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,1,2 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 Walter8

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.

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 infection4 recommends hospital admission for...
severity 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 suggests that the National Early Warning Score version 2 (NEWS2) score 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 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.

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 associated 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 of fever (≥37.8°C) and acute onset of persistent cough (with

---

**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%) |

---

**Table 1 Continued**

| Characteristic         | Statistic/level |
|------------------------|-----------------|
|                        | 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 | Respiratory exhaustion | 292 (1.4%) |
|                        | Severe dehydration | 261 (1.2%) |

AVPU, Alert Verbal Pain Unresponsive.
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 an adverse outcome. In the event, the adverse outcome rate in 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.11 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 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.

### 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     | 20716 | 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      | 20716 | 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       | 20492 | 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) | 20594 | 0.77 (0.76, 0.78) | >>0        | 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) | 20594 | 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) | 19854 | 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) | 20682 | 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 | 20891 | 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.
20,891 in the analysis after excluding 39 who requested withdrawal of their data, 1,530 children, 7 with missing age and 17 with missing outcome data.

Table 1 shows the characteristics of adults in the cohort. Some 13,997 (67.0%) were admitted after ED assessment and 6,521 (31.2%) ultimately tested positive for COVID-19. Overall, 4,611 (22.1%) died or received organ support (primary outcome), with 2,058 (9.9%) receiving organ support and 2,553 (12.2%) dying (22.1%) died or received organ support (primary outcome), with 3,121 (31.2%) ultimately tested positive for COVID-19. 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 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     | 20,716 | 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      | 20,716 | 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       | 20,492 | 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) | 0.97 (0.97 to 0.98) | 0.76 |
| NEWS2 (score) | 20,594 | 0.72 (0.71, 0.73) | >0 | 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) | 20,594 | 0.65 (0.64, 0.66) | >1 | 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 |
| SFAP (6-point rule) | 19,858 | 0.64 (0.63, 0.65) | >0 | 0.68 (0.66 to 0.71) | 0.60 (0.59 to 0.61) | 0.15 (0.15 to 0.17) | 0.95 (0.94 to 0.95) | 0.43 |
| SFAP (7-point rule) | 20,682 | 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 | 20,891 | 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 |

**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     | 20,716 | 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      | 20,716 | 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       | 20,492 | 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) | 20,594 | 0.73 (0.74, 0.76) | >0 | 0.76 (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) | 20,594 | 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 |
| SFAP (6-point rule) | 19,858 | 0.70 (0.70, 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 0.95) | 0.43 |
| SFAP (7-point rule) | 20,682 | 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 | 20,891 | 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 |

**Table 3, Table 4, Note:** NEWS2, National Early Warning Score version 2; PMEWS, Pandemic Modified Early Warning Score; ROC, receiver operating characteristic; SFAP, 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. Recent studies have evaluated NEWS2, CURB-65 and CRB-65 in adult inpatients with confirmed COVID-19. Fan et al (n=654) 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) 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) reported c-statistics of 0.79 for NEWS2 and 0.85 for CURB-65 for predicting death. Satici et al (n=681) 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 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) evaluated NEWS2 as a predictor of intensive care admission and Myrstad et al (n=66) evaluated NEWS2 and CRB-65 as predictors 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.

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/#/ [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-detailed-list/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/swan-haln-coronavirus-initial-investigation-of-possible-cases/investigation-and-initial-clinical-management-of-possible-cases-of-swarm-haln-coronavirus-wcm-cov-infection/criteria [Accessed 27 Aug 2020].
11. Goodacre S, Irving A, Wilson R, et al. The pandemic influenza triage in the emergency department (pilot) cohort study. Health Technol Assess 2015;19:1–70.
12. TataCorp. Tata statistical software: release 16. College Station, TX: TataCorp LLC, 2019.
13. Hirst I, 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. Ferguson NM, 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 Tech Assess 2010;14:337–496.
16 Muller MP, McGeer AI, 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 2020.
18 Bradley P, Frost E, Thamaratnam 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 Satıcı 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 V, 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.