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 Marincowitz1, 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 Walter8

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 20891 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.

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. 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 focus on adults in this study.
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 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.

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. The Pandemic Modified Early Warning Score (PMEWS) 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 of fever (≥37.8°C) and acute onset of persistent cough (with...
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.

### RESULTS

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

<table>
<thead>
<tr>
<th>Triage tool</th>
<th>n</th>
<th>C-statistic</th>
<th>Cut point</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive predictive value (95% CI)</th>
<th>Negative predictive value (95% CI)</th>
<th>Proportion test positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURB-65</td>
<td>20 716</td>
<td>0.75 (0.74, 0.76)</td>
<td>&gt;1</td>
<td>0.71 (0.70 to 0.72)</td>
<td>0.69 (0.69 to 0.70)</td>
<td>0.40 (0.39 to 0.41)</td>
<td>0.89 (0.89 to 0.90)</td>
<td>0.39</td>
</tr>
<tr>
<td>CRB-65</td>
<td>20 716</td>
<td>0.70 (0.69, 0.70)</td>
<td>&gt;0</td>
<td>0.86 (0.85 to 0.87)</td>
<td>0.48 (0.47 to 0.48)</td>
<td>0.32 (0.31 to 0.33)</td>
<td>0.92 (0.92 to 0.93)</td>
<td>0.60</td>
</tr>
<tr>
<td>PMEWS</td>
<td>20 492</td>
<td>0.77 (0.76, 0.77)</td>
<td>&gt;2</td>
<td>0.97 (0.96 to 0.97)</td>
<td>0.30 (0.30 to 0.31)</td>
<td>0.28 (0.27 to 0.29)</td>
<td>0.97 (0.96 to 0.97)</td>
<td>0.76</td>
</tr>
<tr>
<td>NEWS2 (score)</td>
<td>20 594</td>
<td>0.77 (0.76, 0.78)</td>
<td>&gt;3</td>
<td>0.77 (0.76 to 0.78)</td>
<td>0.64 (0.63 to 0.65)</td>
<td>0.38 (0.37 to 0.39)</td>
<td>0.90 (0.90 to 0.91)</td>
<td>0.45</td>
</tr>
<tr>
<td>NEWS2 (rule)</td>
<td>20 594</td>
<td>0.69 (0.68, 0.69)</td>
<td>&gt;0</td>
<td>0.83 (0.82 to 0.84)</td>
<td>0.55 (0.54 to 0.55)</td>
<td>0.34 (0.33 to 0.35)</td>
<td>0.92 (0.91 to 0.92)</td>
<td>0.54</td>
</tr>
<tr>
<td>SFAHP (6-point rule)</td>
<td>19 854</td>
<td>0.70 (0.69, 0.71)</td>
<td>&gt;0</td>
<td>0.74 (0.73 to 0.75)</td>
<td>0.66 (0.65 to 0.67)</td>
<td>0.37 (0.36 to 0.37)</td>
<td>0.90 (0.89 to 0.90)</td>
<td>0.43</td>
</tr>
<tr>
<td>SFAHP (7-point rule)</td>
<td>20 682</td>
<td>0.68 (0.68, 0.69)</td>
<td>&gt;0</td>
<td>0.88 (0.87 to 0.89)</td>
<td>0.48 (0.48 to 0.49)</td>
<td>0.33 (0.32 to 0.33)</td>
<td>0.94 (0.93 to 0.94)</td>
<td>0.60</td>
</tr>
<tr>
<td>WHO algorithm</td>
<td>20 891</td>
<td>0.61 (0.61, 0.62)</td>
<td>&gt;0</td>
<td>0.95 (0.95 to 0.96)</td>
<td>0.27 (0.26 to 0.28)</td>
<td>0.27 (0.26 to 0.28)</td>
<td>0.95 (0.95 to 0.96)</td>
<td>0.78</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>Triage tool</th>
<th>n</th>
<th>C-statistic</th>
<th>Cut point</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive predictive value (95% CI)</th>
<th>Negative predictive value (95% CI)</th>
<th>Proportion test positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURB-65</td>
<td>20,716</td>
<td>0.60 (0.59, 0.61)</td>
<td>(&gt;1)</td>
<td>0.52 (0.50 to 0.54)</td>
<td>0.62 (0.61 to 0.63)</td>
<td>0.13 (0.12 to 0.14)</td>
<td>0.92 (0.92 to 0.93)</td>
<td>0.39</td>
</tr>
<tr>
<td>CRB-65</td>
<td>20,716</td>
<td>0.58 (0.56, 0.59)</td>
<td>(&lt;0)</td>
<td>0.74 (0.72 to 0.76)</td>
<td>0.42 (0.41 to 0.42)</td>
<td>0.12 (0.12 to 0.13)</td>
<td>0.94 (0.93 to 0.94)</td>
<td>0.60</td>
</tr>
<tr>
<td>PMEWS</td>
<td>20,492</td>
<td>0.68 (0.67, 0.69)</td>
<td>(&gt;2)</td>
<td>0.94 (0.93 to 0.95)</td>
<td>0.27 (0.26 to 0.27)</td>
<td>0.12 (0.12 to 0.13)</td>
<td>0.98 (0.97 to 0.98)</td>
<td>0.76</td>
</tr>
<tr>
<td>NEWS2 (score)</td>
<td>20,594</td>
<td>0.72 (0.71, 0.73)</td>
<td>(&lt;0)</td>
<td>0.76 (0.74 to 0.78)</td>
<td>0.58 (0.58 to 0.59)</td>
<td>0.17 (0.16 to 0.17)</td>
<td>0.96 (0.95 to 0.96)</td>
<td>0.45</td>
</tr>
<tr>
<td>NEWS2 (rule)</td>
<td>20,594</td>
<td>0.65 (0.64, 0.66)</td>
<td>(&lt;0)</td>
<td>0.81 (0.79 to 0.83)</td>
<td>0.49 (0.49 to 0.50)</td>
<td>0.15 (0.14 to 0.16)</td>
<td>0.96 (0.96 to 0.96)</td>
<td>0.54</td>
</tr>
<tr>
<td>SFAHP (6-point rule)</td>
<td>19,854</td>
<td>0.64 (0.63, 0.65)</td>
<td>(&lt;0)</td>
<td>0.68 (0.66 to 0.71)</td>
<td>0.60 (0.59 to 0.61)</td>
<td>0.15 (0.15 to 0.17)</td>
<td>0.95 (0.94 to 0.95)</td>
<td>0.43</td>
</tr>
<tr>
<td>SFAHP (7-point rule)</td>
<td>20,682</td>
<td>0.65 (0.64, 0.65)</td>
<td>(&lt;0)</td>
<td>0.86 (0.84 to 0.87)</td>
<td>0.43 (0.42 to 0.44)</td>
<td>0.14 (0.14 to 0.15)</td>
<td>0.97 (0.96 to 0.97)</td>
<td>0.60</td>
</tr>
<tr>
<td>WHO algorithm</td>
<td>20,891</td>
<td>0.57 (0.57, 0.58)</td>
<td>(&lt;0)</td>
<td>0.91 (0.90 to 0.92)</td>
<td>0.24 (0.23 to 0.24)</td>
<td>0.12 (0.11 to 0.12)</td>
<td>0.96 (0.96 to 0.97)</td>
<td>0.78</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Triage tool</th>
<th>n</th>
<th>C-statistic</th>
<th>Cut point</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive predictive value (95% CI)</th>
<th>Negative predictive value (95% CI)</th>
<th>Proportion test positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURB-65</td>
<td>20,716</td>
<td>0.82 (0.82, 0.83)</td>
<td>(&gt;1)</td>
<td>0.86 (0.85 to 0.87)</td>
<td>0.67 (0.66 to 0.68)</td>
<td>0.27 (0.26 to 0.28)</td>
<td>0.97 (0.97 to 0.97)</td>
<td>0.39</td>
</tr>
<tr>
<td>CRB-65</td>
<td>20,716</td>
<td>0.75 (0.75, 0.76)</td>
<td>(&lt;0)</td>
<td>0.96 (0.95 to 0.97)</td>
<td>0.45 (0.44 to 0.46)</td>
<td>0.20 (0.19 to 0.20)</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.60</td>
</tr>
<tr>
<td>PMEWS</td>
<td>20,492</td>
<td>0.78 (0.77, 0.78)</td>
<td>(&gt;2)</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.28 (0.27 to 0.28)</td>
<td>0.16 (0.15 to 0.17)</td>
<td>0.99 (0.99 to 0.99)</td>
<td>0.76</td>
</tr>
<tr>
<td>NEWS2 (score)</td>
<td>20,594</td>
<td>0.73 (0.74, 0.76)</td>
<td>(&gt;4)</td>
<td>0.76 (0.76 to 0.79)</td>
<td>0.60 (0.59 to 0.60)</td>
<td>0.21 (0.20 to 0.22)</td>
<td>0.95 (0.95 to 0.95)</td>
<td>0.45</td>
</tr>
<tr>
<td>NEWS2 (rule)</td>
<td>20,594</td>
<td>0.67 (0.67, 0.68)</td>
<td>(&gt;4)</td>
<td>0.84 (0.83 to 0.86)</td>
<td>0.51 (0.50 to 0.51)</td>
<td>0.19 (0.18 to 0.20)</td>
<td>0.96 (0.95 to 0.96)</td>
<td>0.54</td>
</tr>
<tr>
<td>SFAHP (6-point rule)</td>
<td>19,854</td>
<td>0.70 (0.69, 0.71)</td>
<td>(&gt;4)</td>
<td>0.78 (0.77 to 0.80)</td>
<td>0.62 (0.61 to 0.63)</td>
<td>0.22 (0.21 to 0.23)</td>
<td>0.95 (0.95 to 0.95)</td>
<td>0.43</td>
</tr>
<tr>
<td>SFAHP (7-point rule)</td>
<td>20,682</td>
<td>0.67 (0.67, 0.68)</td>
<td>(&gt;4)</td>
<td>0.90 (0.89 to 0.91)</td>
<td>0.44 (0.44 to 0.45)</td>
<td>0.18 (0.18 to 0.19)</td>
<td>0.97 (0.97 to 0.97)</td>
<td>0.60</td>
</tr>
<tr>
<td>WHO algorithm</td>
<td>20,891</td>
<td>0.62 (0.61, 0.62)</td>
<td>(&gt;4)</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.25 (0.24 to 0.26)</td>
<td>0.15 (0.15 to 0.16)</td>
<td>0.99 (0.99 to 0.99)</td>
<td>0.78</td>
</tr>
</tbody>
</table>
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 al21 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 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 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 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
1SchARR, The University of Sheffield, Sheffield, UK
2Acute intensive Care Unit, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK
3Emergency Department, Lancashire Teaching Hospitals NHS Foundation Trust, Chorley, UK
4Sheffield Children’s Hospital NHS Foundation Trust, Sheffield, UK
5Emergency Department, Sheffield Children’s Hospital NHS Foundation Trust, Sheffield, UK
6Department of Emergency Medicine, Royal London Hospital, London, UK
7Paediatric ED, Imperial College Healthcare NHS Trust, London, UK
8Emergency Department, University Hospital of South Manchester NHS Foundation Trust, Manchester, UK

Twitter Tim Harris @tim harris@resusdocs and Darren Walter @emsdocn

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, AL, IM and DW conceived and designed the study, KB, KB, AL, WS, KS, JS, SC, ES, JM and EY acquired the data, EL, LS, MB, SG, BT, KB and CM analysed the data, SG, AB, KC, CF, TH, FL, AL, IM, DW, EL, LS, 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).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content contains any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Ben Thomas http://orcid.org/0000-0002-6659-6930
Steve Goodacre http://orcid.org/0000-0003-0803-8444
Carl Marincoiwicz http://orcid.org/0000-0003-3433-7564
Sarah Connelly http://orcid.org/0000-0001-9963-5628

REFERENCES
12 StataCorp. Stata statistical software: release 16. College Station, TX: StataCorp LLC, 2019.


