdolescHealth*. 2020;4:653-661.
19. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19 — preliminary report. *N Engl J Med*. 2020. https://doi.org/10.1056/nejmoa2007764
20. Horby P, Lim WS, Emberson J, et al. Effect of dexamethasone in hospitalized patients with COVID-19: preliminary report. 2020. https://doi.org/10.1101/2020.06.22.20137273

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.