COVID

MDPI

Article

Performance of the CURB65, NEWS2, qSOFA, SOFA, REDS, ISARIC 4C, PRIEST and the Novel COVID-19 Severity Scores, Used to Risk-Stratify Emergency Department Patients with COVID-19, on Mortality—An Observational Cohort Study

Thomas Sheerin, Prerak Dwivedi, Adil Hussain and Narani Sivayoham *

Department of Emergency Medicine, St George's University Hospitals NHS Foundation Trust, Blackshaw Road, London SW17 0QT, UK

* Correspondence: narani.sivayoham@stgeorges.nhs.uk; Tel.: +44-208-6721255 (ext. 1577) or +44-208-7251577

Abstract: Objective: To compare the performance of established scoring systems (CURB65, NEWS2, qSOFA, SOFA and REDS) to the newly developed scores (ISARIC 4C, PRIEST and novel COVID-19 severity scores) in mortality prediction for patients with confirmed COVID-19 infection in the emergency department (ED). **Method:** A retrospective observational cohort study of adult patients attending a teaching hospital ED who fulfilled the criteria for suspected sepsis and tested positive for COVID-19. The scores were calculated for each patient. The primary outcome measure was all-cause in-hospital mortality. Receiver operator characteristic (ROC) curves were generated for each score. The area under the ROC (AUROC) curves were compared to that of the ISARIC 4C score. The cut-off points were determined by the statistical software package. The test characteristics at the cut-off scores were noted. **Results:** Of the 504 patients studied, 153 (30.5%) died in hospital. The AUROC of the ISARIC 4C score was similar to all of the scores except for the NEWS2 score and qSOFA scores, which were significantly lower. The test characteristics of the different scores were similar. **Conclusions:** In this single-centre study, the newly developed COVID scores outperformed the NEWS2 and qSOFA scores but did not perform better than the other scores studied.

Keywords: COVID-19; clinical prediction rule; emergency department; prognosis; discrimination; sepsis

Citation: Sheerin, T.; Dwivedi, P.; Hussain, A.; Sivayoham, N. Performance of the CURB65, NEWS2, qSOFA, SOFA, REDS, ISARIC 4C, PRIEST and the Novel COVID-19 Severity Scores, Used to Risk-Stratify Emergency Department Patients with COVID-19, on Mortality—An Observational Cohort Study. *COVID* **2023**, *3*, 555–566. https://doi.org/10.3390/covid3040040

Academic Editor: Luca Quartuccio

Received: 23 February 2023
Revised: 5 April 2023
Accepted: 6 April 2023
Published: 10 April 2023

1. Introduction

In early 2020, clinicians around the world were faced with COVID-19, a new disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). First identified in Wuhan in December 2019 [1], the virus quickly spread within the population, presenting a unique challenge to healthcare providers. On the 11 March 2020, the WHO declared COVID-19 to be a pandemic. The virus affected people of all ages, both fit and frail, and predicting who would have severe disease was difficult. By the end of March 2023, over three years into the pandemic, there have been over half a billion reverse transcriptase polymerase chain reaction (TR-PCR) tests, over 22 million positive cases and over 221,000 deaths attributed to COVID-19 in the United Kingdon (UK) [2].

Scores that effectively predict disease severity and mortality, whilst not a substitute for clinical judgement, are useful tools for clinicians. They aim to support clinical decision making, enabling the timely initiation of appropriate management and better allocation of resources, hopefully leading to safer patient care and improved patient outcomes. It is therefore essential that these tools perform well as they can directly impact patient management. A score that inaccurately predicts severity or mortality could result in more harm to our patients than benefit.

When faced with this new disease, clinicians turned to previously developed severity scoring and mortality prediction tools to help guide decisions on hospital admission and

predict the requirement for advanced support. Established scores such as the CURB65 [3], NEWS2 [4], qSOFA [5] and APACHEII [6] were applied to patients with COVID-19. However, a subsequent evaluation of these scores found that they underestimated mortality and placed too much emphasis on haemodynamic instability [7]. This is perhaps unsurprising when the pathogenesis of COVID as a respiratory virus is considered. In fact, shock was only a feature in severe COVID-19, which accounted for 5% of cases [8].

As the COVID-19 pandemic developed, new prognostic scores were developed to aid severity and mortality prediction. By January 2021, one year on from the first case of COVID-19, a systematic review identified 232 new models, 107 of which were for predicting mortality, severe disease or the level of support likely to be required. However, all scores were found to have a high or unclear risk of bias and many lacked independent external validation [9].

We selected three scores developed and validated during the COVID-19 pandemic that predicted severity or mortality: the ISARIC 4C mortality score [10], PRIEST [11] and the novel severity score [12]. We selected these three COVID scores as they were developed with the intention of predicting the mortality or risk of severe disease at the first point of access of healthcare using information readily available and routinely collected in the emergency department (ED) combined with an objective functional assessment. We selected five pre-existing comparator scores that we felt were the most relevant to disease caused by COVID-19. The CURB-65 score is commonly used and is recommended by the British Thoracic Society as part of the assessment of community-acquired pneumonia [13]. The NEWS2 score is recommended by the Royal College of Physicians for the monitoring of hospital patients [4] and has good evidence for its use as a triage tool in the ED [14–16]. The qSOFA and Sequential Organ Failure Assessment (SOFA) score [17] are internationally recognised tools for assessing sepsis. The REDS score [18] has been developed and is in use in our ED to risk-stratify patients with suspected sepsis.

Our aim was to compare the performance of these pre-existing scores (CURB-65, NEWS2, qSOFA, SOFA and the REDS scores) to the newly developed scores (ISARIC 4C, PRIEST and novel COVID-19 severity scores) in a cohort of patients presenting to a single ED who tested positive for SARS-CoV-2 by RT-PCR. The primary end-point was in-hospital all-cause mortality.

2. Methods

2.1. Setting, Study Design and Population

This study was carried out in the ED of a university teaching hospital in London, UK. The department sees approximately 130,000 adult patients annually. This is a retrospective study of patient data from the ED suspected sepsis database, which is a convenience sample. Patients meeting the following criteria are entered on the suspected sepsis database: adult patients who received intravenous antibiotics for the treatment of suspected sepsis in the ED prior to admission and met any one of the following criteria: (i) a minimum 3 points on the NEWS2 score on arrival, (ii) have a SBP < 100 mmHg on arrival or (iii) have suspected neutropenic sepsis. The latter group is formed of patients having had chemotherapy in the preceding 6 weeks or other conditions that may cause bone marrow suppression, such as drug-induced agranulocytosis or aplastic anaemia. Patients presenting with COVID-19 confirmed by RT-PCR were extracted from the dataset and studied. Those with a negative test were excluded.

2.2. Study Period and Data Collection

Adult patients who were admitted between 1 January and 31 December 2020 who satisfied the above criteria were studied. The patient's initial vital signs, white cell count (WCC), platelet count, INR, use of warfarin or a direct oral anticoagulant (DOAC), urea, creatinine, bilirubin, serum albumin, C-reactive protein, lactate, presence of refractory hypotension, chronic obstructive pulmonary disease (COPD) and other comorbidities such as dementia, malignancy, obesity, diabetes, neurological disorders, the presence of

cardiovascular, renal or liver impairment, the ability to live independently, use of long-term oxygen therapy (LTOT) and previous do-not-attempt-resuscitation (DNAR) orders (community or in-hospital) are routinely collected. Dependence on care (a minimum three-times-a-day care package or nursing home residency) equated to a Rockwood Clinical Frailty Score (CFS) of six of more [19]. The patients were followed up to discharge from the hospital and their outcome was recorded. The method of arrival, admission to the intensive care unit (ICU) and hospital length of stay were also noted.

The clinical notes were reviewed by researchers (the authors) who are doctors. The data on patients meeting the inclusion criteria were collected for the purpose of continuous audit. The researchers were trained to extract the data from the contemporaneous clinical notes. The data were entered on to an electronic spreadsheet (Excel). All data outside the normal range were rechecked by a second researcher against the original data and corrected where necessary, prior to being anonymised.

The CURB-65, NEWS2, qSOFA, SOFA, REDS, ISARIC 4C, PRIEST and novel COVID-19 severity scores were calculated retrospectively using the initial vital signs, blood tests and required co-morbidities identified on admission. Receiver operator characteristic (ROC) curves were constructed for all scores and the cut-off points were identified for each scoring system by the statistical software programme. Patients who had scores above the cut-off point were deemed to have a high risk of mortality. The test-characteristics were calculated for the high-risk populations. The sensitivities for mortality and the AUROC curves of the ISARIC 4C score were compared to all of the other scores.

When calculating the SOFA score, SaO_2 (peripheral oxygen saturation)/FiO_2 ratio [20] was used in place of the PaO_2/FiO_2 ratio because arterial blood gases were not available for most patients. Where available, SaO_2 on 0.21 FiO_2 (normal atmospheric oxygen) was used when calculating the SaO_2/FiO_2 ratio. With regard to the mean arterial pressure (MAP), a score of 1 point was allocated if the initial MAP or the MAP after a fluid bolus was <70 mmHg and a score of 3 points if refractory hypotension was present. The baseline SOFA score and the admission SOFA scores were calculated. The change in SOFA (ΔSOFA) score was also calculated. The test characteristics of a minimum increase of two points was also calculated. Baseline performance status was calculated using the categories defined by the PRIEST score study.

Finally, the mortality rates between 'wave one' (13 March 2020 to 31 August 2020) and 'wave two' (1 September 2020 to 31 December 2021) of the pandemic were compared.

2.3. Sample Size

The minimum sample size required to perform this study would be a population with 10 deaths per variable in the score [21]. The PRIEST score had the largest number of variables at 10 variables. Thus, the minimum deaths in the population studied should be greater than 100.

2.4. Missing Variables

Missing variables were assigned a score of 0 when calculating all of the scores. This decision was made based on the practical use of the scores. This does not imply that the missing variable was normal. For patients on warfarin or a DOAC, a score of 0 was allocated for INR in the REDS score.

2.5. Statistics

MedCalc Statistical Software version 19.7 (MedCalc Software Ltd., Ostend, Belgium) was used for statistical analysis. The baseline variables were checked for normality. If normally distributed, they were described as mean with standard deviation. When normality was rejected, the data were described as a median together with its interquartile range (IQR). Where normality was rejected for continuous variables, univariate analysis was carried out using the Mann–Whitney test. Differences in categorical variables were analysed using the chi-square test. Statistical significance was defined as $p < 0.05$. The

AUROCs were reported together with their 95% confidence interval (CI). The difference in AUROC curves was assessed by the DeLong method [22].

3. Results

Of the 3097 patients in the sepsis database, 504 patients with COVID-19 were studied as shown in Figure 1. Of the 504 patients studied, 153 died in hospital: a mortality rate of 30.5%. Our cohort of patients satisfied the required sample size.

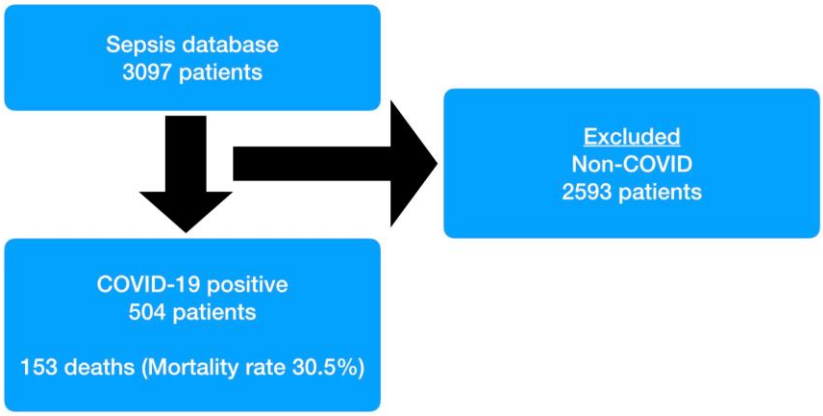

Figure 1. Patient flow diagram.

All physiological variables were available. The missing point of care and laboratory variables were as follows: lactate 20 (4%), INR 46 (9.1%), bilirubin 44 (8.7%), albumin 38 (7.5%), WCC 3 (0.6%), neutrophil count 3 (0.6%), platelets 3 (0.6%) and urea and creatinine 1 (0.2%).

Baseline data for the population studied are presented in Table 1. None of the data were normally distributed. For all scores studied, the difference in score between survivors and non-survivors was significant ($p < 0.0001$). The median age of patients who died was 77 [IQR 65.75–85] years compared to 63 [50.25–75] years, which was statistically significant ($p < 0.0001$). There was no difference in mortality between males and females ($p = 0.32$). There was a statistically significant difference between survivors and non-survivors for all physiological parameters recorded, apart from heart rate and temperature. Similarly, all blood tests showed a statistically significant difference between survivors and non-survivors, apart from the neutrophil count and the total WCC. Of the comorbidities studied, the presence of dementia, a CFS of 6 or more and a previous DNAR order were all found more commonly in non-survivors.

Table 1. The baseline characteristics of the study population.

Patient Characteristics	All (n = 504) Median [IQR] or Number (Percentages)	Survivors (n = 351) Median [IQR] or Number (Percentages)	Non-Survivors (n = 153) Median [IQR] or Number (Percentages)	Difference between Survivors and Non-Survivors
Age (years)	68 [55–80]	63 [50.25–75]	77 [65.75–85]	$p < 0.0001$
Sex (male)	319 (63.3%)	217 (61.8%)	102(66.7%)	$p = 0.32$
Respiratory rate (breaths/minute)	26 [22–32]	26 [22–32]	28 [23–36]	$p = 0.0031$
Oxygen saturation (%)	91.5 [85–95]	92 [88–95]	89 [77.75–94]	$p < 0.0001$
Fraction of inspired oxygen (FiO_2)	0.21 [0.21–0.21]	0.21 [0.21–0.21]	0.21 [0.21–0.3625]	$p < 0.0001$
Heart rate (beats/minute)	98 [84–111]	98 [83–111]	95 [84.75–111]	$p = 0.91$

Table 1. *Cont.*

Patient Characteristics	All (n = 504) Median [IQR] or Number (Percentages)	Survivors (n = 351) Median [IQR] or Number (Percentages)	Non-Survivors (n = 153) Median [IQR] or Number (Percentages)	Difference between Survivors and Non-Survivors
Systolic blood pressure	127 [115–143]	128 [116–143]	89 [77.75–94]	$p < 0.0001$
Temperature (degrees centigrade)	37.6 [36.9–38.4]	37.6 [37–38.6]	37.5 [36.8–38.3]	$p = 0.053$
Glasgow Coma Score	15 [14–15]	15 [15–15]	15 [14–15]	$p < 0.0001$
Altered mental state (new)	151 (30%)	76 (21.7%)	75 (49%)	$p < 0.0001$
Refractory hypotension	8	1	7	$p < 0.001$
Blood results				
WCC ($\times 10^9$/L)	7.6 [5.6–10.7]	7.5 [5.6–10.7]	7.9 [5.7–11]	$p = 0.36$
Neutrophil count	6 [4.2–8.9]	5.75 [4.1–8.5]	6.3 [4.4–9.35]	$p = 0.10$
Platelets ($\times 10^9$/L)	210 [161–288]	216 [164–294.5]	188 [143.75–250.25]	$p = 0.0009$
International normalised ratio (INR)	1.2 [1.1–1.3]	1.2 [1.1–1.3]	1.2 [1.1–1.4]	$p = 0.002$
On warfarin or a DOAC	50 (9.9%)	24 (6.8%)	26 (17%)	$p = 0.001$
Urea (mmol/L)	6.7 [4.5–11.2]	5.8 [4.2–8.875]	9.7 [6.5–14.95]	$p < 0.0001$
Creatinine (micromol/L)	97 [76–138.5]	91 [74–121.75]	124.5 [84–164]	$p < 0.0001$
Bilirubin (micromol/L)	9 [6–12]	8 [6–12]	9 [6–14]	$p = 0.02$
Albumin (g/L)	31 [27–33]	31 [28–34]	29 [26–32]	$p < 0.0001$
C-reactive protein (mg/L)	103 [59–172]	93 [53.75–160.25]	115 [70–211.25]	$p = 0.005$
Lactate (mmol/L)	1.5 [1.1–2.2]	1.4 [1.0–2.0]	1.85 [1.2–2.6]	$p < 0.0001$
Co-morbidities				
Dementia	63 (12.5%)	30 (8.5%)	33 (21.6%)	$p = 0.0001$
Malignancy	25 (5%)	15 (4.3%)	10 (6.5%)	$p = 0.27$
NH residency or live-in carer or minimum TDS care package	105 (20.8%)	51 (14.5%)	54 (35.3%)	$p < 0.0001$
Long-term oxygen therapy (LTOT)	4 (0.7%)	1 (0.3%)	3 (2%)	$p = 0.086$
Community or previous in-hospital DNAR orders	33 (6.5%)	12 (3.4%)	21 (13.7%)	$p < 0.0001$
Any of the above co-morbidities	140 (27.8%)	72 (20.5%)	68 (44.4%)	$p < 0.0001$
Obesity	96 (19%)	69 (19.7%)	27 (17.6%)	$p = 0.62$
Diabetes	153 (30.4%)	99 (28.2%)	54 (35.3%)	$p = 0.11$
Chronic pulmonary disease	53 (10.5%)	33 (9.4%)	20 (13.1%)	$p = 0.27$
Chronic kidney disease	31 (6.2%)	17 (4.8%)	14 (9.2%)	$p = 0.07$
Scores				
ISARIC 4C	12 [8–15]	10 [6–13]	14 [12–16]	$p < 0.0001$
NOVEL	4 [2–5]	3 [2–4]	5 [4–6]	$p < 0.0001$
PRIEST	11.5 [9–14]	10 [8–13]	14 [12–17]	$p < 0.0001$
REDS	2 [1–3]	2 [1–3]	3 [2–4]	$p < 0.0001$
qSOFA	1 [1–2]	1 [1–1]	1 [1–2]	$p < 0.0001$
NEWS2	7 [5–9]	6 [5–8]	8 [6–10]	$p < 0.0001$
CURB65	2 [1–3]	1 [1–2]	3 [2–3]	$p < 0.0001$
SOFA	2 [1–3]	1 [0–2]	3 [2–5]	$p < 0.0001$

Figure 2 shows the ROC curves for the COVID scores and the SOFA score and Figure 3 shows the ROC curves for the ISARIC 4C score and the other scores. The AUROC curve and the test characteristics at the cut-off scores determined by the statistical software package are presented in Table 2. The difference in AUROC curve compared to the ISARIC 4C score and the other scores are presented together with the significance of that difference. The AUROC curve of the NEWS2 and qSOFA scores are significantly smaller compared to the ISARIC 4C score. There was no significant difference between the ISARIC 4C score and the remaining scores. Similarly, there was little difference in the test characteristics.

Of the 284 patients in the first wave, 98 died, which is a mortality rate of 34.5%. Of the 221 patients in the second wave, 56 died, which is a mortality rate of 25.3%. The difference in mortality rates between the first and second wave was significant: $p = 0.03$. The two cohorts were not analysed separately as they did not individually reach the required sample size.

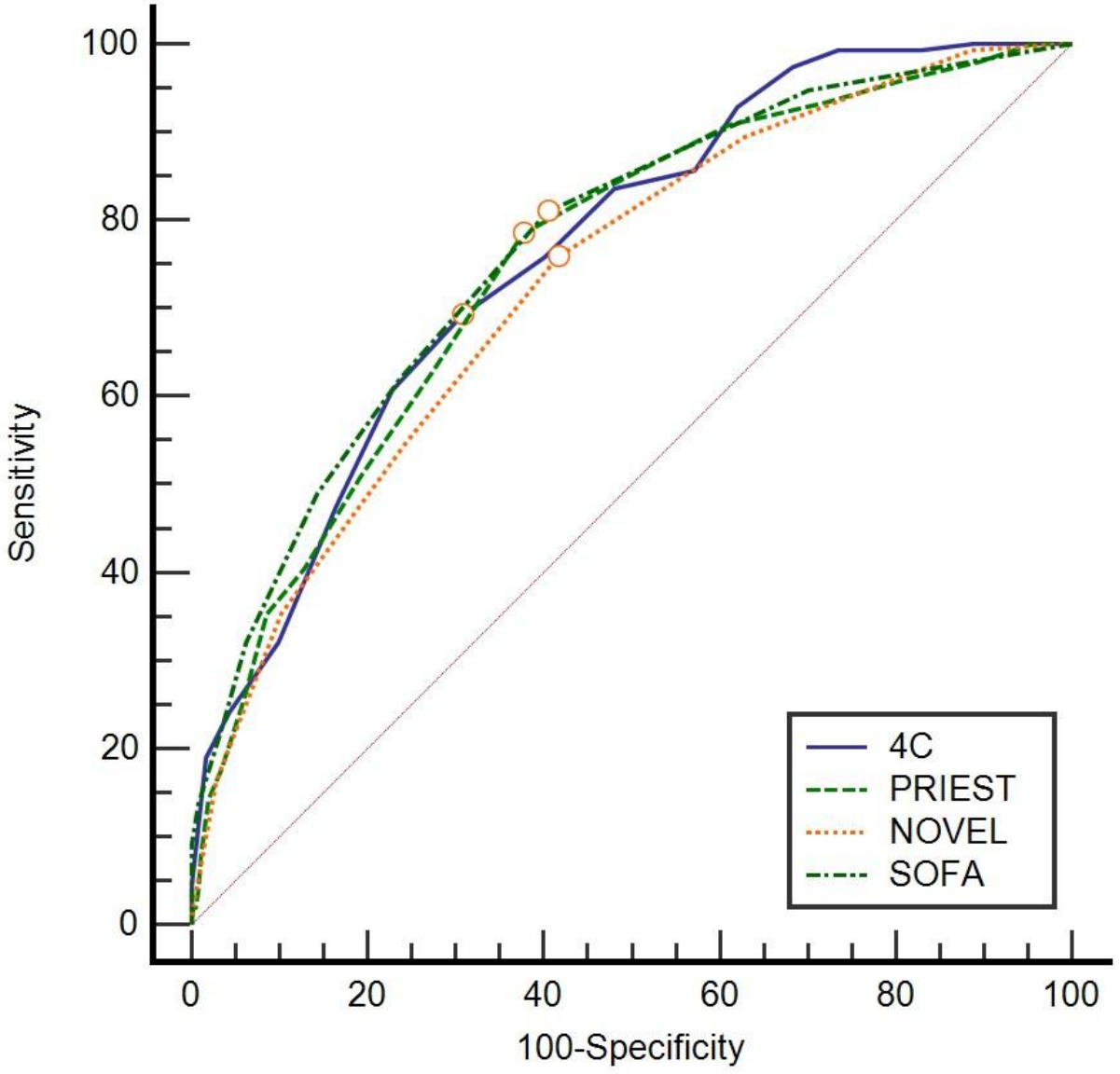

Figure 2. Receiver operator characteristic (ROC) curves for the ISARIC 4C, PRIEST, NOVEL and SOFA scores.

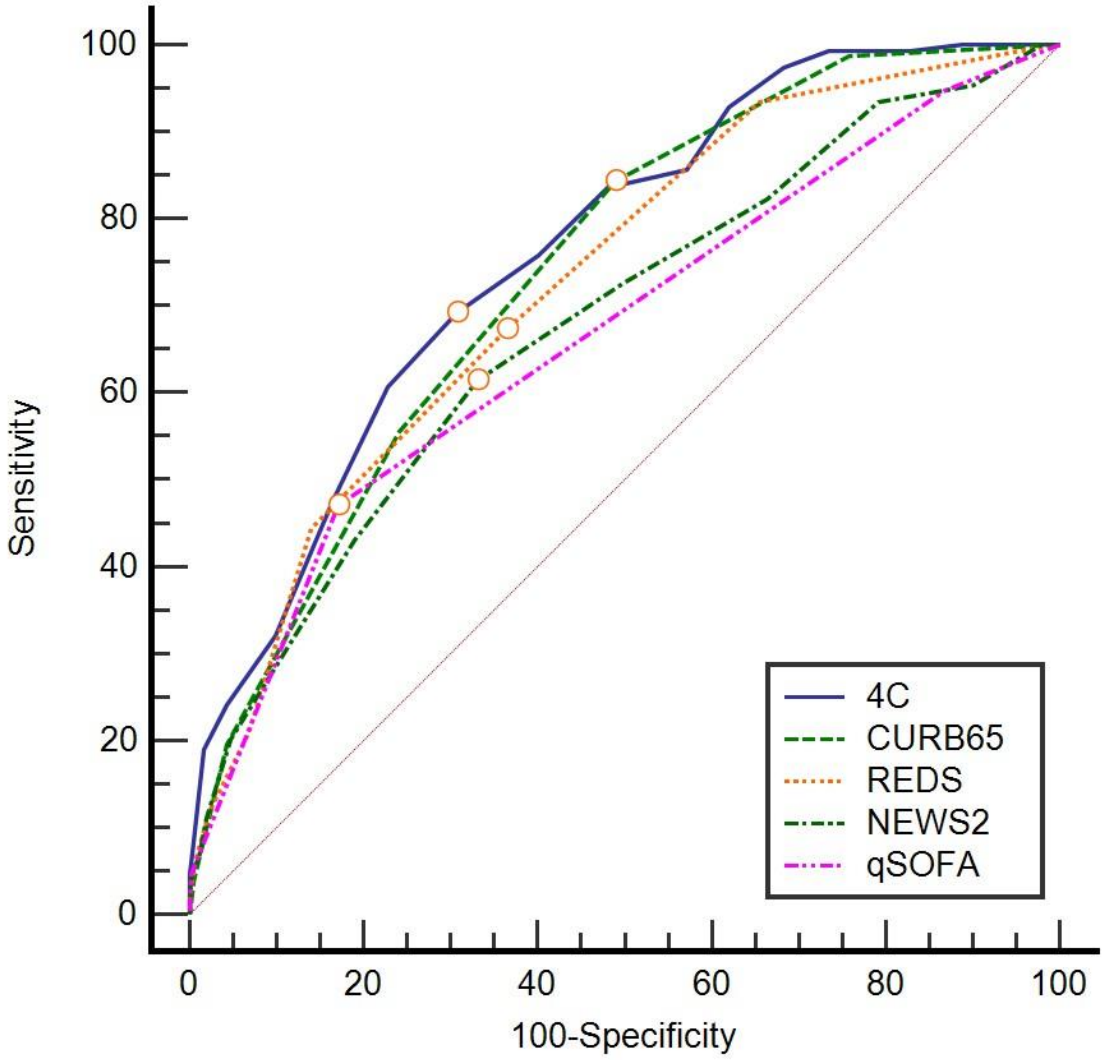

Figure 3. Receiver operator characteristic (ROC) curves for the ISARIC 4C, CURB65, REDS, NEWS2 and qSOFA scores.

Table 2. The AUROC curves, cut-off points and test characteristics of the scores.

Score	AUROC (95% CI)	Significance of Difference in AUROC Curve Compared with 4C Score	Cut-off Score	Sensitivity Percentage (95% CI)	Specificity Percentage (95% CI)	Positive Predictive Percentage (95% CI)	Negative Predictive Value Percentage (95% CI)	Accuracy Percentage (95% CI)
ISARIC 4C	0.76 (0.72–0.8)	NA	≥13	69.3 (61.3–76.5)	69.2 (64.1–74.0)	49.5 (44.8–54.3)	83.8 (80.1–86.9)	69.3 (65.0–73.3)
PREIST	0.75 (0.71–0.79)	$p = 0.61$	≥12	78.4 (71.1–84.7)	62.4 (57.1–67.5)	47.6 (43.7–51.6)	86.9 (82.9–90.1)	67.2 (63.0–71.4)
NOVEL	0.73 (0.69–0.77)	$p = 0.05$	≥4	75.8 (68.2–82.4)	58.4 (53.1–63.6)	44.3 (40.5–48.1)	84.7 (80.5–88.1)	63.7 (59.3–67.9)
SOFA	0.77 (0.73–0.81)	$p = 0.75$	≥2	81.1 (73.9–86.9)	59.5 (54.2–64.7)	46.6 (43.0–50.3)	87.8 (83.7–91.0)	66.1 (61.8–70.2)
CURB65	0.74 (0.7–0.78)	$p = 0.12$	≥2	84.3 (77.6–89.7)	51.0 (45.6–56.3)	42.9 (39.8–46.0)	88.2 (83.6–91.6)	61.1 (56.7–65.4)

Table 2. *Cont.*

Score	AUROC (95% CI)	Significance of Difference in AUROC Curve Compared with 4C Score	Cut-off Score	Sensitivity Percentage (95% CI)	Specificity Percentage (95% CI)	Positive Predictive Percentage (95% CI)	Negative Predictive Value Percentage (95% CI)	Accuracy Percentage (95% CI)
REDS	0.73 (0.68–0.76)	$p = 0.11$	≥3	67.3 (59.3–74.7)	63.5 (58.3–68.6)	44.6 (40.3–49.0)	81.7 (77.8–85.0)	64.7 (60.3–68.9)
NEWS2	0.68 (0.64–0.72)	$p = 0.004$	≥8	61.4 (53.2–69.2)	67.0 (61.8–71.9)	44.8 (40.0–49.6)	79.9 (76.3–83.1)	65.3 (60.9–69.4)
qSOFA	0.67 (0.62–0.71)	$p = 0.0002$	≥2	47.1 (39.0–55.3)	82.9 (78.6–86.7)	54.6 (47.4–61.5)	78.2 (75.4–80.8)	72.0 (67.9–75.9)

AUROC Area Under Receiver Operator Characteristic curve; *CI* Confidence Interval; *REDS* Risk-stratification of Emergency Department Suspected Sepsis; *NEWS2* National Early Warning Score 2, *CURB65* Confusion Urea Respiratory rate Blood pressure 65(years), *SOFA* Sequential Organ Failure Assessment, *qSOFA* quick Sequential Organ Failure Assessment.

4. Discussion

In this retrospective single-centre cohort study, we compared the performance of three prognostic models developed since the start of the COVID-19 pandemic with five pre-existing scores used to assess patients admitted to hospital with the primary endpoint of all-cause mortality during in-patient hospital stay. We found that the NEWS2 and qSOFA scores were not as effective in predicting mortality in our population, with a significantly smaller AUROC curve compared to that of the ISARIC 4C mortality score. However, there was no difference in AUROC curve or the test characteristics between the ISARIC 4C mortality score and the remaining scores studied.

The population studied fulfilled the above-described definition of suspected sepsis. They are therefore likely to be the more unwell COVID-19 patients. This is reflected in the baseline characteristics, with a median NEWS2 score of 7 [IQR 5–9]. Unsurprisingly, the baseline observations of the non-survivor cohort were further from the normal range and, for most parameters, reached a statistically significant difference when compared with survivors. There was no statistically significant difference in the incidence of fever in survivors compared with the non-survivor cohort. This likely reflects that fever is a clinical feature with high prevalence in COVID-19 infection regardless of severity, and was 79.43% in one systematic review [23]. Our data report that the baseline median FiO_2 for both survivors and non-survivor cohorts was 0.21. This reflects the fact that, where available, SaO_2 on 0.21 FiO_2 was used (either recorded at the time of presentation to the ED or documented by pre-hospital providers) to allow for a more accurate calculation of the SaO_2/FiO_2 ratio for the SOFA score [20].

We found that, in our study population, all scores discriminated for mortality, with the AUROC curve values and the 95% CI > 0.5. Of the three newly developed scores, the ISARIC 4C score achieved the highest AUROC curve value of 0.76 (95% CI 0.72–0.8). This AUROC curve value is similar to that published by Knight et al.: 0.79 (95% CI 0.78–0.79) and 0.77 (95% CI 0.76–0.77) in the derivation and validation cohorts, respectively [10]. In addition, the cut-off point of ≥13 in our population was similar to the cut-off point of 15 found by Knight et al.

The PRIEST score had an AUROC of 0.75 (95% CI 0.71–0.79). This is marginally lower than the AUROC curve of 0.8 (95% CI 0.79–0.81) found by Goodacre et al. in their validation cohort [11]. The novel COVID-19 severity score had an AUROC curve of 0.73 (95% CI 0.69–0.77) compared to the AUROC curve published by Aitschui et al. of 0.82 (95% CI 0.81–0.85) in the derivation cohort and 0.8 (95%CI 0.79–0.82) for the validation cohort. Overall, our results are similar to the published validation cohorts.

A review of pre-existing scores applied to patients with COVID-19 published by Bradley et al. [7] provided a useful comparison for our data. This review evaluated the performance of these scores using clinical data routinely recorded at presentation to hospital

with the primary end-point of 30-day mortality. We found that the CURB-65 score had an AUROC curve of 0.74 (95% CI 0.70–0.78), which is comparable to that found by Bradley et al. (0.75) and the AUROC curve of 0.78 for 30-day mortality found by Elmoheen et al. [24]. We found the AUROC curve for qSOFA to be 0.67 (95% CI 0.62–0.71), which is a value similar to 0.62 found by Bradley et al. Other studies found a slightly higher AUROC curve for qSOFA but not significantly so: 0.73 (95% CI 0.72–0.74) [25] and 0.74 (95% CI 0.66–0.82) [26].

The NEWS2 score had an AUROC curve of 0.68 (95% CI 0.64–0.72) in our cohort. This is similar to the AUROC curve of 0.67 found by Bradley et al. [7] and 0.65 (95% CI 0.61–0.68) found by Scott et al. [27]. However, other studies have found an AUROC curve of 0.82 (95% CI0.69–0.95) [28] and 0.84 (95% CI 0.79–0.90) [29]. We found the AUROC curve for the SOFA score to be 0.77 (95% CI 0.73–0.81). This is in the range of other studies. However, unlike the other scores studied, previous research has found a large variability in AUROC curve values for the SOFA score of between 0.59–0.99 [26,30,31].

The AUROC curves for the NEWS2 and qSOFA scores are relatively low (0.68 and 0.67, respectively) compared to the other scores. This may be because they rely solely on the physiological variables. The other scores utilise biochemistry results and incorporate a measure of the baseline function (clinical frailty score and performance status) or consider patient comorbidities. It has been shown that certain comorbidities are risk factors for increasing the severity of disease and mortality [32–35].

It may therefore be expected that scores that include co-morbidity or baseline function might perform better. This does not, however, mean that the physiological-value-only scores are not useful. They may be more suited as triage tools at initial assessment—for example the NEWS2 score at ED triage—rather than a tool for risk stratification or mortality prediction.

The REDS score is currently used in the ED of this centre for patients with suspected sepsis receiving intravenous (IV) antibiotics. It is not routinely calculated for COVID-19 patients unless they are receiving IV antibiotics. The REDS score achieved a similar AUROC curve for mortality when compared to the three COVID scores. Overall, we found that, while the scores developed specifically for assessing patients with COVID-19 perform well for predicting mortality in ED patients, they do not outperform pre-existing scores.

It is important to consider the application of these scores in the ED. Despite the development of more rapid PCR testing for COVID-19, there often remains some diagnostic uncertainty at initial assessment, and there are often multiple acute pathologies present. In these situations, a more general score that has been validated for use in both patients with COVID and with an alternative diagnosis may be more suitable.

Having an accurate and reliable score for predicting mortality that can help to support and direct clinical decision making is invaluable. In our cohort, we found that new models do not outperform pre-existing scoring systems. In the context of a respiratory infection, the CURB65 and NEWS2 scores are of relevance. These are two scores that many clinicians will be familiar with: NEWS2 is recommended by the Royal College of Physicians for the routine monitoring of hospital in-patients [4] and has strong evidence for its use in the ED [14–16], although it may be of more use in tracking clinical course rather than as an initial score for predicting mortality. The CURB65 score is used widely in the assessment of patients presenting with suspected pneumonia [3] and is recommended by the British Thoracic Society as part of the assessment of community-acquired pneumonia [13]. They are already established in current practice, which means that clinicians are likely to find it easier to incorporate them when caring for patients with COVID-19.

A further evaluation of the performance of these scores is warranted both in larger and more diverse patient cohorts. It would also be of benefit to assess these scores in patients with suspected COVID-19 who are discharged from hospital or not treated with antibiotics. This will hopefully reinforce our findings that the scores can predict mortality in patients with COVID-19 and would provide an evidence base for their use in supporting decisions to discharge patients. Our results also find that existing scores perform as well as newly developed scores for predicting mortality and that perhaps future research should focus on

further understanding how these scores relate to disease caused by COVID-19 alongside the development of novel scores.

5. Limitation

Firstly, this is a retrospective single-centre study. Secondly, the cohort selected for this study all fit the previously discussed criteria for suspected sepsis and received intravenous antibiotics. This may have impacted the patient population that was audited. Clinicians may also have become more comfortable with withholding IV antibiotics in patients who were more unwell as they gained more experience managing the disease. Third, the data studied were not normally distributed. We do not know what effect this may have had on the performance of the scores. Finally, our endpoint of mortality at hospital discharge does not account for patients who may have died from an alternative pathology during the admission.

6. Conclusions

In this single-centre observational cohort study, we found that the new scores developed during the COVID-19 pandemic did not outperform some pre-existing scores (CURB-65, SOFA and the REDS scores) in predicting mortality in our population. The NEWS2 and qSOFA scores were less effective in predicting mortality. We have demonstrated that the application of pre-existing scores, such as the CURB-65, is effective for predicting mortality, and suggest that they have benefits over the newly developed scores, including clinician familiarity and application in situations of diagnostic uncertainty.

Author Contributions: Conceptualisation, N.S.; Methodology N.S. and T.S.; Validation T.S., P.D., A.H. and N.S.; Formal analysis N.S. and T.S.; Writing—original draft preparation T.S.; Writing—review and editing, T.S., P.D., A.H. and N.S.; Supervision, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study of routinely collected anonymised data did not include an intervention and did not change the normal process of care. It is a retrospective observational study from a single centre; such studies, in accordance with national Health Research Authority guidance, do not require formal ethics approval [36]. In accordance with the national Health Research Authority guidance around the General Data Protection Regulation and the Data Protection Act of 2018, patient consent is not required for the analysis of anonymised data [37]. This study has been registered with the Clinical Effectiveness and audit office of St George's University Hospital under the registration code AUDI000933.

Data Availability Statement: All data used is included in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef] [PubMed]
2. Public Health England. Coronavirus (COVID-19) in the UK. Available online: https://coronavirus.data.gov.uk/ (accessed on 4 April 2023).
3. Lim, W.S.; Van Der Eerden, M.M.; Laing, R.; Boersma, W.G.; Karalus, N.; Town, G.I.; Lewis, S.A.; Macfarlane, J.T. Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. *Thorax* **2003**, *58*, 377–382. [CrossRef] [PubMed]
4. Royal College of Physicians, UK. National Early Warning Score (NEWS) 2. 2017. Available online: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 (accessed on 3 February 2023).
5. Seymour, C.W.; Liu, V.X.; Kahn, J.M.; Shankar-Hari, M.; Singer, M.; Deutschman, C.S.; Escobar, G.J.; Angus, D.C.; Iwashyna, T.J.; Brunkhorst, F.M.; et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **2016**, *315*, 762–774. [CrossRef]
6. Knaus, W.A.; Draper, E.A.; Wagner, D.P.; Zimmerman, J.E. APACHE II: A severity of disease classification system. *Crit. Care Med.* **1985**, *13*, 818–829. [CrossRef] [PubMed]

7. Bradley, P.; Frost, F.; Tharmaratnam, K.; Wooton, D.G. Utility of established prognaostic scores in COVID-19 hospital admissions: Multicentre prospective evaluation of CURB-65, NEWS2 and qSOFA. *BMJ Open Respir. Res.* **2020**, *7*, e000729. [CrossRef] [PubMed]
8. Siordia, J.A., Jr. Epidemiology and clinical features of COVID-19: A review of current literature. *J. Clin. Virol.* **2020**, *127*, 104357. [CrossRef] [PubMed]
9. Wynants, L.; Van Calster, B.; Collins, G.S.; Riley, R.D.; Heinze, G.; Schuit, E.; Bonten, M.M.J.; Dahly, D.L.; Damen, J.A.; Debray, T.P.A.; et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* **2020**, *369*, m1328. [CrossRef]
10. Knight, S.R.; Ho, A.; Pius, R.; Buchan, I.; Carson, G.; Drake, T.M.; Dunning, J.; Fairfield, C.J.; Gamble, C.; Green, C.A.; 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. [CrossRef]
11. Goodacre, S.; Thomas, B.; Sutton, L.; Burnsall, M.; Lee, E.; Bradburn, M.; Loban, A.; Waterhouse, S.; Simmonds, R.; Biggs, K.; et al. Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort study. *PLoS ONE* **2021**, *16*, e0245840. [CrossRef]
12. Altschul, D.J.; Unda, S.R.; Benton, J.; de la Garza Ramos, R.; Cezayirli, P.; Mehler, M.; Eskandar, E.N. A novel severity score to predict inpatient mortality in COVID-19 patients. *Sci. Rep.* **2020**, *10*, 16726. [CrossRef]
13. Lim, W.S.; Baudouin, S.V.; George, R.C.; Hill, A.T.; Jamieson, C.; Le Jeune, I.; Macfarlane, J.T.; Read, R.C.; Roberts, H.J.; Levy, M.L.; et al. BTS guidelines for the management of community acquired pneumonia in adults: Update 2009. *Thorax* **2009**, *64*, iii1–iii55. [CrossRef] [PubMed]
14. Keep, J.W.; Messmer, A.S.; Sladden, R.; Burrell, N.; Pinate, R.; Tunnicliff, M.; Glucksman, E. National early warning score at Emergency Department triage may allow earlier identification of patients with severe sepsis and septic shock: A retrospective observational study. *Emerg. Med. J.* **2015**, *33*, 37–41. [CrossRef]
15. Corfield, A.R.; Lees, F.; Zealley, I.; Houston, G.; Dickie, S.; Ward, K.; McGuffie, C.; Scottish Trauma Audit Group Sepsis Steering Group. Utility of a single early warning score in patients with sepsis in the emergency department. *Emerg. Med. J.* **2013**, *31*, 482–487. [CrossRef] [PubMed]
16. Bilben, B.; Grandal, L.; Søvik, S. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient—A prospective observational study. *Scand. J. Trauma Resusc. Emerg. Med.* **2016**, *24*, 80. [CrossRef]
17. Vincent, J.L.; Moreno, R.; Takala, J.; Willatts, S.; De Mendoca, A.; Bruining, H.; Reinhart, C.K.; Suter, P.M.; Thijs, L.G. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med.* **1996**, *22*, 707–710. [CrossRef]
18. Sivayoham, N.; Blake, L.A.; Tharimoopantavida, S.E.; Chughtai, S.; Hussain, A.N.; Cecconi, M.; Rhodes, A. The REDS score: A new scoring system to risk-stratify emergency department suspected sepsis: A derivation and validation study. *BMJ Open* **2019**, *9*, e030922. [CrossRef] [PubMed]
19. Rockwood, K.; Song, X.; MacKnight, C.; Bergman, H.; Hogan, D.B.; McDowell, I.; Mitnitski, A. A global clinical measure of fitness and frailty in elderly people. *CMAJ* **2005**, *173*, 489–495. [CrossRef]
20. Grissom, C.K.; Brown, S.M.; Kuttler, K.G.; Boltax, J.P.; Jones, J.; Jephson, A.R.; Orme, J.F., Jr. A Modified Sequential Organ Failure Assessment Score for Critical Care Triage. *Disaster Med. Public Health Prep.* **2010**, *4*, 277–284. [CrossRef] [PubMed]
21. Steill, I.G.; Wells, G.A. Methodological standards for the development of clinical decision rules in emergency medicine. *Ann. Emerg. Med.* **1999**, *33*, 437–447. [CrossRef]
22. Delong, E.R.; Delong, D.M.; Clarke-Pearson, D.L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **1988**, *44*, 837–845. [CrossRef]
23. Islam, M.A.; Kundu, S.; Alam, S.S.; Hossan, T.; Kamal, M.A.; Hassan, R. Prevalence and characteristics of fever in adult and paediatric patients with coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis of 17515 patients. *PLoS ONE* **2021**, *16*, e0249788. [CrossRef] [PubMed]
24. EL Moheen, A.; Abdelhafez, I.; Salem, W.; Bahgat, M.; Elkandow, A.; Tarig, A.; Arshad, N.; Mohamed, K.; Al-Hitmi, M.; Saad, M.; et al. External validation and recalibration of the CURB-65 and PSI for predicting 30-day mortality and critical care intervention in multi-ethnic patients with COVID-19. *Int. J. Infect. Dis.* **2021**, *111*, 108–116. [CrossRef] [PubMed]
25. Artero, A.; Madrazo, M.; Fernandez-Garces, M.; Miguez, A.M.; Garica, A.G.; Vieitez, A.C.; Guijarro, E.G.; Aizpuru, E.M.F.; Gómez, M.G.; Manrique, M.A.; et al. Severity scores in COVID-19 pneumonia: A multicentre, retrospective cohort study. *J. Gen. Int. Med.* **2021**, *36*, 1338–1345. [CrossRef] [PubMed]
26. Liu, S.; Yao, N.; Qiu, Y.; He, C. Predictive performance of SOFA and qSOFA for in-hospital mortality in severe novel coronavirus disease. *Am. J. Emerg. Med.* **2020**, *38*, 2074–2080. [CrossRef]
27. Scott, L.J.; Tavaré, A.; Hill, E.M.; Jordan, L.; Juniper, M.; Srivastava, S.; Redfern, E.; Little, H.; Pullyblank, A. Prognostic value of National Early Warning Scores (NEWS2) and component physiology in hospitalised patients with COVID-19: A multicentre study. *Emerg. Med. J.* **2022**, *39*, 589–594. [CrossRef]
28. Myrstad, M.; Ihle-Hansen, H.; Tveita, A.A.; Andersen, E.L.; Nygård, S.; Tveit, A.; Berge, T. National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19—A prospective cohort study. *Scand. J. Trauma Resusc. Emerg. Med.* **2020**, *28*, 66. [CrossRef] [PubMed]

29. Rigoni, M.; Torri, E.; Nollo, G.; Donne, L.D.; Cozzio, S. NEWS2 is a valuable tool for appropriate clinical management of COVID-19 patients. *Eur. J. Intern. Med.* **2020**, *85*, 118–120. [CrossRef]
30. Raschke, R.A.; Agarwal, S.; Rangan, P.; Heise, C.W.; Curry, S.C. Discriminant Accuracy of the SOFA Score for Determining the Probable Mortality of Patients With COVID-19 Pneumonia Requiring Mechanical Ventilation. *JAMA* **2021**, *325*, 1469. [CrossRef]
31. Yang, Z.; Hu, Q.; Huang, F.; Xiong, S.; Sun, Y. The prognostic value of the SOFA score in patients with COVID-19: A retrospective, observational study. *Medicine* **2021**, *100*, e26900. [CrossRef]
32. Christensen, D.M.; Strange, J.E.; Gislason, G.; Torp-Pedersen, C.; Gerds, T.; Fosbøl, E.; Phelps, M. Charlson Comorbidity Index Score and Risk of Severe Outcome and Death in Danish COVID-19 Patients. *J. Gen. Intern. Med.* **2020**, *35*, 2801–2803. [CrossRef]
33. Wang, B.; Li, R.; Lu, Z.; Huang, Y. Does comorbidity increase the risk of patients with COVID-19: Evidence from meta-analysis. *Aging* **2020**, *12*, 6049–6057. [CrossRef] [PubMed]
34. Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; et al. Clinical characteristics of 138 hospitalised patients with 2019 Novel coronavirus-infected pneumonia un Wuhan, China. *JAMA* **2020**, *323*, 1061–1069. [CrossRef] [PubMed]
35. Sanyaolu, A.; Okorie, C.; Marinkovic, A.; Patidar, R.; Younis, K.; Desai, P.; Hosein, Z.; Padda, I.; Mangat, J.; Altaf, M. Comorbidity and its impact on patients with COVID-19. *SN Compr. Clin. Med.* **2020**, *2*, 1069–1076. [CrossRef] [PubMed]
36. Health Research Authority Decision Tool. Available online: https://www.hra-decisiontools.org.uk/research/ (accessed on 7 April 2023).
37. Available online: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/ (accessed on 7 April 2023).