OPEN ACCESS

# Prognostic accuracy of emergency department triage tools for adults with suspected COVID-19: the PRIEST observational cohort study

Ben Thomas ,[1] Steve Goodacre ,[1] Ellen Lee,[1] Laura Sutton,[1] Matthew Bursnall,[1] Amanda Loban,[1] Simon Waterhouse,[1] Richard Simmonds,[1] Katie Biggs,[1] Carl Marincowitz ,[1] José Schutter,[1] Sarah Connelly ,[1] Elena Sheldon,[1] Jamie Hall,[1] Emma Young,[1] Andrew Bentley,[2] Kirsty Challen,[3] Chris Fitzsimmons,[4,5] Tim Harris,[6] Fiona Lecky,[1] Andrew Lee,[1] Ian Maconochie,[7] Darren Walter[8]

**Handling editor** Richard Body

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2020-210783).

For numbered affiliations see end of article.

**Correspondence to**
Dr Ben Thomas, ScHARR, The University of Sheffield, Sheffield S1 4DP, UK;
b.d.thomas@sheffield.ac.uk

Received 8 October 2020
Accepted 14 May 2021
Published Online First
3 June 2021



**To cite:** Thomas B, Goodacre S, Lee E, *et al*. *Emerg Med J* 2021;**38**:587–593.

## ABSTRACT

**Background** The WHO and National Institute for Health and Care Excellence recommend various triage tools to assist decision-making for patients with suspected COVID-19. We aimed to compare the accuracy of triage tools for predicting severe illness in adults presenting to the ED with suspected COVID-19.

**Methods** We undertook a mixed prospective and retrospective observational cohort study in 70 EDs across the UK. We collected data from people attending with suspected COVID-19 and used presenting data to determine the results of assessment with the WHO algorithm, National Early Warning Score version 2 (NEWS2), CURB-65, CRB-65, Pandemic Modified Early Warning Score (PMEWS) and the swine flu adult hospital pathway (SFAHP). We used 30-day outcome data (death or receipt of respiratory, cardiovascular or renal support) to determine prognostic accuracy for adverse outcome.

**Results** We analysed data from 20 891 adults, of whom 4611 (22.1%) died or received organ support (primary outcome), with 2058 (9.9%) receiving organ support and 2553 (12.2%) dying without organ support (secondary outcomes). C-statistics for the primary outcome were: CURB-65 0.75; CRB-65 0.70; PMEWS 0.77; NEWS2 (score) 0.77; NEWS2 (rule) 0.69; SFAHP (6-point rule) 0.70; SFAHP (7-point rule) 0.68; WHO algorithm 0.61. All triage tools showed worse prediction for receipt of organ support and better prediction for death without organ support. At the recommended threshold, PMEWS and the WHO criteria showed good sensitivity (0.97 and 0.95, respectively) at the expense of specificity (0.30 and 0.27, respectively). The NEWS2 score showed similar sensitivity (0.96) and specificity (0.28) when a lower threshold than recommended was used.

**Conclusion** CURB-65, PMEWS and the NEWS2 score provide good but not excellent prediction for adverse outcome in suspected COVID-19, and predicted death without organ support better than receipt of organ support. PMEWS, the WHO criteria and NEWS2 (using a lower threshold than usually recommended) provide good sensitivity at the expense of specificity.

**Trial registration number** ISRCTN56149622.

## Key messages

### What is already known on this subject

- Emergency management of suspected COVID-19 involves predicting the risk of adverse outcome to determine the need for hospital admission.
- A number of triage tools have been recommended to support decision-making in suspected COVID-19, but accuracy for adverse outcome in suspected COVID-19 is not known.

### What this study adds

- CURB-65, Pandemic Modified Early Warning Score (PMEWS) and National Early Warning Score version 2 (NEWS2) provide good but not excellent prediction for adverse outcome in suspected COVID-19.
- Triage tools predict death without organ support better than need for organ support.
- PMEWS, the WHO criteria and NEWS2 (using a lower threshold than usually recommended) can provide good sensitivity at the expense of specificity.

## INTRODUCTION

The ED has a crucial role in the management of patients with suspected COVID-19. ED management involves assessing the risk of adverse outcome and the need for life-saving intervention, and then using this to determine decisions around admission to hospital and inpatient referral.[1 2] Triage tools can assist decision-making by combining information from clinical assessment in a structured manner to predict the risk of adverse outcome. Triage tools may take the form of a score, which allocates points to risk predictors to indicate an increasing risk of adverse outcome, or a rule, which uses risk predictors to determine a clinical decision, such as hospital admission or discharge. Adults and children presenting to the ED with suspected COVID-19 differ markedly in their need for hospital admission and risk of adverse outcome,[3] so they require different triage tools. We focus on adults in this study.

Guidelines have recommended a number of triage tools for adults with suspected COVID-19. The WHO decision-making algorithm for acute respiratory infection[4] recommends hospital admission for

**Table 1** Characteristics of the study population

| Characteristic | Statistic/level | |
|---|---|---|
| Age (years) | n | 20891 |
| | Mean (SD) | 62.4 (19.7) |
| | Median (IQR) | 64 (48, 79) |
| Sex | Missing | 193 |
| | Male | 10201 (49.3%) |
| | Female | 10497 (50.7%) |
| Ethnicity | Missing/prefer not to say | 4198 |
| | UK/Irish/other white | 14243 (85.3%) |
| | Asian | 1044 (6.3%) |
| | Black/African/Caribbean | 640 (3.8%) |
| | Mixed/multiple ethnic groups | 247 (1.5%) |
| | Other | 519 (3.1%) |
| Presenting features | Cough | 12985 (62.2%) |
| | Shortness of breath | 15570 (74.5%) |
| | Fever | 10276 (49.2%) |
| Symptom duration (days) | n | 18877 |
| | Mean (SD) | 7.9 (8.9) |
| | Median (IQR) | 5 (2, 10) |
| HR (beats/min) | n | 20460 |
| | Mean (SD) | 94.9 (21.6) |
| | Median (IQR) | 93 (80, 108) |
| RR (breaths/min) | n | 20346 |
| | Mean (SD) | 23.3 (7) |
| | Median (IQR) | 22 (18, 26) |
| Systolic BP (mm Hg) | n | 20298 |
| | Mean (SD) | 134.6 (24.9) |
| | Median (IQR) | 133 (118, 149) |
| Diastolic BP (mm Hg) | n | 20212 |
| | Mean (SD) | 78.2 (16.1) |
| | Median (IQR) | 78 (68, 88) |
| Temperature (°C) | n | 20231 |
| | Mean (SD) | 37.1 (1.1) |
| | Median (IQR) | 37 (36.4, 37.8) |
| Oxygen saturation (%) | n | 20632 |
| | Mean (SD) | 94.7 (6.8) |
| | Median (IQR) | 96 (94, 98) |
| GCS | n | 15428 |
| | Mean (SD) | 14.6 (1.4) |
| | Median (IQR) | 15 (15, 15) |
| AVPU | Missing | 2387 |
| | Alert | 17568 (94.9%) |
| | Verbal | 640 (3.5%) |
| | Pain | 183 (1%) |
| | Unresponsive | 113 (0.6%) |
| Comorbidities | Hypertension | 6434 (30.8%) |
| | Heart disease | 4700 (22.5%) |
| | Diabetes | 4129 (19.8%) |
| | Other chronic lung disease | 3764 (18%) |
| | Asthma | 3408 (16.3%) |
| | Renal impairment | 1930 (9.2%) |
| | Active malignancy | 1120 (5.4%) |
| | Steroid therapy | 557 (2.7%) |
| | No chronic disease | 5791 (27.7%) |
| Performance status | Missing | 1078 |
| | Unrestricted normal activity | 10536 (53.2%) |
| | Limited strenuous activity, can do light activity | 2371 (12%) |
| | Limited activity, can self-care | 2776 (14%) |
| | Limited self-care | 2649 (13.4%) |
| | Bed/chair bound, no self-care | 1481 (7.5%) |
| Other clinical concern | Severe respiratory distress | 587 (2.8%) |
| | Respiratory exhaustion | 292 (1.4%) |
| | Severe dehydration | 261 (1.2%) |

AVPU, Alert Verbal Pain Unresponsive.

severe pneumonia (RR >30/min, oxygen saturation <90% or signs of respiratory distress) or respiratory infection associated with comorbidities (age >60, hypertension, diabetes, cardiovascular disease, chronic respiratory disease, chronic renal disease or immunocompromising conditions). The UK National Institute for Health and Care Excellence COVID-19 rapid guideline[5] suggests that the National Early Warning Score version 2 (NEWS2) score[6] can be useful for predicting the risk of deterioration. NEWS2 uses HR, RR, systolic BP, oxygen saturation, temperature and conscious level to allocate a score between 0 and 20. The guideline also notes that the CRB-65 tool can determine the need for hospital admission in adults with pneumonia but has not been validated in people with COVID-19. The CURB-65 pneumonia score[7] uses five variables (confusion, urea level, RR, BP and age) to generate a score between 0 and 5. The CRB-65 score allows use without blood testing by dropping urea measurement from the score.

Triage tools developed or recommended for an influenza pandemic could be used for suspected COVID-19. Guidance during the 2009 H1N1 pandemic included a swine flu adult hospital pathway for ED management with seven criteria, any one of which predicts increased risk and the need for hospital assessment.[8] The Pandemic Modified Early Warning Score (PMEWS)[9] uses physiological variables, age, social factors, chronic disease and performance status to generate a score between 0 and 19.

### Aims and objectives

We aimed to compare the accuracy of triage tools recommended for predicting severe illness in adults presenting to the ED with suspected COVID-19 infection.

## METHODS

We developed the Pandemic Influenza Triage in the Emergency Department (PAINTED) study following the 2009 H1N1 pandemic to evaluate triage tools for suspected pandemic influenza. We modified the PAINTED protocol to become the Pandemic Respiratory Infection Emergency System Triage (PRIEST) study in January 2020 to address any pandemic respiratory infection and include triage tools recommended for COVID-19.

We undertook an observational study to collect standardised predictor variables recorded in the ED, which we then used to evaluate triage tools for predicting adverse outcome up to 30 days after initial hospital presentation. The study did not involve any change to patient care. Hospital admission and discharge decisions were made according to usual practice, informed by local and national guidance.

We identified consecutive patients presenting to the ED of participating hospitals with suspected COVID-19 infection. Patients were eligible if they met the clinical diagnostic criteria[10] of fever (≥37.8°C) and acute onset of persistent cough (with

**Table 2** Summary of ROC analysis for existing triage tools predicting adverse outcomes

| Triage tool | n | C-statistic | Cut point | Sensitivity (95% CI) | Specificity (95% CI) | Positive predictive value (95% CI) | Negative predictive value (95% CI) | Proportion test positive |
|---|---|---|---|---|---|---|---|---|
| CURB-65 | 20 716 | 0.75 (0.74, 0.76) | (>1) | 0.71 (0.70 to 0.72) | 0.69 (0.69 to 0.70) | 0.40 (0.39 to 0.41) | 0.89 (0.89 to 0.90) | 0.39 |
| CRB-65 | 20 716 | 0.70 (0.69, 0.70) | (>0) | 0.86 (0.85 to 0.87) | 0.48 (0.47 to 0.48) | 0.32 (0.31 to 0.33) | 0.92 (0.92 to 0.93) | 0.60 |
| PMEWS | 20 492 | 0.77 (0.76, 0.77) | (>2) | 0.97 (0.96 to 0.97) | 0.30 (0.30 to 0.31) | 0.28 (0.27 to 0.29) | 0.97 (0.96 to 0.97) | 0.76 |
| NEWS2 (score) | 20 594 | 0.77 (0.76, 0.78) | (>4) | 0.77 (0.76 to 0.78) | 0.64 (0.63 to 0.65) | 0.38 (0.37 to 0.39) | 0.90 (0.90 to 0.91) | 0.45 |
| NEWS2 (rule) | 20 594 | 0.69 (0.68, 0.69) | (>0) | 0.83 (0.82 to 0.84) | 0.55 (0.54 to 0.55) | 0.34 (0.33 to 0.35) | 0.92 (0.91 to 0.92) | 0.54 |
| SFAHP (6-point rule) | 19 858 | 0.70 (0.69, 0.71) | (>0) | 0.74 (0.73 to 0.75) | 0.66 (0.65 to 0.67) | 0.38 (0.37 to 0.39) | 0.90 (0.89 to 0.90) | 0.43 |
| SFAHP (7-point rule) | 20 682 | 0.68 (0.68, 0.69) | (>0) | 0.88 (0.87 to 0.89) | 0.48 (0.48 to 0.49) | 0.33 (0.32 to 0.33) | 0.94 (0.93 to 0.94) | 0.60 |
| WHO algorithm | 20 891 | 0.61 (0.61, 0.62) | (>0) | 0.95 (0.95 to 0.96) | 0.27 (0.26 to 0.28) | 0.27 (0.26 to 0.28) | 0.95 (0.95 to 0.96) | 0.78 |

NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; ROC, receiver operating characteristic; SFAHP, swine flu adult hospital pathway.

or without sputum), hoarseness, nasal discharge or congestion, shortness of breath, sore throat, wheezing and sneezing. This was determined on the basis of the assessing clinician recording that the patient had suspected COVID-19 or completing a standardised assessment form designed for suspected pandemic respiratory infection.[11]

We planned to evaluate triage tools recommended for use in the COVID-19 pandemic or the 2009 H1N1 influenza pandemic, as outlined in the Introduction section: the WHO algorithm, NEWS2, CURB-65, CRB-65, PMEWS and the swine flu adult hospital pathway (SFAHP). The triage tools are described in online supplemental appendix 1. NEWS2 can be used as a score, with thresholds between 0 and 20 on the total score, or a rule, with a single threshold of a total score greater than 4 or a score of 3 on any parameter. We therefore evaluated the performance of NEWS2 as both a score and a rule. The SFAHP has a criterion (G) that is positive if there is any clinical concern. This is difficult to judge objectively or identify from clinical records, so we evaluated the pathway in two ways: (1) a 6-point rule that did not include parameter G; (2) a 7-point rule in which parameter G was positive if the NEWS2 rule was positive. NEWS2 is widely used in the UK health service to identify clinical concern.

Data collection was both prospective and retrospective. We provided participating EDs with a standardised data collection form that included the predictor variables used in the triage tools.[11] Participating sites could adapt the form to their local circumstances, including integrating it into electronic or paper clinical records to facilitate prospective data collection, or using it as a template for research staff to retrospectively extract data from clinical records. We did not seek consent to collect data but information about the study was provided in the ED and patients could withdraw their data at their request. Patients with multiple presentations to hospital were only included once, using data from the first presentation identified by research staff.

Research staff at participating hospitals reviewed patient records at 30 days after initial attendance to identify any adverse outcomes. Patients who died or required respiratory, cardiovascular or renal support were classified as having an adverse outcome. Patients who survived to 30 days without requiring respiratory, cardiovascular or renal support were classified as having no adverse outcome. Respiratory support was defined as any intervention to protect the patient's airway or assist their ventilation, including non-invasive ventilation or acute administration of continuous positive airway pressure. It did not include supplemental oxygen alone or nebulised bronchodilators. Cardiovascular support was defined as any intervention to maintain organ perfusion, such as inotropic drugs, or invasively monitor cardiovascular status, such as central venous pressure or pulmonary artery pressure monitoring, or arterial BP monitoring. It did not include peripheral intravenous cannulation or fluid administration. Renal support was defined as any intervention to assist renal function, such as haemofiltration, haemodialysis or peritoneal dialysis. It did not include intravenous fluid administration.

The primary outcome was death, or respiratory, cardiovascular or renal support, as defined above. We also planned secondary analyses using the following outcomes: (1) respiratory, cardiovascular or renal support to predict need for life-saving treatment; (2) death without respiratory, cardiovascular or renal support to predict poor prognosis. If triage tools are used to determine treatment decisions, such as referral to critical care, then it is helpful to know how well they predict need for treatment rather than a potentially irremediable poor prognosis.

We retrospectively applied each triage tool to the data, excluding pregnant women from analysis of NEWS2. Online supplemental appendix 1 provides details of scoring and handling missing data for the triage tools. For each tool we plotted the receiver operating characteristic (ROC) curve and calculated the area under the ROC curve (c-statistic) for discriminating between cases with and without adverse outcome. We calculated sensitivity, specificity, positive predictive value and negative predictive value at the following prespecified decision-making thresholds based on recommended or usual use: 0–1 vs 2–5 for CURB-65; 0–2 vs 3+ for PMEWS; 0–4 vs 5–20 for the NEWS2 score. The WHO algorithm and swine flu adult hospital pathway are positive if any criterion is positive. We used STATA (V.16) for analyses.[12]

The sample size was dependent on the size and severity of the pandemic, but based on a previous study in the 2009 H1N1 influenza pandemic we estimated we would need to collect data from 20 000 patients across 40–50 hospitals to identify 200 with an adverse outcome. In the event, the adverse outcome rate in adults was much higher in the COVID-19 pandemic, giving us adequate power to undertake primary and secondary analyses.

## Patient and public involvement

The Sheffield Emergency Care Forum (SECF) is a public representative group interested in emergency care research.[13] Members of SECF advised on the development of the PRIEST study and two members joined the Study Steering Committee. Patients were not involved in the recruitment to and conduct of the study. We are unable to disseminate the findings to study participants directly.

## RESULTS

The PRIEST study recruited 22 484 patients from 70 EDs across 53 sites between 26 March and 28 May 2020. We included

**Table 3** Summary of ROC analysis for existing triage tools predicting organ support

| Triage tool | n | C-statistic | Cut point | Sensitivity (95% CI) | Specificity (95% CI) | Positive predictive value (95% CI) | Negative predictive value (95% CI) | Proportion test positive |
|---|---|---|---|---|---|---|---|---|
| CURB-65 | 20716 | 0.60 (0.59, 0.61) | (>1) | 0.52 (0.50 to 0.54) | 0.62 (0.61 to 0.63) | 0.13 (0.12 to 0.14) | 0.92 (0.92 to 0.93) | 0.39 |
| CRB-65 | 20716 | 0.58 (0.56, 0.59) | (>0) | 0.74 (0.72 to 0.76) | 0.42 (0.41 to 0.42) | 0.12 (0.12 to 0.13) | 0.94 (0.93 to 0.94) | 0.60 |
| PMEWS | 20492 | 0.68 (0.67, 0.69) | (>2) | 0.94 (0.93 to 0.95) | 0.27 (0.26 to 0.27) | 0.12 (0.12 to 0.13) | 0.98 (0.97 to 0.98) | 0.76 |
| NEWS2 (score) | 20594 | 0.72 (0.71, 0.73) | (>4) | 0.76 (0.74 to 0.78) | 0.58 (0.58 to 0.59) | 0.17 (0.16 to 0.17) | 0.96 (0.95 to 0.96) | 0.45 |
| NEWS2 (rule) | 20594 | 0.65 (0.64, 0.66) | (>0) | 0.81 (0.79 to 0.83) | 0.49 (0.49 to 0.50) | 0.15 (0.14 to 0.16) | 0.96 (0.96 to 0.96) | 0.54 |
| SFAHP (6-point rule) | 19858 | 0.64 (0.63, 0.65) | (>0) | 0.68 (0.66 to 0.71) | 0.60 (0.59 to 0.61) | 0.16 (0.15 to 0.17) | 0.95 (0.94 to 0.95) | 0.43 |
| SFAHP (7-point rule) | 20682 | 0.65 (0.64, 0.65) | (>0) | 0.86 (0.84 to 0.87) | 0.43 (0.42 to 0.44) | 0.14 (0.14 to 0.15) | 0.97 (0.96 to 0.97) | 0.60 |
| WHO algorithm | 20891 | 0.57 (0.57, 0.58) | (>0) | 0.91 (0.90 to 0.92) | 0.24 (0.23 to 0.24) | 0.12 (0.11 to 0.12) | 0.96 (0.96 to 0.97) | 0.78 |

NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; ROC, receiver operating characteristic; SFAHP, swine flu adult hospital pathway.

20891 in the analysis after excluding 39 who requested withdrawal of their data, 1530 children, 7 with missing age and 17 with missing outcome data.

Table 1 shows the characteristics of adults in the cohort. Some 13997 (67.0%) were admitted after ED assessment and 6521 (31.2%) ultimately tested positive for COVID-19. Overall, 4611 (22.1%) died or received organ support (primary outcome), with 2058 (9.9%) receiving organ support and 2553 (12.2%) dying without organ support (secondary outcomes). Organ support involved respiratory support for 1944 (9.3%), cardiovascular for 517 (2.5%) and renal support for 218 (1%).

Table 2 shows the results for the primary analysis, table 3 the results for secondary analysis predicting receipt of organ support and table 4 the results for secondary analysis predicting death without organ support. The ROC curves for these analyses are shown in figures 1–3.

In the primary analysis presented in table 2, none of the triage tools showed excellent discrimination (c-statistic >0.8) but CURB-65, PMEWS and the NEWS2 score showed good discrimination (>0.7). This may reflect the use of multiple points across these tools, as opposed to a single decision-making threshold for other tools. At the prespecified threshold, PMEWS and the WHO criteria showed good sensitivity (0.97 and 0.95, respectively) at the expense of specificity (0.30 and 0.27, respectively). The sensitivities of other triage tools at the prespecified threshold were below 0.9, although with higher specificities. A sensitivity analysis of the NEWS2 score including 85 pregnant women who were excluded from the primary analysis produced no change in the c-statistic (and CI).

The triage tools generally showed worse prediction for receipt of organ support and better prediction for death without organ support. This was most marked for CURB-65 and CRB-65, and least marked for the NEWS2 score. Only the NEWS2 score showed good prediction for organ support (c-statistic >0.7).

Online supplemental table S1 shows the sensitivity and specificity at each threshold for the triage tools with multiple potential thresholds for decision-making (CURB-65, CRB-65, PMEWS and NEWS2). These results suggest that NEWS2 score could offer good sensitivity (0.96) at the expense of specificity (0.28), if we use a score greater than 1 to predict adverse outcome. The sensitivity of CURB-65 is 0.90 and CRB-65 is 0.86 at the lowest threshold (any score above 0 predicts adverse outcome).

Online supplemental table S2 shows the proportion with an adverse outcome at each level of each score. This analysis shows that patients with a risk of adverse outcome of 5% or less could be identified using the WHO algorithm, a NEWS2 score of 0–1 or a PMEWS score of 0–2.

## DISCUSSION

ED clinicians usually use triage tools to support decisions, such as admission to hospital, where sensitivity needs to be optimised at the expense of specificity to avoid missed opportunities to predict and prevent adverse outcome. Our analysis suggests that the WHO algorithm or PMEWS greater than 2 provide good sensitivity at the expense of specificity, and could be used to support decision-making where sensitivity needs to be optimised. The NEWS2 score needs to use a lower threshold (any score above 1) than currently recommended to achieve a comparable balance of sensitivity and specificity.

The triage tools predicted death without organ support better than they predicted receipt of organ support. Only the NEWS2 score predicted receipt of organ support with good accuracy. This reflects NEWS2 using only physiological

**Table 4** Summary of ROC analysis for existing triage tools predicting death without organ support

| Triage tool | n | C-statistic | Cut point | Sensitivity (95% CI) | Specificity (95% CI) | Positive predictive value (95% CI) | Negative predictive value (95% CI) | Proportion test positive |
|---|---|---|---|---|---|---|---|---|
| CURB-65 | 20716 | 0.82 (0.82, 0.83) | (>1) | 0.86 (0.85 to 0.87) | 0.67 (0.66 to 0.68) | 0.27 (0.26 to 0.28) | 0.97 (0.97 to 0.97) | 0.39 |
| CRB-65 | 20716 | 0.75 (0.75, 0.76) | (>0) | 0.96 (0.95 to 0.97) | 0.45 (0.44 to 0.46) | 0.20 (0.19 to 0.20) | 0.99 (0.98 to 0.99) | 0.60 |
| PMEWS | 20492 | 0.78 (0.77, 0.78) | (>2) | 0.99 (0.98 to 0.99) | 0.28 (0.27 to 0.28) | 0.16 (0.15 to 0.17) | 0.99 (0.99 to 0.99) | 0.76 |
| NEWS2 (score) | 20594 | 0.75 (0.74, 0.76) | (>4) | 0.78 (0.76 to 0.79) | 0.60 (0.59 to 0.60) | 0.21 (0.20 to 0.22) | 0.95 (0.95 to 0.95) | 0.45 |
| NEWS2 (rule) | 20594 | 0.67 (0.67, 0.68) | (>0) | 0.84 (0.83 to 0.86) | 0.51 (0.50 to 0.51) | 0.19 (0.18 to 0.20) | 0.96 (0.95 to 0.96) | 0.54 |
| SFAHP (6-point rule) | 19858 | 0.70 (0.69, 0.71) | (>0) | 0.78 (0.77 to 0.80) | 0.62 (0.61 to 0.63) | 0.22 (0.21 to 0.23) | 0.95 (0.95 to 0.96) | 0.43 |
| SFAHP (7-point rule) | 20682 | 0.67 (0.67, 0.68) | (>0) | 0.90 (0.89 to 0.91) | 0.45 (0.44 to 0.45) | 0.18 (0.18 to 0.19) | 0.97 (0.97 to 0.97) | 0.60 |
| WHO algorithm | 20891 | 0.62 (0.61, 0.62) | (>0) | 0.99 (0.98 to 0.99) | 0.25 (0.24 to 0.26) | 0.15 (0.15 to 0.16) | 0.99 (0.99 to 0.99) | 0.78 |

NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; ROC, receiver operating characteristic; SFAHP, swine flu adult hospital pathway.

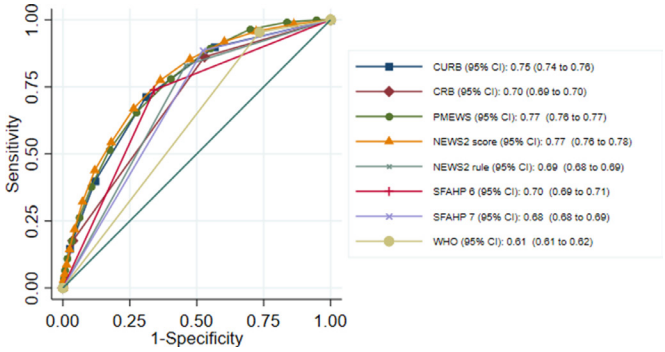


**Figure 1** Overlaid receiver operating characteristic (ROC) curves for triage tools predicting adverse outcome in adults. NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; SFAHP, swine flu adult hospital pathway.

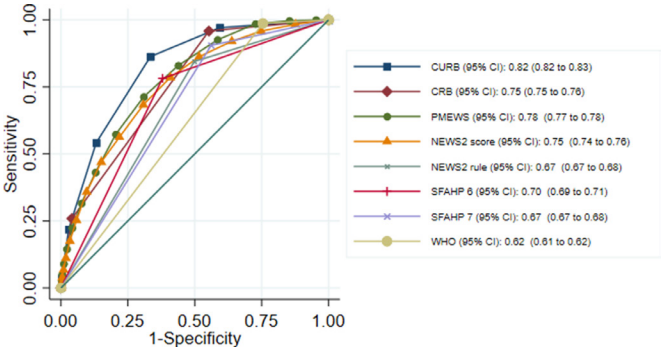


**Figure 3** Overlaid receiver operating characteristic (ROC) curves for triage tools predicting death without support in adults. NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; SFAHP, swine flu adult hospital pathway.

measures, while other triage tools include age, performance status or comorbidities that are more likely to predict death without organ support.

Studies undertaken during the 2009 H1N1 influenza pandemic suggested that existing triage tools have suboptimal accuracy for predicting adverse outcome in acute respiratory infections, with c-statistics below 0.8.[14–16] Recent studies have evaluated NEWS2, CURB-65 and CRB-65 in adult inpatients with confirmed COVID-19. Fan *et al* (n=654)[17] reported c-statistics of 0.81, 0.85 and 0.80, respectively, for NEWS2, CURB-65 and CRB-65 as predictors of in-hospital death. The conventional thresholds for positivity of scores above 4, 1 and 0 offered suboptimal sensitivity (0.79, 0.63 and 0.83), with corresponding specificities of 0.69, 0.91 and 0.69. Bradley *et al* (n=830)[18] reported c-statistics of 0.67 for NEWS2 and 0.74 for CURB-65 as predictors of 30-day mortality, with sensitivities and specificities at conventional thresholds of 0.83 and 0.37 for NEWS2, and 0.80 and 0.59 for CURB-65. Ma *et al* (n=305)[19] reported c-statistics of 0.79 for NEWS2 and 0.85 for CURB-65 for predicting death. Satici *et al* (n=681)[20] reported a c-statistic of 0.79 for predicting 30-day mortality with CURB-65, with sensitivity of 0.73 and specificity of 0.85 at the conventional threshold. Nguyen *et al*[21] reported that 36/171 (21%) patients with CURB-65 scores of 0 or 1 died or received intensive care admission. Gidari *et al* (n=68)[22] evaluated NEWS2 as a predictor of intensive care admission and Myrstad *et al* (n=66)[23] evaluated NEWS2 and CRB-65 as predictors

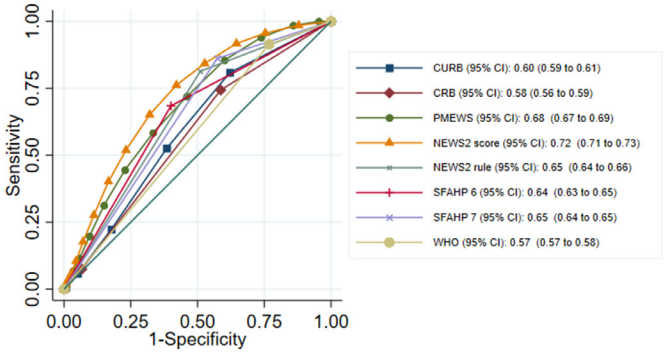


**Figure 2** Overlaid receiver operating characteristic (ROC) curves for triage tools predicting any support in adults. NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; SFAHP, swine flu adult hospital pathway.

of death or intensive care admission, but the small sizes produced imprecise estimates of prognostic parameters.

These studies concur with our findings that the conventional thresholds for NEWS2 and CURB-65 offer inadequate sensitivity to support discharge decisions after ED assessment. The larger studies used 30-day or in-hospital mortality as their outcome. Our analysis suggests that this may overestimate prognostic accuracy if the tools are used to predict need for life-saving treatment rather than simply predicting mortality.

We collected data from a clinically relevant population of patients presenting with suspected COVID-19 across a large and varied range of EDs. The large sample size and high rate of adverse outcome allowed us to estimate parameters with a high degree of precision in primary and secondary analyses. The main limitation is that the triage tools applied to data collected from clinical record review or a standardised data collection form, rather than being applied directly to the patient by the assessing clinician. This may have led to underestimation of the performance of the triage tool, especially when relevant data were missing. Table 1 shows that data were relatively complete for age, physiological variables and performance status, but the recording of other parameters (respiratory distress, respiratory exhaustion, dehydration) was limited by inability to determine whether the feature was not present or not recorded. This is most salient for the swine flu adult hospital pathway and may have led to underestimation of the sensitivity of this triage tool. Another potential limitation is that we may have missed adverse outcomes if patients attended a different hospital after initial hospital discharge. This is arguably less likely in the context of a pandemic, in which movements between regions were curtailed, but cannot be discounted. Finally, although some triage tools can be used in the prehospital or community setting, we recommend caution in extrapolating our findings to other settings, where there may be a lower prevalence of adverse outcome.

The clinical utility of our findings needs careful interpretation. Triage tools should not be used as the sole (or even principal) criteria for decision-making but should be used alongside clinical judgement. Our analysis did not evaluate how triage tools perform alongside or in comparison to clinical judgement. Further research would be helpful to explore this issue and determine how triage tools are best used in practice. Furthermore, although predicting death and need for organ support is clearly important to decision-making,

there are other factors that may determine hospital admission decisions. For example, it would be helpful to predict the need for supplemental oxygen. We excluded this from our outcome definition because use of supplemental oxygen may be poorly recorded and as a simple intervention it may be used when not clearly indicated. However, there is no doubt that some patients in our cohort will have required supplemental oxygen and will not have met our definition of an adverse outcome.

Our findings suggest that the WHO algorithm or PMEWS greater than 2 could be used to support hospital admission decisions, providing good sensitivity at the expense of specificity. The NEWS2 score would need to use a threshold greater than 1 to achieve a similar balance of sensitivity and specificity. If a triage tool is used to select patients for higher levels of treatment, rather than simply predict risk of adverse outcome, then NEWS2 offers better discrimination than other triage tools. Use of triage tools for this purpose may also require a different balance of sensitivity and specificity, with a higher threshold being used to ensure higher levels of care are reserved for those most likely to benefit.

In general, however, the accuracy of the triage tools evaluated was far from optimal, especially for predicting receipt of organ support. This is arguably unsurprising since they were developed for a variety of purposes and none were derived using data from patients presenting to the ED with suspected COVID-19. Research to derive and validate triage tools specific for COVID-19 is therefore an urgent priority.

**Author affiliations**
[1]ScHARR, The University of Sheffield, Sheffield, UK
[2]Acute intensive Care Unit, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK
[3]Emergency Department, Lancashire Teaching Hospitals NHS Foundation Trust, Chorley, UK
[4]Sheffield Children's Hospital NHS Foundation Trust, Sheffield, UK
[5]Emergency Department, Sheffield Children's Hospital NHS Foundation Trust, Sheffield, UK
[6]Department of Emergency Medicine, Royal London Hospital, London, UK
[7]Paediatric ED, Imperial College Healthcare NHS Trust, London, UK
[8]Emergency Department, University Hospital of South Manchester NHS Foundation Trust, Manchester, UK

**Twitter** Tim Harris @tim harris@resusdocs and Darren Walter @emsdocuk

**Acknowledgements** We thank Katie Ridsdale for clerical assistance with the study, Erica Wallis (sponsor representative), all members of the Study Steering Committee (online supplemental appendix 2) and the site research teams who delivered the data for the study (online supplemental appendix 3), and the research team at The University of Sheffield past and present (online supplemental appendix 4).

**Contributors** SG, AB, KC, CF, TH, FL, ALe, IM and DW conceived and designed the study. BT, KB, ALo, SW, RS, JS, SC, ES, JH and EY acquired the data. EL, LS, MB, SG, BT, KB and CM analysed the data. SG, AB, KC, CF, TH, FL, ALe, IM, DW, EL, LS, SG, BT, KB and CM interpreted the data. All authors contributed to drafting the manuscript. BT is the guarantor of the paper.

**Funding** The PRIEST study was funded by the UK National Institute for Health Research Health Technology Assessment (HTA) programme (project reference 11/46/07).

**Disclaimer** The funder played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The North West-Haydock Research Ethics Committee gave a favourable opinion on the PAINTED study on 25 June 2012 (reference 12/NW/0303) and on the updated PRIEST study on 23 March 2020. The Confidentiality Advisory Group of the Health Research Authority granted approval to collect data without patient consent in line with Section 251 of the NHS Act 2006.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open-access repository. Data are available upon reasonable request. Anonymised data are available from the corresponding author upon reasonable request (contact details on first page).



**ORCID iDs**
Ben Thomas http://orcid.org/0000-0002-6659-6930
Steve Goodacre http://orcid.org/0000-0003-0803-8444
Carl Marincowitz http://orcid.org/0000-0003-3043-7564
Sarah Connelly http://orcid.org/0000-0001-9963-5628

## REFERENCES

1 NHS. Clinical guide for the management of emergency department patients during the coronavirus pandemic. Available: https://www.england.nhs.uk/coronavirus/secondary-care/other-resources/specialty-guides/#ae [Accessed 27 Aug 2020].
2 American College of Emergency Physicians. Guide to coronavirus disease (COVID-19). Available: https://www.acep.org/corona/covid-19-field-guide/cover-page/ [Accessed 27 Aug 2020].
3 The PRIEST Research Group. Characterisation of 22446 patients attending UK emergency departments with suspected COVID-19 infection: Observational cohort study [Preprint]. *medRxiv*:2020.08.10.20171496.
4 World Health Organisation. Clinical care of severe acute respiratory infections – tool kit. Available: https://www.who.int/publications-detail/clinical-care-of-severe-acute-respiratory-infections-tool-kit [Accessed 27 Aug 2020].
5 National Institute for Health and Care Excellence. COVID-19 rapid guideline: managing suspected or confirmed pneumonia in adults in the community, 2020. Available: www.nice.org.uk/guidance/ng165 [Accessed 27 Aug 2020].
6 Royal College of Physicians. *National early warning score (news) 2: standardising the assessment of acute-illness severity in the NHS. updated report of a working Party*. London: RCP, 2017.
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;58:377–82.
8 Challen K, Bright J, Bentley A, *et al*. Physiological-social score (PMEWS) vs. CURB-65 to triage pandemic influenza: a comparative validation study using community-acquired pneumonia as a proxy. *BMC Health Serv Res* 2007;7:33.
9 Department of Health. *Swine flu clinical package*. London: Department of Health, 2009.
10 Public Health England. COVID-19: investigation and initial clinical management of possible cases. Available: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-initial-investigation-of-possible-cases/investigation-and-initial-clinical-management-of-possible-cases-of-wuhan-novel-coronavirus-wn-cov-infection#criteria [Accessed 27 Aug 2020].
11 Goodacre S, Irving A, Wilson R, *et al*. The pandemic influenza triage in the emergency department (painted) pilot cohort study. *Health Technol Assess* 2015;19:1–70.
12 StataCorp. *Stata statistical software: release 16*. College Station, TX: StataCorp LLC, 2019.
13 Hirst E, Irving A, Goodacre S. Patient and public involvement in emergency care research. *Emerg Med J* 2016;33:665–70.
14 Goodacre S, Challen K, Wilson R, *et al*. Evaluation of triage methods used to select patients with suspected pandemic influenza for hospital admission: cohort study. *Health Technol Assess* 2010;14:173–263.
15 Rowan KM, Harrison DA, Walsh TS, *et al*. The swine flu triage (swift) study: development and ongoing refinement of a triage tool to provide regular information to guide immediate policy and practice for the use of critical care services during the H1N1 swine influenza pandemic. *Health Technol Assess* 2010;14:337–496.

16 Muller MP, McGeer AJ, Hassan K, *et al*. Evaluation of pneumonia severity and acute physiology scores to predict ICU admission and mortality in patients hospitalized for influenza. *PLoS One* 2010;5:e9563.
17 Fan G, Tu C, Zhou F. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *Eur Respir J*2020.
18 Bradley P, Frost F, Tharmaratnam K, *et al*. The utility of established prognostic scores in COVID-19 hospital admissions: a multicenter prospective evaluation of CURB-65, NEWS2, and qSOFA [Preprint]. *medRxiv*:2020.07.15.20154815.
19 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.
20 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;98:84–9.
21 Nguyen Y, Corre F, Honsel V, *et al*. Applicability of the CURB-65 pneumonia severity score for outpatient treatment of COVID-19. *J Infect* 2020;81:e96–8.
22 Gidari A, De Socio GV, Sabbatini S, *et al*. Predictive value of national early warning score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. *Infect Dis* 2020;52:698–704.
23 Myrstad M, Ihle-Hansen H, Tveita AA, *et al*. 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.

## Appendix 1: Triage tool scoring details

**CURB-65 and CRB-65**

The CURB-65 score uses five parameters, each scoring one point when positive and zero if negative, to give a total score between zero and five. The CRB-65 score does not include urea, so the total score is between zero and four.

Five parameters:

1. Confusion: Glasgow Coma Score (GCS) is less than 15 or AVPU is recorded as anything less than alert (A)
2. Urea: Raised blood urea over 7mmol/litre
3. Respiratory: rate of 30 breaths per minute or more
4. Blood pressure: diastolic BP is 60mmHg or less or systolic BP is 90 mmHg or less
5. Age: 65 years or more

If data for a parameter was missing, we scored the parameter as zero.

**PMEWS**

PMEWS uses six physiological and patient parameters to calculate a score from zero to 19. The score is calculated by taking the score in the table below dependent on each of the six physiological parameters then adding points for two patient parameters after if they are positive.

Physiological:

| Score | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Respiratory Rate | ≤8 | | | 9-18 | 19-25 | 26-29 | ≥30 |
| Oxygen saturation | <89 | 90-93 | 94-96 | >96 | | | |
| Heart Rate | ≤40 | 41-50 | | 51-100 | 101-110 | 111-129 | ≥130 |
| Systolic BP | ≤70 | 71-90 | 90-100 | >100 | | | |
| Temperature | | ≤35.0 | 35.1-36.0 | 36.1-37.9 | 38-38.9 | ≥39 | |
| Neurology | | | | Alert | Confused Agitated* | Voice | Pain Uncon |

* confused/agitated is based on GCS<15, or if confusion is ticked as a presenting feature

Patient:

1. Add 1 point if age>65
2. Add 1 point if either:
    a. Patient lives alone / no fixed abode or
    b. has a co-morbidity (respiratory, cardiac, renal, immunosuppressed, diabetes)
    c. performance status is more than two suggesting limited activity can self-care, limited activity limited self-care, or bed/chair bound no self-care.

If data for a parameter was missing, we scored the parameter as zero. If more than three parameters were missing, we did not calculate a score for the patient. If AVPU was missing and GCS was recorded, we imputed the following AVPU scores using GCS: Alert if GCS=15, Verbal response if GCS=12-14; Pain response if GCS=9-11 and Unconscious if GCS<9.

**Swine Flu Adult Hospital Pathway**

The Swine Flu Adult Hospital Pathway consists of seven criteria operating as a rule, with the rule being positive if any criteria reaches its threshold.

| Criteria Label | Criteria | Threshold |
|---|---|---|
| A | Severe respiratory distress | Severe breathlessness (severe respiratory distress ticked on form) |
| B | Respiratory rate | Over 30 breaths per minute |
| C | Oxygen saturation | ≤92% on pulse oximetry, breathing air or on oxygen |
| D | Respiratory exhaustion | New abnormal breathing pattern (respiratory exhaustion ticked on form) |
| E | Dehydration or shock[a] | Systolic BP <90 mmHg and /or diastolic BP <60 mmHg. Sternal capillary refill time >2 seconds, reduced skin turgor |
| F | Altered conscious level[b] | New confusion, striking agitation or seizures |

| | | |
|---|---|---|
| G | Other clinical concern[c] | Another clinical concern |

a) dehydration or shock was defined if they were recorded as having severe dehydration, or systolic BP <90 mmHg and /or diastolic BP <60 mmHg, or central capillary refill is categorised as abnormal.
b) Altered conscious was positive if GCS was less than 15 or AVPU was anything other than A.
c) Other clinical concern was not recorded in the data, so two scores were calculated, one where point G was ignored and the Swine Flu hospital pathway calculated based on the first 6 items, and another where clinical concern was considered positive if NEWS2 was greater than 4 or any parameter in NEWS2 was given a score of 3.

If data for a parameter was missing, we defined the parameter as being negative. If more than three parameters were missing, we did not calculate a score for the patient.

**NEWS2**

The NEWS2 has seven parameters, each of which are scores from zero to three providing an overall score between zero and 20. The scores for each parameter can be found in the table below.

| Score | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Respiratory Rate | ≤8 | | 9-11 | 12-20 | | 21-24 | ≥25 |
| Oxygen saturation | ≤91 | 92-93 | 94-95 | ≥96 | | | |
| Heart Rate | ≤40 | | 41-50 | 51-90 | 91-110 | 111-130 | ≥131 |
| Systolic BP | ≤90 | 91-100 | 101-110 | 111-219 | | | ≥220 |
| Temperature | ≤35.0 | | 35.1-36.0 | 36.1-38.0 | 38.1-39.0 | ≥39.1 | |
| Neurology | | | | Alert | | | Confusion, Voice, Pain, Unresponsive |
| Air or Oxygen | | Oxygen (based on $FiO_2$>21%, or $FiO_2$>0 L/min) | | Air | | | |

If data for a parameter was missing, we scored the parameter as zero. If more than three parameters were missing, we did not calculate a score for the patient. The scale for patients with confirmed hypercapnic respiratory failure was not used. We analysed NEWS2 in two ways: (1) As a score, with thresholds between zero and 20 on the total score; (2) As a rule, with a single threshold of a total score greater than four or a score of three on any parameter.

**The WHO decision-making algorithm for hospitalisation with pneumonia**

The WHO decision-making algorithm for hospitalisation with pneumonia recommends admission for an adult patient (rule positive) if any of the following are present:

- respiratory rate >30/minute,
- oxygen saturation <90%,
- respiratory distress,
- age >60,
- any of the following comorbidities; hypertension, diabetes, cardiovascular disease, chronic respiratory disease, renal impairment immunosuppression

If data for a parameter was missing, we assumed it was negative. If more than three parameters were missing, we did not calculate a score for the patient



Thomas B, *et al. Emerg Med J* 2021;0:1–7. doi: 10.1136/emermed-2020-210783

## Supplementary Table S1: Sensitivity and specificity of each threshold of CURB, CRB, PMEWS and NEWS2 for predicting the primary and secondary outcomes

| | Primary: Adverse outcome | | Secondary: Death without support | | Secondary: Any Support | |
|---|---|---|---|---|---|---|
| Threshold | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
| **CURB-65** | | | | | | |
| >0 | 0.90 (0.89 to 0.91) | 0.44 (0.43 to 0.45) | 0.97 (0.96 to 0.98) | 0.41 (0.40 to 0.42) | 0.81 (0.79 to 0.82) | 0.38 (0.38 to 0.39) |
| >1 | 0.71 (0.70 to 0.72) | 0.69 (0.69 to 0.70) | 0.86 (0.85 to 0.87) | 0.67 (0.66 to 0.68) | 0.52 (0.50 to 0.54) | 0.62 (0.61 to 0.63) |
| >2 | 0.39 (0.38 to 0.41) | 0.88 (0.88 to 0.89) | 0.53 (0.51 to 0.55) | 0.87 (0.87 to 0.88) | 0.22 (0.20 to 0.24) | 0.83 (0.82 to 0.83) |
| >3 | 0.14 (0.13 to 0.15) | 0.98 (0.97 to 0.98) | 0.21 (0.19 to 0.23) | 0.97 (0.97 to 0.97) | 0.06 (0.05 to 0.07) | 0.95 (0.95 to 0.95) |
| >4 | 0.02 (0.02 to 0.03) | 1.00 (1.00 to 1.00) | 0.03 (0.03 to 0.04) | 1.00 (1.00 to 1.00) | 0.01 (0.00 to 0.01) | 0.99 (0.99 to 0.99) |
| **CRB-65** | | | | | | |
| >0 | 0.86 (0.85 to 0.87) | 0.48 (0.47 to 0.48) | 0.96 (0.95 to 0.97) | 0.45 (0.44 to 0.46) | 0.74 (0.72 to 0.76) | 0.42 (0.41 to 0.42) |
| >1 | 0.49 (0.47 to 0.50) | 0.82 (0.81 to 0.82) | 0.63 (0.61 to 0.65) | 0.80 (0.80 to 0.81) | 0.30 (0.28 to 0.32) | 0.76 (0.75 to 0.76) |
| >2 | 0.17 (0.16 to 0.18) | 0.96 (0.96 to 0.97) | 0.25 (0.23 to 0.27) | 0.96 (0.96 to 0.96) | 0.07 (0.06 to 0.08) | 0.94 (0.93 to 0.94) |
| >3 | 0.03 (0.02 to 0.03) | 1.00 (1.00 to 1.00) | 0.04 (0.04 to 0.05) | 1.00 (1.00 to 1.00) | 0.01 (0.01 to 0.01) | 0.99 (0.99 to 0.99) |
| **PMEWS** | | | | | | |
| >0 | 1.00 (1.00 to 1.00) | 0.05 (0.05 to 0.06) | 1.00 (1.00 to 1.00) | 0.05 (0.05 to 0.05) | 1.00 (1.00 to 1.00) | 0.05 (0.04 to 0.05) |
| >1 | 0.99 (0.99 to 0.99) | 0.17 (0.16 to 0.17) | 1.00 (0.99 to 1.00) | 0.15 (0.14 to 0.15) | 0.99 (0.98 to 0.99) | 0.14 (0.14 to 0.15) |
| >2 | 0.97 (0.96 to 0.97) | 0.30 (0.30 to 0.31) | 0.99 (0.98 to 0.99) | 0.28 (0.27 to 0.28) | 0.94 (0.93 to 0.95) | 0.27 (0.26 to 0.27) |
| >3 | 0.89 (0.88 to 0.90) | 0.45 (0.45 to 0.46) | 0.92 (0.91 to 0.93) | 0.42 (0.41 to 0.43) | 0.85 (0.84 to 0.87) | 0.40 (0.40 to 0.41) |
| >4 | 0.78 (0.76 to 0.79) | 0.60 (0.59 to 0.61) | 0.83 (0.81 to 0.84) | 0.57 (0.56 to 0.57) | 0.72 (0.70 to 0.74) | 0.54 (0.54 to 0.55) |
| >5 | 0.65 (0.64 to 0.67) | 0.73 (0.72 to 0.74) | 0.71 (0.69 to 0.73) | 0.70 (0.69 to 0.70) | 0.58 (0.56 to 0.60) | 0.67 (0.66 to 0.68) |
| >6 | 0.51 (0.50 to 0.53) | 0.83 (0.82 to 0.83) | 0.57 (0.55 to 0.59) | 0.80 (0.79 to 0.80) | 0.44 (0.42 to 0.46) | 0.77 (0.77 to 0.78) |
| >7 | 0.37 (0.36 to 0.39) | 0.90 (0.89 to 0.90) | 0.43 (0.41 to 0.45) | 0.87 (0.87 to 0.88) | 0.31 (0.29 to 0.33) | 0.85 (0.85 to 0.86) |
| >8 | 0.26 (0.25 to 0.27) | 0.94 (0.94 to 0.94) | 0.31 (0.29 to 0.33) | 0.92 (0.92 to 0.93) | 0.20 (0.18 to 0.21) | 0.91 (0.90 to 0.91) |
| >9 | 0.17 (0.16 to 0.18) | 0.97 (0.97 to 0.97) | 0.22 (0.21 to 0.24) | 0.96 (0.96 to 0.96) | 0.11 (0.10 to 0.13) | 0.94 (0.94 to 0.95) |
| >10 | 0.11 (0.10 to 0.12) | 0.98 (0.98 to 0.99) | 0.14 (0.13 to 0.16) | 0.98 (0.98 to 0.98) | 0.07 (0.06 to 0.08) | 0.97 (0.96 to 0.97) |



Thomas B, *et al. Emerg Med J* 2021;0:1–7. doi: 10.1136/emermed-2020-210783

| | | | | | | |
|---|---|---|---|---|---|---|
| >11 | 0.06 (0.06 to 0.07) | 0.99 (0.99 to 0.99) | 0.09 (0.08 to 0.10) | 0.99 (0.99 to 0.99) | 0.03 (0.03 to 0.04) | 0.98 (0.98 to 0.98) |
| >12 | 0.03 (0.03 to 0.04) | 1.00 (1.00 to 1.00) | 0.05 (0.04 to 0.06) | 1.00 (0.99 to 1.00) | 0.01 (0.01 to 0.02) | 0.99 (0.99 to 0.99) |
| >13 | 0.02 (0.01 to 0.02) | 1.00 (1.00 to 1.00) | 0.03 (0.02 to 0.03) | 1.00 (1.00 to 1.00) | 0.01 (0.00 to 0.01) | 1.00 (0.99 to 1.00) |
| >14 | 0.01 (0.01 to 0.01) | 1.00 (1.00 to 1.00) | 0.01 (0.01 to 0.02) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.01) | 1.00 (1.00 to 1.00) |
| >15 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.01) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |
| >16 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |
| >17 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |
| **NEWS2** | | | | | | |
| >0 | 0.98 (0.98 to 0.99) | 0.14 (0.13 to 0.15) | 0.98 (0.98 to 0.99) | 0.13 (0.12 to 0.13) | 0.98 (0.98 to 0.99) | 0.12 (0.12 to 0.13) |
| >1 | 0.96 (0.95 to 0.96) | 0.28 (0.28 to 0.29) | 0.96 (0.95 to 0.96) | 0.26 (0.25 to 0.26) | 0.96 (0.95 to 0.96) | 0.25 (0.24 to 0.26) |
| >2 | 0.92 (0.91 to 0.93) | 0.40 (0.39 to 0.41) | 0.92 (0.91 to 0.93) | 0.37 (0.36 to 0.37) | 0.92 (0.90 to 0.93) | 0.36 (0.35 to 0.37) |
| >3 | 0.85 (0.84 to 0.86) | 0.53 (0.52 to 0.54) | 0.86 (0.84 to 0.87) | 0.49 (0.48 to 0.50) | 0.84 (0.82 to 0.86) | 0.48 (0.47 to 0.48) |
| >4 | 0.77 (0.76 to 0.78) | 0.64 (0.63 to 0.65) | 0.78 (0.76 to 0.79) | 0.60 (0.59 to 0.60) | 0.76 (0.74 to 0.78) | 0.58 (0.58 to 0.59) |
| >5 | 0.66 (0.65 to 0.68) | 0.74 (0.73 to 0.75) | 0.68 (0.66 to 0.69) | 0.70 (0.69 to 0.70) | 0.64 (0.62 to 0.67) | 0.68 (0.68 to 0.69) |
| >6 | 0.54 (0.52 to 0.55) | 0.82 (0.82 to 0.83) | 0.56 (0.54 to 0.57) | 0.78 (0.78 to 0.79) | 0.51 (0.49 to 0.54) | 0.77 (0.77 to 0.78) |
| >7 | 0.43 (0.42 to 0.45) | 0.88 (0.88 to 0.89) | 0.46 (0.44 to 0.48) | 0.85 (0.85 to 0.86) | 0.40 (0.38 to 0.42) | 0.84 (0.83 to 0.84) |
| >8 | 0.32 (0.30 to 0.33) | 0.93 (0.92 to 0.93) | 0.35 (0.33 to 0.37) | 0.91 (0.90 to 0.91) | 0.27 (0.25 to 0.29) | 0.89 (0.89 to 0.89) |
| >9 | 0.21 (0.20 to 0.23) | 0.96 (0.95 to 0.96) | 0.25 (0.23 to 0.26) | 0.94 (0.94 to 0.95) | 0.17 (0.16 to 0.19) | 0.93 (0.93 to 0.93) |
| >10 | 0.14 (0.13 to 0.15) | 0.98 (0.97 to 0.98) | 0.17 (0.16 to 0.19) | 0.97 (0.96 to 0.97) | 0.10 (0.09 to 0.12) | 0.96 (0.95 to 0.96) |
| >11 | 0.08 (0.08 to 0.09) | 0.99 (0.98 to 0.99) | 0.11 (0.10 to 0.12) | 0.98 (0.98 to 0.98) | 0.05 (0.05 to 0.07) | 0.97 (0.97 to 0.98) |
| >12 | 0.05 (0.04 to 0.06) | 0.99 (0.99 to 0.99) | 0.07 (0.06 to 0.08) | 0.99 (0.99 to 0.99) | 0.03 (0.03 to 0.04) | 0.99 (0.98 to 0.99) |
| >13 | 0.03 (0.03 to 0.04) | 1.00 (1.00 to 1.00) | 0.04 (0.03 to 0.05) | 1.00 (0.99 to 1.00) | 0.02 (0.01 to 0.03) | 0.99 (0.99 to 0.99) |
| >14 | 0.01 (0.01 to 0.02) | 1.00 (1.00 to 1.00) | 0.02 (0.01 to 0.02) | 1.00 (1.00 to 1.00) | 0.01 (0.01 to 0.01) | 1.00 (1.00 to 1.00) |
| >15 | 0.01 (0.00 to 0.01) | 1.00 (1.00 to 1.00) | 0.01 (0.01 to 0.01) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.01) | 1.00 (1.00 to 1.00) |
| >16 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.01) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |
| >17 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |
| >18 | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) | 0.00 (0.00 to 0.00) | 1.00 (1.00 to 1.00) |

**Appendix 2: Study Steering Committee**

| Title | First Name | Last Name | Job Title | Name of employing institution, and any institutions where this nominee holds an Honorary Contract | Membership Type: | Independent | Expertise |
|---|---|---|---|---|---|---|---|
| Mrs | Shan | Bennett | PPI | | PPI Member | Yes | PPI, Sheffield Emergency Care Forum |
| Prof (Associate Professor) | Paul | Baxter | Senior Lecturer in Biostatistics | University of Leeds. Honorary contract with Leeds Teaching Hospitals NHS Trust. | Member | Yes | Biostatistics |
| Prof | Tim | Coats | Professor of Emergency Medicine | University of Leicester | Chair | Yes | Clinician Emergency Medicine Research |
| Mrs | Enid | Hirst | Co-ordinator of Sheffield Emergency Care Forum (PPI) | | PPI Member | Yes | PPI, Sheffield Emergency Care Forum |
| Mrs | Beryl | Darlison | PPI | | PPI Member | Yes | PPI, Sheffield Emergency Care Forum |
| Dr | Kavin | Smith | Deputy Director Healthcare Public Health England, Yorkshire and the Humber (Replaced Will Morton as the PHE representative) | Public Health England | Member | Yes | Public health |
| Dr | Will | Morton | Consultant in Health Protection at Public Health England | Public Health England. Honorary contract with the University of Manchester. | Member | Yes | Health protection specialist |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Dr | Nazir | Lone | Senior Clinical Lecturer in Critical Care | The University of Edinburgh, Honorary Consultant in Critical Care at the Royal Infirmary of Edinburgh. | Member | No | Cliniciain Critical Care, Critical Care Epidemiology |
| Dr | Graham | McClelland | Research paramedic | North East Ambulance Service NHS Trust | Member | Yes | Clinician Paramedic |
| Prof | Steve | Goodacre | PRIEST CI | The University of Sheffield | Member | No | Clinician Emergency Medicine Research |
| Mrs | Rachel | Robinson | Chief Nurse | Integrated Care 24 Ltd | Member | Yes | Clinician, 111 Knowledge |
| Dr | Mathew | Beattie | Medical Director North East Ambulance Service Foundation Trust | North East Ambulance Service Foundation Trust | Member | Yes | Clinician, North East Ambulance Service |

**Appendix 3: Site Research Staff**

Anna Wilson
Bethan Holroyd-Hind
Chris Fitzsimmons
Julie Morcombe
Mrs Claire Brookes
Mrs Heidi Redfearn
Dr Ben Bloom
Imogen Skene
Raine Astin-Chamberlain
Laura Barman
Janine Mallison
Lisa Baldwin
Anneli James
Lucy Newton
Mark D Lyttle
Rebecca Hoskins
Sally Melson
Bristol Clinical Trials Team
Mrs Emma Storr
Mr Martin Walton
Dr Felix Wood
Annie Rose RGN
Francesca Wright RGN
Aarzoo Khan
Liza Keating
Emma Craig
Elizabeth Taylor
Andrew Rees MBA RICR
Simon Sharpe
Heather Sellers
Dr Tanya de Weymarn
Gloucestershire Cancer Clinical Trials Team
Jagtar Pooni
Sara Simmons
Dani Steward
Dr Adrian Marsh
Sr Mandy Carnahan
Sr Lucy Price
Dr Mark Harrison
Rebecca Emmonds
Jane Luke
Northumbria Clinical Trials Team
Mrs Katrina Parkinson
Miss Georgia Thomasson
Mrs Alda Remegoso
Bernard Hadebe
Arianna Bellini
Chloe Lyons
Tracy Marsden
Mr Paul Brittain
Joanne Galliford
Prisca Gondo
Paula Harman
Melanie Darwent
Ross Downes
Sally Beer
Dr Jennifer Lockwood
Harrogate Research team
Alasdair Gray
Polly Black
Amanda Lyle
Yvonne Lester
Dane Goodere-Bennett
Dr Huw Steven Jenkins
Dr John Wright
Kimberley Webster
Professor Richard Body
Dr Eloise Cook
Non-clinical COVID-19 Research Delivery Team at Manchester
Abdo Sattout
Melanie Harrison
Sarah Stevenson
Dr Adrian Boyle
Susie Hardwick
Debbie Read
Frank Coffey
Megan Meredith
Helen Navarra
Mrs Judith Ratcliff
Fiona Thompson
Amanda Adamson
Dr Gareth Hampton
Dr Sarah Wilson
Mrs Joana Da Rocha
Dr Charlotte Griffiths
Dr Nam Tong
Mrs Tracy Fuller
Mrs Hannah Bloxham
Alastair Richards
Debra Barnett
Lindianne Aitken

Suzannah Pegler
Maggie Walton
Tim Slade
Fleur Cantle
Hannah Cotton
Maeve Cockrell
Jessica Law
Ava Williams
Janet Mills
Janice Birt
Cassandra Gleeson
Dr Recebba Macfarlane
Mrs Lisa Evans
Ms Eloise Van Vuren
Dr Amelia Gruber
Dr Ignacio Cardona
Laura O'Rourke
Julie Quigley
Mohammad Zubair Ahmad
Daniella Hydes
Suzanne Mason
Mishel Cunningham
Nicola Lancaster
Amanda Cowton
Sarah Clark
Jane Varin
Karl Ward
Ella Sykes
Heather Jarman
Desislava Baramova
Marta Pizzorusso
Dr Sarah Essex
Mrs Andrea Watson
Mr Craig Mower
Sara Bennett
Judith Bell
Abigail Pemberton
Dr Jill Woodhead
Sherwood Forest Clinical Trials Team
Dr Amber Nocher
Dr Henrietta Morton King
Mrs Jo-Ann Taylor
Dr Shayma Habeeb
Wojciech Sawicki
Kate Martin
Nicola Charnley
Mr Matthew Edward Ryan
Dr Shrouk Messahel
Dr Daniel B Hawcutt
Miss Laura Purandare
Mr Daniel Griffiths
Miss Rebecca Miln
Robert Hull
Laura Robertson
Michaela Sutherland
Bolton Clinical Trials Team
Christine Dixon
Ellen Jessup-Dunton
Reina Layug
Dr. Rajendar Garlapati
Farzana Masters
Yvonne Grimes
Joseph Dykes
Katharine Gantert
Favour Chukwunonyerem
Dilara Arslan

**Appendix 4: Supporting Research Staff**

Marie Hyslop

Dan Beever

Samuel Keating

Kerry Wilson

Heather Dakin

Edwin Burkinshaw

Kirsty Pemberton

Tim Chater

Chris Turtle

Emily Turton

Matthew Bursnall

Mike Bradburn

Jennifer Petrie

Lizzie Swaby

Gemma Hackney

Judith Cohen

## Supplementary Table S2: Proportion of patients with an adverse outcome for each level of each triage tool

| | CURB-65 | | CRB-65 | | PMEWS | | NEWS2 score | | NEWS2 rule | | SFAHP-6 | | SFAHP-7 | | WHO | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Score** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** | **N** | **Prop AO** |
| 0 | 7538 | 0.06 | 8320 | 0.08 | 872 | 0 | 2331 | 0.03 | 9538 | 0.08 | 11361 | 0.1 | 8337 | 0.06 | 4624 | 0.05 |
| 1 | 5008 | 0.17 | 7224 | 0.24 | 1821 | 0.02 | 2427 | 0.05 | 11056 | 0.34 | 5309 | 0.29 | 4769 | 0.19 | 16267 | 0.27 |
| 2 | 4487 | 0.32 | 3824 | 0.38 | 2326 | 0.05 | 2095 | 0.08 | | | 2210 | 0.47 | 4436 | 0.33 | | |
| 3 | 2647 | 0.43 | 1173 | 0.55 | 2738 | 0.12 | 2389 | 0.13 | | | 723 | 0.62 | 2160 | 0.48 | | |
| 4 | 901 | 0.61 | 175 | 0.74 | 2873 | 0.18 | 2171 | 0.17 | | | 205 | 0.76 | 725 | 0.62 | | |
| 5 | 135 | 0.73 | | | 2606 | 0.22 | 2075 | 0.23 | | | 43 | 0.86 | 205 | 0.76 | | |
| 6 | | | | | 2204 | 0.29 | 1896 | 0.3 | | | 7 | 0.86 | 43 | 0.86 | | |
| 7 | | | | | 1702 | 0.36 | 1448 | 0.33 | | | | | 7 | 0.86 | | |
| 8 | | | | | 1216 | 0.43 | 1254 | 0.43 | | | | | | | | |
| 9 | | | | | 845 | 0.47 | 941 | 0.5 | | | | | | | | |
| 10 | | | | | 549 | 0.53 | 624 | 0.53 | | | | | | | | |

AO: adverse outcome