BMJ Open

External validation of the 4C Mortality Score for patients with COVID-19 and pre-existing cardiovascular diseases/ risk factors

Shunsuke Kuroda [iD],[1] Shingo Matsumoto,[2] Takahide Sano,[3] Takeshi Kitai,[4] Taishi Yonetsu [iD],[5] Shun Kohsaka,[6] Sho Torii,[7] Takuya Kishi,[8] Issei Komuro,[9] Ken-ichi Hirata,[10] Koichi Node,[11] Yuya Matsue[12]

For numbered affiliations see end of article.

Correspondence to
Dr Yuya Matsue;
yuya8950@gmail.com

ABSTRACT

Objectives Predictive algorithms to inform risk management decisions are needed for patients with COVID-19, although the traditional risk scores have not been adequately assessed in Asian patients. We aimed to evaluate the performance of a COVID-19-specific prediction model, the 4C (Coronavirus Clinical Characterisation Consortium) Mortality Score, along with other conventional critical care risk models in Japanese nationwide registry data.

Design Retrospective cohort study.

Setting and participants Hospitalised patients with COVID-19 and cardiovascular disease or coronary risk factors from January to May 2020 in 49 hospitals in Japan.

Main outcome measures Two different types of outcomes, in-hospital mortality and a composite outcome, defined as the need for invasive mechanical ventilation and mortality.

Results The risk scores for 693 patients were tested by predicting in-hospital mortality for all patients and composite endpoint among those not intubated at baseline (n=659). The number of events was 108 (15.6%) for mortality and 178 (27.0%) for composite endpoints. After missing values were multiply imputed, the performance of the 4C Mortality Score was assessed and compared with three prediction models that have shown good discriminatory ability (RISE UP score, A-DROP score and the Rapid Emergency Medicine Score (REMS)). The area under the receiver operating characteristic curve (AUC) for the 4C Mortality Score was 0.84 (95% CI 0.80 to 0.88) for in-hospital mortality and 0.78 (95% CI 0.74 to 0.81) for the composite endpoint. It showed greater discriminatory ability compared with other scores, except for the RISE UP score, for predicting in-hospital mortality (AUC: 0.82, 95% CI 0.78 to 0.86). Similarly, the 4C Mortality Score showed a positive net reclassification improvement index over the A-DROP and REMS for mortality and over all three scores for the composite endpoint. The 4C Mortality Score model showed good calibration, regardless of outcome.

Conclusions The 4C Mortality Score performed well in an independent external COVID-19 cohort and may enable appropriate disposition of patients and allocation of medical resources.

Trial registration number UMIN000040598.

Strengths and limitations of this study

► In order to compensate for the original research where the 4C (Coronavirus Clinical Characterisation Consortium) Mortality Score was derived and tested only in the cohort of the same nation, the Japanese nationwide cohort of COVID-19 was used for the external validation of the score.

► The precision of the calculated scores was considerably high because the registry data form in this study was built with reference to ISARIC, where the 4C Mortality Score was originally developed.

► The study focused not only on in-hospital mortality but also on a composite outcome of the need for invasive mechanical ventilation and mortality, which has not been evaluated in previous study.

► The applicability of the 4C Mortality Score to various clinical settings needs to be further tested, especially outside the UK and Japan.

INTRODUCTION

The novel coronavirus identified at the end of 2019 in Wuhan, China has spread worldwide and became a pandemic in March 2020. The number of patients diagnosed with COVID-19 exceeds 150 000 000 according to the WHO and continues to grow.[1]

To use our limited resources efficiently, it is important for healthcare practitioners to identify patients at high risk and to allocate medical resources for them appropriately. This highlights the need for a reliable and practical prognostic prediction tool; however, a universally accepted tool has yet to be established. Over 50 prognostic models have been proposed for predicting either mortality or deterioration of disease, but most studies have suffered from poorly designed validation processes, inappropriate statistical methodologies and high bias in patient recruitment.[2] More recently, the 4C (Coronavirus Clinical

Characterisation Consortium) Mortality Score, a scoring system for prediction of in-hospital mortality, was developed based on data from a large population that contains more than 30 000 subjects with COVID-19 in the UK[3] and showed good discrimination and calibration; however, its generalisability remains to be investigated. Specifically, (1) it remains unclear how well the 4C Mortality Score works in the population outside the UK because healthcare systems vary considerably among countries, which might affect the accuracy of prediction. Although differences in ethnicity were reportedly associated with worse outcomes in patients with COVID-19,[4 5] Asians accounted for only 0.7% of all subjects in the derivation cohort. (2) It was unclear whether the 4C Mortality Score also predicts other clinically relevant outcomes, such as the need for invasive mechanical ventilation (IMV). Approximately one-third of hospitalised patients develop acute respiratory distress syndrome, which requires support with a ventilator.[6] This puts a heavy burden on healthcare systems, resulting in shortages of medical resources during the pandemic.[7]

Therefore, the aim of this study was to evaluate the external validity of the 4C Mortality Score in an independent cohort, that is, one with a different ethnicity and healthcare systems, in predicting poor outcomes, defined as mortality and IMV.

METHODS

Study population

This study was a retrospective post-hoc analysis of the Clinical Outcomes of COVID-19 Infection in Hospitalized Patients with Cardiovascular Diseases and/or Risk Factors (CLAVIS-COVID), a Japanese nationwide registry that was endorsed by the Japanese Circulation Society. Briefly, this registry was designed to investigate the clinical features and outcomes of patients with COVID-19 with pre-existing or developing cardiovascular disease or coronary risk factors (CVDRF) from January 2020 to May 2020 in 49 acute care hospitals in Japan. The design and primary outcome have been described elsewhere.[8] Even though there are numerous hospitals in Japan, this study focused on major acute care hospitals that accommodated patients with COVID-19 during that time period, resulting in an enrolment of approximately 9.0% (1518 of 16 851) of all Japanese cases of COVID-19. Diagnosis was confirmed with PCR test using oropharyngeal swab specimens in all subjects.

Before the first patient was enrolled, information on the registry was published as an abstract in the University Hospital Medical Information Network Clinical Trial Registry, in accordance with the International Committee of Medical Journal Editors.Written informed consent from each patient was waived under Japanese law as this was a retrospective observational study.

The prevalence of pre-existing CVDRF is high in patients with COVID-19; these factors are associated with poor outcomes,[9] and management of such patients is

likely to be problematic. Therefore, we focused on hospitalised patients with COVID-19 and pre-existing CVDRF in this analysis. Cardiovascular disease was defined as heart failure, coronary artery disease, myocardial infarction, valvular heart disease, arrhythmia, stroke/transient ischaemic attack, deep venous thrombosis, pulmonary embolism, peripheral arterial disease, aortic aneurysm, aortic dissection, cardiopulmonary arrest, heart transplantation, left ventricular assist device, cardiac implantable electronic device, pericarditis, myocarditis, congenital heart disease and pulmonary hypertension. Risk factors were defined as diabetes mellitus, hypertension and dyslipidaemia.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Study outcomes

We evaluated the predictive ability of the 4C Mortality Score for two different types of outcomes: in-hospital mortality and a composite outcome defined as the need for IMV and mortality. Therefore, patients intubated prior to admission were excluded from the analysis of the composite endpoint.

Calculation of the 4C mortality risk score

As the data form in CLAVIS-COVID was built with reference to International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC),[10] we calculated each individual 4C Mortality Score from the data set in which variables basically followed the data collection form in ISARIC-UK. This score is the sum of the points assigned to eight independent parameters, including age, sex, number of comorbidities, respiratory rate, peripheral oxygen saturation level, Glasgow Coma Scale score, blood urea nitrogen level and C reactive protein level, as described in online supplemental table 1. The number of comorbidities was determined by counting the presence of each of the following diseases or conditions: chronic cardiac disease, chronic respiratory disease, chronic renal disease, liver disease, dementia, chronic neurological conditions, autoimmune disease, diabetes mellitus, malignancy and clinician-defined obesity. Because dementia and HIV infection were not collected in the registry data, they were assumed to be absent when counting comorbidities. In the original 4C Mortality Score, 2 points were given if peripheral oxygen saturation at room air was below 92%, but not all patients had saturation data at room air. Thus, we considered patients with <92% saturation, regardless of oxygen therapy, to have a score of 2 points in this study.

Selection of the scoring systems

We reviewed the studies comparing multiple prediction models in patients with COVID-19 and selected three scores that have previously shown good discriminatory ability in comparison with the 4C score: the Risk

Stratification in the EmergencyDepartment in Acutely Ill Older Patients (RISE UP) score,[11] the Japan Respiratory Society community associated pneumonia severity index (A-DROP scoring system)[12] and the Rapid Emergency Medicine Score (REMS).[13] These three scores were not COVID-19-specific risk scores but were reported to have the highest discriminatory ability in predicting short-term mortality in patients with COVID-19, with an area under the receiver operating characteristic curve (AUC) of 0.83 (RISE UP score),[14] 0.87 (A-DROP)[15] and 0.76 (REMS).[16] We calculated the RISE UP score based on age, vital signs, serum albumin, blood urea nitrogen, lactate dehydrogenase and bilirubin, as proposed in the original score. Likewise, the A-DROP was calculated based on age, dehydration, respiratory failure, orientation disturbance and low systolic blood pressure, and the REMS accounted for six variables, namely heart rate, blood pressure, respiratory rate, consciousness, age and level of oxygen saturation.

Statistical analysis

Continuous variables were expressed as mean and SD if normally distributed, or as median and IQR if not normally distributed. To describe the heterogeneity between the two studies, CLAVIS-COVID and ISARIC-UK, the derivation cohort also served as a comparison (online supplemental table 2). In patients who had missing elements required for calculation of the risk scores, missing values were multiply imputed with the assumption of missing at random. We generated 20 data sets with imputation using the variables presented in table 1. As we used two different inclusion criteria corresponding to two types of outcomes, multiple imputations were performed for in-hospital death and for the composite endpoint separately. After the scores were calculated, all statistical analyses were applied to the generated data sets. We calculated the AUC, the continuous net reclassification index (NRI) and the integrated discrimination improvement (IDI) index, and subsequently averaged across 20 pooled data sets to retrieve the corrected value and its range using Rubin's rule.[17]

To assess the calibration of the models, we constructed logistic regression models of each risk score for predicting outcomes and created a calibration plot for the predicted versus observed risk. The Hosmer-Lemeshow goodness-of-fit test was also performed. Statistical analyses were performed using R (V.4.0.3; R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org).

RESULTS
Study population

Among 1518 registry patients, 693 with pre-existing or known CVDRF were included in the analysis for in-hospital mortality. The mean age of the patients was 68±15 years (64.8% male), and almost all patients were Japanese (96.1%). Four hundred and twenty-two patients (60.8%) had at least one comorbidity, and diabetes was the most

prevalent among all designated comorbidities. During hospitalisation, 108 patients died (15.6%).

After the exclusion of 34 patients who were already intubated at the time of the index admission, the composite endpoint was evaluated in 659 patients. In this cohort, IMV was initiated for 119 patients during hospitalisation and 41 (34.4%) patients died after a median IMV duration of 9 (IQR: 6–16) days. Overall, 178 patients met the composite endpoint of death or IMV (27.0%). There were no missing data on prognosis during the index admission.

We compared patient characteristics between cases with or without missing 4C Mortality Score in online supplemental table 2. At least one element for calculating the 4C Mortality Score was missing in 217 cases (31.3%), namely respiratory rate in 153 cases, peripheral oxygen saturation in 4 cases, Glasgow Coma Scale score in 68 cases, blood urea nitrogen level in 14 cases and C reactive protein level in 26 cases. There was no significant difference in the event rate between patients with or without missing 4C Mortality Score (in-hospital mortality: 16.1% vs 15.3%, respectively, p=0.88; composite endpoint of mortality and IMV: 26.3% vs 32.4%, respectively, p=0.13). The number of cases in which any elements of the A-DROP and the REMS were missing was the same as that for the 4C Mortality Score, but the RISE UP score was missing in 290 cases (41.8%).

Table 1 summarises the differences in characteristics among tertiles based on the 4C Mortality Score after imputation. Tertile 1 comprised patients with the lowest 4C Mortality Score. Tertile 3, the group with the highest 4C Mortality Score, had a significantly higher risk of in-hospital mortality and composite outcomes. Most components of the 4C Mortality Score, except for sex, systolic blood pressure, body temperature, comorbidities of liver disease, hypertension and obesity, showed significant differences among the groups. Lymphocyte count was significantly lower while neutrophil count was significantly higher in tertile 3.

When compared with the ISARIC-UK cohort, patients in the CLAVIS-COVID cohort had fewer comorbidities and lower levels of blood urea nitrogen, creatinine and C reactive protein, which were suggestive of lower in-hospital mortality (online supplemental table 2).

Figure 1A shows the distribution of the 4C Mortality Score. The median 4C Mortality Score was 9 points (IQR: 6–11). No in-hospital deaths or composite events occurred in patients with a 4C Mortality Score of <7 points or <4 points, respectively. The distribution of other scores was skewed to the left in the A-DROP score and REMS and to the right in the RISE UP score (figure 1B–D).

Discrimination of the 4C Mortality Score, RISE UP, A-DROP and REMS

A high 4C Mortality Score was associated with poor outcomes (OR per 1 point: 1.54, 95% CI 1.42 to 1.68 for in-hospital death; OR: 1.36, 95% CI 1.28 to 1.45 for the composite endpoint). As shown in table 2, the models'

Table 1 Patient characteristics

Variables	Overall N=693	4C Mortality Score			P value
		Tertile 1 (0–6) n=217	Tertile 2 (7–10) n=252	Tertile 3 (11–19) n=224	
In-hospital mortality	108 (15.6)	1 (0.5)	23 (9.1)	84 (37.5)	<0.001
Composite outcome	211 (30.4)	21 (9.7)	68 (27.0)	122 (54.5)	<0.001
Age	68.3±14.9	53.5±9.8	71.3±11.9	79.3±9.4	<0.001
Female	244 (35.2)	80 (36.9)	96 (38.1)	68 (30.4)	0.17
Japanese	666 (96.1)	202 (93.1)	241 (95.6)	223 (99.6)	0.002
Comorbidities					
Chronic cardiac disease*	131 (18.9)	9 (4.1)	50 (19.8)	72 (32.1)	<0.001
Chronic renal disease	52 (7.5)	3 (1.4)	9 (3.6)	40 (17.9)	<0.001
Malignancy	15 (2.2)	1 (0.5)	3 (1.2)	11 (4.9)	0.002
Moderate or severe liver disease	1 (0.1)	0 (0.0)	0 (0.0)	1 (0.4)	0.35
Obesity	47 (6.8)	22 (10.1)	15 (6.0)	10 (4.5)	0.05
Chronic pulmonary disease	35 (5.1)	1 (0.5)	12 (4.8)	22 (9.8)	<0.001
Diabetes	266 (38.4)	62 (28.6)	105 (41.7)	99 (44.2)	0.001
Hypertension	513 (74.0)	154 (71.0)	187 (74.2)	172 (76.8)	0.38
Dyslipidaemia	269 (38.8)	104 (47.9)	100 (39.7)	65 (29.0)	<0.001
Number of designated comorbidities†					<0.001
0	271 (39.1)	129 (59.4)	97 (38.5)	45 (20.1)	
1	272 (39.2)	75 (34.6)	106 (42.1)	91 (40.6)	
≥2	150 (21.6)	13 (6.0)	49 (19.4)	88 (39.3)	
Respiratory rate (breaths/min)	20.55±5.93	18±5	20±6	23±6	<0.001
Peripheral oxygen saturation (%)	96 (94, 98)	97 (96, 98)	96 (94, 98)	95 (91, 97)	<0.001
Systolic blood pressure (mm Hg)	133±22	135±20	132±21	132±23	0.26
Diastolic blood pressure (mm Hg)	79±15	84±15	77±14	75±16	<0.001
Temperature (°C)	38.0±0.9	38.0±0.9	37.9±0.9	38.1±0.8	0.15
Heart rate (beats per minute)	86±18	86±17	84±16	89±20	0.03
Glasgow Coma Scale score	15 (15, 15)	15 (15, 15)	15 (15, 15)	15 (13, 15)	<0.001
Haemoglobin (g/dL)	13.4 (11.7, 14.7)	14.4 (13.3, 15.4)	13.15 (11.5, 14.5)	12.5 (10.6, 14.0)	<0.001
White cell count (×10^9/L)	5.70 (4.40, 7.52)	5.30 (4.10, 6.60)	5.45 (4.40, 7.40)	6.60 (4.70, 9.02)	<0.001
Neutrophil count (×10^9/L)	4.18 (2.93, 5.76)	3.69 (2.64, 4.71)	4.13 (2.98, 5.63)	5.35 (3.52, 7.50)	<0.001
Lymphocyte count (×10^9/L)	0.92 (0.66, 1.24)	1.12 (0.83, 1.52)	0.93 (0.68, 1.23)	0.78 (0.52, 1.02)	<0.001
Platelet count (×10^9/L)	19.0 (14.6, 24.8)	20.0 (15.9, 25.6)	19.1 (14.5, 25.5)	17.1 (13.1, 22.9)	<0.001
Sodium (mmol/L)	138 (135, 141)	139 (136, 141)	137 (135, 140)	137 (134, 141)	0.004
Potassium (mmol/L)	4 (3.70, 4.30)	4 (3.60, 4.22)	4 (3.70, 4.20)	4.10 (3.70, 4.50)	0.04
Bilirubin (mg/dL)	0.65±0.47	0.65±0.59	0.64±0.35	0.66±0.45	0.85
Lactate dehydrogenase (U/L)	305±189	251±139	301±181	360±221	<0.001
Blood urea nitrogen (mmol/L)	7.58±5.44	4.87±2.02	6.56±3.87	11.37±6.95	<0.001
Creatinine (µmol/L)	72.49 (57.46, 93.70)	67.18 (54.81, 77.79)	71.60 (57.46, 88.40)	86.63 (63.43, 135.25)	<0.001
C reactive protein (mg/L)	56.80 (19, 114.60)	25.40 (7.10, 62.10)	55.25 (19.73, 105.55)	103.55 (50.70, 156.85)	<0.001

*Chronic cardiac disease was defined as heart failure, coronary artery disease, myocardial infarction and valvular heart disease.
†Number of designated comorbidities was defined using the Charlson Comorbidity Index, with the addition of clinician-defined obesity.
4C, Coronavirus Clinical Characterisation Consortium.

discrimination was verified for each risk score. As seen in figure 2A, the 4C Mortality Score predicted in-hospital death better than the other scores (0.84 in 4C Mortality Score vs 0.82, 0.78 and 0.74 in RISE UP, A-DROP and REMS, respectively), but the difference in AUC with the RISE UP score was modest (delta AUC: 0.024) (table 2).
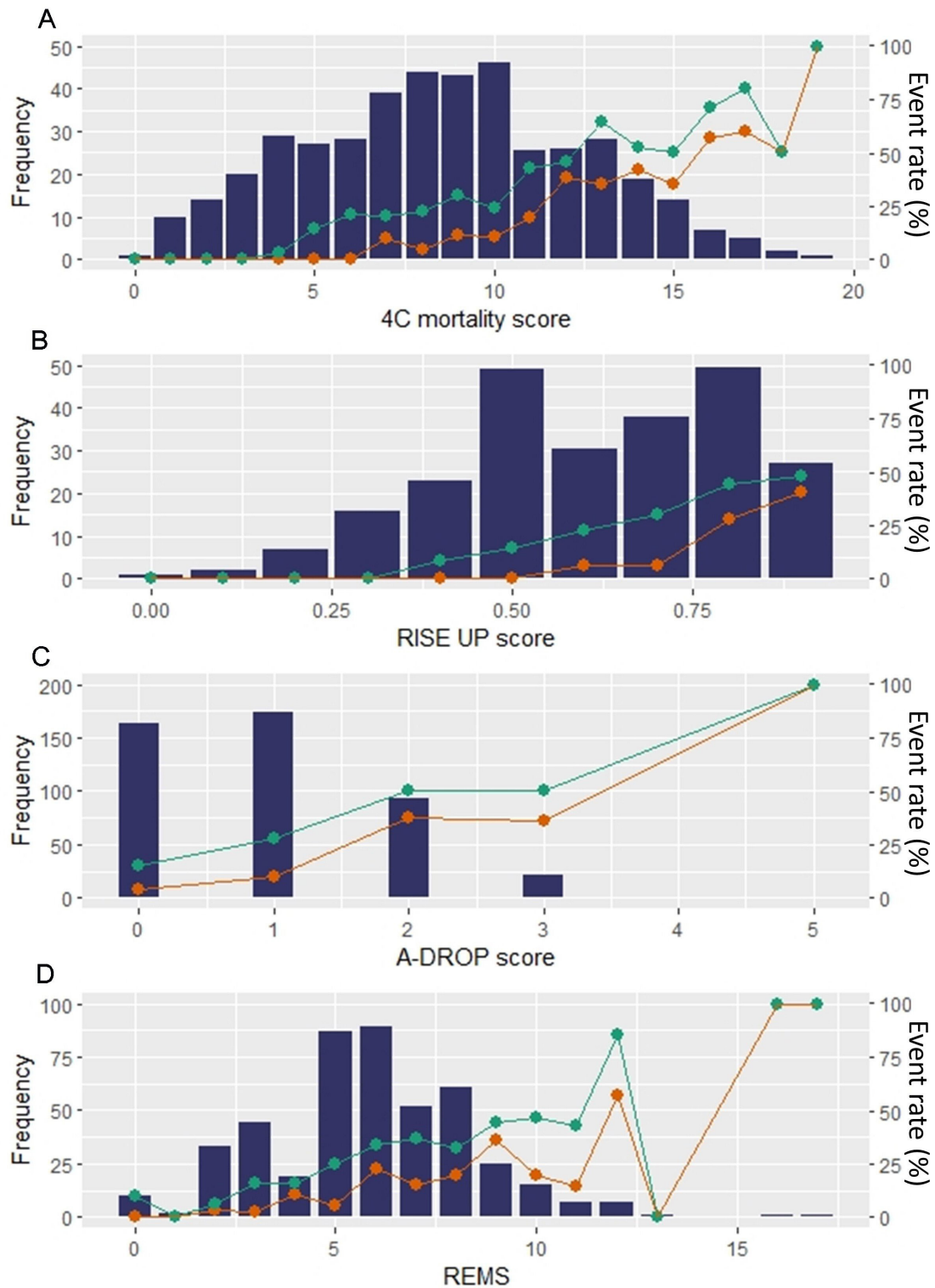
Figure 1 Distribution and event rates among scores. Bar plots represent the number of patients across the range of each risk model; (A) 4C mortality score, (B) RISE UP score, (C) A-DROP score, and (D) REMS, respectively. Line plots indicate the rate of outcomes. Orange lines indicate mortality, and green lines indicate composite events. 4C, Coronavirus Clinical Characterisation Consortium; REMS, Rapid Emergency Medicine Score.

The 4C Mortality Score improved the net risk classification for in-hospital mortality compared with the RISE UP score (NRI=16.2%), A-DROP score (NRI=54.6%) and REMS (NRI=83.1%), respectively (table 2).

Similar results were observed for the prediction of the composite endpoint. The 4C Mortality Score yielded the best discrimination among the three risk scores (figure 2B), with an AUC of 0.78, 0.72, 0.70 and 0.69,

Table 2 Comparison of area under the curve and net reclassification improvement of the 4C Mortality Score in predicting outcomes

	Receiver operating characteristic			
	AUC (95% CI; lower and upper bound)	Difference in AUC (AUC 4C Mortality Score–AUC other models, 95% CI)	Continuous NRI (95% CI)	IDI (95% CI)
In-hospital death				
4C Mortality score	0.84 (0.80 to 0.88)	Reference	Reference	Reference
RISE UP score	0.82 (0.78 to 0.86)	0.024 (−0.006 to 0.055)	0.162 (−0.041 to 0.366)	0.068 (0.033 to 0.102)
A-DROP score	0.78 (0.73 to 0.82)	0.065 (0.035 to 0.096)	0.546 (0.347 to 0.746)	0.081 (0.041 to 0.122)
REMS	0.74 (0.69 to 0.78)	0.103 (0.064 to 0.142)	0.831 (0.644 to 1.017)	0.132 (0.093 to 0.171)
Composite outcome				
4C Mortality Score	0.78 (0.74 to 0.81)	Reference	Reference	Reference
RISE UP score	0.72 (0.68 to 0.76)	0.052 (0.026 to 0.079)	0.442 (0.274 to 0.610)	0.074 (0.052 to 0.096)
A-DROP	0.70 (0.65 to 0.74)	0.079 (0.051 to 0.106)	0.608 (0.444 to 0.772)	0.088 (0.065 to 0.111)
REMS	0.69 (0.64 to 0.73)	0.088 (0.053 to 0.124)	0.596 (0.431 to 0.762)	0.101 (0.074 to 0.128)

AUC, area under the receiver operating characteristic curve; 4C, Coronavirus Clinical Characterisation Consortium; IDI, integrated discrimination improvement; NRI, net reclassification index; REMS, Rapid Emergency Medicine Score.

respectively. We also found a 44.2% improvement over the RISE UP score, 60.8% improvement over A-DROP and 59.6% improvement over REMS in predicting the composite endpoint (table 2).

The IDI for the 4C Mortality Score was consistently positive compared with the other three scores for in-hospital mortality (vs RISE UP score, A-DROP score and REMS: 0.068, 0.081 and 0.132, respectively) and composite outcome (vs RISE UP score, A-DROP score and REMS: 0.074, 0.088 and 0.101, respectively) (table 2).

Calibration of the model

We constructed calibration plots for the assessment of the models. For the 4C Mortality Score model, the calibration curve for both mortality and composite outcome almost coincided with the ideal slope (figure 3A,B),

while deviation of points from an ideal line was observed in the calibration plot for both outcomes for the REMS model, indicating a lack of calibration, and those for both outcomes for the RISE UP and A-DROP models were well calibrated (online supplemental figure 1). The Hosmer-Lemeshow goodness-of-fit test had a p value of 0.001 for the RISE UP score model for in-hospital mortality, while it had a p value greater than 0.05 for the other models for both outcomes.

DISCUSSION

In this study, we validated the 4C Mortality Score for two outcomes (all-cause in-hospital mortality and the combined endpoint of all-cause death and introduction
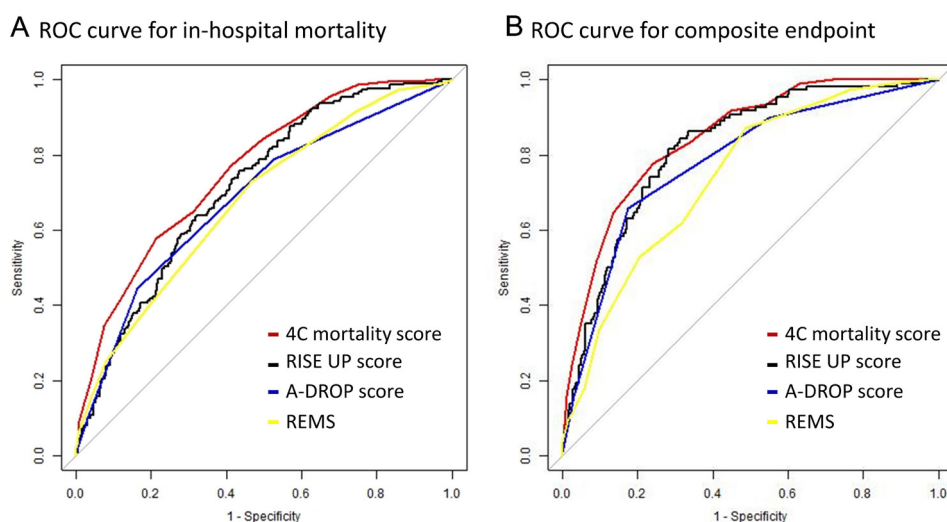

A ROC curve for in-hospital mortality

B ROC curve for composite endpoint

Figure 2 Receiver operating characteristic (ROC) curve analysis for in-hospital mortality (A) and for composite outcome (B) in all models. 4C, Coronavirus Clinical Characterisation Consortium; REMS, Rapid Emergency Medicine Score.

A 4C mortality model for mortality
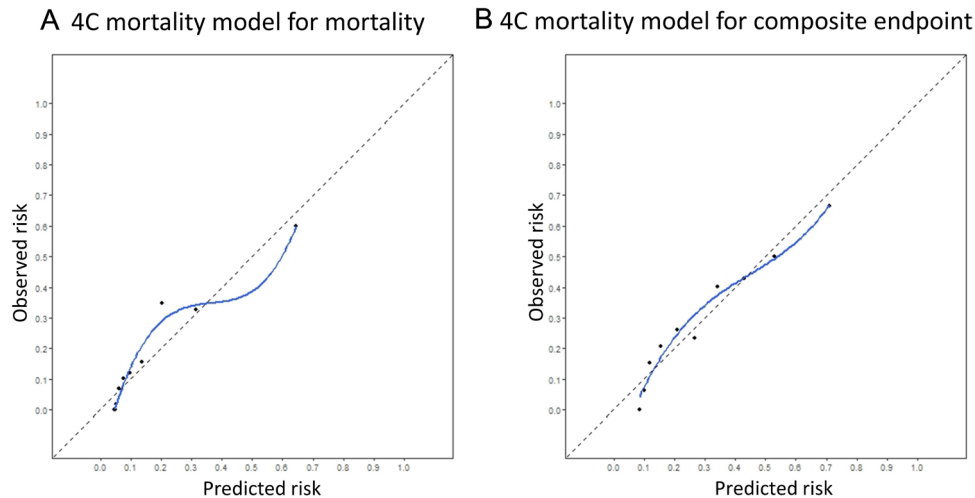
B 4C mortality model for composite endpoint

Figure 3 Calibration plot of the 4C Mortality Score models for in-hospital mortality (A) and a composite outcome (B). The blue line represents a Loess-smoothed regression line in each calibration plot, being close to the diagonal reference (the dashed line). 4C, Coronavirus Clinical Characterisation Consortium.

of IMV) in patients with COVID-19 complicated with CVDRF. The 4C Mortality Score performed well, as shown in previous studies. Of note, we demonstrated that this model is applicable to predicting not only in-hospital death but also the composite endpoint of the need for IMV and death. Furthermore, the 4C Mortality Score was superior to the pre-existing prediction models that have also been validated in patients with COVID-19 for predicting mortality. Our study results showed that (1) the 4C Mortality Score was generalisable to other clinically relevant events (composite outcome of death and need for IMV) and was useful in different ethnicities and healthcare settings; and (2) the 4C Mortality Score was superior to pre-existing risk scores originally developed for patients with COVID-19 and patients visiting the emergency department, except for the model of RISE UP score for mortality.

Importance of external validation of the risk models

One of the clinical challenges of scoring systems is the applicability of the model to the population in which it is actually used in a clinical setting. In general, there is a certain risk of overfitting to the derivation cohort when developing risk models, and the predictive ability is prone to overestimation. Therefore, it is important to externally validate the developed risk model in an independent cohort, although this is rarely performed, despite the fact that numerous risk prediction models are proposed in the medical field. Indeed, Siontis et al[18]. studied newly proposed risk models in any medical field and reported that only 16% of them were subsequently validated after the first publication. Amidst the pandemic, many prediction models have been proposed; however, their external validity has rarely been examined. In addition, a recent study reported that most prediction models developed for patients with COVID-19 failed to reproduce a similar ability to discriminate high-risk patients.[16] This implies that it is difficult to expect the same ability for a prediction

model outside the cohort in which the model was derived. In fact, the AUC of the receiver operating characteristics of both A-DROP and REMS fell short in the present study compared with those previously reported as 0.87 and 0.84, respectively.[15 19]

As for the 4C Mortality Score, the primary study tested the external validity within the ISARIC-UK cohort, yet both the derivation and validation cohorts were developed in the same way and in the same country. More recently, the performance of the 4C Mortality Score was tested in a single medical centre in the Netherlands; however, there was lack of generalisability and missing information on ethnicity.[14] At this point, it is noteworthy that the 4C Mortality Score performed well in the cohort of patients from the Japanese nationwide registry which has different features, including ethnicity, comorbidities and healthcare system. Despite these differences between the two cohorts, the 4C Mortality Score showed similar discriminatory ability to that in the previous study (AUC: 0.786).[3] A previous study reported that the RISE UP score was comparable with the 4C Mortality Score in predicting mortality.[14] Our results were in line with the previous study; however, the 4C mortality model fitted better than the RISE UP score based on the results of the goodness-of-fit test. These findings imply that the 4C Mortality Score is generalisable to patients with COVID-19 regardless of ethnicity and healthcare system.

Identification of patients at high risk of requiring mechanical ventilation

Our study also evaluated a composite endpoint that included the need for an IMV. We found that the 4C Mortality Score could predict this composite endpoint well and was better than the other scores.

Even though mortality is one of the most clinically relevant outcomes, it does not directly project the burden on healthcare systems. Patients who survive but require intensive care are not accounted for if only death is used

as the outcome. Considering that the main purpose of risk stratification tools is to allow clinicians to determine the disposition of patients and the allocation of limited medical resources appropriately, it is important to identify not only those likely to die but also those likely to require intensive care. In this context, our findings on the predictability of the 4C Mortality Score for the composite endpoint incorporating IMV may expand the clinical utility of the 4C risk score. Although several studies have already developed a model for predicting those at high risk for progression to severe COVID-19,[16 20 21] the definitions of severe COVID-19 used in those studies were heterogeneous and there is no universally accepted definition of severe COVID-19, which make interpretation of the study results difficult. On the other hand, a recent study adopted a composite endpoint of mortality and admission to the medium/intensive care unit instead of IMV.[14] In this study, the 4C Mortality Score was not superior to the RISE UP score in predicting the composite endpoint. Although admission to the medium/intensive care unit generally reflects deterioration of general condition, including respiratory status, the difference in outcomes may explain the inconsistency of the results with ours. Given that patients who require IMV therapy are likely to stay on a ventilator for about 2 weeks and nearly one-third of them die,[22–24] a combined endpoint of IMV and in-hospital mortality appears to be a reasonable outcome of interest in terms of healthcare burden. Moreover, a recent study has shown that the COVID-19 intensive care unit case load was associated with increased mortality for critical patients[25]; appropriate and early disposition according to the risk model capable of identifying those at high risk for intensive care should therefore be considered.

Limitations

This retrospective post-hoc analysis has several limitations that must be addressed. The patients in our study cohort were almost all Japanese and it is not clear whether this model is applicable to different cohorts; however, our study results were consistent with those of the previous study in terms of model discrimination and calibration, and superior to other risk models in performance. Second, the practicality of the 4C Mortality Score was not evaluated in our study; however, it can be readily calculated, either manually or online, even if its calculation is more complicated than the A-DROP. Third, there were several missing values in our data set, and we analysed only complete cases in this study, which may have led to potential selection bias. Fourth, one of the most important limitations of our study is that we did not have two of the variables included in the original 4C Mortality Score, namely HIV infection and dementia. However, we assumed that a limited number of patients with HIV were included, given that the prevalence of HIV is less than 0.2% in Japan.[26] No clear definition of dementia was proposed in the ISARIC data set and it is not clinically feasible to diagnose dementia in the acute setting

of COVID-19, as the disease itself can possibly affect the status of consciousness.

CONCLUSION

The 4C Mortality Score performed well in a cohort with CVDRF, independent of the original cohort, and showed consistent results with the former study in terms of discrimination and calibration. The score also predicts the composite outcome of IMV and in-hospital mortality, and its predictive ability is superior to those of other prediction models previously proposed for patients with COVID-19. Our study results showed the generalisability of risk stratification of patients with COVID-19 based on the 4C Mortality Score, and it may aid in the disposition of patients and the allocation of limited medical resources during the pandemic.

Author affiliations
[1]Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, Ohio, USA
[2]Department of Cardiovascular Medicine, Toho University Faculty of Medicine, Ota-ku, Japan
[3]Department of Cardiovascular Medicine, Toho University Graduate School of Medicine, Ota-ku, Japan
[4]Department of Cardiovascular Medicine, Kobe City Medical Center General Hospital, Kobe, Japan
[5]Department of Cardiovascular Medicine, Tokyo Medical and Dental University, Bunkyo-ku, Japan
[6]Department of Cardiology, Keio University School of Medicine, Shinjuku-ku, Japan
[7]Department of Cardiology, Tokai University School of Medicine, Isehara, Japan
[8]Department of Graduate School of Medicine (Cardiology), International University of Health and Welfare, Fukuoka, Japan
[9]Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Japan
[10]Division of Cardiovascular Medicine, Kobe University Graduate School of Medicine, Kobe, Japan
[11]Department of Cardiovascular Medicine, Saga University, Saga, Japan
[12]Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, Bunkyo-ku, Japan

ORCID iDs
Shunsuke Kuroda http://orcid.org/0000-0002-7547-891X
Taishi Yonetsu http://orcid.org/0000-0002-1798-5008

REFERENCES

1 WHO coronavirus (COVID-19) dashboard. Available: https://covid19.who.int
2 Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:18.
3 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.
4 Zelner J, Trangucci R, Naraharisetti R. Racial disparities in COVID-19 mortality are driven by unequal infection risks. *Clin Infect Dis* 2020.
5 Williamson EJ, Walker AJ, Bhaskaran K, *et al*. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.
6 Tzotzos SJ, Fischer B, Fischer H, *et al*. Incidence of ARDS and outcomes in hospitalized patients with COVID-19: a global literature survey. *Crit Care* 2020;24:516.
7 Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. *N Engl J Med* 2020;382:e41.
8 Matsumoto S, Kuroda S, Sano T, *et al*. Clinical and biomarker profiles and prognosis of elderly patients with coronavirus disease 2019 (COVID-19) with cardiovascular diseases and/or risk factors. *Circ J* 2021;85:921–8.
9 Dhakal BP, Sweitzer NK, Indik JH, *et al*. SARS-CoV-2 infection and cardiovascular disease: COVID-19 heart. *Hear Lung Circ* 2020;29:973–87.
10 Docherty AB, Harrison EM, Green CA, *et al*. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.
11 Zelis N, Buijs J, de Leeuw PW, *et al*. A new simplified model for predicting 30-day mortality in older medical emergency department patients: the rise up score. *Eur J Intern Med* 2020;77:36–43.
12 Miyashita N, Matsushima T, Oka M, *et al*. The JRS guidelines for the management of community-acquired pneumonia in adults: an update and new recommendations. *Intern Med* 2006;45:419–28.
13 Olsson T, Terent A, Lind L. Rapid emergency medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med* 2004;255:579–87.
14 van Dam PMEL, Zelis N, van Kuijk SMJ, *et al*. Performance of prediction models for short-term outcome in COVID-19 patients in the emergency department: a retrospective study. *Ann Med* 2021;53:402–9.
15 Fan G, Tu C, Zhou F, *et al*. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *Eur Respir J* 2020;56:2002113.
16 Gupta RK, Marks M, Samuels THA, *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;56:2003498.
17 Rubin DB, ed. *Multiple imputation for nonresponse in surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 1987.
18 Siontis GCM, Tzoulaki I, Castaldi PJ, *et al*. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68:25–34.
19 Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critically ill patients with novel coronavirus disease. *Acad Emerg Med* 2020;27:461–8.
20 Gong J, Ou J, Qiu X, *et al*. A tool for early prediction of severe coronavirus disease 2019 (covid-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *Clin Infect Dis* 2020;71:833–40.
21 Kaeuffer C, Ruch Y, Fabacher T, *et al*. The BAS^2IC Score: A Useful Tool to Identify Patients at High Risk of Early Progression to Severe Coronavirus Disease 2019. *Open Forum Infect Dis* 2020;7:ofaa405.
22 Ma Q, Qi D, Deng X-Y, *et al*. Corticosteroid therapy for patients with severe novel coronavirus disease 2019. *Eur Rev Med Pharmacol Sci* 2020;24:8194–201.
23 Tomazini BM, Maia IS, Cavalcanti AB, *et al*. Effect of dexamethasone on days alive and Ventilator-Free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: the Codex randomized clinical trial. *JAMA* 2020;324:1307–16.
24 Ziehr DR, Alladina J, Petri CR, *et al*. Respiratory pathophysiology of mechanically ventilated patients with COVID-19: a cohort study. *Am J Respir Crit Care Med* 2020;201:1560–4.
25 Bravata DM, Perkins AJ, Myers LJ, *et al*. Association of intensive care unit patient load and demand with mortality rates in US Department of Veterans Affairs hospitals during the COVID-19 pandemic. *JAMA Netw Open* 2021;4:e2034266.
26 GBD 2015 HIV Collaborators. Estimates of global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2015: the global burden of disease study 2015. *Lancet HIV* 2016;3:e361–87.