Utility of severity assessment tools in COVID-19 pneumonia: a multicentre observational study

Authors: Asim Ahmed,[A] Sayed A Alderazi,[B] Rumaisa Aslam,[C] Barooq Barkat,[D] Bethan L Barker,[E] Rahul Bhat,[D] Samuel Cassidy,[F] Louise E Crowley,[G] Davinder PS Dosanjh,[H] Hussain Ebrahim,[A] Najla Elndari,[I] Claudia Gardiner,[J] Atena Gogokhia,[C] Frances S Grudzinska,[K] Megha T Gurung,[B] Terry Hughes,[L] Iyad Ismail,[M] Natasha Iredale,[B] Sannaan Irshad,[D] Sarah Johnson,[N] Diana Kavanagh,[O] Thomas Knight,[B] Alana Livesey,[J] Sebastian T Lugg,[P] Manoj Marathe,[D] Andrew McDougall,[A] Wasim Nawaz,[C] Kimberly Nettleton,[A] Lauren O'Flynn,[J] Kelvin Okoth,[Q] Dhruv Parekh,[R] Rita Perry,[S] Elizabeth J Pudney,[T] Ambreen Sadiq,[C] Olutobi Soge,[U] Rhania Soloman,[B] Marina Soltan,[K] Martin Strecker,[A] Onn S Thein,[K] David Thickett,[V] Ajit Thomas[W] and Riah Thornton[J]

ABSTRACT

Background

Severity scores in pneumonia and sepsis are being applied to SARS-CoV-2 infection. We aimed to assess whether these severity scores are accurate predictors of early adverse outcomes in COVID-19.

Methods

We conducted a multicentre observational study of hospitalised SARS-CoV-2 infection. We assessed risk scores (CURB65, qSOFA, Lac-CURB65, MuLBSTA and NEWS2) in relation to admission to intensive care or death within 7 days of admission, defined as early severe adverse events (ESAE). The 4C Mortality Score was also assessed in a sub-cohort of patients.

Findings

In 2,387 participants, the overall mortality was 18%. In all scores examined, increasing score was associated with increased risk of ESAE. Area under the curve (AUC) to predict ESAE for CURB65, qSOFA, Lac-CURB65, MuLBSTA and NEWS2 were 0.61, 0.62, 0.59, 0.59 and 0.68, respectively. AUC to predict ESAE was 0.60 with ISARIC 4C Mortality Score.

Conclusion

None of the scores examined accurately predicted ESAE in SARS-CoV-2 infection. Non-validated scores should not be used to inform clinical decision making in COVID-19.

KEYWORDS: COVID-19, pneumonia, severity score, collaborative

DOI: 10.7861/clinmed.2020-1107

Authors: [A]specialist registrar in respiratory medicine, Royal Stoke University Hospital, Stoke on Trent, UK; [B]specialist registrar in respiratory medicine, George Eliot Hospital, Nuneaton, UK; [C]specialist registrar in respiratory medicine, Good Hope Hospital, Birmingham, UK; [D]specialist registrar in respiratory medicine, Russells Hall Hospital, Dudley, UK; [E]consultant respiratory physician, Worcestershire Acute Hospitals NHS Trust, Worcester, UK; [F]foundation trainee doctor, Queen Elizabeth Hospital Birmingham, Birmingham, UK; [G]academic clinical fellow in respiratory medicine, Institute of Inflammation and Ageing, Birmingham, UK and Queen Elizabeth Hospital Birmingham, Birmingham, UK; [H]consultant respiratory physician, Institute of Inflammation and Ageing, Birmingham, UK and Birmingham Lung Research Unit, Birmingham, UK; [I]internal medicine trainee, George Eliot Hospital, Nuneaton, UK; [J]specialist registrar in respiratory medicine, Worcestershire Acute Hospitals NHS Trust, Worcester, UK; [K]academic clinical fellow in respiratory medicine, Institute of Inflammation and Ageing, Birmingham, UK and Queen Elizabeth Hospital Birmingham, Birmingham, UK; [L]statistician, Birmingham Centre for Observational and Prospective Studies (BiCOPS), Birmingham, UK; [M]consultant respiratory physician, Good Hope Hospital, Birmingham, UK; [N]respiratory research nurse, Queen Elizabeth Hospital Birmingham, Birmingham, UK; [O]consultant respiratory physician, Sandwell and West Birmingham NHS Trust, Birmingham, UK; [P]clinical lecturer in respiratory medicine, Institute of Inflammation and Ageing, Birmingham, UK and Queen Elizabeth Hospital Birmingham, Birmingham, UK; [Q]research administrator, Birmingham Centre for Observational and Prospective Studies (BiCOPS), Birmingham, UK; [R]consultant in critical care and respiratory medicine and senior clinical lecturer in acute and critical care, Institute of Inflammation and Ageing, Birmingham, UK and Queen Elizabeth Hospital Birmingham, Birmingham, UK; [S]research manager, Birmingham Centre for Observational and Prospective Studies (BiCOPS), Birmingham, UK; [T]consultant respiratory physician, Russells Hall Hospital, Dudley, UK; [U]internal medicine trainee, Good Hope Hospital, Birmingham, UK; [V]professor of respiratory medicine, Institute of Inflammation and Ageing, Birmingham, UK and Queen Elizabeth Hospital Birmingham, Birmingham, UK; [W]consultant respiratory physician, Royal Stoke University Hospital, Stoke on Trent, UK

Introduction

The COVID-19 pandemic is a global health emergency that has resulted in over 172 million infections and more than 3.7 million deaths.[1] In a proportion of cases, SARS-CoV-2 infection causes pneumonia (C-CAP).[2] This can progress to development of the acute respiratory distress syndrome (ARDS) precipitating the need for invasive mechanical ventilation. Mortality of patients with COVID-19-related ARDS is similar to usual ARDS.[3]

Repurposed treatments (such as steroids and anti-interleukin-6) have been shown to reduce mortality; but are only effective in a subset of hospitalised patients (such as those requiring oxygen therapy or those with elevated inflammatory markers).[4,5] This highlights the heterogeneity of SARS-CoV-2 infection and demonstrates the need for tools that accurately stratify risk of adverse outcomes.

Early targeted treatment for those most at risk may attenuate progression to more severe disease.[6] It is, therefore, vital that we identify COVID-19 patients at high risk of progression to severe disease as soon as possible after hospital admission. Ideally this would be done with easily measurable markers already in routine clinical practice. This could enable early triage of patients to appropriate supportive care and guide risk stratification into ongoing clinical trials.

Studies have developed their own risk stratification tools to enable identification of high-risk patients who may benefit from therapies such as immunosuppression.[7] However, this tool requires external validation and additional training for clinicians prior to use.[8] The rush to understand and be better prepared to treat COVID-19 has resulted in a number of novel prognostic tools, which may have a high risk of bias and inaccurate reporting.[9] More recently, the ISARIC 4C (Coronavirus Clinical Characteristic Consortium) Mortality Score has been developed for patients hospitalised with COVID-19 to aid categorisation of those at highest risk of death.[10]

In practice, prognostic tools developed for community-acquired pneumonia (CAP) and other diseases have continued to be used for SARS-CoV-2 infection. CAP is a leading cause of sepsis worldwide. Prompt identification of those at high risk of adverse outcomes improves survival by enabling early escalation of care.[11] There are multiple severity assessment tools recommended for risk stratification in CAP and sepsis. Typically, these scores combine early admission physiology with initial investigations to derive a score.[12,13] However, there is little quantitative information about how these scores perform in patients hospitalised with COVID-19, despite their continued clinical use.[13] Anecdotal reports suggested that established scoring systems were not providing additional benefit to patient management. However, the ability to use a familiar tool to stratify risk is appealing to clinicians.

The aim of this study, therefore, was to determine how currently used scores for CAP and sepsis (confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score (CURB65); lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score (Lac-CURB65); quick Sequential (Sepsis-related) Organ Failure Assessment (q-SOFA); multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score (MuLBSTA); and National Early Warning Score 2 (NEWS2)) perform as risk stratification tools for early severe outcomes in SARS-CoV-2 infection. Following the development of the 4C Mortality Score, we also assessed it as a risk stratification for early severe outcomes in patients with COVID-19.

Methods

Study institution and subjects

The study included all adults admitted to eight NHS hospitals within the West Midlands, UK, with confirmed SARS-CoV-2 infection between March 2020 and July 2020, during the first wave of the pandemic. Each institution developed a method of identifying cases: cases were collected in a consecutive manner or by random selection of one-third of cases for institutions with >1,000 admissions during the study period. Infection diagnosis was based upon polymerase chain reaction (PCR) analysis of a combined nose and throat swab as recommended by Public Health England. PCR negative patients were excluded from the analysis.

C-CAP was defined as those with confirmed SARS-CoV-2 infection who also met the criteria for CAP using British Thoracic Society guidelines.[14]

Ethical approval was deemed not to be required based on the UK Health Research Authority decision tool, local approval were granted at each site by research governance teams.[15] Patient identifiable data were accessed by direct care teams only. Study data were collected and managed using the Research Electronic Data Capture (REDCap) tool hosted by the University of Birmingham.[16,17] REDCap is a secure, web-based software platform designed to support data capture for research studies, providing an intuitive interface for validated data capture; audit trails for tracking data manipulation and export procedures; automated export procedures for seamless data downloads to common statistical packages; and procedures for data integration and interoperability with external source.

Patient management

Patients were admitted and treated initially according to BTS guidelines for CAP with antibiotics, fluids and controlled oxygen.[14] In general, presence of radiographic pneumonia and/or a raised procalcitonin were indicators for antibiotic therapy. CURB65 was used to guide antibiotic therapy.

Across the study sites, the following interventional trials were recruiting during the period of data collection: RECOVERY (NCT04381936), RECOVERY RS (ISRCTN16912075), REMAP-CAP (NCT02735707) and inhaled interferon-beta-1α in COVID-19 (NCT04385095).

All patients received venous thromboembolic disease prophylaxis unless it was contraindicated. Target oxygen saturations were ≥92%, unless at risk of hypercapnic respiratory failure.

Provision of ward-based continuous positive airway pressure (CPAP) non-invasive ventilation varied between sites, some only offered CPAP as part of the RECOVERY RS trial, others used CPAP for patients who were not for escalation to the intensive care unit (ICU). Bi-level ward-based non-invasive ventilation was not widely used. Use of awake or conscious proning was common across all sites. At the time of data collection, dexamethasone and anti-interleukin-6 were not recognised as standard care and, therefore, were not used outside the auspices of the recruiting trials.

Data collection and scoring analysis

In addition to the first set of physiological observations recorded on admission to hospital (level of alertness, respiratory rate,
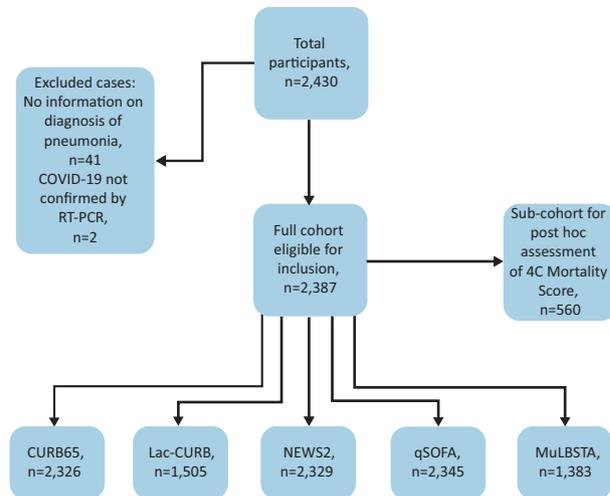
Fig 1. Modified consort diagram demonstrating inclusion of cases. For a detailed overview of missing data, see supplementary material S1, Table S1. CURB65 = confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; Lac-CURB65 = lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; MuLBSTA = multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score; NEWS2 = National Early Warning Score 2; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; RT-PCR = reverse transcription – polymerase chain reaction.

temperature, oxygen saturations, blood pressure and heart rate), the first set of venous or arterial blood gas, biochemical and haematological laboratory results were also collected from electronic patient records.

CURB65, Lac-CURB65, NEWS2, MuLBSTA and qSOFA scores were calculated as per their validation studies.[18–22] 4C Mortality Scores were subsequently calculated in patients recruited from a single centre (termed sub-cohort), as per availability of variable data.[10]

To assess for confusion in qSOFA, CURB65 and Lac-CURB65, we reviewed the admission documents and scored for confusion if any of the following were documented: abnormal AVPU score (alert, response to voice, pain or unresponsive), Glasgow coma scale score ≤14, abnormal mental state examination, or documentation of confusion or delirium. See Fig 1 for modified consort diagram for completeness of score data.

Lac-CURB65 score and NEWS2 were grouped into 'low', 'moderate' and 'high' risk categories, as previously described.[19,20] In the sub-cohort, 4C Mortality Scores were grouped into low (0–3), intermediate (4–8), high (9–14) and very high (≥15) risk, as previously described.[10]

The primary composite outcome measures assessed were death within 7 days of admission or admission to ICU. A composite endpoint of these two outcomes was used, termed early severe adverse event (ESAE). We adopted a composite outcome because it was clear that, for many patients, it was not deemed appropriate to be admitted to the ICU, but they still had severe and often fatal infection. Seven-day fixed period mortality was chosen for multi-centre data collection as this would not be influenced by delays in hospital discharge and could easily be captured within the short period of data collection compared with in-hospital mortality. Additional in-hospital mortality data were captured for the sub-cohort analysis.

Patient and public involvement

There was no patient or public involvement in the development of this study. This study was undertaken in accordance with the STROBE guidelines for cohort studies.

Statistical analysis

Differences in parameters between groups were assessed using the Mann–Whitney *U* test. The performance of severity assessment scores was assessed in all COVID-19 patients, as well as in the sub-group who had evidence of C-CAP. Comparison of proportions was performed using the chi-squared test for trend. Sensitivity, specificity, receiver operating characteristic (ROC) curve analysis, and positive and negative predictive values (PPV; NPV) were calculated for each scoring system. For missing data, we undertook multiple imputation. DeLong's test was used to compare areas under the curves (AUCs). Statistical analysis was carried out using Stata SE version 16.1 and SPSS version 27. Data are presented as mean for normally distributed data and median (interquartile range (IQR)) for non-parametric data.

Results

Demographics and outcomes

Two-thousand, four-hundred and thirty patients were screened for inclusion. Forty-one were excluded as there were no data regarding presence of CAP, two were excluded as COVID-19 was not confirmed by reverse transcription – PCR testing. Two-thousand, three-hundred and eighty-seven cases were included in the final analysis (Fig 1). Missing data and frequency of imputed data are described in supplementary material S1, Table S1.

The median age of patients included was 73 years (IQR 58–82). Fifty-four per cent (1,293) of cases were men. Seventy-three per cent (1,750) of cases fulfilled the criteria for C-CAP. Comorbidities were similar between C-CAP and non-C-CAP cases. Overall, 7-day mortality was 18%; with an ICU admission rate of 12%. The composite outcome of ESAE occurred in 28% of cases. ICU admission, 7-day mortality and the composite outcome occurred more commonly in the C-CAP group. Full demographic and outcome information is available in supplementary material S1, Table S2.

Ability of the severity scoring systems to stratify risk

In the overall SARS-CoV-2 infection group, increasing score value was associated with increased risk of 7-day mortality and increased likelihood of ESAE for all five scores examined (supplementary material S1, Table S3). In the C-CAP alone group, again increasing score value was associated with increased risk of adverse outcome for all five scores tested (Table 1).

Overall accuracy of the different scoring systems to predict ESAE

Receiver operator curve analysis was performed to assess overall accuracy of the scores to predict 7-day mortality and ESAE in both SARS-CoV-2 infection and C-CAP. For all SARS-CoV-2 infection, AUC to predict the combined ESAE for CURB65, qSOFA, Lac-CURB65, MuLBSTA and NEWS2 were 0.61, 0.62, 0.59, 0.59 and 0.68, respectively (Fig 2).

Table 1. Ability of severity scoring systems to stratify risk for 7-day mortality or early severe adverse event for COVID-19 pneumonia only

CURB65, n=1,708	0	1	2	3	4	5	p value
Number of cases, n	331	431	470	366	99	11	
ESAE, n (%)	67 (20)	133 (31)	158 (34)	154 (42)	52 (53)	7 (64)	<0.001
Lac-CURB65, n=1,155	Low	Moderate	High				
Number of cases, n	372	407	376				
ESAE, n (%)	95 (26)	130 (32)	163 (43)				<0.001
NEWS2, n=1,708	Low	Moderate	High				
Number of cases, n	637	385	686				
ESAE, n (%)	106 (17)	131 (34)	337 (49)				<0.001
MuLBSTA score, n=1,042	Low	High					
Number of cases, n	446	596					
ESAE, n (%)	112 (25)	210 (35)					<0.001
qSOFA score, n=1,718	0	1	2	3			
Number of cases, n	554	883	258	23			
ESAE, n (%)	111 (20)	336 (35)	115 (45)	13 (57)			<0.001

Comparison of proportions were done using chi-squared test for trend. Lac-CURB65 cut-offs: low = CURB65 ≤1 and/or lactate <2.0 mmol/L; moderate = CURB65 2 and/or lactate 2.0–4.0 mmol/L; high = CURB65 ≥3 and/or lactate >4.0 mmol/L. NEWS2 score cut-offs: low = score 0–4; medium = score 5–6; high = aggregate score ≥7. CURB65 = confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; ESAE = early severe adverse event; Lac-CURB65 = lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; MuLBSTA = multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score; NEWS2 = National Early Warning Score 2; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment.

For C-CAP, AUC to predict ESAE for CURB65, qSOFA, Lac-CURB65, MuLBSTA and NEWS2 were 0.61, 0.61, 0.59, 0.56 and 0.67, respectively (Fig 3). NEWS2 was significantly better at predicting ESAE than CURB65, LacCURB65, qSOFA and MuLBSTA (supplementary material S1, Table S4).

Performance characteristics of severity assessment tools to predict ESAE

Using ESAE as the outcome of interest, we calculated performance characteristics using pre-defined cut-offs. NEWS2 ≥medium had
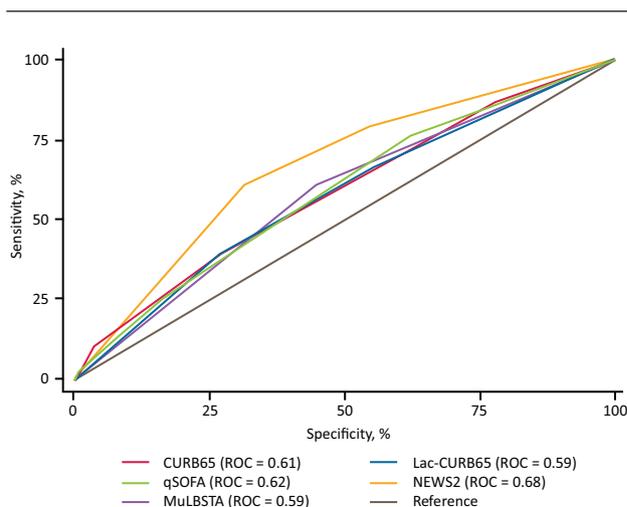

Fig 2. Receiver operator curve for early severe adverse event in all SARS-CoV-2 infection. CURB65 = confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; Lac-CURB65 = lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; MuLBSTA = multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score; NEWS2 = National Early Warning Score 2; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; ROC = area under the receiver operator curve.
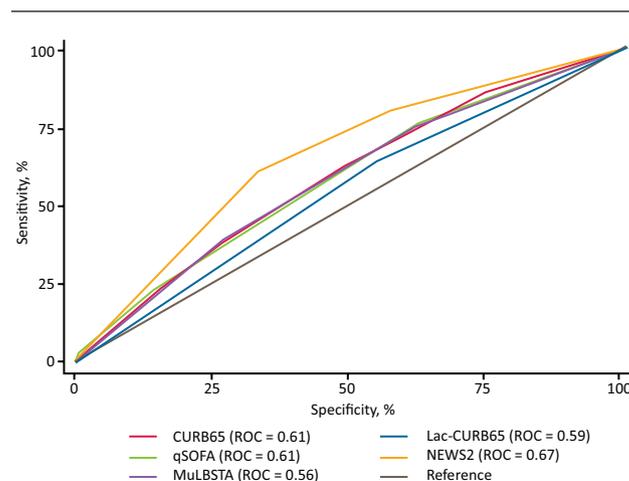

Fig 3. Receiver operator curve for early severe adverse event in COVID-19-associated community-acquired pneumonia only. CURB65 = confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; Lac-CURB65 = lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; MuLBSTA = multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score; NEWS2 = National Early Warning Score 2; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; ROC = area under the receiver operator curve.

Table 2. Performance characteristics of severity assessment tools to predict early severe adverse outcome in SARS-CoV-2 infection

	Sensitivity, %	Specificity, %	PPV, %	NPV, %
CURB65 ≥2, n=1,353	66.00	48.51	33.20	78.63
CURB65 ≥3, n=700	38.00	76.61	38.65	76.11
Lac-CURB-65 ≥moderate, n=1,035	76.22	34.20	31.59	78.30
Lac-CURB-65 high, n=489	43.36	71.84	38.04	76.08
qSOFA ≥2, n=431	23.82	86.21	40.10	74.49
NEWS2 ≥medium, n=1,349	80.12	50.75	38.84	86.73
NEWS2 high, n=850	57.80	71.82	44.47	81.34
MuLBSTA high, n=666	62.01	56.68	33.33	81.03

CURB65 = confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; Lac-CURB65 = lactate, confusion, urea nitrogen, respiratory rate, blood pressure and age of 65 years or older score; MuLBSTA = multilobular infiltration, hypo-lymphocytosis, bacterial coinfection, smoking history, hyper-tension and age score; NEWS2 = National Early Warning Score 2; NPV = negative predictive value; PPV = positive predictive value; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment.

the best overall performance with sensitivity of 80.12% and a negative predictive value of 86.73%, overall qSOFA ≥2 had the poorest sensitivity and negative predictive value (23.82% and 74.49%, respectively; Table 2).

Assessment of the impact of missing values on the analysis

For most variables, data was missing in <1% (supplementary material S1, Table S1). For variables with significant missing data (lactate, smoking status, evidence of multi-lobar consolidation and bacterial co-infection), multiple imputation was undertaken. Complete data could not be imputed, however, with multiple imputation, data were complete in >95% of cases. The full analysis was repeated using multiple imputation and there were no significant differences from the data presented.

Assessment of the of 4C Mortality Score

A total of 560 eligible patients with SARS-CoV-2 infection from the largest single centre were subsequently assessed for assessment of 4C Mortality Scores, of those, 462 patients had C-CAP (supplementary material S1, Table S5). In both C-CAP (n=407) and SARS-CoV-2 infection (n=493), increasing 4C Mortality Scores were associated with increased risk of ESAE, 7-day mortality and in-hospital mortality (supplementary material S1, Table S6 and Table S7). The overall accuracy of the 4C Mortality Score for ESAE in SARS-CoV-2 infection was 0.60, compared with CURB65 (0.58), qSOFA (0.60) and NEWS2 (0.66; supplementary material S1, Fig S1 and Fig S2). Using ESAE, the 4C Mortality Score of ≥intermediate had a sensitivity of 96.96%, specificity of 10.53%, PPV of 32.30% and NPV of 87.8%, comparing with sensitivity of 85.53% and NPV of 88.02% in NEWS2 ≥medium (supplementary material S1, Table S8). The overall accuracy of 4C Mortality Score in SARS-CoV-2 infection for 7-day mortality and in-hospital mortality was 0.78 and 0.76, respectively, comparing with CURB65 (0.78 and 0.75) qSOFA (0.68 and 0.67) and NEWS2 (0.68 and 0.68; supplementary material S1, Fig S3, Fig S4, Fig S5, Fig S6).

Discussion

This study describes a large cohort of hospitalised patients with SARS-CoV-2 infection, in a diverse, multi-centre setting, in an area of the UK heavily impacted during the first wave of the pandemic in Spring 2020. This work confirms that commonly used severity assessment tools can stratify risk according to increasing risk of adverse outcomes, however, none of the examined tools demonstrated clinically useful discrimination in ESAE.

A key strength of this study is use of established severity assessment tools, which have been adapted into clinical practice during the COVID-19 pandemic. We also included MuLBSTA, an established score validated for viral pneumonia, but not specifically COVID-19.[21] In addition, we were able to compare the performance of these generic assessment scores with the new validated COVID-19-specific ISARIC 4C Mortality Score in a subset of the cohort.

CURB65 is a well-established score for prediction of mortality in CAP, the benefit of CURB65 is easy bedside scoring and most clinicians have a good understanding of its limitations.[18] CURB65 is validated to predict 30-day mortality, and the disease course in COVID-19 is different to conventional CAP. Even in CAP, CURB65 is not a perfect score with AUC varying from 0.65–0.83 to predict 30-day mortality.[23,24]

In this study, CURB65 was not an accurate predictor of ESAE in either SARS-CoV-2 infection or C-CAP with AUC 0.61 for both. Adjusting the cut-off used to ≥2 rather than ≥3 improved sensitivity but not specificity although NPV was higher at 81%. The limitation of CURB65 in COVID-19 may reflect the absence of oxygenation parameters, in contrast to CAP, haemodynamic instability is less common in COVID-19 while hypoxia is more common. Additionally, mortality and ESAE in each of the CURB65 groups was higher than seen in previously stated CAP cohorts; typically in CAP, a CURB65 score of 0–1 would indicate that a patient could be considered for ambulatory management, but in this study, a CURB65 of 0–1 had between 2%–11% mortality, and 16%–31% ESAE, demonstrating that these scores should not be used in isolation.[18]

Lac-CURB65 was developed to refine CURB65 and including lactate does improve the prognostic ability in CAP.[19,25,26] In this

study, addition of lactate does not seem to improve the prognostic ability of CURB65, either in the overall infection group or in the C-CAP group.

MuLBSTA has been validated in a cohort of viral pneumonia, with an AUC of 0.811 (superior to CURB65) in the validation study, however, in this cohort, MuLBSTA was inferior to CURB65 in the overall prognostic ability but also demonstrated a low sensitivity and NPV.[21] The MuLBSTA derivation study defines bacterial infection as bacterial growth in either sputum or blood, however, this definition was not suitable to apply at the point of admission, so we adapted the score to focus less on microbiological sampling and took the view of the treating clinician at the point of admission.[21] This change in definition may have impacted the accuracy of MuLBSTA in this cohort, but clearly waiting for culture results is not feasible when rapid decision making about severity assessment is warranted. The use of MuLBSTA in clinical practice is also limited by the need for smoking history for score calculation, which was missing in >40% of the patients in our study.

We also opted to include severity assessment tools validated for use in sepsis, and all acute adult hospital admissions (qSOFA and NEWS2). qSOFA has been shown to predict increasing risk of mortality in sepsis.[27] In this work, qSOFA had a very low sensitivity to predict ESAE that includes mortality when calculated at the point of admission. It may be that qSOFA is more representative when scored serially rather than in isolation, as demonstrated in other CAP studies of qSOFA.[28]

NEWS2 is the recommended screening tool in use in UK hospitals to detect deterioration of any aetiology.[20] In this study, NEWS2 had the best overall performance in both the overall COVID-19 infection group, and C-CAP group with AUC 0.68 and 0.67. NEWS2 with a cut-off of ≥medium also had high sensitivity and NPV to predict ESAE. NEWS2 is routinely recorded, and the composite score used as an indicator for senior medical review. It has the advantage of including a wide range of physiological observations but does not need any investigation results, making it easily applicable at the point of admission. Though it performed the best of the scores assessed, we would caution against its sole use, and clinical assessment and acumen remain essential.

The 4C Mortality Score was developed using UK data from greater than 35,000 cases. As described by this study, they found that established scores (such as qSOFA, CURB65 and NEWS) did not have good discrimination, and that their novel tool (4C Mortality Score) outperformed other COVID-19-specific tools in addition to more established tools.[10] When using 7-day or in-hospital mortality, the AUC for 4C Mortality Score performed similarly in our sub-cohort for all SARS-CoV-2 infection (0.78 and 0.76, respectively) to that seen in the original validation cohort for inpatient mortality (0.77). However, our data suggest that the use of this score for prediction of ESAE is poor.

It is not clear why these severity scoring systems have performed poorly in this patient cohort. Many of the established severity scoring systems are likely ineffective as they were derived and validated in very different patient cohorts (ie deterioration or death in sepsis and/or bacterial pneumonia). In addition, most of the scores assessed were derived to assess mortality at different time points (CURB65 at 30-days, MuLBSTA at 90-days, and qSOFA and 4C Mortality Score in hospital), while we were investigating their utility in predicting early adverse outcomes (a composite outcome measure including death and admission to ICU within 7

days).[10,21,23,27] The subsequent ISARIC 4C Mortality Deterioration Score goes some way to address this gap and aid clinical decision making.[29]

Despite the formulation and validation of COVID-19-specific severity assessment tools, clinicians and organisations continue to use/recommend the use of pre-existing severity assessment tools. Our study highlights the risks of relying on such tools, but also suggests that even the COVID-19-specific tools may not perform particularly well. Overall, clinical judgment remains essential and the development of tools that include other factors (such as deprivation indices) may prove to be more useful.[30]

This study uses a widely representative sample of participants from across the West Midlands, UK, where COVID-19 had significant impact during spring 2020. This multi-centre approach included data from small district general hospitals through to a large tertiary unit with Europe's largest ICU. This is a diverse area of the UK and the inclusion of a significant Black, Asian and minority ethnic population (20%) is crucial to validating these scores in populations who are often not involved in research yet have disproportionately high mortality from COVID-19.

Across the recruiting units, the standard of care varied, the main variance was in the provision of CPAP at ward level. Most centres were actively recruiting to RECOVERY and REMAP-CAP, two of the major interventional platform studies open during the period of study. This variation in standard of care increases our confidence in the findings. In addition, mortality and rates of ICU admission are in keeping with the ISARIC observational study further supporting this being a representative sample of patients.[31]

Additionally, this study was delivered by a respiratory specialist trainee collaborative during a period of unprecedented demand on respiratory trainees, where many other research and training opportunities were put on hold. We believe that trainee collaboratives are a sustainable way to deliver patient-focused research that is relevant to our clinical practice and aims to meet the challenge of the Royal College of Physicians' 'Delivering research for all' project.

Each of the tools examined here is in use clinically, although some are more widespread than others. All are easily computed within 4 hours of admission to hospital and will provide a basis for clinicians to make treatment escalation plans at the point of admission.

Limitations

The limitations of this study are the retrospective nature of data collection and missing data. However, missing data was only a significant issue for MuLBSTA and Lac-CURB65. This limits the usage of these scoring systems in clinical practice and potentially demonstrating the benefit of using established scores (such as CURB65) where clinicians routinely record these data as part of their assessments. We used multiple imputation and found that this did not alter outcomes significantly. We also only have very short-term data at 7 days. It would be beneficial to have longer-term outcomes (such as 30-day mortality) when assessing scores designed for use of this timepoint. However, severity scores (such as CURB65 and 4C Mortality Score) demonstrated predictive value in mortality alone at 7-day and in-hospital mortality outcomes. Finally, our cohort was from the first wave of the pandemic and, therefore, we have not assessed the accuracy of these severity scoring systems in patients from the second or third waves in the UK.

In future, more nuanced scores that also predict risk of complications and longer-term outcomes would be beneficial to inpatient care, but also predicting who may benefit from intensive follow-up.

Conclusion

We have demonstrated that, in a large and diverse cohort of hospitalised adults with SARS-CoV-2 infection, established severity scores are able to predict increasing risk of adverse outcomes, but their overall accuracy is low. Clinicians should be wary of applying established tools to a new disease and should understand the limitations of each. Validation of severity assessment tools in COVID-19 is crucial to improve clinical care but will also support development of research, where accurate stratification of severity will influence study protocols and development of new treatments. ■

Summary

What is known?
COVID-19 often causes severe disease with high mortality, yet there are limited treatment options. SARS-CoV-2 infection can be heterogenous, the ability to stratify risk will inform patient care and clinical trial stratification.

What is the question?
Do commonly used risk stratification tools aid in predicting early adverse outcomes in SARS-CoV-2 infection?

What was found?
Established severity scores for pneumonia and sepsis do not accurately predict risk of early death or admission to the intensive care unit in SARS-CoV-2 infection.

What is the implication for practice?
Clinicians should be wary of using non-validated tools to inform clinical decisions in COVID-19. Further validation of COVID-19-specific risk stratification tools is urgently warranted.

Supplementary material

Additional supplementary material may be found in the online version of this article at www.rcpjournals.org/clinmedicine:
S1 – Fuller data tables and figures.

References

1 World Health Organization. *Coronavirus disease (COVID-19) pandemic*. WHO, 2020. www.who.int/emergencies/diseases/novel-coronavirus-2019 [Accessed 27 August 2020].

2 Wu C, Chen X, Cai Y *et al*. Risk Factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934–43.

3 Mahida RY, Chotalia M, Alderman J *et al*. Characterisation and outcomes of ARDS secondary to pneumonia in patients with and without SARS-CoV-2: a single-centre experience. *BMJ Open Respiratory Research* 2020;7:e000731.

4 RECOVERY Collaborative Group; Horby P, Lim WS *et al*. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N Engl J Med* 2021;384:693–704.

5 RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform study. *Lancet* 2021;397:1637–45.

6 Synairgen. Press release: Synairgen announces positive results from trial of SNG001 in hospitalised COVID-19 patients. *GlobalNewswire* 2020.

7 Galloway JB, Norton S, Barker RD *et al*. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study. *Journal of Infection* 2020;81:282–8.

8 Cambridge Clinical Trials Unit. *TACTIC-R: Training videos*. CCTU, 2020. https://cctu.org.uk/portfolio/COVID-19/TACTIC/TACTIC-R [Accessed 28 August 2020].

9 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:m1328.

10 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.

11 Dellinger RP. The Surviving Sepsis Campaign: Where have we been and where are we going? *Cleve Clin J Med* 2015;82:237–44.

12 Goulden R, Hoyle MC, Monis J *et al*. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg Med J* 2018;35:345–9.

13 Grudzinska FS, Aldridge K, Hughes S *et al*. Early identification of severe community-acquired pneumonia: a retrospective observational study. *BMJ Open Respir Res* 2019;6:e000438.

14 Lim WS, Baudouin SV, George RC *et al*. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009;64 Suppl 3:iii1–55.

15 NHS Health Research Authority. *Do I need NHS REC review?* HRA. www.hra-decisiontools.org.uk/ethics

16 Harris PA, Taylor R, Minor BL *et al*. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019;95:103208.

17 Harris PA, Taylor R, Thielke R *et al*. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009;42:377–81.

18 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;58:377–82.

19 Chen YX, Li CS. Lactate on emergency department arrival as a predictor of mortality and site-of-care in pneumonia patients: a cohort study. *Thorax* 2015;70:404–10.

20 Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. RCP, 2017.

21 Guo L, Wei D, Zhang X *et al*. Clinical features predicting mortality risk in patients with viral pneumonia: the MuLBSTA score. *Front Microbiol* 2019;10:2752.

22 Singer M, Deutschman CS, Seymour CW *et al*. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315:801–10.

23 Barlow G, Nathwani D, Davey P. The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia. *Thorax* 2007;62:253–9.

24 Grudzinska FS, Aldridge K, Hughes S *et al*. Early identification of severe community-acquired pneumonia: a retrospective observational study. *BMJ Open Respiratory Research* 2019;6:e000438.

25 Borregaard N, Herlin T. Energy metabolism of human neutrophils during phagocytosis. *J Clin Invest* 1982;70:550–7.

26 Frenzen FS, Kutschan U, Meiswinkel N *et al*. Admission lactate predicts poor prognosis independently of the CRB/CURB-65 scores in community-acquired pneumonia. *Clin Microbiol Infect* 2018;24:306.e1-306.e6.

27 Seymour CW, Liu VX, Iwashyna TJ *et al*. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315:762–74.

28 Ranzani OT, Prina E, Menendez R *et al*. New sepsis definition (Sepsis-3) and community-acquired pneumonia mortality. a validation and clinical decision-making study. *Am J Respir Crit Care Med* 2017;196:1287–97.

29 Gupta RK, Harrison EM, Ho A *et al*. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: as prospective cohort study. *Lancet Respir Med* 2021;9:349–59.

30 Soltan MA, Varney J, Sutton B *et al*. COVID-19 risk stratification tools should incorporate multi-ethnic age structures, multimorbidity and deprivation metrics for air pollution, household overcrowding, housing quality and adult skills. *BMJ Open Respir Res* 2021;8:e000951.

31 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.

Address for correspondence: Dr Davinder Dosanjh, Department of Respiratory Medicine, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Foundation Trust, Mindelsohn Way, Edgbaston, Birmingham B15 2GW, UK.
Email: davinder.dosanjh@uhb.nhs.uk
Twitter: @Dr_D_Dosanjh