

Performance of prediction models for short-term outcome in COVID-19 patients in the emergency department: a retrospective study

Paul M. E. L. van Dam^a, Noortje Zelis^a, Sander M. J. van Kuijk^b, Aimée E. M. J. H. Linkens^c, Renée A. G. Brüggemann^c, Bart Spaetgens^c, Iwan C. C. van der Horst^d and Patricia M. Stassen^a

^aDepartment of Internal Medicine, Division of General Internal Medicine, Section Acute Medicine, Maastricht University Medical Center, Maastricht, The Netherlands; ^bDepartment of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Center, Maastricht, The Netherlands; ^cDepartment of Internal Medicine, Division of General Internal Medicine, Section Geriatric Medicine, Maastricht University Medical Center, Maastricht, The Netherlands; ^dDepartment of Intensive Care Medicine, Maastricht University Medical Center, Maastricht, The Netherlands

ABSTRACT

Introduction: Coronavirus disease 2019 (COVID-19) has a high burden on the healthcare system. Prediction models may assist in triaging patients. We aimed to assess the value of several prediction models in COVID-19 patients in the emergency department (ED).

Methods: In this retrospective study, ED patients with COVID-19 were included. Prediction models were selected based on their feasibility. Primary outcome was 30-day mortality, secondary outcomes were 14-day mortality and a composite outcome of 30-day mortality and admission to medium care unit (MCU) or intensive care unit (ICU). The discriminatory performance of the prediction models was assessed using an area under the receiver operating characteristic curve (AUC).

Results: We included 403 patients. Thirty-day mortality was 23.6%, 14-day mortality was 19.1%, 66 patients (16.4%) were admitted to ICU, 48 patients (11.9%) to MCU, and 152 patients (37.7%) met the composite endpoint. Eleven prediction models were included. The RISE UP score and 4C mortality scores showed very good discriminatory performance for 30-day mortality (AUC 0.83 and 0.84, 95% CI 0.79–0.88 for both), significantly higher than that of the other models.

Conclusion: The RISE UP score and 4C mortality score can be used to recognise patients at high risk for poor outcome and may assist in guiding decision-making and allocating resources.

ARTICLE HISTORY

Received 6 December 2020
Revised 9 February 2021
Accepted 12 February 2021

KEYWORDS

COVID-19; prediction; prognosis; mortality; emergency department



Background


To mitigate the burden on the healthcare system caused by the Coronavirus disease 2019 (COVID-19) pandemic, it is necessary to identify patients who are at high risk of poor outcomes early in the course of the disease [1–3]. Although most patients with COVID-19 develop only mild symptoms, some develop severe and potentially fatal complications [1,2,4,5]. Prediction models could help forecast outcomes when patients present to the emergency department (ED) and may assist in triaging patients when allocating healthcare resources.

Several triage and prediction models have been developed to identify ED patients with a high risk of adverse outcome [6–10]. Some of these models were specifically designed for patients with pneumonia

(CURB-65) and sepsis (abbreviated Mortality Emergency Department Sepsis (abMEDS) and sepsis-related organ failure assessment (SOFA)) or for older patients (Risk Stratification in the Emergency Department in Acutely Ill Older Patients (RISE UP)) [6–9]. These models may be useful in patients with COVID-19 as well, as they often present with pneumonia and sepsis, and patients older than 65 years have a higher risk of poor outcome [11–14]. A recent systematic review reported on several new prediction models specifically designed for patients with COVID-19 [15]. Some models were found to have a good discriminatory performance with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.84.

The present retrospective study aims to validate several previously developed prediction models in patients with COVID-19 in the ED [6–10,15–18].

CONTACT Paul M. E. L. van Dam  paul.van.dam@mumc.nl  Department of Internal Medicine, Maastricht University Medical Center, Maastricht, The Netherlands

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Materials and methods

Study design and setting

This retrospective cohort study was performed at the ED of the Maastricht University Medical Centre + (MUMC+). This is a combined secondary/tertiary care centre in the Netherlands, with 22,000 ED visits every year. The medical ethics committee of the MUMC+ approved this study (METC 2020-1572). Informed consent was obtained from all individual participants. This study was conducted and reported in line with the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) guidelines [19].

Study sample

The study sample consisted of consecutive adult (18 years or older) medical ED patients diagnosed with COVID-19 during the first wave of the COVID-19 pandemic in the period from March 11th until May 8th 2020. Patients were included if they met the following criteria: (1) symptoms compatible with COVID-19 (i.e. coughing, common cold, sore throat, dyspnoea, acute diarrhoea, vomiting, fever or an unexpectedly discovered oxygen saturation below 92%); and (2) positive result of the polymerase chain reaction (PCR) for SARS-CoV-2 in respiratory specimens or (3) (very) high suspicion of COVID-19 according to the chest computed tomography (CT) scan (CO-RADS 4 or CO-RADS 5) [20]. We excluded patients who revisited the ED after an earlier ED presentation during the study period. In order to perform external validation of prediction models in our sample, we aimed to comply with the rule of thumb to include approximately 100 patients who met the primary outcome, similar to other studies [21].

Data collection

Data collection was performed by medical students and resident doctors, who were blinded to the study hypotheses. We collected data on age, sex and information regarding comorbidity according to the Charlson Comorbidity Index (CCI) from electronic medical records [22]. We also retrieved the following vital signs: heart rate (HR), systolic blood pressure (SBP), mean arterial blood pressure (MAP), respiratory rate (RR), oxygen saturation, temperature and Glasgow Coma Scale (GCS). For each vital sign, we used the initial (i.e. first recorded) value during the ED visit. The Alert Verbal Pain Unresponsive (AVPU) scale was derived from the GCS [23]. If RR or GCS were missing,

we used paCO_2 and descriptions in the medical records to deduce these values, similar to other studies [6,18,24]. In addition, we collected routinely assessed laboratory tests: haemoglobin, haematocrit, leukocytes, thrombocytes, lymphocytes, D-dimer, blood gas analysis, bicarbonate, sodium, potassium, blood urea nitrogen (BUN), creatinine, lactate dehydrogenase (LDH), bilirubin, albumin and C-reactive protein (CRP). If haematocrit and pO_2 values were missing, we used haemoglobin and oxygen saturation to calculate these values, similar to other studies [25,26].

Furthermore, we collected the results of the PCR for SARS-CoV-2 in respiratory specimens and the results of the chest CT scan [20]. The results of the chest CT scan were determined by a radiologist. Finally, we retrieved data on length of hospital stay, admission to the medium care unit (MCU) or intensive care unit (ICU), and 30-day and 14-day mortality. Data on mortality were verified using the medical records. In the Netherlands, all deaths are registered by the municipal administration office, and these data are linked to the medical records.

Prediction models

We searched PubMed for studies on prediction models focussing on patients with COVID-19 using a combination of methodological search terms (prognostic, prediction model, score, regression) and COVID-19 search terms (COVID-19, SARS-CoV-2, coronavirus). In addition, we checked reference lists of manuscripts we identified this way. The search was performed on June 17th and repeated on September 11th to check for more recent publications.

We selected prediction models based on the inclusion of readily available variables in the ED and the aim to predict the risk of mortality or progression to severe illness (i.e. tachypnoea, hypoxia and ICU admission with shock, mechanical ventilation, or organ failure). We excluded models that were not clearly described or were not feasible in our ED setting. Prediction models were also excluded if the included variables or the risk calculation were unclear. Models developed using machine learning techniques other than regression and radiologic models were excluded because these could not be reproduced in our setting.

Outcomes

The primary outcome was all-cause mortality within 30 days of ED presentation. The secondary outcomes were all-cause mortality within 14 days and a composite outcome of 30-day mortality and admission to the

MCU/ICU. In our hospital, all patients admitted to the ICU were mechanically ventilated.

Statistical analysis

Baseline characteristics were analysed using descriptive statistics on the observed data. For each patient, we completed variables of the included prediction models. When the score could be completed in less than 95% of patients due to missing values, data were imputed using stochastic regression imputation. We calculated the AUC under the ROC curve to quantify the discriminatory performance of the included prediction models. An AUC of 0.5 corresponds with very poor discriminatory performance, whereas an AUC of 1.0 means perfect accuracy. We compared the AUCs of the included models using the method of DeLong. All data were analysed using IBM SPSS Statistics for Windows, IBM Corporation, Armonk NY, version 25.0.

Results

Study sample

During the study period, 415 ED patients met the inclusion criteria. After the exclusion of 12 patients because of refusal of informed consent, we included 403 patients for analysis (Table 1). The median age of patients was 71 years (IQR 60-78), and 255 patients (63.2%) were older than 65 years. Most patients

Table 1. Characteristics of the study sample.

	Study sample (<i>n</i> = 403)
Age, median (IQR), years	71 (60–78)
Male, <i>n</i> %	266 (66.0)
Symptoms upon ED visit, <i>n</i> %	
Fever	249 (61.8)
Cough	287 (71.2)
Dyspnoea	278 (69.0)
General malaise	187 (46.4)
Headache	52 (12.9)
Thoracic pain	61 (15.1)
Nausea	92 (22.8)
Diarrhoea	90 (22.3)
No respiratory symptoms	55 (13.6)
PCR positive, <i>n</i> %	323 (80.1)
Chest CT scan positive, <i>n</i> %	325 (80.6)
Outcomes	
Admission to hospital, <i>n</i> %	307 (76.2)
Length of hospital stay, median (IQR), days	6 (3–12)
Treatment restrictions, <i>n</i> %	142 (35.2)
Mortality within 14 days, <i>n</i> %	77 (19.1)
Mortality within 30 days, <i>n</i> %	95 (23.6)
MCU admission, <i>n</i> %	48 (11.9)
ICU admission, <i>n</i> %	66 (16.4)
Composite endpoint (mortality and/or admission to MCU/ICU), <i>n</i> %	152 (37.7)

CT: computed tomography; ED: emergency department; ICU: intensive care unit; IQR: interquartile range; MCU: medium care unit; PCR: polymerase chain reaction; SD: standard deviation.

(66.0%) were male. The PCR for SARS-CoV-2 was positive in 323 patients (80.1%) and the chest CT scan was positive in 325 patients (80.6%). A total of 307 patients (76.2%) were admitted to the hospital, whereas the other patients were discharged home for further recovery. The median length of hospital stay was 6 days (IQR 3–12).

In our sample, 66 patients (16.4%) were admitted to the ICU, 48 patients (11.9%) to the MCU, and 95 patients died during follow up, yielding a 30-day mortality of 23.6% and a 14-day mortality of 19.1%. The survival curve is shown in Figure 1. A total of 152 patients (37.7%) met the composite endpoint of 30-day mortality and admission to MCU/ICU.

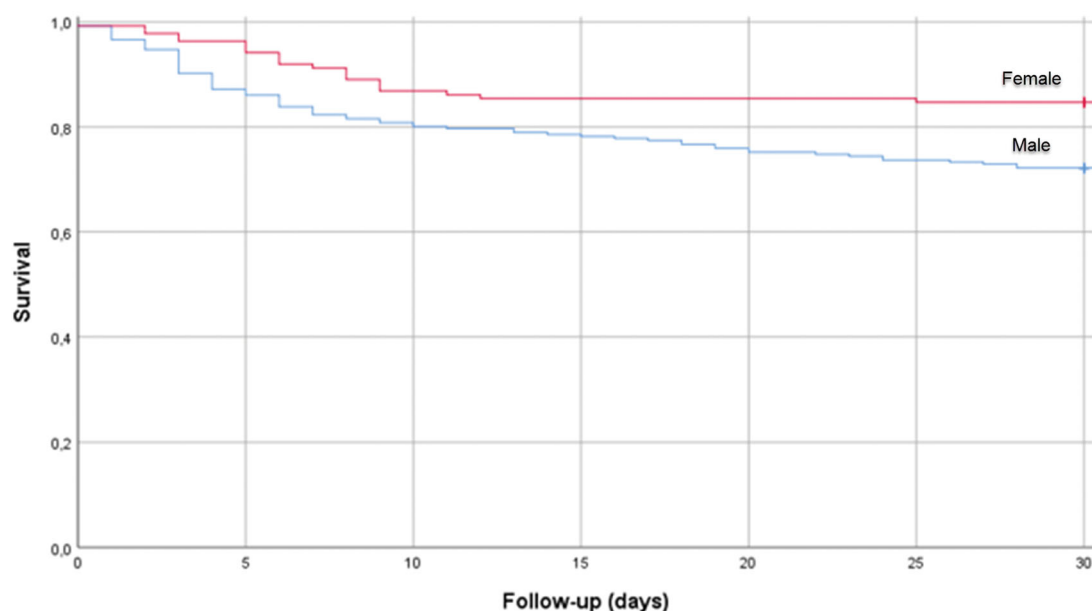
Prediction models

We included 11 prediction models (Table 2), of which seven prediction models were not explicitly developed for patients with COVID-19: RISE UP, CURB-65, Modified Early Warning Score (MEWS), Rapid Emergency Medicine Score (REMS), abbMEDS, SOFA and Acute Physiology And Chronic Health Evaluation II (APACHE II) [6–10,16–18]. Furthermore, in a recent systematic review, 16 prediction models specifically designed for patients with COVID-19 were identified [15]. Of these models, eight estimated mortality risk in patients with suspected or confirmed COVID-19, five aimed to predict progression to severe disease, and three estimated length of hospital stay. We excluded 14 of these models for the following reasons: no clear description of the variables or risk calculation (*n* = 5), not compatible with our setting because of the use of machine learning (*n* = 5), or inclusion of radiologic characteristics (*n* = 4). We included two prognostic models from the systematic review (ACP score and Host risk factor score) [29,30]. Additionally, we included two more recently published prediction models (CALL score and the Coronavirus Clinical Characterisation Consortium (4C) mortality score) not included in the systematic review [27,28].

A total of six prediction models (RISE UP, 4C mortality, CURB-65, SOFA, APACHE II and CALL) could be calculated in less than 95% of the patients because of missing values (vital signs and laboratory tests, Supplementary Table 2). Therefore, missing data were imputed using stochastic regression imputation.

Validation of the prediction models

The prediction models were used to calculate the risk of an adverse outcome (Table 3; Figure 2). The RISE



No at risk	0	5	10	15	20	25	30
Male	266	231	214	208	201	195	192
Female	137	132	119	117	117	116	116

Figure 1. Survival curve and table of the study sample.

Table 2. Overview of included prediction models.

Prediction model	Original population	Items	Outcome
RISE UP [6]	Older patients (>65 years)	Age, HR, MAP, RR, O ₂ saturation, GCS, BUN, bilirubin, albumin, LDH	Mortality (30 days)
4C mortality score [27]	Patients with COVID-19	Age, sex, comorbidity, RR, GCS, O ₂ saturation, BUN, CRP	Mortality (in hospital)
CURB-65 [7]	Patients with pneumonia	Age, confusion, BUN, RR, blood pressure	Mortality (30 days)
MEWS [10]	Patients in the ED	HR, blood pressure, O ₂ saturation, RR, temperature, AVPU	Mortality and ICU admission ^a
REMS [18]	Patients in the ED	Age, MAP, HR, RR, GCS, O ₂ saturation	Mortality ^a
abbMEDS [9]	Sepsis patients	Age, comorbidity, confusion, septic shock, thrombocytes, pneumonia, respiratory symptoms, confusion, nursing home	Mortality (28 days)
SOFA [16]	ICU patients	pO ₂ /FiO ₂ ratio, GCS, MAP, thrombocytes, bilirubin, creatinine	Mortality ^a
APACHE II [17]	ICU patients	Age, HR, RR, MAP, GCS, temperature, haematocrit, pH, pO ₂ , leucocytes, sodium, potassium, creatinine, comorbidity	Mortality ^a
CALL score [28]	Patients with COVID-19	Age, comorbidity, lymphocytes, LDH	Progression to severe disease ^{a,b}
ACP index [29]	Patients with COVID-19	Age, CRP	Mortality (12 days)
Host risk factor score [30]	Patients with COVID-19	Age, sex, hypertension	Progression to severe disease ^{a,b}

4C: Coronavirus Clinical Characterisation Consortium; abbMEDS: abbreviated Mortality Emergency Department Sepsis; ACP: Age C-reactive Protein; APACHE II: Acute Physiology and Chronic Health Evaluation II; AVPU: Alert Verbal Pain Unresponsive; BUN: blood urea nitrogen; CALL: Comorbidity Age Lymphocyte LDH; CRP: C-reactive protein; CURB-65: Confusion Urea Respiration Blood pressure; ED: emergency department; GCS: Glasgow coma scale; HR, heart rate; ICU: intensive care unit; LDH: lactate dehydrogenase; MAP: mean arterial pressure; MEWS: Modified Early Warning Score; REMS: Rapid Emergency Medicine Score; RISE UP: Risk Stratification in the Emergency Department in Acutely Ill Older Patients; RR: respiratory rate; SOFA: Sepsis-related Organ Failure Assessment.

^aTimespan not specified in the original article.

^bProgression to severe disease was not specified in the original article.

UP score and 4C mortality score showed the best discriminatory performance and respectively yielded an AUC of 0.83 (95% CI: 0.79–0.88) and 0.84 (95% CI: 0.79–0.88) for 30-day mortality, an AUC of 0.83 (95% CI: 0.79–0.88) and 0.83 (95% CI: 0.79–0.88) for 14-day mortality, and an AUC of 0.79 (95% CI: 0.74–0.84) and 0.77 (95% CI: 0.72–0.82) for the composite endpoint.

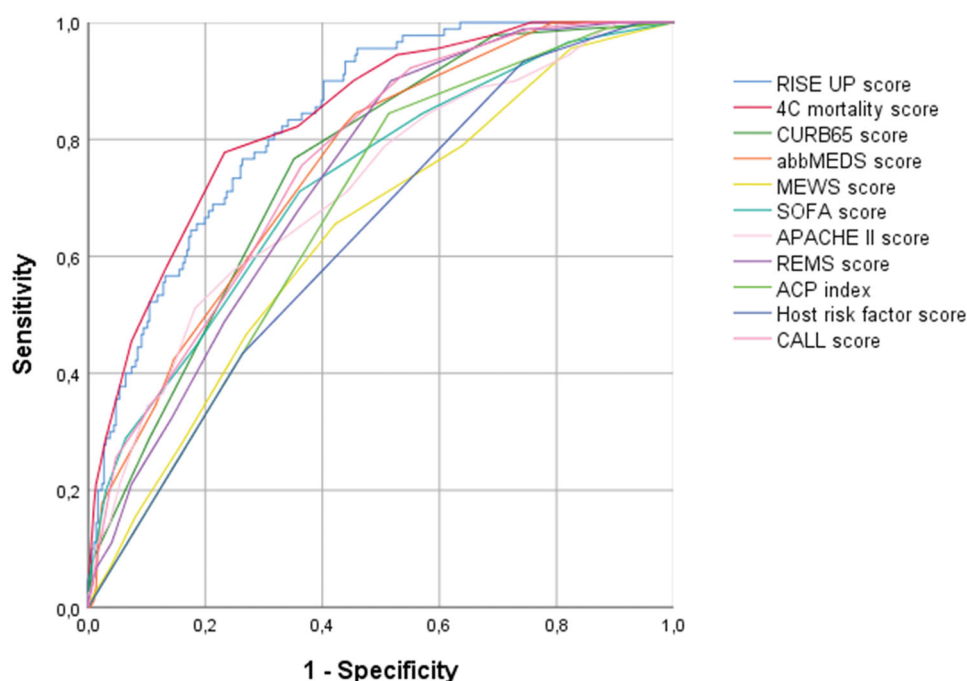
In comparison, the CURB-65, MEWS, REMS, abbMEDS, SOFA, APACHE II, CALL, ACP and Host risk factor score yielded AUCs ranging from 0.64 to 0.76 for 30-day mortality, AUCs ranging from 0.62 to 0.76 for 14-day mortality, and AUCs ranging from 0.68 to 0.76 for the composite endpoint. The discriminatory performance of the RISE UP score and 4C mortality

Table 3. Comparison of the AUCs of included prediction models.

Prediction model	30-day mortality (AUC, 95% CI)	Difference with RISE UP (p value) ^a	14-day mortality (AUC, 95% CI)	Difference with RISE UP (p value) ^a	Composite endpoint (mortality and/or MCU/ICU admission) (AUC, 95% CI)	Difference with RISE UP (p value) ^a
RISE UP	0.83 (0.79–0.88)	Reference	0.83 (0.79–0.88)	Reference	0.79 (0.75–0.84)	Reference
4C mortality score	0.84 (0.79–0.88)	.914	0.83 (0.79–0.88)	.926	0.77 (0.72–0.82)	.170
CURB-65	0.75 (0.70–0.80)	<.001	0.75 (0.70–0.81)	.005	0.68 (0.62–0.73)	<.001
MEWS	0.64 (0.58–0.70)	<.001	0.62 (0.56–0.69)	<.001	0.73 (0.68–0.78)	.037
REMS	0.73 (0.68–0.78)	<.001	0.74 (0.69–0.79)	.001	0.72 (0.66–0.77)	<.001
abbMEDS	0.75 (0.70–0.81)	<.001	0.75 (0.69–0.81)	.003	0.71 (0.66–0.76)	<.001
SOFA	0.72 (0.67–0.78)	<.001	0.72 (0.65–0.78)	<.001	0.76 (0.71–0.81)	.200
APACHE II	0.71 (0.65–0.78)	<.001	0.73 (0.67–0.79)	.004	0.69 (0.63–0.74)	<.001
CALL score	0.76 (0.71–0.81)	.002	0.76 (0.70–0.81)	.012	0.70 (0.65–0.75)	<.001
ACP index	0.67 (0.61–0.73)	<.001	0.67 (0.61–0.73)	<.001	0.66 (0.61–0.72)	<.001
Host risk factor score	0.64 (0.57–0.70)	<.001	0.62 (0.56–0.69)	<.001	0.63 (0.57–0.68)	<.001

4C: Coronavirus Clinical Characterisation Consortium; abbMEDS: abbreviated Mortality Emergency Department Sepsis; ACP: age C-reactive protein; APACHE II: Acute Physiology and Chronic Health Evaluation II; AUC: area under the curve; CALL: Comorbidity Age Lymphocyte LDH; CI: confidence interval; CURB-65: Confusion Urea Respiration Blood pressure; ICU: intensive care unit; MCU: medium care unit; MEWS: Modified Early Warning Score; REMS: Rapid Emergency Medicine Score; RISE UP: Risk Stratification in the Emergency Department in Acutely Ill Older Patients; SOFA: Sepsis-related Organ Failure Assessment.

^aComparison of the AUC of the prediction model with the AUC of the RISE UP score.

**Figure 2.** ROC curves for predicting 30-day mortality.

score was significantly higher than that of the other models using the DeLong method.

Discussion

In this retrospective study, we externally validated 11 prediction models for their ability to predict mortality or admission to MCU/ICU in ED patients with COVID-19. We found that both the RISE UP score and 4C mortality score had very good discriminatory performance, which was the highest of the models we analysed. The models yielded high AUCs for both 14-day

mortality (both AUC of 0.83) and 30-day mortality (AUC of 0.83 and 0.84). The nine other models showed significantly lower discriminatory performance. The CURB-65, REMS, abbMEDS, SOFA, APACHE II and CALL score had a good discriminatory performance (AUC ranging from 0.71 to 0.76). In contrast, the ACP index and Host risk factor score had a moderate to poor performance (AUC of 0.67 and 0.64, respectively). Most prediction models had a higher discriminatory performance for predicting mortality than for predicting the composite outcome of mortality and MCU/ICU admission.

Prediction models

The RISE UP score was recently developed to predict 30-day all-cause mortality in older medical ED patients and consists of easily and readily available items during the ED visit [6]. It is not unexpected that the model works well for admitted patients with COVID-19, since many of these patients in our cohort (63.2%) were 65 years or older. High mortality in older patients with COVID-19 was shown previously [11–14]. The 4C mortality score was recently developed to predict in-hospital mortality in a very large cohort of COVID-19 patients in the UK [27]. The good discriminatory performance of both the RISE UP and 4C mortality scores can be explained because these models include items that reflect the severity of illness in ED patients and are indicative of sepsis, organ failure and/or shock (i.e. abnormal vital signs, LDH, BUN, Bilirubin). The items of the RISE UP and 4C mortality score are quite similar. Elevated levels of LDH were found to predict adverse outcomes in patients with COVID-19 [31]. The prognosis of ED patients is reflected by the presentation of the patients at the ED, which results from both the severity of the current disease and pre-existing factors (i.e. age and comorbidities) [1,4]. Regarding feasibility, the probability of a poor outcome can be predicted in the first two hours of the ED visit by both models. One disadvantage of the 4C mortality score may be that it contains the number of comorbidities of the ED patients, which is not always available in the ED. This is a disadvantage compared to the RISE UP score, which consists of six items readily available in the ED. Moreover, the RISE UP score can easily be implemented with an online calculator (<https://jscalc.io/calc/o1vzp36bIDGQUCYI>). To guide clinical decision-making, prediction models that can be computed easily and quickly are of great value.

The CURB-65 is commonly used to assess the severity and mortality in patients with community-acquired pneumonia [7]. In our cohort, we found that the score had a moderate to good ability to discriminate between mortality and survival (AUC of 0.75). In other studies in patients with COVID-19, the CURB-65 score was found to have a very good discriminatory performance for mortality and progression to severe disease with AUCs ranging from 0.81 to 0.88 [32–35]. The highest AUC (0.88) was found in a Turkish study [33]. Their high AUC may be explained by the inclusion of patients with less severe COVID-19 (more often lower CURB scores and lower mortality) compared to our patients. The MEWS and REMS were designed for early detection of high-risk patients by assigning points to vital signs and can both be easily applied in the ED. In

our cohort, the MEWS score showed only reasonable discriminatory performance for 30-day mortality (AUC of 0.64), while the REMS score yielded moderate to good performance (AUC of 0.73). In one Chinese study, the MEWS score and REMS score were analysed in 138 patients with COVID-19 [36]. The MEWS showed an AUC of 0.68, similar to the AUC in our sample. The REMS score was found to have an AUC of 0.84. Our patients were older than the patients in the Chinese study (median 71 versus 58 years), which probably explains the higher AUC, as the AUC was 0.77 in the 50 Chinese patients older than 65 years.

APACHE II and SOFA scores are used to predict mortality in ICU patients. The discriminatory performance for 30-day mortality of these scores in our cohort was moderate to good (AUC of 0.71 and 0.72, respectively). These findings were comparable to those reported in other studies with patients with COVID-19 [31,32,37]. In one Chinese study in ICU patients with COVID-19, the AUC of the APACHE II score was 0.97, and the AUC of the SOFA score was 0.87, which is much higher than the AUCs we found [32]. However, our patients were less frequently admitted to the ICU (only 16.4%). Consequently, our population is more heterogeneous and mortality is probably more difficult to predict. The APACHE II score was less feasible in an ED setting, because in our ED, an arterial blood gas is measured on indication only (in 37.5% of our patients, no arterial blood gas was measured).

The three other prediction models specifically designed for patients with COVID-19 had varying predictive performances in our cohort. The CALL score had a good predictive value (AUC of 0.76). This CALL score was developed to predict progression to severe disease in the first 5 to 10 days in a cohort of 208 Chinese patients with COVID-19 [28]. The AUC in the Chinese study was 0.86, which was higher than the AUC we found. Application of a new model in an independent cohort usually results in a lower AUC. In addition, the patients in the Chinese cohort were much younger than our patients (mean 44 versus 71 years), and their follow-up period was shorter. The ACP index was developed to predict 12-day mortality in patients with COVID-19 in Wuhan [29]. The Host risk factor score was developed to predict mortality or progression to severe disease [30]. The discriminatory performance of these two scores was not reported by the authors. In our external validation, both scores had poor discriminatory performance (AUC of 0.67 (ACP index) and 0.64 (host risk factor score)). In a recent Spanish study in nursing home residents, the ACP and host risk factor score yielded comparable low

AUCs (AUC of 0.60 and 0.55, respectively) [35]. The difference between our study and the original Chinese studies may also be explained by the different phase of the COVID-19 pandemic in which the studies took place, as in Europe, physicians were already slightly more prepared, and outcomes may therefore differ.

Study limitations

Our study had several limitations. First, our study was performed in a single medical centre, limiting the generalizability of the results. However, our cohort of patients with COVID-19 was relatively large and has been recruited in one of the most heavily affected areas of the Netherlands. Furthermore, by validating all prediction models in the same cohort, there were no differences in the patient sample, and we could truly compare the scores [38]. Second, the process of selecting prediction models for our analysis might have been incomplete. We chose prediction models that were feasible in our ED setting, which may be different for other EDs. Last, in a subgroup of patients with pre-existing frailty or severe comorbidity, it was decided to initiate conservative care only (35.2% had treatment restrictions). As these decisions affect mortality and likelihood of going to the ICU on the one hand, and may differ in other countries on the other hand, we decided to study MCU/ICU admissions as a composite outcome only. In addition, we decided to perform a subgroup analysis in the 261 patients without treatment restrictions ([Supplementary Table 1](#)). We found comparable AUCs for 30-day mortality, 14-day mortality and the composite outcome (AUC of 0.84, 0.82 and 0.81 for the RISE UP, respectively). We found some differences in the performance of the models between patients with and without treatment restrictions, which may be due to the smaller number of patients and the smaller number of events.

Conclusion

In conclusion, the RISE UP and 4C mortality score had the highest discriminatory performance for short term mortality in ED patients with COVID-19. Prediction models like the RISE UP and 4C mortality score are useful for identifying patients at high risk for adverse outcomes and may be a first step in guiding clinical decision-making and allocating healthcare resources in this pandemic in which we have to deal with scarcity of clinical facilities and materials. However, this needs to be the subject of further investigation.

Author contributors

PD, NZ and PMS collected clinical data. PD, NZ and SMJK performed the statistical analysis. All authors interpreted data. PD drafted the first version of the manuscript. NZ, IH SMJK, RB, AL, BS and PMS critically reviewed the manuscript. All authors have read and approved the final version of the manuscript.

Disclosure statement

All authors have no conflicts of interest to disclose.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability statement

Additional data are available upon reasonable request.

References

- [1] Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507–513.
- [2] Sohrabi C, Alsafi Z, O'Neill N, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J Surg*. 2020;76:71–76.
- [3] World Health Organization. Coronavirus disease (COVID-19): situation report - 134 2020. [updated 2 June 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200602-covid-19-sitrep-134.pdf?sfvrsn=cc95e5d5_2.
- [4] Guan WJ, Ni ZY, Hu Y, et al.; China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382(18):1708–1720.
- [5] Bruggemann R, Gietema H, Jallah B, et al. Arterial and venous thromboembolic disease in a patient with COVID-19: a case report. *Thromb Res*. 2020;191:153–155.
- [6] 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.
- [7] Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377–382.
- [8] Shapiro NI, Wolfe RE, Moore RB, et al. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. *Crit Care Med*. 2003;31(3):670–675.
- [9] Roest AA, Tegtmeier J, Heyligen JJ, et al. Risk stratification by abbMEDS and CURB-65 in relation to

- treatment and clinical disposition of the septic patient at the emergency department: a cohort study. *BMC Emerg Med.* 2015;15:29
- [10] Subbe CP, Kruger M, Rutherford P, et al. Validation of a modified Early Warning Score in medical admissions. *QJM.* 2001;94(10):521–526.
- [11] Wang L, He W, Yu X, et al. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect.* 2020;80(6):639–645.
- [12] Shahid Z, Kalayanamitra R, McClafferty B, et al. COVID-19 and older adults: what we know. *J Am Geriatr Soc.* 2020;68(5):926–929.
- [13] Feng Y, Ling Y, Bai T, et al. COVID-19 with different severities: a multicenter study of clinical features. *Am J Respir Crit Care Med.* 2020;201(11):1380–1388.
- [14] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395(10229):1054–1062.
- [15] Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ.* 2020;369:m1328
- [16] Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707–710.
- [17] Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med.* 1985;13(10):818–829.
- [18] Hilderink MJ, Roest AA, Hermans M, et al. Predictive accuracy and feasibility of risk stratification scores for 28-day mortality of patients with sepsis in an emergency department. *Eur J Emerg Med.* 2015;22(5):331–337.
- [19] Cuschieri S. The STROBE guidelines. *Saudi J Anaesth.* 2019;13(Suppl 1):S31–S4.
- [20] Prokop M, van Everdingen W, van Rees Vellinga T, for the COVID-19 Standardized Reporting Working Group of the Dutch Radiological Society, et al. CO-RADS - A categorical CT assessment scheme for patients with suspected COVID-19: definition and evaluation. *Radiology.* 2020;296(2):E97–E104.
- [21] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med.* 2016;35(2):214–226.
- [22] Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373–383.
- [23] Salottolo K, Carrick M, Johnson J, et al. Bar-Or D. A retrospective cohort study of the utility of the modified early warning score for interfacility transfer of patients with traumatic injury. *BMJ Open.* 2017;7(5):e016143.
- [24] Howell MD, Donnino MW, Talmor D, et al. Performance of severity of illness scoring systems in emergency department patients with infection. *Acad Emerg Med.* 2007;14(8):709–714.
- [25] Kokholm G. Simultaneous measurements of blood pH, pCO₂, pO₂ and concentrations of hemoglobin and its derivatives—a multicenter study. *Scand J Clin Lab Invest Suppl.* 1990;203:75–86.
- [26] Madan A. Correlation between the levels of SpO₂ and PaO₂. *Lung India.* 2017;34(3):307–308.
- [27] Knight SR, Ho A, Pius R, et al.; ISARIC4C investigators. 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.
- [28] Ji D, Zhang D, Xu J, et al. Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. *Clin Infect Dis* 2020;71:1393–1399.
- [29] Lu JH, Fan R. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. 2020.
- [30] Shi Y, Yu X, Zhao H, et al. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care.* 2020;24(1):108.
- [31] Han Y, Zhang H, Mu S, et al. Lactate dehydrogenase, an independent risk factor of severe COVID-19 patients: a retrospective and observational study. *Aging (Albany NY).* 2020;12(12):11245–11258.
- [32] Zou X, Li S, Fang M, et al. Acute physiology and chronic health evaluation II score as a predictor of hospital mortality in patients of coronavirus disease 2019. *Crit Care Med.* 2020;48(8):e657–e665–e665.
- [33] Satici C, Demirkol MA, Sargin Altunok E, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. *Int J Infect Dis.* 2020;98:84–89.
- [34] Ma X, Ng M, Xu S, et al. Development and validation of prognosis model of mortality risk in patients with COVID-19. *Epidemiol Infect.* 2020;148:e168.
- [35] Bernabeu-Wittel M, Ternero-Vega JE, Diaz-Jimenez P, et al. Death risk stratification in elderly patients with covid-19. A comparative cohort study in nursing homes outbreaks. *Arch Gerontol Geriatr.* 2020;91:104240.
- [36] 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(6):461–468.
- [37] Assaf D, Gutman Y, Neuman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med.* 2020;15(8):1435–1443.
- [38] Keuning BE, Kaufmann T, Wiersema R, et al.; HEALICS consortium. Mortality prediction models in the adult critically ill: a scoping review. *Acta Anaesthesiol Scand.* 2020;64(4):424–442.