OPEN

Impact of Surge Strain and Pandemic Progression on Prognostication by an Established COVID-19–Specific Severity Score

Christina Yek, MD[1]

Jing Wang, MS[2]

Jonathan Fintzi, PhD[3]

Alex G. Mancera, MS[1]

Michael B. Keller, MD[1]

Sarah Warner, MPH[1]

Sameer S. Kadri, MD[1]

IMPORTANCE: Many U.S. State crisis standards of care (CSC) guidelines incorporated Sequential Organ Failure Assessment (SOFA), a sepsis-related severity score, in pandemic triage algorithms. However, SOFA performed poorly in COVID-19. Although disease-specific scores may perform better, their prognostic utility over time and in overcrowded care settings remains unclear.

OBJECTIVES: We evaluated prognostication by the modified 4C (m4C) score, a COVID-19–specific prognosticator that demonstrated good predictive capacity early in the pandemic, as a potential tool to standardize triage across time and hospital-surge environments.

DESIGN: Retrospective observational cohort study.

SETTING: Two hundred eighty-one U.S. hospitals in an administrative healthcare dataset.

PARTICIPANTS: A total of 298,379 hospitalized adults with COVID-19 were identified from March 1, 2020, to January 31, 2022. m4C scores were calculated from admission diagnosis codes, vital signs, and laboratory values.

MAIN OUTCOMES AND MEASURES: Hospital-surge index, a severity-weighted measure of COVID-19 caseload, was calculated for each hospital-month. Discrimination of in-hospital mortality by m4C and surge index-adjusted models was measured by area under the receiver operating characteristic curves (AUC). Calibration was assessed by training models on early pandemic waves and measuring fit (deviation from bisector) in subsequent waves.

RESULTS: From March 2020 to January 2022, 298,379 adults with COVID-19 were admitted across 281 U.S. hospitals. m4C adequately discriminated mortality in wave 1 (AUC 0.779 [95% CI, 0.769–0.789]); discrimination was lower in subsequent waves (wave 2: 0.772 [95% CI, 0.765–0.779]; wave 3: 0.746 [95% CI, 0.743–0.750]; delta: 0.707 [95% CI, 0.702–0.712]; omicron: 0.729 [95% CI, 0.721–0.738]). m4C demonstrated reduced calibration in contemporaneous waves that persisted despite periodic recalibration. Performance characteristics were similar with and without adjustment for surge.

CONCLUSIONS AND RELEVANCE: Mortality prediction by the m4C score remained robust to surge strain, making it attractive for when triage is most needed. However, score performance has deteriorated in recent waves. CSC guidelines relying on defined prognosticators, especially for dynamic disease processes like COVID-19, warrant frequent reappraisal to ensure appropriate resource allocation.

KEYWORDS: COVID-19; healthcare rationing; mortality; surge capacity; triage

The COVID-19 pandemic necessitated crisis standards of care (CSC) across the U.S. States, underscoring the need for better guidance on whom to prioritize for life-sustaining treatment (1). During the height of the pandemic, most CSC guidelines (2, 3) identified clinical risk scores to

KEY POINTS

Question: Did performance of the modified 4C score, an established COVID-19 prognosticator, change during the pandemic and in overcrowded hospitals?

Findings: In this retrospective cohort study of 281 U.S. hospitals, the modified 4C score offered adequate discrimination and calibration of mortality risk in COVID-19 that was robust to hospital-surge strain. However, performance worsened in subsequent waves and periodic recalibration was only transiently successful.

Meaning: Prognostic score performance in dynamic disease states may wane over time, warranting their frequent reappraisal to guide application during evolving crisis conditions.

prioritize resource allocation for patients with higher short-term survival odds. Several issues with this approach have since emerged: score selection was often arbitrary and not evidence-based, whereas seemingly objective models designed without equity in mind may have inadvertently discriminated against individuals representing racial/ethnic minorities, lower socioeconomic classes, and advanced ages among other vulnerable groups (1). Notably, the widely adopted Sequential Organ Failure Assessment (SOFA) score has been found to be not much better at predicting postintubation COVID-19 mortality than flipping a coin (3, 4), and its poor calibration in racial minority groups (5) may have exacerbated existing healthcare inequities during the pandemic.

A potential solution involves the targeted use of prognostic scores developed and tested within the populations they are meant to serve. Among COVID-19–specific scores, the 4C score (6) is one of the more rigorously studied (7), with external validation in cohorts across multiple countries (8–10). The 4C score was developed in a United Kingdom cohort of COVID-19 inpatients during the first pandemic wave, with the purpose of predicting disease mortality using seven readily available variables from admission (age, sex, comorbidities, respiratory rate [RR], oxygen saturation, Glasgow Coma Scale score [GCS], urea, and C-reactive protein [CRP]) (6). It has been used to guide

clinical triage (11) and antiviral (remdesivir) usage in the United Kingdom (12), and for risk adjustment in studies evaluating COVID-19 therapeutics (13, 14). A recent study proposed removing CRP and replacing GCS with presence/absence of acute encephalopathy to minimize data missingness in a large U.S. Registry (9). This modified 4C score ("m4C") demonstrated comparable performance to the original 4C score, providing an attractive alternative for large-scale population studies that use electronic datasets, and potential for wider uptake globally. However, the contemporary performance of these scores remains unstudied. Further, score performance in surge-strained care settings (i.e., where triage is most urgently needed), has not been evaluated.

There is a glaring need to critically appraise CSC policies and identify strategies to improve their creation. Here, we leverage a contemporary cohort of hospitalized adults with COVID-19 in the United States to examine performance of the m4C score across time, and with and without the addition of a novel predictor representing hospital-surge strain.

MATERIALS AND METHODS

Study Design

This was a multicenter, retrospective cohort study to validate the m4C score in a contemporary U.S. cohort of adults hospitalized with COVID-19. Given the deidentified nature of the data, institutional ethics board review was not necessary based on the policy of the National Institutes of Health (NIH) Office of Human Subjects Research Protections. Study design and reporting followed Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines (**Supplemental Data 1**, http://links.lww.com/CCX/B282). Data were obtained from the COVID-19 release of the Premier, Inc. Healthcare Database. Adults (≥ 18 yr old) admitted as inpatients between March 1, 2020, and January 31, 2022, were identified at U.S. hospitals that continuously reported encounter data. COVID-19 encounters were defined as those with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) polymerase chain reaction positivity or any encounter diagnosis of COVID-19 (**Table S1**, http://links.lww.com/CCX/B282). Only the first encounter for each patient during the study period was included. Patients with do-not-resuscitate

status present-on-admission ($n = 28,990$) and those without binary sex assignment ($n = 9$) were excluded (**Fig. S1**, http://links.lww.com/CCX/B282). Pandemic waves were defined as: wave 1 (from March 1 to May 31, 2020), wave 2 (from June 1, 2020, to September 30, 2020), wave 3 (from October 1, 2020 to June 30, 2021), delta (dominated by variant B.1.617.2; from July 1, 2021, to December 31, 2021), and early omicron (dominated by variant B.1.1.529; January 1, 2022, to January 31, 2022) based on national estimates of disease and variant prevalence, with rounding to the nearest month to accommodate monthly data reporting in Premier, Inc. (15). The primary outcome was in-hospital mortality, defined as death during hospitalization or discharge to hospice. Outcomes were captured from discharge disposition.

m4C Score Calculation

Age, sex, comorbidities, RR, peripheral oxygen saturation on room air (Spo_2), GCS or acute encephalopathy (neurologic function) and urea were used to calculate m4C (**Table S2**, http://links.lww.com/CCX/B282); CRP was additionally used to calculate 4C (**Table S3**, http://links.lww.com/CCX/B282) (6). Missingness of score components (**Table S4**, http://links.lww.com/CCX/B282) was handled by: 1) assigning highest scores for specific organ systems where other measures of organ failure were present at admission and 2) imputation using multiple imputation by chained equations (**Supplemental Data 2** and **Table S5**, http://links.lww.com/CCX/B282). A sensitivity analysis was conducted using only complete case records for which all m4C components were available. Two additional comparators were derived: the original 4C score, as a benchmark against which to evaluate m4C performance, and a modification of m4C excluding age, to evaluate a potential age-agnostic score that would address concerns of ageism in CSC (Table S3, http://links.lww.com/CCX/B282) (1).

Surge Index Calculation

Surge index, a severity-weighted measure of COVID-19 caseload relative to prepandemic bed capacity, was calculated for each hospital-month during the study period (**Supplemental Data 3**, http://links.lww.com/CCX/B282) (16). This measure of surge was selected due to its established association (both in categorical and continuous forms) with COVID-19 mortality (16) and the hypothesis that a COVID-19 triage tool would be most applicable in a disease-specific hospital-surge scenario. To reaffirm the surge-mortality relationship in this study cohort, surge categories were defined by examining the distribution of surge indices for every hospital-month in which there were greater than 1 COVID-19 inpatient admissions, then stratifying hospital-months by percentile. Categories of low, moderate, and high surge were defined for hospital-months within less than 50th, greater than or equal to 50th to less than 75th, and greater than or equal to 75th percentiles, respectively. Monthly crude mortality of COVID-19 patients across all hospitals was stratified by surge category.

External Validation of m4C Score

Discriminative function of score models was tested with performance of each assessed by area under the receiver operating characteristic curve (AUC). First, AUC of 4C was compared against that of m4C and m4C without age, using data from wave 1 to best approximate the period in which 4C was originally derived. Next, we explored performance of m4C over time by comparing AUC of m4C for each pandemic wave. Point estimates and CIs were generated from multiple imputed datasets using pooling and bootstrap methods, respectively (**Supplemental Data 4**, http://links.lww.com/CCX/B282). For 4C, a cutoff of greater than or equal to 9 was selected to optimally rule in mortality (6). The m4C score excludes CRP and, therefore, represents a different range of values than 4C. Hence, a cutoff of greater than or equal to 6 was selected to approximate performance metrics of 4C (**Table S6**, http://links.lww.com/CCX/B282). AUC differences between waves and corresponding SEs were computed using Delong's algorithm. Rubin's rules were applied to obtain the 95% CIs and p values of AUC differences.

m4C Score Calibration

To examine score calibration across pandemic waves, we had to first define model coefficients, which were not reported in either original development cohorts for 4C or m4C (6, 9). To do this, we developed a generalized additive model with mortality as the outcome and m4C as the primary exposure variable and trained this model on wave 1 data. We performed 10-fold

cross-validation to obtain the predicted probabilities of the model. Coefficients extracted from wave 1 models were applied to subsequent waves to compute the predicted probabilities of mortality, and the above procedure was repeated for each of the 100 imputed datasets. To generate calibration belts, predicted probabilities were averaged across the 100 imputations per pandemic wave and plotted against the observed mortality with Locally Weighted Scatterplot Smoothing. This process was then repeated by training the model on 1) wave 3 data and 2) delta wave data to test whether updated model coefficients demonstrated improved performance in early data from the omicron wave.

Surge Index Model

To test whether m4C performance was impacted by surge-strain, a generalized additive model with mortality as the outcome and m4C and log-surge index as primary exposure variables was defined ("surge-adjusted model") and compared with the model of m4C alone. Log-surge index was selected due to its linear relationship with COVID-19 mortality (16) and applied as a continuous variable in this spline model to allow for greater power over using a categorical version of surge. As with m4C alone, the surge-adjusted model was cross validated on wave 1 data and applied to 100 imputed datasets for subsequent waves.

Statistical Tools

Analyses were performed using *R* software (version 4.1.3) and the following packages: "mice" for multiple imputations, "pROC" to calculate AUCs, "mgcv" for construction of generalized additive models. This work used the computational resources of the NIH high-performance computing cluster.

RESULTS

Cohort Characteristics

From March 2020 to January 2022, 298,379 adults with COVID-19 were admitted across 281 U.S. healthcare facilities in Premier, Inc. (**Table S7**, http://links.lww.com/CCX/B282). Median age was 62 years, and racial/ethnic makeup was mixed with the majority of patients identifying as non-Hispanic White (58%). Most hospitals were urban (83%) and located in the South (66%). Crude in-hospital mortality (death or discharge to

hospice) was 12.0%, versus 32.2% and 16.9% in the original 4C and m4C development cohorts, respectively (**Table S8**, http://links.lww.com/CCX/B282). We subdivided the cohort into five "waves" using national estimates of COVID-19 prevalence and distinction of delta and early omicron variant-predominant periods, on the basis that this would mirror trends in hospital COVID-19 caseloads and therefore surge strain. Crude mortality in hospitalized patients with COVID-19 was highest during wave 1 (14.7%) followed by delta (13.0%), and lowest during early omicron (10.3%). When stratified by m4C, mortality rates were highest during wave 1 for m4C greater than or equal to 9 and highest during delta for m4C less than or equal to 8 (**Fig. S2** and **Table S9**, http://links.lww.com/CCX/B282).

Relationship Between Surge and Mortality

There were greater than 1 COVID-19 inpatient admissions in 5155 hospital-months across the 23-month study period. Surges were greatest during early omicron, with 170 of 215 (79.1%) hospital-months categorized as high surge (surge index greater than or equal to 75th percentile), and lowest during wave 1, with 416 of 528 (78.8%) hospital-months categorized as low surge (surge index < 50th percentile) (**Fig. 1*A***; and **Table S10**, http://links.lww.com/CCX/B282). Unadjusted mortality was higher in moderate-surge and high-surge hospital-months (11.3% and 13.1%, respectively) compared with low-surge hospital-months (9.07%) (**Table S11**, http://links.lww.com/CCX/B282). This relationship was maintained across all waves apart from early omicron (**Fig. 1*B***).

COVID-19 Mortality Prediction Using m4C

We first compared performance of m4C to the original 4C score and m4C without age using receiver operating characteristic curves of the three scores generated with wave 1 data (**Fig. 2*A***; and **Table S12**, http://links.lww.com/CCX/B282). AUC of m4C (0.779 [95% CI, 0.769–0.789]) in this cohort was similar to that in the original m4C development cohort (0.786 [95% CI, 0.773–0.799]) (8) as well as that of 4C in this cohort (0.783 [95% CI, 0.773–0.793]) and in the original 4C development cohort (0.79 [95% CI, 0.78–0.79]) (6). Although m4C without age demonstrated modest discriminative ability (AUC 0.734 [95% CI, 0.723–0.744]), this was significantly reduced compared with that of m4C and 4C.
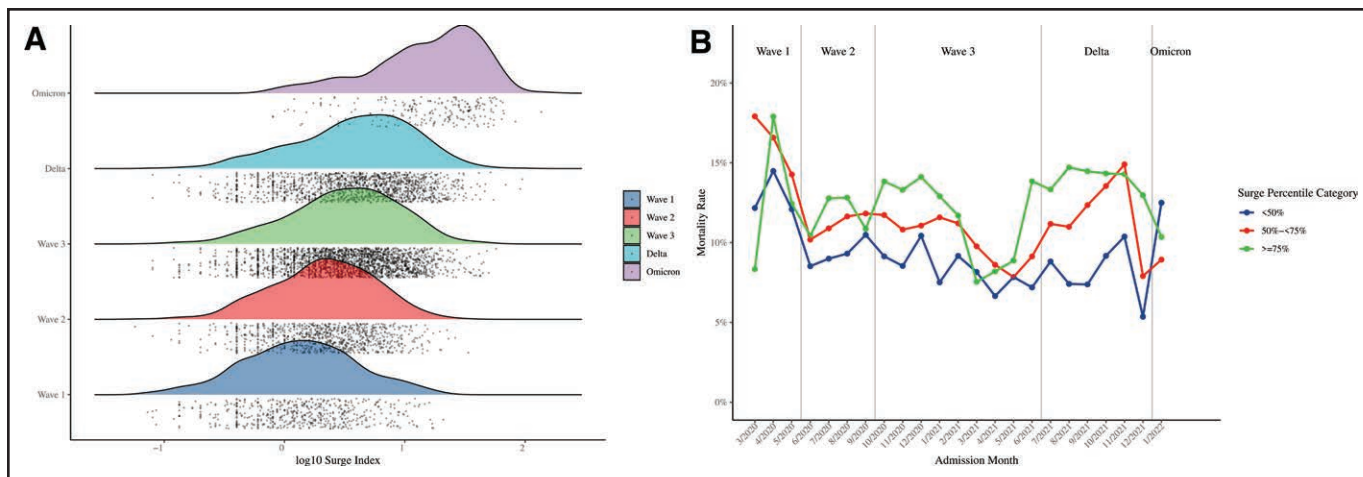
Figure 1. A, Distribution of log-transformed hospital-month surge index by pandemic wave. Individual hospital-months and cumulative distributions are represented by *dots* and *colored shaded* areas, respectively. **B**, Crude mortality rates by hospital-month surge index percentile category across pandemic waves.
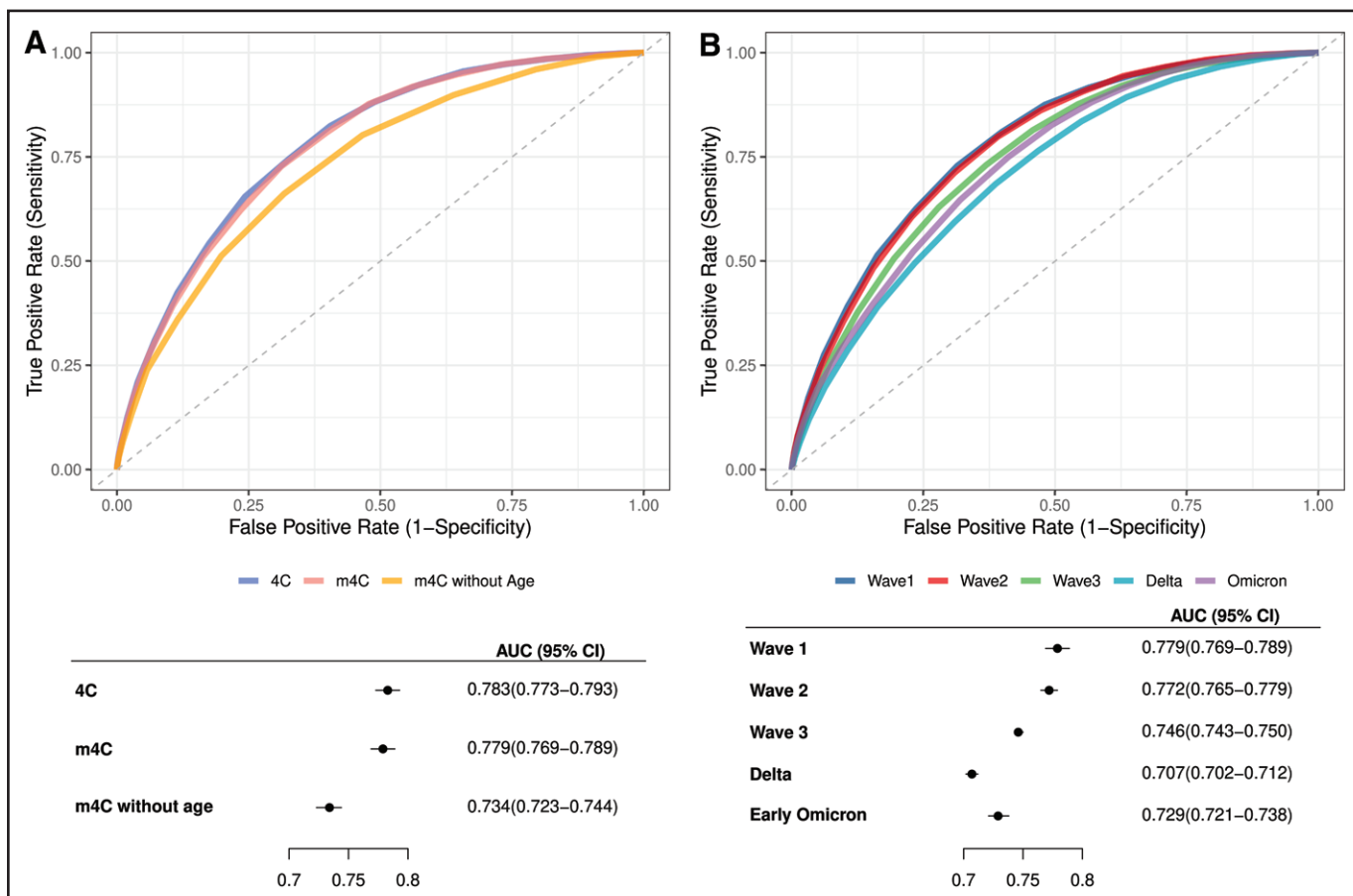

Figure 2. Mortality discrimination by 4C, m4C, and derivative scores. **A**, Receiver operating characteristic curves and forest plot of AUCs of 4C, modified 4C score (m4C), and m4C without age in wave 1. **B**, Receiver operating characteristic curves and forest plot of the area under the receiver operating characteristic curves (AUCs) of m4C by pandemic wave.

Next, the cohort was subdivided by pandemic wave to examine m4C performance over time. When compared with wave 1, AUC was significantly lower in all subsequent waves other than wave 2 (**Fig. 2B**; and **Table S13**, http://links.lww.com/CCX/B282). AUC was lowest during delta (0.707 [95% CI,

0.702–0.712]), followed by early omicron (0.729 [95% CI, 0.721–0.738]). m4C sensitivity was lowest during delta (0.836 [95% CI, 0.830–0.843]) and highest during early omicron (0.923 [95% CI, 0.913–0.932]), whereas specificity was lowest during early omicron (0.355 [95% CI, 0.349–0.361]) and highest during delta (0.448 [95% CI, 0.445–0.451]) and wave 2 (0.448 [95% CI, 0.442–0.453]) (**Tables S14** and **S15**, http://links.lww.com/CCX/B282). These trends were largely preserved in a sensitivity analysis of m4C performance using complete case records (**Tables S16** and **S17**, http://links.lww.com/CCX/B282).

A generalized additive model predicting COVID-19 mortality from m4C fit to wave 1 data showed poor calibration across subsequent waves denoted by significant deviation from the bisector, which represents perfect calibration (**Fig. S3**, http://links.lww.com/CCX/B282). Specifically, calibration belts were largely below the bisector, suggesting over-prediction of mortality by m4C after wave 1; an exception was delta, during which m4C under-predicted mortality across a narrow range of low predicted probabilities of death (2–11%) denoted by the calibration belt lying above the bisector.

Even after updating model coefficients by training the model on wave 3 data or delta wave data, calibration in subsequent waves remained poor.

Impact of Surge on m4C

A generalized additive model predicting COVID-19 mortality from m4C and log-transformed surge index demonstrated nominally higher, but not significantly different, AUCs from that of m4C alone (**Fig. 3**), suggesting that m4C score performance remained robust regardless of hospital-surge strain. The surge-adjusted model demonstrated poor calibration across waves after wave 1, with calibration not significantly different from that of m4C alone (**Fig. S4**, http://links.lww.com/CCX/B282).

DISCUSSION

This study offers insights into the feasibility of applying a disease-specific prognosticator for patient triage as part of CSC. Using the modified 4C score as an example, we found that while the score was robust to hospital-surge strain, its predictive performance declined over time with worse discrimination and calibration in recent (i.e., delta and omicron variant-predominant) waves. Score recalibration with contemporaneous data was a transient solution that did not improve performance in subsequent waves. These findings emphasize the need for responsive, prospective validation of risk scores and the CSC policies that rely on them to guide resource allocation for rapidly evolving diseases.

The COVID-19 pandemic saw unprecedented strain on health systems around the world. Relative shortages of diagnostic tests, vaccines, therapeutics, and hospital beds contributed to excess mortality
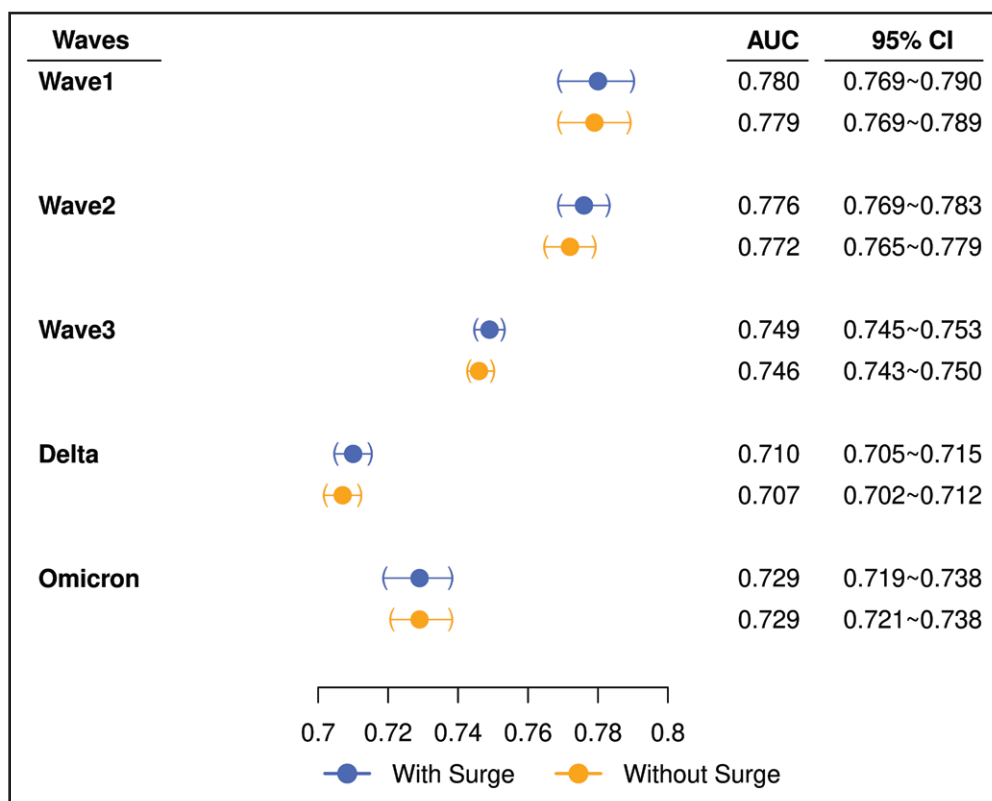

Figure 3. Area under the receiver operating characteristic curve (AUCs) of modified 4C score (m4C) ("without surge") and surge-adjusted models ("with surge") by pandemic wave. *Filled circles* and *whiskers* represent point estimates of AUC and 95% CIs, respectively.

Waves	AUC	95% CI
Wave1	0.780	0.769~0.790
	0.779	0.769~0.789
Wave2	0.776	0.769~0.783
	0.772	0.765~0.779
Wave3	0.749	0.745~0.753
	0.746	0.743~0.750
Delta	0.710	0.705~0.715
	0.707	0.702~0.712
Omicron	0.729	0.719~0.738
	0.729	0.721~0.738

With Surge   Without Surge

that might have been avoided, had adequate systems existed for their distribution (17). In the wake of the pandemic, the contingency frameworks that failed us must be scrutinized to ensure tragedies are not repeated. Among the fundamental values guiding allocation of finite resources is the need to maximize benefits across a population (17). Surveys of community and clinician attitudes (18, 19) demonstrate a preference for utilitarian approaches prioritizing patients that are most likely to survive. But how can this be operationalized? In this regard, the use of clinical risk scores is appealing—a robust score could provide an objective measure of survival probability to help individual triage including in the absence of medical expertise, and to guide central resource allocation through large-scale extraction of risk probabilities over a population.

For entities such as myocardial infarction that display at best mild fluctuations in host, etiologic, treatment and care setting distributions over time, prognosticators are likely to perform consistently. In contrast, SARS-CoV-2 has displayed major evolution in virulence, organ tropism, transmissibility, population susceptibility, and available treatments (20–22). This led to our hypothesis that prognostic performance of COVID-19–specific scores likely decayed during the pandemic, which our study results confirmed. These rapid changes in prognostication by an otherwise robust score may be explained by dramatic shifts in baseline population risk (e.g., due to widespread vaccination/immunity) and reduced effects of predictor variables (e.g., due to effective therapeutics and management strategies "leveling the playing field"). Our findings highlight the need to update COVID-19 risk stratification tools to reflect changes in the disease landscape. The broader implication: triage policies for rapidly evolving diseases such as COVID-19 must be periodically validated if they are to guide provision of care.

Reassuringly, we found that controlling for surge index did not improve m4C score discrimination or calibration, suggesting that the m4C score performs consistently across varying levels of hospital strain. This is particularly relevant as high-surge settings in which CSC are triggered are precisely where prognostic tools are needed. Ventilator triage was frequent headline fodder during COVID-19 (23), whereas moral distress from being forced to prioritize one patient over another was a major contributor to emotional

burnout among healthcare workers (24). By critiquing risk scores in extraordinary surge conditions, we may assuage unease around their use in crises. A final lesson learned from the pandemic postmortem has been the importance of equity (1). The SOFA score, adopted by many triage guidelines, was widely criticized for disadvantaging racial and ethnic minorities during COVID-19 (1). However, proponents supported its age-agnostic approach to short-term risk stratification. Although 4C has demonstrated adequate calibration in different racial/ethnic groups (6), patients with advanced age, and those with underlying comorbidities (25–27), it relies heavily on age as the single most heavily weighted component. Here, we examined discriminative function of the m4C score by subtracting the age component and found that score AUC was diminished, but still better than that of the SOFA score reported elsewhere (3, 4). Rather than modifying key predictors that could compromise prognostication, we suggest that CSC policies encourage nuanced decision-making balancing ethical considerations with absolute mortality risk to optimize equitable care.

Our study has several strengths. External validation studies of the original and modified 4C scores have been limited to the first few months of the pandemic; (8, 28) here, we used a large multicenter cohort to provide the only updated validation of m4C and in doing so revealed decay in performance at later stages of the pandemic. After confirming comparable performance of m4C and 4C in wave 1, we elected to study m4C over the original score given the latter's adaptability for use in electronic administrative datasets. The development of such pragmatic "e-versions" of prognosticators that have been validated against their analog equivalents is an increasingly popular practice in epidemiological research, with applications in large-scale surveillance and policy evaluation (29). Finally, our selected metric of hospital-surge strain has been previously validated and found to directly correlate with incremental COVID-19 mortality (16).

We also recognize the limitations of our study. Cases of COVID-19, severity of acute illness, and presence of comorbidities might not be fully captured by administrative coding. Outside of the pandemic wave (viral variant, time) and hospital surge, other factors (e.g., vaccination and therapeutic availability, social mobility restrictions) not considered here may have impacted m4C score performance in our cohort. We

encountered significant missingness of certain m4C components (vital signs and urea). These data were considered missing at random as, unlike CRP, they are commonly ordered upon hospitalization, and were therefore handled using multiple imputation, as in the original 4C derivation study (6). The exception was the handling of missing binary sex assignments, which (although an imperfect solution emblematic of wider issues in the scientific literature) we elected to exclude rather than impute. Given that performance of our 4C score was equivalent to that in the original derivation cohort, and analyses using complete case data versus multiply imputed datasets yielded comparable results for m4C, we believe this strategy to be robust. Notably, data missingness is a common challenge to large epidemiologic studies particularly in the context of a pandemic; the original 4C derivation study allowed for up to 33% missingness of evaluated score components (6). Although large and diverse, the study cohort was not nationally representative, and we did not evaluate score calibration across specific patient groups including individuals representing racial/ethnic minorities, disadvantaged socioeconomic classes, and those living with disabilities and/or comorbidities, who may have been disproportionately impacted by the pandemic. Further study is required before the m4C score is adopted in clinical guidelines. Finally, the use of large datasets to evaluate risk stratification models is not a substitute for impact studies, which would evaluate outcomes of their real-world implementation against current standards of care.

COVID-19 is a moving target: novel interventions, divergent strains, and impacts of hospital environments have fluctuated and influenced mortality over time. Scores may be an important component of CSC policies facilitating objective triage. Their performance across time, populations, and settings should be rigorously studied and periodically reappraised in any evolving public health crisis.

ACKNOWLEDGMENTS

1 Critical Care Medicine Department, National Institutes of Health Clinical Center, Bethesda, MD.

2 Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD.

3 Biostatistics Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD.

REFERENCES

1. Hick JL, Hanfling D, Wynia M, et al: Crisis standards of care and COVID-19: What did we learn? How do we ensure equity? What should we do?. NAM Perspectives. Discussion, National Academy of Medicine, Washington, DC. 2021. Available at: https://nam.edu/crisis-standards-of-care-and-covid-19-what-did-we-learn-how-do-we-ensure-equity-what-should-we-do/. Accessed November 16, 2022

2. Chelen JSC, White DB, Zaza S, et al: US ventilator allocation and patient triage policies in anticipation of the COVID-19 surge. *Health Secur* 2021; 19:459–467

3. Keller MB, Wang J, Nason M, et al: Preintubation Sequential Organ Failure Assessment score for predicting COVID-19 mortality: External validation using electronic health record from 86 US healthcare systems to appraise current ventilator triage algorithms. *Crit Care Med* 2022; 50:1051–1062

4. Raschke RA, Agarwal S, Rangan P, et al: Discriminant accuracy of the SOFA score for determining the probable mortality of patients with COVID-19 pneumonia requiring mechanical ventilation. *JAMA* 2021; 325:1469–1470

5. Ashana DC, Anesi GL, Liu VX, et al: Equitably allocating resources during crises: racial differences in mortality prediction models. *Am J Respir Crit Care Med* 2021; 204:178–186

6. Knight SR, Ho A, Pius R, et al: Risk stratification of patients admitted to hospital with COVID-19 using the ISARIC WHO clinical characterisation protocol: Development and validation of the 4C mortality score. *BMJ* 2020; 370:m3339

7. Appel KS, Geisler R, Maier D, et al: A systematic review of predictor composition, outcomes, risk of bias, and validation of COVID-19 prognostic scores. *Clin Infect Dis* 2023:ciad618

8. Ong SWX, Sutjipto S, Lee PH, et al: Validation of ISARIC 4C mortality and deterioration scores in a mixed vaccination status cohort of hospitalized COVID-19 patients in Singapore. *Clin Infect Dis* 2022; 75:e874–e877

9. Gordon AJ, Govindarajan P, Bennett CL, et al: External validation of the 4C mortality score for hospitalised patients with COVID-19 in the RECOVER network. *BMJ Open* 2022; 12:e054700

10. de Jong VMT, Rousset RZ, Antonio-Villa NE, et al: Clinical prediction models for mortality in patients with COVID-19: External validation and individual participant data meta-analysis. *BMJ* 2022; 378:e069881

11. Blunsum AE, Perkins JS, Arshad A, et al: Evaluation of the implementation of the 4C mortality score in United Kingdom hospitals during the second pandemic wave. *medRxiv* 2022:2021.12.18.21268003. Available at: https://www.medrxiv.org/content/10.1101/2021.12.18.21268003v2. Accessed September 7, 2022

12. NHS England: Interim clinical commissioning policy: Remdesivir for patients hospitalised due to COVID-19, 2022. Available at: https://www.england.nhs.uk/coronavirus/documents/interim-clinical-commissioning-policy-remdesivir-for-patients-hospitalised-due-to-covid-19-adults-and-adolescents-12-years-and-older/. Accessed January 26, 2023

13. PharmaMar: A Phase 3, multicentre, randomised, controlled trial to determine the efficacy and safety of two dose levels of plitidepsin versus control in adult patient requiring hospitalisation for management of moderate COVID-19 infection. clinicaltrials.gov, 2022. Available at: https://clinicaltrials.gov/ct2/show/NCT04784559. Accessed September 6, 2022

14. Petty H, Baines L, Danielsson-Waters E, et al: Clinical outcomes and use of 4C mortality scoring in COVID-19 patients treated with remdesivir in a large UK centre. *Eur Respir J* 2021; 58:PA3683. Available at: https://erj.ersjournals.com/content/58/suppl_65/PA3683. Accessed September 7, 2022

15. CDC: COVID data tracker. Centers for Disease Control and Prevention, 2020. Available at: https://covid.cdc.gov/covid-data-tracker. Accessed November 15, 2021

16. Kadri SS, Sun J, Lawandi A, et al: Association between caseload surge and COVID-19 survival in 558 US hospitals, March to August 2020. *Ann Intern Med* 2021; 174:1240–1251

17. Emanuel EJ, Persad G: The shared ethical framework to allocate scarce medical resources: A lesson from COVID-19. *Lancet* 2023; 401:1892–1902

18. Dowling A, Lane H, Haines T: Community preferences for the allocation of scarce healthcare resources during the COVID-19 pandemic: A review of the literature. *Public Health* 2022; 209:75–81

19. Fjølner J, Haaland OA, Jung C, et al: Who gets the ventilator? A multicentre survey of intensivists' opinions of triage during the first wave of the COVID-19 pandemic. *Acta Anaesthesiol Scand* 2022; 66:859–868

20. Nyberg T, Ferguson NM, Nash SG, et al: Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: A cohort study. *Lancet* 2022; 399:1303–1312

21. Qi H, Liu B, Wang X, et al: The humoral response and antibodies against SARS-CoV-2 infection. *Nat Immunol* 2022; 23:1008–1020

22. Robinson PC, Liew DFL, Tanner HL, et al: COVID-19 therapeutics: Challenges and directions for the future. *Proc Natl Acad Sci USA* 2022; 119:e2119893119

23. Baker M, Fink S: At the top of the COVID-19 curve, how do hospitals decide who gets treatment?. The New York Times, 2020. Available at: https://www.nytimes.com/2020/03/31/us/coronavirus-covid-triage-rationing-ventilators.html. Accessed May 18, 2023

24. Denning M, Goh ET, Tan B, et al: Determinants of burnout and other aspects of psychological well-being in healthcare workers during the COVID-19 pandemic: A multinational cross-sectional study. *PLoS One* 2021; 16:e0238666

25. Covino M, De Matteis G, Burzo ML, et al: Predicting in-hospital mortality in COVID-19 older patients with specifically developed scores. *J Am Geriatr Soc* 2021; 69:886–887

26. Kuroda S, Matsumoto S, Sano T, et al: External validation of the 4C mortality score for patients with COVID-19 and pre-existing cardiovascular diseases/risk factors. *BMJ Open* 2021; 11:e052708

27. Albai O, Frandes M, Sima A, et al: Practical applicability of the ISARIC-4C score on severity and mortality due to SARS-CoV-2 infection in patients with type 2 diabetes. *Medicina* 2022; 58:848

28. Knight SR, Gupta RK, Ho A, et al: Prospective validation of the 4C prognostic models for adults hospitalised with COVID-19 using the ISARIC WHO clinical characterisation protocol. *Thorax* 2022; 77:606–615

29. Rhee C, Zhang Z, Kadri SS, et al: Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 SOFA criteria. *Crit Care Med* 2019; 47:307–314