

## ORIGINAL CONTRIBUTION

# Inter-rater reliability and prospective validation of a clinical prediction rule for SARS-CoV-2 infection

Adam E. Nevel MD | Jeffrey A. Kline MD 

Department of Emergency Medicine,  
Indiana University School of Medicine,  
Indianapolis, Indiana, USA

**Correspondence**

Jeffrey A. Kline, MD, Department of  
Emergency Medicine, Indiana University  
School of Medicine, Indianapolis, IN  
46033, USA.

Email: jefkline@iu.edu

**Supervising Editor:** Mark B. Mycyk.

**Abstract**

**Objectives:** Accurate estimation of the risk of SARS-CoV-2 infection based on bedside data alone has importance to emergency department (ED) operations and throughput. The 13-item CORC (COVID [or coronavirus] Rule-out Criteria) rule had good overall diagnostic accuracy in retrospective derivation and validation. The objective of this study was to prospectively test the inter-rater reliability and diagnostic accuracy of the CORC score and rule (score  $\leq 0$  negative,  $> 0$  positive) and compare the CORC rule performance with physician gestalt.

**Methods:** This noninterventive study was conducted at an urban academic ED from February 2021 to March 2021. Two practitioners were approached by research coordinators and asked to independently complete a form capturing the CORC criteria for their shared patient and their gestalt binary prediction of the SARS-CoV-2 test result and confidence (0%–100%). The criterion standard for SARS-CoV-2 was from reverse transcriptase polymerase chain reaction performed on a nasopharyngeal swab. The primary analysis was from weighted Cohen's kappa and likelihood ratios (LRs).

**Results:** For 928 patients, agreement between observers was good for the total CORC score,  $\kappa = 0.613$  (95% confidence interval [CI] = 0.579–0.646), and for the CORC rule,  $\kappa = 0.644$  (95% CI = 0.591–0.697). The agreement for clinician gestalt binary determination of SARS-CoV-2 status was  $\kappa = 0.534$  (95% CI = 0.437–0.632) with median confidence of 76% (first–third quartile = 66–88.5). For 425 patients who had the criterion standard, a negative CORC rule (both observers scored CORC  $< 0$ ), the sensitivity was 88%, and specificity was 51%, with a negative LR (LR $^-$ ) of 0.24 (95% CI = 0.10–0.50). Among patients with a mean CORC score of  $> 4$ , the prevalence of a positive SARS-CoV-2 test was 58% (95% CI = 28%–85%) and positive LR was 13.1 (95% CI = 4.5–37.2). Clinician gestalt demonstrated a sensitivity of 51% and specificity of 86% with a LR $^-$  of 0.57 (95% CI = 0.39–0.74).

**Conclusion:** In this prospective study, the CORC score and rule demonstrated good inter-rater reliability and reproducible diagnostic accuracy for estimating the pretest probability of SARS-CoV-2 infection.

**KEYWORDS**

COVID-19, decision making, diagnosis, multivariable analysis, probability, prognosis, registries, risk, SARS-CoV-2

## INTRODUCTION

Emergency care clinical operations require accurate risk stratification for SARS-CoV-2 infection. This concern starts at the triage area in which practitioners (often nurses) must make time-sensitive decisions regarding safe placement of patients in the emergency department (ED). This necessity amplifies during times of “surge capacity” in which the number of active patients exceeds the number of ED rooms, forcing the consideration of nonideal patient care spaces, such as ED hallway beds, for infectious patients. Aimed at addressing this urgent need, the large body of recent literature addressing COVID-19 has included many decision rules for the prognosis of patients with suspected SARS-CoV-2 as well as decision rules that estimate of probability of SARS-CoV-2 infection using knowledge of laboratory and radiological data.<sup>1-8</sup> However, far fewer studies have reported structured prediction rules designed to estimate the pretest probability of SARS-CoV-2 at the bedside without the use of laboratory or radiological imaging.<sup>9-11</sup> Additionally, only a handful of studies have accompanied these SARS-CoV-2 prediction rules with specific cutoffs that represent low-probability scenarios for which practitioners can forego diagnostic testing.<sup>12-16</sup> Fewer studies have examined the diagnostic accuracy of physician gestalt for SARS-CoV-2 infection.<sup>17,18</sup> The studies that have been published on prediction have been criticized for high bias, and none have undergone full stages of validation. This includes (a) inter-rater reliability testing of the components; (b) overall categorization of pretest probability; (c) independent, prospective diagnostic performance in real practice; and (d) comparison with physician gestalt.<sup>19-21</sup>

To help address this gap, this report prospectively tests the inter-rater reliability and diagnostic accuracy of the CORC rule (COVID Rule-out Criteria) and compares its performance with physician gestalt. The CORC rule was derived and validated from data collected retrospectively from 19,850 patients, from 116 hospitals in 22 states using the RECOVER network.<sup>22</sup> The CORC rule consists of 13 component criteria, each scored with either a +1 (age > 50 years, Black race, Latin/Hispanic ethnicity, residential exposure to COVID-19, history of fever, history of myalgias, history of cough, history of loss of taste or smell, triage SpO<sub>2</sub> reading < 95%, triage temperature ≥ 38°C) or -1 (White race, no known exposure to COVID-19, current smoker). When compared against a criterion standard of reverse transcriptase polymerase chain reaction (rt-PCR) testing on a same-day nasopharyngeal swab sample, a CORC score of 0 or less produced a negative likelihood ratio (LR-) of 0.22 (95% confidence interval [CI] = 0.19-0.26) in the validation subset (*n* = 9,925). In this report, the primary goal is to assess the inter-rater reliability of a CORC score of 0 from prospectively assessed data in one ED in the United States in the spring of 2021.

## MATERIALS AND METHODS

This was a noninterventional study conducted at Indiana University Health Methodist Hospital ED in Indianapolis, Indiana, from

February 2021 to March 2021. The Indiana University School of Medicine Institutional Review Board reviewed and deemed the protocol exempted as non-human subjects research. Part of this requirement was that all data had to be collected on the same day and protected health information could not be recorded. These requirements precluded any follow-up beyond the day of enrollment. The primary objective of the study was to test inter-rater reliability, with a secondary objective to test the diagnostic accuracy of the CORC rule. The authors anticipate that the greatest value of CORC is as a screening tool in high-throughput areas, administered by any member of a health care team. Accordingly, the only exclusion criteria for patients was inability to communicate answers regarding symptoms required for CORC (recent fever, myalgia, or loss of taste or smell). For the purpose of testing diagnostic accuracy, we also included patients undergoing testing for SARS-CoV-2 as part of usual care during the times when a coordinator was present. A trained research coordinator was positioned in the ED during daytime hours, 7 a.m. to 6 p.m., 6 days per week, and approached two practitioners about completing a 15-item paper research form. The data collection form is shown in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1111/acem.14309/full>). This data collection form included all elements of the CORC rule, as well as questions investigating any known SARS-CoV-2-positive test for the patient within 14 days. For cases in which diagnoses were not yet known, the form asked for the evaluator's gestalt estimate (yes or no) whether the patient was SARS-CoV-2 positive as well as their confidence in this estimate on a 0%-100% scale. The forms could be completed by physicians (including residents and faculty), medical students, or nurses, or if two ED team members could not reliably complete the form, the second observer was a member of the research team. All observers were approached as soon as both had seen the patient and always before the results of diagnostic testing for SARS-CoV-2. A staff assistant transferred data from paper forms into a REDCap electronic form. A random sample of 10% (*n* = 90 patients, or 180 forms) of electronic entries in REDCap were compared against the paper form by an independent research coordinator to ensure fidelity of transfer for CORC components.

The criterion standard for the subgroup with SARS-CoV-2 testing was the result from an rt-PCR performed on a nasopharyngeal swab. The test platform was the Cobas Liat system (Roche Molecular Diagnostics).

## Sample size

The sample size was predicated on the primary outcome of the kappa ( $\kappa$ ) coefficient, applied to the binary CORC rule negative or positive, which is defined as a CORC score of <0 or >0 points, respectively; we expected approximately a 50% distribution in each category. Therefore, by iterative calculation, to narrow the 95% CI around  $\kappa$  to <10% would require approximately 900 patients. Within this 900 we also sought to enroll approximately 400 patients who

were tested for SARS-CoV-2 and expected 10% (40) to be positive, and based on the previous validation population, we expected 90% diagnostic sensitivity for the CORC rule, thus allowing the 95% CI around this proportion to be <10%.

## Data analysis

The analysis refers to the CORC score as the total numeric value from the 13 items and the CORC rule as the binary result of CORC rule positive (>0) or negative ( $\leq 0$ ). The primary analysis was the weighted Cohen's kappa for the binary output of the CORC rule and the total CORC score, each component of the CORC score, and the binary gestalt prediction of SARS-CoV-2 status by providers. For calculation of  $\kappa$  for gestalt, we only used encounters where both observers were either physicians or advanced practice providers (nurse practitioners or physician assistants). Diagnostic performance was assessed from the area under the receiver operating characteristic curve (AUROC) and from  $2 \times 2$  contingency tables, using the result of rtPCR test done on the same day as the criterion standard. Patients who were not tested for SARS-CoV-2 or who were known positive within 14 days were excluded from diagnostic performance testing. Statistical analyses were performed with SPSS software, version 27.

## RESULTS

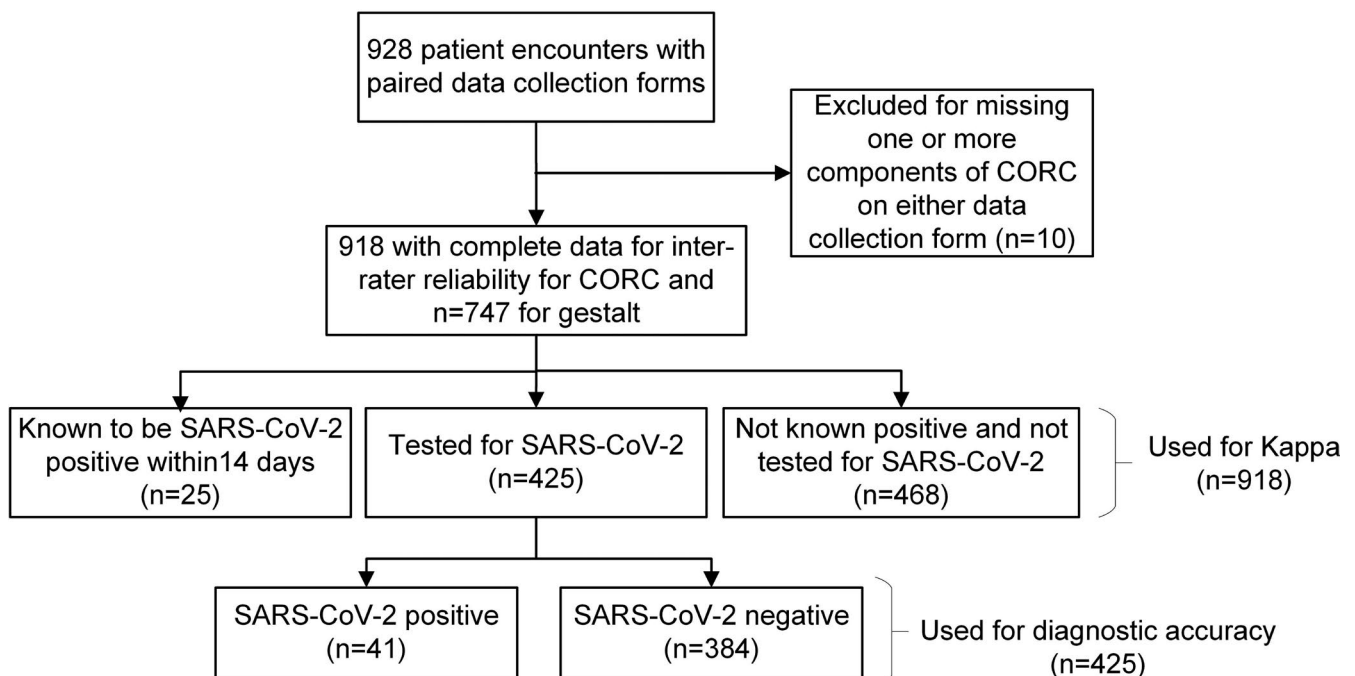
Paired forms from two providers were obtained from 928 patient encounters, completed by faculty (39%), residents (29%), nurses

(15%), advanced practitioners (11%), research personnel (11%), and medical students (2%). Data from 10 encounters were excluded because one or more component of CORC score was missing. Figure 1 shows the breakdown of patients into four categories: patients known to be positive for SARS-CoV-2 within 14 days ( $n = 25$ ), patients who tested positive for SARS-CoV-2 ( $n = 41$ , 9.6% prevalence), patients who tested negative for SARS-CoV-2 ( $n = 384$ ), and those not tested but used for assessment of agreement ( $n = 468$ ). Gestalt estimates from two clinicians were available for 741 encounters, including 415 with the criterion standard for SARS-CoV-2.

## Inter-rater reliability

The raw agreement for the numeric CORC score was 53% and for the binary CORC rule was 84%. Table 1 presents the details of the main findings for agreement. Figure 2 shows a Bland-Altman plot for the difference between total CORC score versus the average of two observers. The 95% limits of agreement were  $-2.4$  to  $+2.5$  and the intraclass correlation coefficient was 0.75 (95% CI = 0.72–0.78).

The range of  $\kappa$  values for each component of was a low of  $\kappa = 0.45$ , (95% CI = 0.375–0.525), for symptoms of muscle aches, to a high of  $\kappa = 0.921$  (95% CI = 0.896–0.945) for Black race. Overall, the agreement for the total CORC score was good ( $\kappa = 0.613$ , 95% CI = 0.579–0.646) as was the agreement for CORC rule negative ( $\kappa = 0.644$ , 95% CI = 0.591–0.697). The agreement for clinician gestalt binary determination of SARS-CoV-2 status yielded  $\kappa = 0.534$  (95% CI = 0.437–0.632).

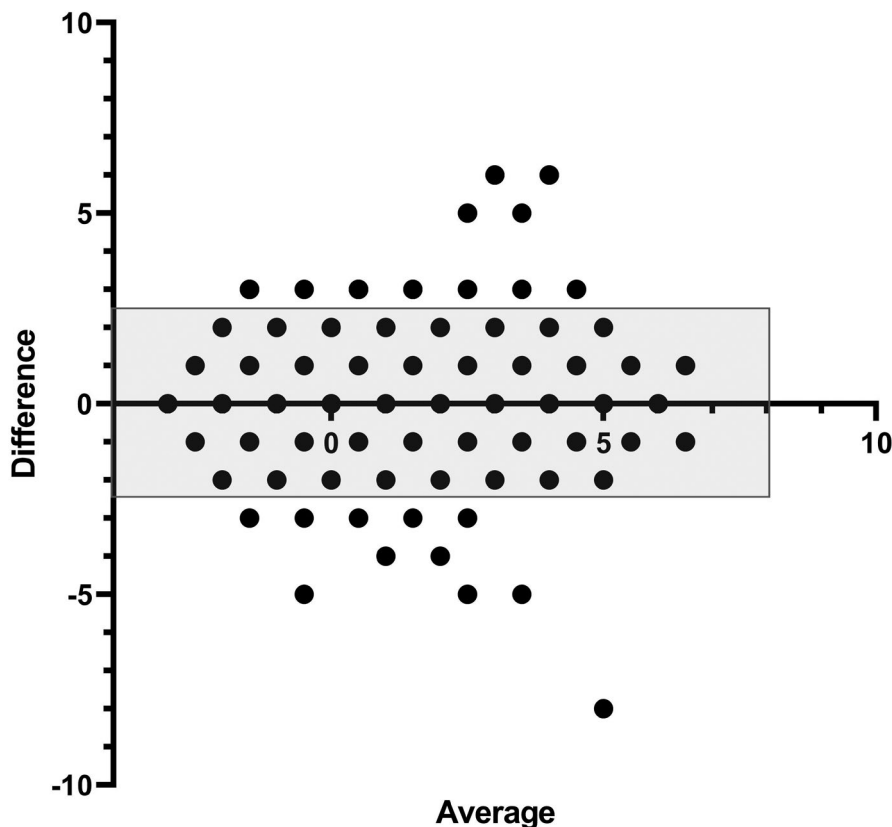


**FIGURE 1** Flow diagram of patient encounters

Variable	Observer 1	Observer 2	Weighted kappa	Lower bound	Upper bound
Black race	44%	45%	0.921	0.896	0.945
White race	51%	50%	0.891	0.863	0.920
Hispanic or Latino ethnicity	4%	4%	0.712	0.604	0.819
Age > 50 years	49%	49%	0.865	0.833	0.896
Symptom					
Loss of sense of taste or smell	3%	3%	0.579	0.459	0.699
Nonproductive cough	19%	14%	0.590	0.524	0.657
Fever	11%	11%	0.654	0.580	0.728
Muscle aches	16%	14%	0.450	0.375	0.525
Exposure to COVID-19	8%	7%	0.522	0.429	0.614
Residential contact with known or suspected COVID-19 infection	3%	3%	0.499	0.372	0.626
Pulse oximetry at triage < 95%	13%	14%	0.638	0.566	0.709
Temperature > 37.5°C	5%	5%	0.651	0.540	0.761
Current smoker	25%	26%	0.589	0.532	0.647
Total CORC score (mean)	0.0	0.0	0.613	0.579	0.646
CORC rule (score ≤ 0)	67%	69%	0.644	0.591	0.697
Gestalt negative	12%	10%	0.534	0.437	0.632

**TABLE 1** Inter-rater reliability for the CORC score and rule

Abbreviation: CORC, COVID-19 (coronavirus) rule out criteria.

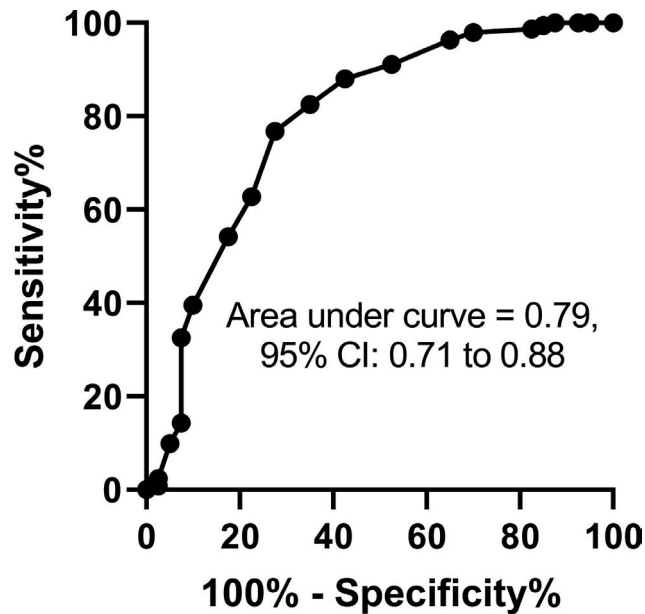


**FIGURE 2** Bland-Altman plot showing the difference in the score from the CORC score between two independent observers on the Y-axis, plotted as a function of the average of the two reviewer scores on the X-axis. The shaded rectangle contains 95% confidence limits for the Y-axis data (-2.4 to +2.5). CORC, COVID-19 Rule-out Criteria

### Diagnostic performance of the CORC score and rule

In the 425 patients who had the criterion standard, the overall diagnostic accuracy of the mean total CORC score from two observers

as indicated by the AUROC (Figure 3) was 0.79 (95% CI = 0.70–0.88). Regarding the binary CORC rule performance, Table 2 shows the 2 × 2 contingency table and associated diagnostic indexes. For the CORC rule to be considered negative in Table 2, both observers



**FIGURE 3** Receiver operating characteristic curve for the mean of the total CORC score from two observers for the criterion standard of a positive same-day nucleic acid test on a nasopharyngeal swab for SARS-CoV-2. CORC, COVID-19 Rule-out Criteria

**TABLE 2** Diagnostic performance of the CORC rule

	SARS test same day	
	+	-
CORC rule result		
+ (score > 0)	36	189
- (score ≤ 0)	5	195
Index	Value	95% CI
Prevalence	9.6%	6.9%–12.8%
False negative rate	2.5%	0.8%–5.6%
Sensitivity	88%	74%–95%
Specificity	51%	46%–57%
LR-	0.24	0.10–0.50

Abbreviation: CORC, Coronavirus rule out criteria.

had to have a CORC score of 0 or less. With this requirement, the LR- was 0.24 (95% CI = 0.10–0.50). Using the alternative standard of either observer negative, the sensitivity of the CORC rule decreased to 73% (95% CI = 57%–85%), and the specificity increased to 68% (95% CI = 63%–72%), yielding a likelihood ratio of 0.40 (95% CI = 0.23–0.62).

Among patients with a CORC score of >4 (mean of two observers), the prevalence (i.e., predictive value positive or posterior probability) of a positive SARS-CoV-2 rtPCR was 58% (95% CI = 28%–85%) and the positive LR was 13.1 (95% CI = 4.5–37.2). For the 25 patients with known SARS-CoV-2 infection within the previous 14 days, the CORC rule was positive (>0) in 15 of 20, (sensitivity = 80%, 95% CI = 61%–92%). Data Supplement S1 shows the sensitivity,

specificity, and LR- for the 11 CORC components. Individually, none had a LR- lower than 0.60.

### Diagnostic performance of gestalt

Table 3 shows the same diagnostic indexes for gestalt as in Table 2 for the CORC rule result. While clinician gestalt clearly had higher specificity (86%, 95% CI = 82%–89%) than the CORC rule (51%, 95% CI = 46%–57%), clinician gestalt sensitivity was significantly lower at 51% (95% CI = 35%–68%), resulting in a significantly higher LR- (0.57, 95% CI = 0.39–0.74). Clinicians' written confidence in their gestalt estimate was similar for patients who tested negative for SARS-CoV-2 (75%, first-third quartile = 65–87.5) as for patients who tested positive (77.5%, first-third quartile = 65%–90%). For the 358 patients not tested for SARS-CoV-2 and not known to be positive within 14 days, and with usable gestalt for two clinicians, one or both clinicians thought the patient was positive in 12 of 358 (9.5%), and clinicians were more confident in their gestalt estimates than for patients who underwent testing (median confidence = 90%, first-third quartile = 82.5–95).

## DISCUSSION

This work tested the inter-rater reliability and diagnostic performance of a 13-component clinical prediction rule designed to assess the pretest probability for SARS-CoV-2 in undifferentiated patients using data commonly available at the bedside. From a total of 918 ED patients scored by two independent observers, we found good overall agreement for the raw CORC score ( $\kappa = 0.61$ , 95% CI = 0.58–0.65) and for the binary CORC rule ( $\kappa = 0.644$ , 95% CI = 0.591–0.697). Among the 425 patients who had the criterion standard of an rtPCR test for SARS-CoV-2 performed on a same-day nasopharyngeal swab, the CORC rule negative (i.e., a CORC score of 0 or less) had a sensitivity of 88%, specificity of 52%, and LR- of 0.24 (95% CI = 0.10–0.50), which is similar to the LR- found in the initial validation, 0.22 (95% CI = 0.19–0.26).<sup>15</sup> Thus, assuming this finding is further validated by others, the CORC rule negative can lead to a very

**TABLE 3** Diagnostic performance of clinician gestalt

	SARS-Cov-2 test same day	
	+	-
Gestalt estimate		
Yes	20	53
No	19	323
Index	Value	95% CI
Prevalence	9.4%	6.7%–12.4%
False negative rate	5.6%	3.3%–8.4%
Sensitivity	51%	35%–68%
Specificity	86%	82%–89%
LR-	0.57	0.39–0.74

Abbreviation: LR-, negative likelihood ratio.

low posterior probability (e.g., <1%), if the underlying prevalence of infection is <5%. The CORC rule outperformed clinician gestalt, both in terms of inter-rater reliability ( $\kappa = 0.534$ , 95% CI = 0.437–0.632) and in terms of diagnostic performance, as evidenced by the diagnostic sensitivity of only 51% (95% CI = 35%–68%), and a LR<sup>-</sup> of 0.57 (95% CI = 0.39–0.74). Perhaps suggestive of the value of the CORC rule is the fact that clinicians had 75% or greater confidence in their estimates, including patients who tested positive and for whom they were wrong in almost half of cases.

To the best of our knowledge, this is the first clinical prediction rule to be tested prospectively, live in the ED setting. The CORC score originated from a national sample of 19,850 patients, randomly split in half for derivation then validation procedures.<sup>15,22</sup> Other clinical prediction rules that were restricted to use of data available at the bedside have proposed three to 11 criteria in the rule, and all were derived on retrospective samples, but none have yet been tested for inter-rater reliability or diagnostic accuracy in an independently and retrospectively collected validation sample.<sup>12–16</sup>

The CORC rule was designed for use in high-throughput areas where personnel could have access to a pulse oximeter and a thermometer, for example, in the triage area of the ED, but also possibly in intake areas for homeless shelters or extended-care facilities. As more of the U.S. population becomes vaccinated against SARS-CoV-2, the subset of persons to which this will be most applicable will likely be those who remain unvaccinated. Whether or not this rule has any applicability to patients who are vaccinated remains uncertain. We did not collect information about vaccination status; this work was completed on March 22, 2021, at which time the Centers for Disease Control and Prevention (CDC) estimated that 48,988,582 Americans were vaccinated.<sup>23</sup> The LR<sup>-</sup> of 0.24 suggests that the CORC rule could produce a posterior probability below 1% in populations with less than a 5% prevalence. In late May 2021, the CDC estimates indicate the 7-day average of percent positivity from tests at 3.4% in the United States.<sup>23</sup> As the proportion of the population fully vaccinated increases, it is likely this positive rate decreases, which could increase the utility of the CORC rule in the general population.

## LIMITATIONS

Limitations include the single-center design and convenience sample. Whether the CORC score, or any other clinical prediction rule for SARS-CoV-2 infection, will remain relevant and accurate after mass vaccination represents an obvious potential limitation to the importance of this work. For example, the CORC score contains 13 components, all weighted equally, and it remains likely that some of these predictors will change in significance, as the underlying prevalence of infection decreases in response to vaccinations. Furthermore, these data were collected in early 2021 and the frequency with which observers marked yes to “household exposure to COVID-19” was only 3%, significantly lower than the 7.2% rate (95% CI = 6.8%–7.5%) that was recorded in the initial derivation and validation sample from 2020.

Likewise, we speculate that the driving forces behind the increased weighting of persons of color higher than that of White patients in the original sample from 2020 may diminish with higher vaccination rates and increased access to care. While children under 18 years of age were included in the original derivation and first validation of CORC, no children were represented in the present sample.

## CONCLUSION

In conclusion, prospectively collected data indicate good inter-rater reliability and reproducible diagnostic accuracy of the CORC score and CORC rule for estimating the pretest probability of SARS-CoV-2 infection.

## ACKNOWLEDGMENTS

The authors thank Patti Hogan and Amanda Klimeck for administrative oversight of the RECOVER Network.

## CONFLICT OF INTEREST

The authors have no potential conflicts to disclose.

## ORCID

Jeffrey A. Kline  <https://orcid.org/0000-0001-7190-3109>

## REFERENCES

1. Mudatsir M, Fajar JK, Wulandari L, et al. Predictors of COVID-19 severity: a systematic review and meta-analysis. *F1000Res*. 2020;9:1107.
2. Gupta RK, Marks M, Samuels TH, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. *Eur Respir J*. 2020;56(6):2003498.
3. Chew WM, Loh CH, Jalali A, et al. A risk prediction score to identify patients at low risk for COVID-19 infection [online ahead of print]. *Singapore Med J*. 2021. <https://doi.org/10.11622/smedj.2021019>
4. Xia Y, Chen W, Ren H, et al. A rapid screening classifier for diagnosing COVID-19. *Int J Biol Sci*. 2021;17(2):539–548.
5. Kurstjens S, van der Horst A, Herpers R, et al. Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med*. 2020;58(9):1587–1593.
6. Fink DL, Khan PY, Goldman N, et al. Development and internal validation of a diagnostic prediction model for COVID-19 at time of admission to hospital [online ahead of print]. *QJM*. 2020. <https://doi.org/10.1093/qjmed/hcaa305>
7. AlJame M, Ahmad I, Imtiaz A, Mohammed A. Ensemble learning model for diagnosing COVID-19 from routine blood tests. *Inform Med Unlocked*. 2020;21:100449.
8. McDonald SA, Medford RJ, Basit MA, Diercks DB, Courtney DM. Derivation with internal validation of a multivariable predictive model to predict COVID-19 test results in emergency department patients. *Acad Emerg Med*. 2021;28:206–214.
9. Carpenter CR, Mudd PA, West CP, Wilber E, Wilber ST. Diagnosing COVID-19 in the emergency department: a scoping review of clinical examinations, laboratory tests, imaging accuracy, and biases. *Acad Emerg Med*. 2020;27:653–670.
10. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.

11. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical predictive models for COVID-19: systematic study. *J Med Internet Res*. 2020;22(10):e21439.
12. Raberahona M, Rakotomalala R, Rakotomijoro E, et al. Clinical and epidemiological features discriminating confirmed COVID-19 patients from SARS-CoV-2 negative patients at screening centres in Madagascar. *Int J Infect Dis*. 2021;103:6-8.
13. Zimmerman RK, Nowalk MP, Bear T, et al. Proposed clinical indicators for efficient screening and testing for COVID-19 infection using Classification and Regression Trees (CART) analysis. *Hum Vaccin Immunother*. 2021;17(4):1109-1112.
14. Trubiano JA, Vogrin S, Smibert OC, et al. COVID-MATCH65-A prospectively derived clinical decision rule for severe acute respiratory syndrome coronavirus 2. *PLoS One*. 2020;15(12):e0243414.
15. Kline JA, Camargo CA, Courtney DM, et al. Clinical prediction rule for SARS-CoV-2 infection from 116 U.S. emergency departments 2-22-2021. *PLoS One*. 2021;16(3):e0248438.
16. van Walraven C, Manuel DG, Desjardins M, Forster AJ. Derivation and internal validation of a model to predict the probability of severe acute respiratory syndrome coronavirus-2 infection in community people. *J Gen Intern Med*. 2021;36(1):162-169.
17. Peyrony O, Marbeuf-Gueye C, Truong VY, et al. Accuracy of emergency department clinical findings for diagnosis of coronavirus disease 2019. *Ann Emerg Med*. 2020;76(4):405-412.
18. Nazerian P, Morello F, Prota A, et al. Diagnostic accuracy of physician's gestalt in suspected COVID-19: prospective bicentric study. *Acad Emerg Med*. 2021;28:404-411.
19. Collins GS, van Smeden M, Riley RD. COVID-19 prediction models should adhere to methodological and reporting standards. *Eur Respir J*. 2020;56(3):2002643.
20. Shamsoddin E. Can medical practitioners rely on prediction models for COVID-19? A systematic review. *Evid Based Dent*. 2020;21(3):84-86.
21. Green SM, Schriger DL, Yealy DM. Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. *Ann Emerg Med*. 2014;64(3):286-291.
22. Kline JA, Pettit KL, Kabrhel C, Courtney DM, Nordenholz KE, Camargo CA. Multicenter registry of United States emergency department patients tested for SARS-CoV-2. *J Am Coll Emerg Physicians Open*. 2020;1(6):1341-1348.
23. Trends in Number of COVID-19 Vaccinations in the US. Centers for Disease Control and Prevention. 2021. Accessed May 22, 2021. <https://covid.cdc.gov/covid-data-tracker/#vaccination-trends>

## SUPPORTING INFORMATION

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

**How to cite this article:** Nevel AE, Kline JA. Inter-rater reliability and prospective validation of a clinical prediction rule for SARS-CoV-2 infection. *Acad Emerg Med*. 2021;28:761-767. <https://doi.org/10.1111/acem.14309>