OPEN ACCESS

Prehospital early warning scores for adults with suspected sepsis: retrospective diagnostic cohort study

Steve Goodacre ,[1] Laura Sutton,[1] Ben Thomas ,[1] Olivia Hawksworth ,[1] Khurram Iftikhar,[2] Susan Croft,[2] Gordon Fuller,[1] Simon Waterhouse,[1] Daniel Hind,[1] Mike Bradburn,[1] Michael Anthony Smyth ,[3] Gavin D Perkins,[3] Mark Millins,[4] Andy Rosser,[5] Jon M Dickson ,[1] Matthew Joseph Wilson[1]

Handling editor Edward Carlton

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2023-213315).

[1]Sheffield Centre for Health and Related Research (SCHARR), The University of Sheffield, Sheffield, UK
[2]Emergency Department, Northern General Hospital, Sheffield, UK
[3]Clinical Trials Unit, University of Warwick, Coventry, UK
[4]Yorkshire Ambulance Service NHS Trust, Wakefield, UK
[5]West Midlands Ambulance Service, West Midlands, UK

Correspondence to
Professor Steve Goodacre, The University of Sheffield, Sheffield S10 2TN, UK;
s.goodacre@sheffield.ac.uk

Received 28 April 2023
Accepted 18 August 2023
Published Online First 6 September 2023

To cite: Goodacre S, Sutton L, Thomas B, *et al*. *Emerg Med J* 2023;**40**:768–776.

ABSTRACT

Background Ambulance services need to identify and prioritise patients with sepsis for early hospital assessment. We aimed to determine the accuracy of early warning scores alongside paramedic diagnostic impression to identify sepsis that required urgent treatment.

Methods We undertook a retrospective diagnostic cohort study involving adult emergency medical cases transported to Sheffield Teaching Hospitals ED by Yorkshire Ambulance Service in 2019. We used routine ambulance service data to calculate 21 early warning scores and categorise paramedic diagnostic impressions as sepsis, infection, non-specific presentation or other presentation. We linked cases to hospital records and identified those meeting the sepsis-3 definition who received urgent hospital treatment for sepsis (reference standard). Analysis determined the accuracy of strategies that combined early warning scores at varying thresholds for positivity with paramedic diagnostic impression.

Results We linked 12 870/24 955 (51.6%) cases and identified 348/12 870 (2.7%) with a positive reference standard. None of the strategies provided sensitivity greater than 0.80 with positive predictive value greater than 0.15. The area under the receiver operating characteristic curve for the National Early Warning Score, version 2 (NEWS2) applied to patients with a diagnostic impression of sepsis or infection was 0.756 (95% CI 0.729, 0.783). No other early warning score provided clearly superior accuracy to NEWS2. Paramedic impression of sepsis or infection had sensitivity of 0.572 (0.519, 0.623) and positive predictive value of 0.156 (0.137, 0.176). NEWS2 thresholds of >4, >6 and >8 applied to patients with a diagnostic impression of sepsis or infection, respectively, provided sensitivities and positive predictive values of 0.522 (0.469, 0.574) and 0.216 (0.189, 0.245), 0.447 (0.395, 0.499) and 0.274 (0.239, 0.313), and 0.314 (0.268, 0.365) and 0.333 (0.284, 0.386).

Conclusion No strategy is ideal but using NEWS2 alongside paramedic diagnostic impression of infection or sepsis could identify one-third to half of sepsis cases without prioritising unmanageable numbers. No other score provided clearly superior accuracy to NEWS2.

Trial registration number researchregistry5268, https://www.researchregistry.com/browse-the-registry#home/registrationdetails/5de7bbd97ca5b50015041c33/

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Guidelines for sepsis recommend urgent treatment within 1 hour for people with suspected sepsis who are at highest risk. Ambulance services can use early warning scores alongside paramedic diagnostic impression to identify and prioritise people with suspected sepsis.

WHAT THIS STUDY ADDS

⇒ This retrospective diagnostic cohort study of 12 870 patients showed that no combination of early warning score alongside diagnostic impression provides sensitivity greater than 0.80 with positive predictive value greater than 0.15. Using the National Early Warning Score, version 2 (NEWS2) at thresholds of >4 to >8 in patients with a diagnostic impression of infection or sepsis could identify one-third to half of sepsis cases without prioritising unmanageable numbers. No alternative early warning score provided clearly superior accuracy to NEWS2.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Ambulance services and hospitals can use the estimates of NEWS2 sensitivity and positive predictive value to identify an appropriate NEWS2 threshold score to guide the use of prealerts for patients with suspected sepsis.

INTRODUCTION

Sepsis is a life-threatening response to a severe infection, which can lead to tissue damage, organ failure and death.[1] Guidelines for sepsis highlight the importance of early recognition and treatment, with treatment recommended within 1 hour of presentation for those at highest risk.[1–4] The emergency care system can only achieve this if sepsis is recognised and prioritised. This may involve ambulance services prealerting the ED that they are transporting a patient with suspected sepsis. However, prioritising too many patients with suspected sepsis may delay assessment of other urgent cases or may result in a lack of meaningful prioritisation.

Ambulance services can use prehospital early warning scores to identify people with a high risk of sepsis.[5] Early warning scores use clinical observations to determine a score, with a higher score indicating a higher risk of adverse outcome. They may be generic (applicable to a range of conditions) or specific to sepsis. Clinicians need to determine a threshold value of the score for decision-making that balances the risks of missing sepsis against prioritising too many patients. Sepsis may present with non-specific symptoms,[1] so clinicians need to decide whether to suspect sepsis and apply an early warning score to all medical cases, non-specific presentations, suspected infection or just suspected sepsis.

Systematic reviews have identified many potential prehospital early warning scores for sepsis but supporting evidence has substantial weaknesses and reports inconsistent findings.[6–8] This may be explained by differences in study populations, reference standard definitions, the threshold score used or whether the score was applied to all medical cases or just those with evidence of infection.[5]

Evaluating the accuracy of an early warning score or diagnostic assessment for sepsis involves determining the sensitivity (to reflect the risk of missing sepsis) and the specificity (to reflect the risk of prioritising cases without sepsis). A score with apparently high specificity may prioritise an unmanageable number of cases if the prevalence of sepsis is low, such as when the score is applied to all medical cases. Furthermore, specificity (the proportion of patients without sepsis who have a score below the threshold) may be difficult to interpret in clinical practice. We therefore use positive predictive value (the proportion of patients with a score above the threshold who have sepsis) rather than specificity to interpret the risk of prioritising too many patients.

We aimed to determine the accuracy of prehospital early warning scores, used alongside paramedic diagnostic impression, for identifying sepsis requiring urgent treatment in adult medical cases transported to hospital by emergency ambulance.

METHODS

This study is the main component of the Prehospital Early Warning Scores for Sepsis study. Full details of the project will be reported in the National Institute for Health Research library.[9] We planned to undertake a retrospective observational cohort study across two ambulance services and four hospitals using the UK NHS Data Access Request Service from NHS Digital to link ambulance service to hospital data. However, NHS Digital was unable to provide this service, so we implemented an alternative process using NHS numbers (a unique number for each NHS patient) to link Yorkshire Ambulance Service data to Sheffield Teaching Hospitals data.

We used routine ambulance service data to identify all adult emergency medical cases transported to the Sheffield Teaching Hospitals ED from 1 January to 31 December 2019. We excluded cases with injury, mental health problems, cardiac arrest or direct transfer to specialist services (including maternity, cardiac or stroke services). We also excluded cases with no NHS number and patients who had opted out of allowing use of their data for research. Individuals can inform NHS Digital or their general practice that they wish to opt out of having their NHS data used for research and planning purposes. Yorkshire Ambulance Service checked cases against the national data opt-out service and removed records from the data set if they were identified as belonging to individuals who have opted out.

We evaluated any early warning score that prehospital professionals could use and that we could calculate from the ambulance service electronic patient report form (ePRF). We included dichotomous scores (ie, rules) that simply categorise into high and low-risk groups, but for simplicity refer collectively to early warning scores. We searched the Embase, CINAHL, PubMed, ClinicalTrials.gov, the ISRCTN registry and Research Registry for relevant studies and selected 21 scores for evaluation.[3 10–29] Online supplemental table 1 outlines the scores and compares their constituent variables. The scores used combinations of age, temperature, HR, RR, peripheral oxygen saturation, conscious level and BP, along with a small number of other variables. During the study period, Yorkshire Ambulance Service used an electronic patient record that calculated the National Early Warning Score, version 2 (NEWS2)[10] from constituent variables so paramedics would have been aware of this score.

We calculated each score for each case using ePRF data. We used the first recorded measurement for each variable. If the variable was not recorded in the first set of observations, then the first recorded measurement was used from a subsequent set of observations. We inferred conscious level or ACVPU (alert, confused, voice, pain, unresponsive) from the GCS, assuming 15 equals alert, 14 equals confused, 12–13 equal voice, 9–11 equal pain and 3–8 equal unresponsive. We modified scores that included variables that would not be available in routine practice or were not recorded on the ePRF. For example, we removed lactate, oliguria and recent chemotherapy from the UK Sepsis Trust red flag criteria.[3] Online supplemental appendix 1 provides details of how each score is calculated, any modifications or assumptions in calculating the score from routine data and the threshold for decision-making.

The ePRF recorded a paramedic diagnostic impression from a list of options. We categorised the options as sepsis, infection (excluding sepsis), non-specific diagnostic impression in which sepsis could be suspected or other diagnostic impression in which sepsis would not usually be suspected (see online supplemental appendix 2 for details). We then applied each early warning score alongside diagnostic impression as follows:

1. Score applied to cases with impression of sepsis. Cases with impression of infection, non-specific or other were categorised as score negative.
2. Score applied to cases with impression of sepsis or infection. Cases with impression of non-specific or other were categorised as score negative.
3. Score applied to cases with impression of sepsis, infection or non-specific. Cases with impression of other were categorised as score negative.
4. Score applied to all cases regardless of diagnostic impression.

We defined the reference standard (sepsis requiring urgent treatment) as being positive if the patient met the sepsis-3 definition of sepsis and received treatment for sepsis within 4 hours of initial assessment at hospital.[30] We planned a secondary analysis using just the sepsis-3 definition as the reference standard but 95% of cases meeting the sepsis-3 definition received urgent treatment, so the results of the secondary analysis matched the primary analysis. We therefore only report the primary analysis.

We used routine hospital data to select those with a primary or secondary International Classification of Diseases 10 admission code or cause of death compatible with sepsis, or an ED code for sepsis. Research nurses briefly reviewed the ED records of these cases and selected patients for expert review if they had any diagnosis or treatment for sepsis recorded in the ED notes or sepsis as an admission diagnosis on the hospital discharge summary.

Two experts independently reviewed hospital records for the selected patients and determined whether there was: (1) evidence of infection and life-threatening organ dysfunction (according to

the sepsis-3 definition[30]) within 4 hours of initial assessment; and (2) treatment for sepsis given within 4 hours. Evidence of infection could include microbiology reports identifying organisms, radiology reports identifying infective changes or other markers strongly suggesting infection. Organ dysfunction was defined as a Sequential (sepsis-related) Organ Failure Assessment (SOFA) score of 2 or more points worse than normal. We estimated the SOFA score using the ED observations chart and first blood results after admission. In accordance with the sepsis-3 definition,[30] we assumed the normal SOFA score would be 0 unless there was evidence in the hospital records to suggest otherwise. Treatment for sepsis was based on relevant guidelines[1 2] and typically involved intravenous antibiotic therapy. One of the experts also estimated the Clinical Frailty Score using information in the hospital records.[31]

If the two reviewers disagreed on the overall sepsis-3 judgement or whether urgent treatment for sepsis was given, then a consensus decision was reached through discussion. Disagreements over an element of the sepsis-3 definition (evidence of infection or change in SOFA score) were left unresolved if they did not affect the overall judgement.

We used the patient as the unit of analysis and only included the first eligible episode per patient. Kappa scores were calculated to determine the agreement between reference standard adjudicators. We constructed receiver operating characteristic (ROC) curves to evaluate sensitivity and specificity over the range of each score. We calculated the area under the ROC curve and sensitivities, specificities and positive and negative predictive values at key cut-points, each with a 95% CI.

We anticipated a low prevalence of reference standard positive cases, based on data from Smyth *et al*,[27] so we based the sample size on identifying at least 200 reference standard positive cases. Collins *et al*[32] recommend basing external validation studies on a minimum of 100–200 events.[32] Our sample size would allow us to estimate the sensitivity of an early warning score with an SE of 2.1% assuming sensitivity of 90%, and the area under the ROC curve with an SE of 2% assuming an area under the ROC curve of at least 0.75.[33]

Clinical experts in the research team reviewed ED attendance data and determined that a positive predictive value of 0.15 or lower would result in too many positive cases for meaningful prioritisation and that sensitivity exceeding 0.8 would be considered good.

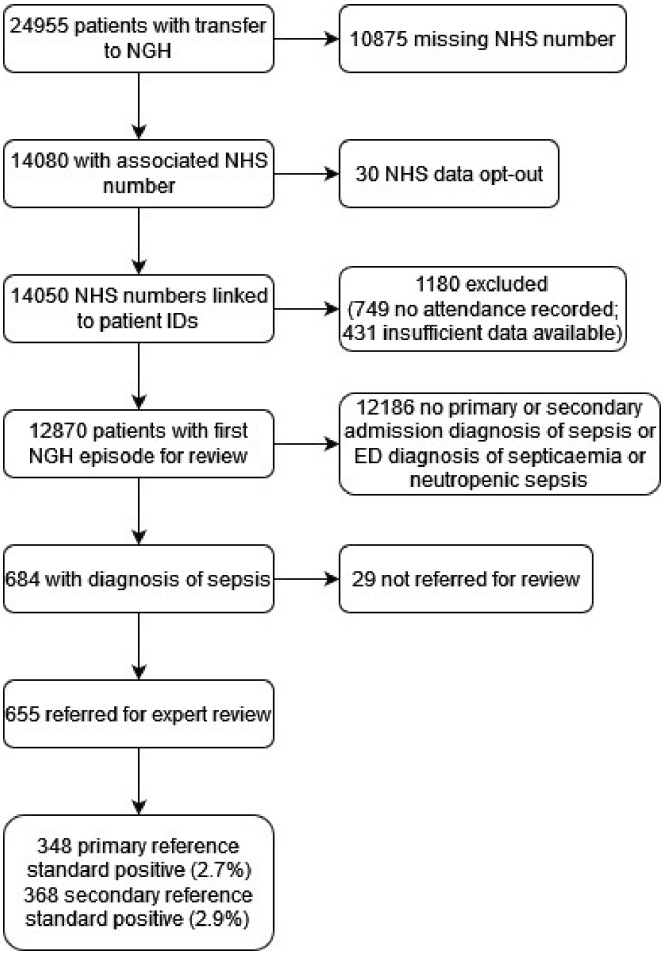

Figure 1 Participant flow through the study. NGH, Northern General Hospital.

Patient and public involvement

The Sheffield Emergency Care Forum (SECF) is a public representative group interested in emergency care research.[34] Two members of SECF joined the project management group and helped develop and deliver the project. Public representatives supported the use of patient data without consent and reviewed the early warning scores to determine patient and public acceptability, resulting in one score being modified to remove care home residence as a variable. 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

Figure 1 shows the flow of eligible cases. We identified 24 955 cases transported to Sheffield Teaching Hospitals ED in 2019, of whom 14 050 (56.3%) had NHS numbers and no opt-out. Table 1 shows the characteristics of the 14 050 patients and compares them to those unavailable for linkage. Included patients were markedly older (median age 71 vs 55 years) and more likely to be female (54.7% vs 53.0%) and white ethnicity (95.7% vs 91.8%). We linked 12 870/14 050 cases (91.6%) with a hospital attendance or admission, which comprised the study cohort.

There were 684/12 870 episodes with an admission or ED coding for sepsis. The research nurses referred 655/684 (95.8%) for expert review. The experts judged that 368/655 (56.2%) episodes met the sepsis-3 definition and 348/368 (94.6%) of these received urgent treatment for sepsis. Therefore, 348/12 870 (2.7%) met the reference standard definition. Online supplemental table 2 shows the agreement between the reference standard adjudicators. Agreement was moderate (kappa=0.62) for evidence of infection but disagreements tended to occur in cases that did not meet the SOFA score criterion, so overall judgement on the sepsis-3 definition was good (kappa=0.89), as was agreement for whether urgent treatment was given (kappa=0.87).

There was radiological evidence of infection in 175/348 (50.1%) cases, microbiological evidence in 171 (49.0%) and other clinical evidence in 328 (94.0%). The sites of suspected infection were chest (155, 44.4%), urine (78, 22.3%), biliary (43, 12.3%), abdominal (16, 4.6%), skin (25, 7.2%), other (6, 1.7%)and unknown (26, 7.4%). Mean Clinical Frailty Score was 5.6 (median 6.0, range 2.0–9.0) and mean SOFA score was 3.9

Table 1 Characteristics of patients available for linkage with hospital data

	Not linked (n=10 905)	Linked (n=14 050)	Total (n=24 955)
Age (years)			
Mean (SD)	55.2 (23.3)	65.3 (21.2)	60.9 (22.7)
Median (IQR)	55.0 (34.0, 76.0)	71.0 (51.0, 82.0)	65.0 (42.0, 80.0)
Range	16.0–102.0	16.0–105.0	16.0–105.0
Sex			
Missing	0	22	22
Female	5484 (50.3%)	7672 (54.7%)	13 156 (52.8%)
Male	5421 (49.7%)	6356 (45.3%)	11 777 (47.2%)
Ethnicity			
Missing	5290	6880	12 170
White	5153 (91.8%)	6860 (95.7%)	12 013 (94.0%)
Asian	136 (2.4%)	122 (1.7%)	258 (2.0%)
Black	73 (1.3%)	55 (0.8%)	128 (1.0%)
Mixed	49 (0.9%)	32 (0.4%)	81 (0.6%)
Other	204 (3.6%)	101 (1.4%)	305 (2.4%)
ACVPU			
Missing	0	0	0
Alert	9754 (89.4%)	13 232 (94.2%)	22 986 (92.1%)
Confusion	341 (3.1%)	387 (2.8%)	728 (2.9%)
Voice	386 (3.5%)	257 (1.8%)	643 (2.6%)
Pain	192 (1.8%)	107 (0.8%)	299 (1.2%)
Unresponsive	232 (2.1%)	67 (0.5%)	299 (1.2%)
GCS			
Mean (SD)	14.4 (2.0)	14.7 (1.2)	14.5 (1.6)
Median (IQR)	15.0 (15.0, 15.0)	15.0 (15.0, 15.0)	15.0 (15.0, 15.0)
Range	3.0–15.0	3.0–15.0	3.0–15.0
Diastolic BP (mm Hg)			
Mean (SD)	83.1 (17.5)	82.1 (17.2)	82.6 (17.4)
Median (IQR)	83.0 (72.0, 94.0)	82.0 (71.0, 93.0)	82.0 (71.0, 93.0)
Range	0.0–190.0	5.0–195.0	0.0–195.0
Systolic BP (mm Hg)			
Mean (SD)	139.0 (26.5)	142.1 (27.4)	140.8 (27.1)
Median (IQR)	138.0 (122.0, 153.0)	140.0 (124.0, 158.0)	139.0 (123.0, 156.0)
Range	53.0–257.0	43.0–285.0	43.0–285.0
HR (beats/min)			
Mean (SD)	89.5 (22.8)	88.7 (21.9)	89.1 (22.3)
Median (IQR)	87.0 (74.0, 103.0)	86.0 (73.0, 102.0)	86.0 (74.0, 102.0)
Range	0.0–218.0	0.0–216.0	0.0–218.0
Oxygen saturation (%)			
Mean (SD)	96.0 (4.9)	95.6 (4.9)	95.8 (4.9)
Median (IQR)	97.0 (95.0, 98.0)	97.0 (95.0, 98.0)	97.0 (95.0, 98.0)
Range	18.0–100.0	10.0–100.0	10.0–100.0
Supplemental oxygen			
Missing	18	27	45
No	10 345 (95.0%)	13 252 (94.5%)	23 597 (94.7%)
Yes	542 (5.0%)	771 (5.5%)	1313 (5.3%)
Respiration (breath/min)			
Mean (SD)	19.7 (6.0)	20.5 (6.1)	20.1 (6.0)
Median (IQR)	18.0 (16.0, 20.0)	18.0 (16.0, 22.0)	18.0 (16.0, 22.0)
Range	0.0–93.0	0.0–91.0	0.0–93.0
Temperature (°C)			
Mean (SD)	36.8 (1.0)	37.0 (1.0)	36.9 (1.0)
Median (IQR)	36.8 (36.2, 37.3)	36.9 (36.4, 37.4)	36.8 (36.4, 37.4)
Range	26.0–41.3	27.1–41.8	26.0–41.8
Glucose (mmol/L)			
Mean (SD)	7.1 (3.2)	7.4 (3.4)	7.2 (3.3)

Continued

Table 1 Continued

	Not linked (n=10 905)	Linked (n=14 050)	Total (n=24 955)
Median (IQR)	6.2 (5.4, 7.6)	6.4 (5.5, 8.0)	6.3 (5.5, 7.8)
Range	0.5–36.6	0.9–49.0	0.5–49.0
Prealerted			
No	10 307 (94.5%)	13 419 (95.5%)	23 726 (95.1%)
Yes	598 (5.5%)	631 (4.5%)	1229 (4.9%)
Impression			
1—Sepsis	222 (2.0%)	407 (2.9%)	629 (2.5%)
2—Infection	471 (4.3%)	912 (6.5%)	1383 (5.5%)
3—Non-specific	3494 (32.0%)	4962 (35.3%)	8456 (33.9%)
4—Other	6718 (61.6%)	7769 (55.3%)	14 487 (58.1%)

(median 3.0, range 2.0–14.0). Some 28 (8.0%) were admitted to critical care and 261 (74.8%) survived to hospital discharge or 30 days after attendance, whichever was sooner.

Paramedic diagnostic impression of sepsis had sensitivity (95% CI) of 0.328 (0.28, 0.379) and positive predictive value of 0.285 (0.243, 0.331); infection or sepsis had sensitivity of 0.572 (0.519, 0.623) and positive predictive value of 0.156 (0.137, 0.176); and non-specific, infection or sepsis had sensitivity of 0.897 (0.86, 0.924) and positive predictive value of 0.053 (0.048, 0.059). Online supplemental table 3 shows the full details.

Figures 2–5 show the ROC curves for each score alongside diagnostic impression of sepsis, infection, non-specific and all cases. Online supplemental table 4 reports the areas under each ROC curve and online supplemental tables 5–12 show the accuracy parameters behind the ROC curves. The area under the ROC curve is greater when the scores are used less selectively with paramedic diagnostic impression. However, the accuracy parameters in the online supplemental tables show that the positive predictive value is low (<0.15) if specificity is below 0.9. The area under the ROC curve is therefore a poor reflection of accuracy at the thresholds that yield acceptable positive predictive value (ie, specificity >0.9). Figures 2–5 show that none of the alternative scores had superior accuracy to NEWS2. The possible exception is the Screening to Enhance Prehospital Identification of Sepsis (SEPSIS) score that has a higher area under the ROC curve when applied to non-specific or all cases, but has similar accuracy to NEWS2 at thresholds that provide specificity greater than 0.9.

Table 2 shows the accuracy parameters (reproduced from online supplemental tables 5–12) for early warning scores at specified thresholds, selected on the basis of their use in sepsis guidelines[3 4 17 30] in patients with an impression of infection or sepsis. NEWS2>4, NEWS2>6 and quick SOFA (qSOFA)>1 provide a range of options with varying sensitivity and positive predictive value. qSOFA>1 provides similar accuracy to NEWS2>8 (also included in the table). The modified NHS prealert criteria[17] provide slightly inferior accuracy to NEWS2>6. The modified UK Sepsis Trust criteria[3] provide similar accuracy to NEWS2>4.

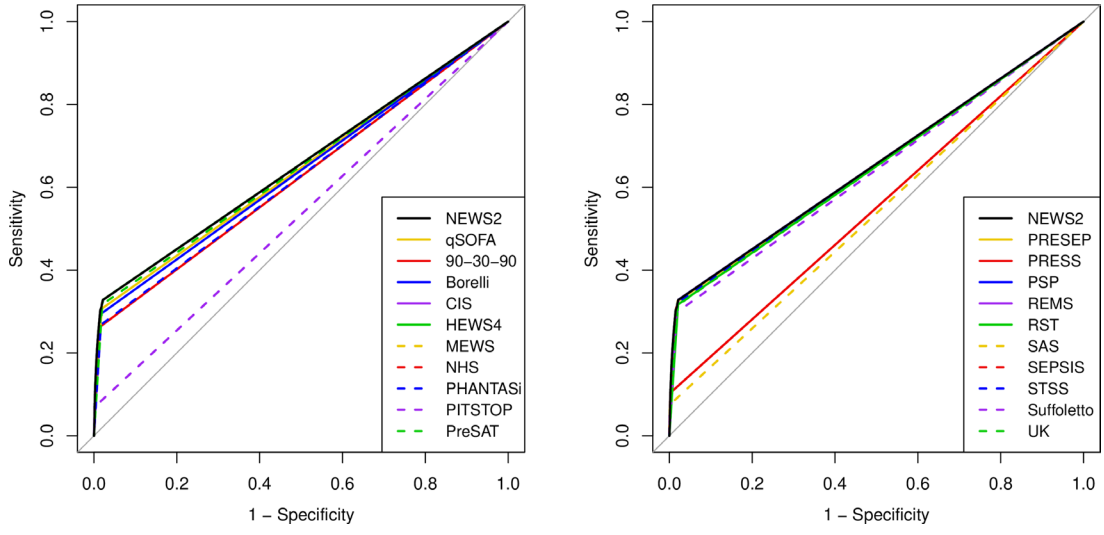

Figure 2 Receiver operating characteristic (ROC) curves for early warning scores applied to diagnostic impression of sepsis. CIS, Critical Illness Score; HEWS, Hamilton Early Warning Score; MEWS, Modified Early Warning Score; NEWS2, National Early Warning Score, version 2; PHANTASi, Prehospital Antibiotics Against Sepsis; PITSTOP, Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients; PreSAT, Prehospital Sepsis Assessment Tool; PRESEP, Prehospital Early Sepsis Detection; PRESS, Prehospital Severe Sepsis; PSP, Prehospital Sepsis Project; qSOFA, quick Sequential Organ Failure Assessment; REMS, Rapid Emergency Medicine Score; RST, Robson Screening Tool; SEPSIS, Screening to Enhance Prehospital Identification of Sepsis; STSS, Simple Triage Scoring System.

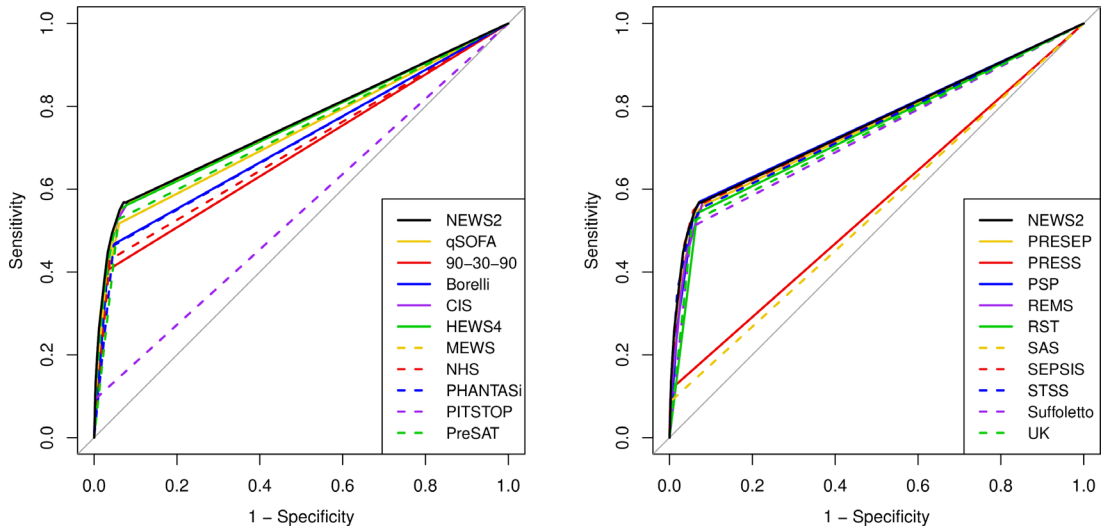

Figure 3 Receiver operating characteristic (ROC) curves for early warning scores applied to diagnostic impression of sepsis or infection. CIS, Critical Illness Score; HEWS, Hamilton Early Warning Score; MEWS, Modified Early Warning Score; NEWS2, National Early Warning Score, version 2; PHANTASi, Prehospital Antibiotics Against Sepsis; PITSTOP, Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients; PreSAT, Prehospital Sepsis Assessment Tool; PRESEP, Prehospital Early Sepsis Detection; PRESS, Prehospital Severe Sepsis; PSP, Prehospital Sepsis Project; qSOFA, quick Sequential Organ Failure Assessment; REMS, Rapid Emergency Medicine Score; RST, Robson Screening Tool; SEPSIS, Screening to Enhance Prehospital Identification of Sepsis; STSS, Simple Triage Scoring System.

DISCUSSION

We found that no combination of early warning score alongside paramedic diagnostic impression provided sensitivity greater than 0.8 and positive predictive value greater than 0.15 for sepsis. The appropriate trade-off between sensitivity and positive predictive value will depend on the consequences of prioritisation. However, prioritising more than five people for each case of sepsis (which would be the consequence of using a strategy with positive predictive value of 0.15 or lower) risks overstretching ED capacity and a loss of meaningful prioritisation.

No score had superior accuracy to NEWS2. The only possible exception was the SEPSIS score[27] when thresholds were used that optimised sensitivity at the expense of low positive predictive value. NEWS2 is widely used in the UK NHS, so any alternative score would need to demonstrate clear superiority to justify the additional training and documentation required in this setting.

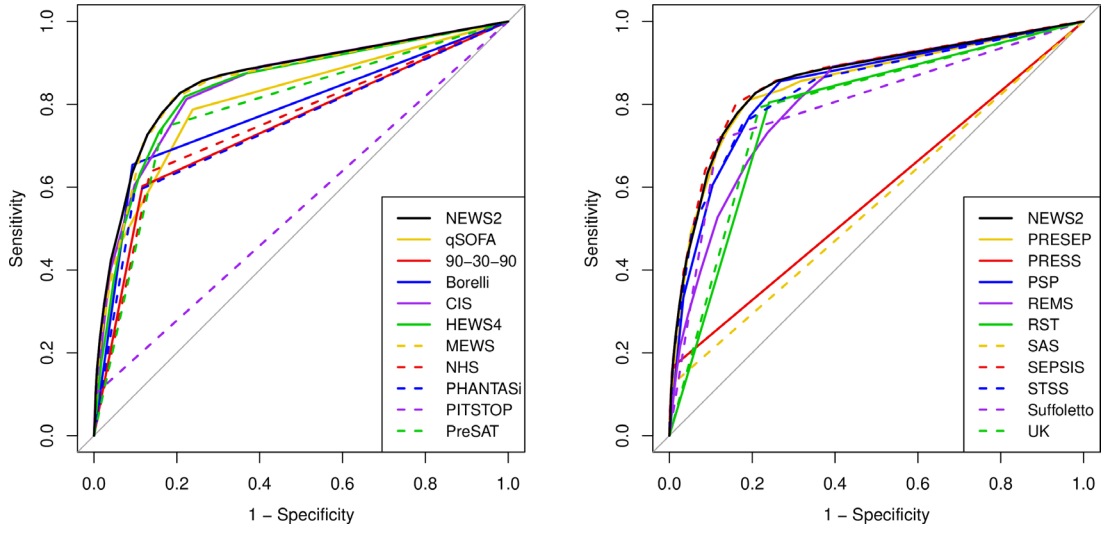

Figure 4 Receiver operating characteristic (ROC) curves for early warning scores applied to diagnostic impression of sepsis, infection or non-specific presentation. CIS, Critical Illness Score; HEWS, Hamilton Early Warning Score; MEWS, Modified Early Warning Score; NEWS2, National Early Warning Score, version 2; PHANTASi, Prehospital Antibiotics Against Sepsis; PITSTOP, Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients; PreSAT, Prehospital Sepsis Assessment Tool; PRESEP, Prehospital Early Sepsis Detection; PRESS, Prehospital Severe Sepsis; PSP, Prehospital Sepsis Project; qSOFA, quick Sequential Organ Failure Assessment; REMS, Rapid Emergency Medicine Score; RST, Robson Screening Tool; SEPSIS, Screening to Enhance Prehospital Identification of Sepsis; STSS, Simple Triage Scoring System.

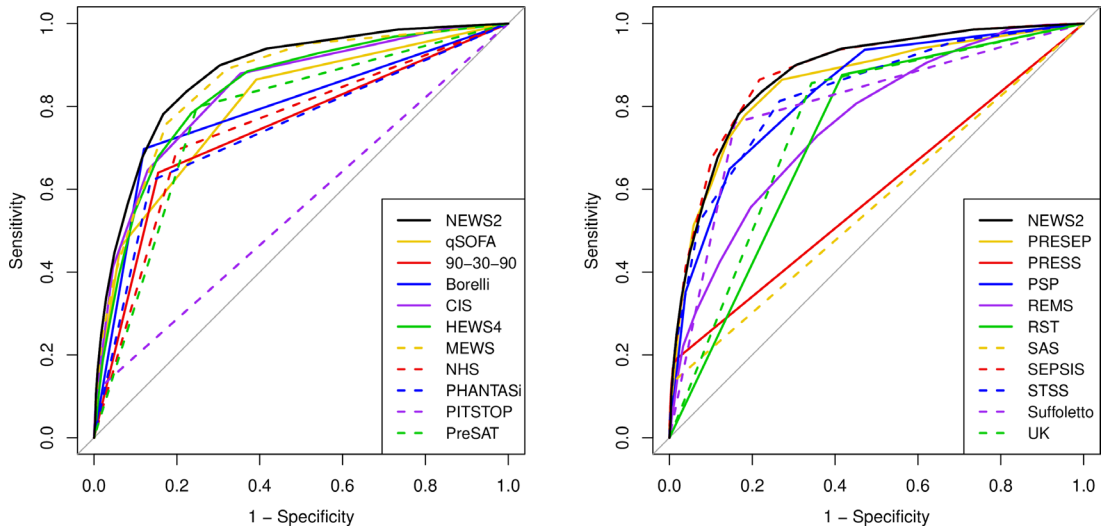

Figure 5 Receiver operating characteristic (ROC) curves for early warning score applied to all diagnostic impressions. CIS, Critical Illness Score; HEWS, Hamilton Early Warning Score; MEWS, Modified Early Warning Score; NEWS2, National Early Warning Score, version 2; PHANTASi, Prehospital Antibiotics Against Sepsis; PITSTOP, Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients; PreSAT, Prehospital Sepsis Assessment Tool; PRESEP, Prehospital Early Sepsis Detection; PRESS, Prehospital Severe Sepsis; PSP, Prehospital Sepsis Project; qSOFA, quick Sequential Organ Failure Assessment; REMS, Rapid Emergency Medicine Score; RST, Robson Screening Tool; SEPSIS, Screening to Enhance Prehospital Identification of Sepsis; STSS, Simple Triage Scoring System.

Using NEWS2 at thresholds of >4 and >6 to prioritise patients with suspected infection, as recommended by the Academy of Medical Royal Colleges clinical decision support framework,[4] would provide sensitivities of 0.522 and 0.447, respectively, and positive predictive values of 0.216 and 0.274. To prioritise fewer patients, NEWS2 could be used with a threshold of >8, which would provide similar sensitivity and positive predictive value (0.314 and 0.333) to using qSOFA with a threshold of >1 (0.305 and 0.356). Using NEWS2 alongside paramedic diagnostic impression improves positive predictive value at the expense of sensitivity, compared with paramedic diagnostic impression alone.

We recently searched for studies validating the accuracy of early warning scores for suspected sepsis in a prehospital population and identified 13 studies evaluating the scores included in this study.[5] There was substantial variation in the reported results, with no consistent evidence that any score was superior to the others. Variations in study populations, outcomes and the thresholds used make comparisons difficult. A systematic review of hospital studies found that at established thresholds NEWS tended to have higher sensitivity while qSOFA tended to have higher specificity.[35] Our study suggests that this difference reflects the chosen threshold. The sensitivity and specificity of NEWS2 at a higher threshold than usually recommended (>8) are similar to the sensitivity and specificity of qSOFA>1.

Our findings are similar to other studies evaluating multiple scores in a large cohort. Lane *et al* found that no single strategy had high sensitivity and specificity for classifying sepsis, but the Critical Illness Prediction score, NEWS and qSOFA showed good prediction for sepsis.[36] Smyth *et al* identified three strategies offering an acceptable balance between sensitivity and positive predictive value: SEPSIS>2, Systemic Inflammatory Response Syndrome (SIRS) criteria >1 and NEWS>4.[27] These studies did not identify any early warning score with clearly superior accuracy to NEWS2.

Key strengths of our study include the large sample size including sufficient cases with sepsis to estimate sensitivity with reasonable precision. The reference standard was based on an internationally recognised definition of sepsis that was adjudicated by two independent clinicians with acceptable interobserver agreement. The main limitation is that we were only able to link around half the eligible cases with hospital records. Those linked tended to be much older, possibly reflecting more frequent contact with health services. Sepsis is associated with

Table 2 Accuracy of selected early warning scores alongside paramedic impression of sepsis or infection for identifying sepsis receiving urgent treatment

Early warning score	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Paramedic impression alone	0.572 (0.519, 0.623)	0.914 (0.909, 0.919)	0.156 (0.137, 0.176)	0.987 (0.985, 0.989)
NEWS2>4	0.522 (0.469, 0.574)	0.947 (0.943, 0.951)	0.216 (0.189, 0.245)	0.986 (0.984, 0.988)
NEWS2>6	0.447 (0.395, 0.499)	0.967 (0.964, 0.97)	0.274 (0.239, 0.313)	0.984 (0.982, 0.986)
NEWS2>8	0.314 (0.268, 0.365)	0.983 (0.98, 0.985)	0.333 (0.284, 0.386)	0.981 (0.978, 0.983)
qSOFA>1	0.305 (0.259, 0.355)	0.985 (0.982, 0.987)	0.356 (0.304, 0.412)	0.981 (0.978, 0.983)
NHS prealert	0.429 (0.378, 0.482)	0.962 (0.959, 0.966)	0.24 (0.208, 0.275)	0.984 (0.981, 0.986)
UK Sepsis Trust	0.522 (0.469, 0.574)	0.945 (0.941, 0.949)	0.209 (0.183, 0.237)	0.986 (0.984, 0.988)

NEWS2, National Early Warning Score, version 2; NPV, negative predictive value; PPV, positive predictive value; qSOFA, quick Sequential Organ Failure Assessment.

age and comorbidity, but our findings may not be generalisable to younger patients with little comorbidity. The single-centre design limits the generalisability of the findings. The predominantly white ethnicity of our population may limit generalisability to patients of other ethnicities. We collected data over a year to mitigate the effects of seasonality and used data from 2019 as we felt that this was a typical year in terms of the prevalence of respiratory pathogens (if such a thing exists), but rates of presentations requiring prioritisation may show marked seasonality and variation according to the prevalence of respiratory pathogens. We may have misclassified cases as reference standard negative if they had sepsis but the ED or hospital discharge codes did not include sepsis.

Other limitations relate to the sepsis-3 definition.[30] While adjudicating the reference standard we noticed that the change in SOFA score often reflected the local effects of infection (eg, respiratory failure in pneumonia or raised bilirubin in biliary infection) or an exacerbation of underlying comorbidity, rather than organ failure likely to reflect a dysregulated host response to infection. The sepsis-3 definition was based on evidence that the SOFA score predicts mortality,[11] but this may not translate into potential to benefit from treatment.[37 38] Our reference standard may therefore include many patients who do not have a dysregulated response to infection and are unlikely to benefit from early treatment. We tried to address this issue by including receipt of urgent treatment for sepsis in our definition, but 95% of presentations received early treatment for sepsis.

Paramedic awareness of the NEWS2 score may have influenced their assessment of diagnostic impression, particularly in terms of differentiating sepsis from other infections. This may mean that paramedic diagnostic impression and NEWS2 scores are correlated to a degree. Use of NEWS2 in the ED may have prompted greater investigation for infection in patients with a higher NEWS2 score. However, NEWS2 scores were not routinely recorded in the hospital records used in reference standard assessment, so the reference standard adjudicators were not aware of the patient's NEWS2 (or any other) score.

The implications of our findings are that any combination of diagnostic impression and early warning score is likely to result in too many cases being prioritised or cases of sepsis being missed. EDs must therefore either ensure capacity to handle large numbers of cases being prioritised or ensure that missed cases do not suffer excessive delays. Ambulance services could use NEWS2 in patients with evidence of infection at thresholds of >4, >6 or >8, depending on the capacity of EDs to handle prioritised cases or avoid excessive delay for missed cases.

Future research is required to improve prehospital identification of sepsis but new scores based on currently measured physiological parameters are unlikely to improve on NEWS2. Prehospital biomarkers could improve early warning scores but future research needs to address the limitations of the sepsis-3 definition. Until we are able to measure the dysregulated host response that characterises sepsis, we will risk developing methods that identify patients with infection and organ failure, but do not have a dysregulated host response.

In summary, we found no ideal strategy but using NEWS2 alongside paramedic diagnostic impression of infection or sepsis could identify one-third to half of sepsis cases without prioritising unmanageable numbers.

Twitter Khurram Iftikhar @K_Ifti, Susan Croft @drsusiec and Jon M Dickson @Dr_J_M_Dickson

Acknowledgements We thank Anna Wilson and Sarah Bird (Research Nurses, Sheffield Teaching Hospitals NHS Foundation Trust) for screening cases for reference standard adjudication; Erica Wallis (Research Coordinator, Sheffield Teaching Hospitals NHS Foundation Trust) for her support and advice on ethical and governance issues; Linda Abouzeid, Lynn Simmons and Enid Hirst (Public Representatives, Sheffield Emergency Care Forum) and Peter Hewkin (Independent Public Representative) for their insight and guidance; Katie Ridsdale and Jack Rose (Trial Support Officers, The University of Sheffield CTRU) for providing central administrative support; Martin Bayley, Steven Wood and Thomas Butterfield (Healthcare Computer Scientists, Sheffield Teaching Hospitals NHS Foundation Trust) for linking ambulance service to hospital data; Richard Pilbery (Research Paramedic Fellow, Yorkshire Ambulance Service) for providing the ambulance service data; Janette Turner (Reader in Urgent & Emergency Care, The University of Sheffield) and Catheryn James (Clinical Manager, Yorkshire Ambulance Service) for their help with developing the proposal and involvement in the project management group; the NIHR Sheffield Clinical Research Facility for supporting the study; the members of the Study Steering Committee: Tim Coats (Professor in Emergency Medicine, University of Leicester), Alice Riddell (Patient Representative, Sheffield Emergency Care Forum), Shan Bennett (Patient Representative, Sheffield Emergency Care Forum), Duncan Robertson (Consultant Paramedic and Regional Clinical Lead, Welsh Ambulance Service), James Fullerton (Specialist Registrar in Clinical Pharmacology and Therapeutics, University of Oxford), Graham McClelland (Paramedic Research Fellow, North East Ambulance Service), Fenella Wrigley (Chief Medical Officer, London Ambulance Service), Joie Ensor (Associate Professor in Biostatistics, University of Birmingham).

Contributors SG conceived the study. SG, GF, DH, MB, MAS, GDP, MM, AR, JMD and MJW designed the study. BT, OH, SW and MM acquired the data. SG, KI, SC and GF undertook reference standard adjudication. LS analysed the data. All authors interpreted the data and contributed to drafting the manuscript. SG is the guarantor of the paper.

Funding The PHEWS study was funded by the UK National Institute for Health Research Health Technology Assessment (HTA) programme (project reference 17/136/10).

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 All authors declare grant funding to their employing institutions from the National Institute for Health Research.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the London Stanmore Research Ethics Committee, the Health Research Authority and the Confidentiality Advisory Group (reference number 19/LO/1443). The Confidentiality Advisory Group gave section 251 approval for the use of patient data without consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Anonymised data are available from the corresponding author upon reasonable request (contact s.goodacre@sheffield.ac.uk).

ORCID iDs
Steve Goodacre http://orcid.org/0000-0003-0803-8444
Ben Thomas http://orcid.org/0000-0002-6659-6930
Olivia Hawksworth http://orcid.org/0000-0001-6513-100X
Michael Anthony Smyth http://orcid.org/0000-0003-0220-2223
Jon M Dickson http://orcid.org/0000-0002-1361-2714

REFERENCES

1 National Institute for Health and Care Excellence. NICE Guideline 51 [NG51]. Sepsis: recognition, diagnosis and early management. Available: https://www.nice.org.uk/guidance/ng51 [Accessed 14 Apr 2023].
2 Evans L, Rhodes A, Alhazzani W, *et al*. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47:1181–247.
3 UK Sepsis Trust. Sepsis screening tool prehospital. Available: https://sepsistrust.org/wp-content/uploads/2020/08/Sepsis-Prehospital-12-231219.pdf [Accessed 14 Apr 2023].
4 Bion JBG, Boyle A, Carrol E, *et al*. *Academy of Medical Royal Colleges Statement on the Initial Antimicrobial Treatment of Sepsis*. Academy of Medical Royal Colleges, 2022.
5 Goodacre S, Thomas B, Smyth M, *et al*. Should prehospital early warning scores be used to identify which patients need urgent treatment for sepsis? *BMJ* 2021;375:2432.
6 Hamilton F, Arnold D, Baird A, *et al*. Early warning scores do not accurately predict mortality in sepsis: a meta-analysis and systematic review of the literature. *J Infect* 2018;76:241–8.
7 Lane D, Ichelson RI, Drennan IR, *et al*. Prehospital management and identification of sepsis by emergency medical services: a systematic review. *Emerg Med J* 2016;33:408–13.
8 Smyth MA, Brace-McDonnell SJ, Perkins GD. Identification of adults with sepsis in the prehospital environment: a systematic review. *BMJ Open* 2016;6:e011218.
9 Accuracy, impact and cost-effectiveness of prehospital clinical early warning scores for adults with suspected sepsis. Available: https://www.fundingawards.nihr.ac.uk/award/17/136/10 [Accessed 14 Apr 2023].
10 Royal College of Physicians, London. National early warning score (NEWS) 2. 2017. Available: http://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [Accessed 14 Apr 2023].
11 Seymour CW, Liu VX, Iwashyna TJ, *et al*. Assessment of clinical criteria for sepsis: for the third International consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016;315:762–74.
12 Wallgren UM, Castrén M, Svensson AEV, *et al*. Identification of adult septic patients in the prehospital setting: a comparison of two screening tools and clinical judgment. *Eur J Emerg Med* 2014;21:260–5.
13 Borrelli G, Koch E, Sterk E, *et al*. Early recognition of sepsis through emergency medical services pre-hospital screening. *Am J Emerg Med* 2019;37:1428–32.
14 Seymour CW. Prediction of critical illness during out-of-hospital emergency care. *JAMA* 2010;304:747.
15 Tam BHSc MD B, Xu BHSc M, Kwong BHSc Cand. M, *et al*. The admission Hamilton early warning score (HEWS) predicts the risk of critical event during hospitalization. *Can Journ Gen Int Med* 2017;11.
16 Jouffroy R, Saade A, Ellouze S, *et al*. Prehospital triage of septic patients at the SAMU regulation: comparison of qSOFA, MRST, MEWS and PRESEP scores. *Am J Emerg Med* 2018;36:820–4.
17 Association of Ambulance Chief Executives and the Royal College of Emergency Medicine. UK NHS ambulance services pre-alert guideline for the deteriorating adult patient. Available: https://rcem.ac.uk/wp-content/uploads/2021/10/NHS_Ambulance_Services_pre_alert_criteria_guideline.pdf [Accessed 14 Apr 2023].
18 Alam N, Oskam E, Stassen PM, *et al*. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. *Lancet Respir Med* 2018;6:40–50.
19 Scales DD. Paramedic initiated treatment of sepsis targeting out-of-hospital patients (PITSTOP). 2023. Available: clinicaltrials.gov
20 Walchok JG, Pirrallo RG, Furmanek D, *et al*. Paramedic-initiated CMS sepsis core measure bundle prior to hospital arrival: a stepwise approach. *Prehosp Emerg Care* 2017;21:291–300.
21 Bayer O, Schwarzkopf D, Stumme C, *et al*. An early warning scoring system to identify septic patients in the prehospital setting: the PRESEP score. *Acad Emerg Med* 2015;22:868–71.
22 Polito CC, Isakov A, Yancey AH, *et al*. Prehospital recognition of severe sepsis: development and validation of a novel EMS screening tool. *Am J Emerg Med* 2015;33:1119–25.
23 Baez AA, Cochon L. Acute care diagnostics collaboration: assessment of a Bayesian clinical decision model integrating the prehospital sepsis score and point-of-care lactate. *Am J Emerg Med* 2016;34:193–6.
24 Olsson T, Terent A, Lind L. Rapid emergency medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med* 2004;255:579–87.
25 Robson W, Nutbeam T, Daniels R. Sepsis: a need for prehospital intervention *Emerg Med J* 2009;26:535–8.
26 Guerra WF, Mayfield TR, Meyers MS, *et al*. Early detection and treatment of patients with severe sepsis by prehospital personnel. *J Emerg Med* 2013;44:1116–25.
27 Smyth MA, Gallacher D, Kimani PK, *et al*. Derivation and internal validation of the screening to enhance prehospital identification of sepsis (SEPSIS) score in adults on arrival at the emergency department. *Scand J Trauma Resusc Emerg Med* 2019;27:67.
28 Talmor D, Jones AE, Rubinson L, *et al*. Simple triage scoring system predicting death and the need for critical care resources for use during epidemics. *Crit Care Med* 2007;35:1251–6.
29 Suffoletto B, Frisch A, Prabhu A, *et al*. Prediction of serious infection during prehospital emergency care. *Prehosp Emerg Care* 2011;15:325–30.
30 Singer M, Deutschman CS, Seymour CW, *et al*. The third International consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016;315:801.
31 Rockwood K, Song X, MacKnight C, *et al*. A global clinical measure of fitness and frailty in elderly people. *CMAJ* 2005;173:489–95.
32 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
33 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
34 Hirst E, Irving A, Goodacre S. Patient and public involvement in emergency care research. *Emerg Med J* 2016;33:665–70.
35 Sabir L, Ramlakhan S, Goodacre S. Comparison of qSOFA and hospital early warning scores for prognosis in suspected sepsis in emergency department patients: a systematic review. *Emerg Med J* 2022;39:284–94.
36 Lane DJ, Wunsch H, Saskin R, *et al*. Screening strategies to identify sepsis in the Prehospital setting: a validation study. *CMAJ* 2020:E230–9.
37 Goodacre S. Using clinical risk models to predict outcomes: what are we predicting and why? *Emerg Med J* 2023;40:728–30.
38 Goodacre S, Fuller G, Conroy S, *et al*. Diagnosis and management of sepsis in the older adult. *BMJ* 2023;382:e075585.

Supplementary Appendix 1: Categorisation of diagnostic impressions

Category	Diagnostic impression
1	Sepsis
2	Cold & flu, Febrile illness, Meningitis, Pyrexia of unknown origin, Chest infection-pneumonia, UTI
3	Catheter problems, COPD, Convulsion/Fitting, Collapse-reason unknown, Confused/distressed/upset, Diarrhoea/Constipation, Dizzy/near faint/ loss of coordination, Shortness of Breath, Generally unwell, Haematuria, Headache, Hypotension, Other medical condition, Urinary Retention, Shock (hypovolemic), Transient Loss of Consciousness, Unconscious, Vomiting
4	Abdominal pain, Alcohol related, Allergic reaction/rash, Anaphylactic shock, AAA, Asthma, Bite/sting, Bleeding PR, Bleeding PV, Cardiac arrest, Cardiac STEMI, Cardiac chest pain (ACS), Cardiac NSTEMI, Choking, Carbon monoxide poisoning, Dental, Drug overdose, End of life care / Palliative, Epileptic fit, Epistaxis, Eye injury/eye problem, Falls, Gynaecological, Haematemesis, Haemoptysis, Haemorrhage/lacerations, Hyperglycaemia, Hypertension, Hypoglycaemia, Neurological problems, Pain – back non-traumatic, Pain – other, Panic attack, Poisoning, Pulmonary embolism, Rape/sexual assault, Pneumothorax (spontaneous), Renal problems/colic, Respiratory arrest, Seizures (non-EP), Smoke inhalation, Solvent related, Stroke – FAST positive, Unable to cope, No injury or illness, Cardiac Arrhythmia, Vascular Emergency (Non AAA), Dead on EMS arrival - signs inconsistent with life, Resuscitation unsuccessful
Excluded	Burns, Drowning, Electrocution, Fracture/possible fracture, Hanging, Head injury, Major trauma, Minor cuts & bruising, Minor injuries – other, Multiple injuries Obstetric – BBA, Obstetric – birth imminent, Obstetric – miscarriage, Obstetric – normal labour, Obstetric – premature labour, Obstetric emergency (other), Psychiatric problems, Spinal injury, Sprain/strain/dislocation, Stabbed/shot/weapon wound, Wound Closure, Non accidental injury

Supplementary Appendix 2: Details of each early warning score

90-30-90

Dichotomous assessment, positive if any of the following criteria are met:

1. Systolic BP < 90mmHg
2. Respiratory rate > 30/minute
3. Oxygen saturation < 90%

Modification:
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 90% on air (i.e. the criteria is positive).*

Missing data:
Assume any missing criterion is negative/normal.

The Borelli strategy

Dichotomous assessment, positive if three or more criteria are met:

- Respiratory rate > 20/minute
- Heart rate > 90/minute
- Systolic BP < 90 mmHg
- Documented fever or temperature >38.3°C or <36°C
- New onset of mental status change
- O2 saturation < 90%
- Suspected infection

Modification:
Documented fever or temperature >38.3°C or <36°C, is effectively just temperature >38.3°C or <36°C
New onset of mental status change assumed if the GCS verbal scale is <5. If the GCS verbal scale is missing, then mental status change is assumed if GCS<15 or AVPU<A.
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 90% on air (i.e. the criteria is positive).

Missing data:
Assume any missing variable is negative

Critical illness score (CIS)

Score	0	1	2
Respiratory Rate	12-23	<12 or 24-35	>35
Heart Rate	<120	≥120	
Systolic BP	>90	≤90	
Age	<45	≥45	
SpO2	≥88	<88	
GCS	15	8-14	<8

Thresholds of >4 or >0 are suggested, depending upon whether specificity or sensitivity are to be optimised

Modification:
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 88% on air (i.e. scores 1 point).

Missing data:
Assume any missing variable scores zero.

Hamilton Early Warning Score (HEWS)

Score	3	2	1	0	1	2	3
Respiratory Rate	<8	8-13		14-20		21-30	>30
Oxygen saturation	<85		85-91	>91			
Heart Rate		<40	40-50	51-100	101-110	111-130	>130
Systolic BP	<70	71-90		91-170		171-200	>200
Temperature	<35.0		35.1-36.4	36.5-38.0	38.1-39.0	>39.0	
Neurology				Alert	Voice	Pain	Unresponsive
Air or Oxygen				Air	≤ 5 L/min or ≤ 50% by mask		>5 L/min or >50% by mask

*CAM positive removed as not routinely recorded

Threshold > 4

Modification:
If AVPU is missing, infer from GCS.
If on oxygen but amount unknown, score 2 points

Missing data:
Assume any missing variable scores zero

Modified Early Warning Score (MEWS)

MEWS has five parameters, each of which are scored from zero to two or three providing an overall score between zero and 14.

Score	3	2	1	0	1	2	3
Respiratory Rate		<9		9-14	15-20	21-29	≥30
Heart Rate		≤40	41-50	51-100	101-110	111-129	≥130
Systolic BP	≤70	71-80	81-100	101-199		≥200	
Temperature		<35.0		35.0-38.4		≥38.5	
AVPU				Alert	Voice	Pain	Unresponsive

A threshold of 5 or more has been shown to be associated with an increased risk of death

Modification:
If AVPU missing, infer AVPU from GCS

Missing data:
Assume any missing variable scores zero.

National Early Warning Score, version 2 (NEWS2)

The NEWS2 has seven parameters, each of which are scored from zero to three providing an overall score between zero and 20.

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			

We will not use the scale for patients with confirmed hypercapnic respiratory failure.

Modification:
If AVPU is missing, infer AVPU from GCS

Missing data:
Assume any missing variable scores zero.

NHS pre-alert

Pre-alert if any of the following are present:

- Respiratory rate ≤8 or ≥25
- O2 saturations on oxygen <92% (Patients usually running normal oxygen saturations) <84% (Patients with chronic hypercapnic respiratory failure)
- Systolic <90mmHg OR downward-trending systolic where symptomatic
- Tachycardia ≥131
- GCS motor <4

Sepsis red flag criteria evaluated as part of UK Sepsis Trust criteria

Modification:
Drop <84% oxygen saturation threshold for patients with chronic hypercapnic respiratory failure
Drop downward-trending systolic where symptomatic

Missing data:
Assume any missing criterion is negative

Prehospital ANTibiotics Against Sepsis (PHANTASi)

Dichotomous assessment, positive if both the following criteria are met:

1. Temperature >38 °C or < 36 °C
2. Heart rate > 90 beats per minute or respiratory rate > 20 per minutes

Modification:
None required

Missing data:
Assume any missing variable is negative/normal.

Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients clinical trial (PITSTOP)

Dichotomous assessment, positive if all the following three criteria are met:

1. Paramedic suspects possible infection
2. Temperature ≥ 38.0 °C
3. Systolic BP < 100mmHg

Modification:
None required

Missing data:
Assume any missing criterion is negative/normal.

Prehospital Sepsis Assessment Tool (PreSAT)

Dichotomous assessment, positive if both the following criteria are met:

1. Presentation suggestive of infection
2. Any two from (a) temperature >38°C or <36°C, (b) heart rate > 90/min, (c) respiratory rate > 20/min, (d) systolic BP < 90mmHg

Modification:
None required

Missing data:
Assume any missing variable is negative/normal.

Prehospital Early Sepsis Detection (PRESEP)

Parameter	Score
Temperature > 38°C	4
Temperature < 36°C	1
SaO2 < 92%	2
RR > 22 breaths/min	1
HR > 90 beats/min	2
BP < 90 mm Hg	2

Recommended threshold >3

Modification:
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 92% on air (i.e. score 2 points).

Missing data:
Assume any missing variable scores zero.

Prehospital Severe Sepsis (PRESS)

The score is only applied to patients meeting all three of the following criteria, so patients not meeting these criteria should score zero:

- Heart rate >90/min
- Respiratory rate >20/min
- Systolic blood pressure BP <110 mmHg

Score	0	1	2	3	4	5
Age	<40		≥60		40-59	
SpO2	≥90	80-89		70-79	60-69	<60
Systolic BP	100-109	90-99	80-89	70-79	60-69	<60
Hot tactile temperature				X		
ED chief concern: sick person				X		
Nursing home transport					X	

Threshold > 1

Modification:
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 90% on air and scored 2 points.
Infer hot tactile temperature from recorded temperature > 38°C
Drop ED chief concern sick person – address this through diagnostic impression
Drop nursing home transport

Missing data:
Assume any missing variable scores zero

Prehospital Sepsis Project (PSP)

Parameter	Score
Temperature > 38°C	1
Heart rate / systolic BP ≥ 0.7	2
Respiratory rate > 22/min	1

Low risk = 0-1 point, moderate risk = 2 points, high risk = 3-4 points

Modification:
None required

Missing data:
Assume missing temperature or respiratory rate scores zero.
Assume heart rate / systolic BP scores zero unless either (a) heart rate is > 100 and systolic BP is missing, or (b) systolic BP is <100 and heart rate is missing.

qSOFA

Parameter	Score
GCS <15	1
Respiratory rate ≥22	1
Systolic BP ≤100	1

Total score 0-3
Low risk = 0 or 1
High risk = 2 or 3

Modification:
None required

Missing data:
Assume any missing variable scores zero

Rapid Emergency Medicine Score (REMS)

Score	0	1	2	3	4	5	6
Age	<45		45-54	55-64		65-74	>65
MAP	70-109		50-69 or 110-129	130-159	>159 or <50		
Heart rate	70-109		55-69 or 110-139	40-54 or 140-179	<179 or <40		
Respiratory rate	12-24	10-11 or 25-34	6-9	35-49	<6 or >49		
SpO2	>89%	86-89%		75-85%	<75%		
GCS	>13	11-13	8-10	5-7	3-4		

High risk (REMS ≥3): patient may need aggressive treatment
Low risk (REMS <3): patient may be appropriate to triage for routine treatment

Modification:
If the oxygen saturation is measured on supplemental oxygen, it is assumed to be < 89% on air and scored 2 points.

Missing data:
Assume any missing variable scores zero

Robson Screening Tool (RST)

Dichotomous assessment, positive if presentation suggestive of infection and any two of:

1. Temperature > 38.3° or < 36°
2. Heart rate > 90 beats/min
3. Respiratory rate > 20 breaths/min
4. Acutely altered mental status
5. Plasma glucose > 6.6 mmol/l (unless diabetic)

Modification:
Criterion 5 applies regardless of whether they are diabetic
Acutely altered mental status change assumed if the GCS verbal scale is <5. If the GCS verbal scale is missing, then acutely altered mental status is assumed if GCS<15 or AVPU=A.

Missing data:
Assume any missing criterion is negative

Screening to Enhance Prehospital Identification of Sepsis (SEPSIS)

Parameter	-1	0	1	2
Age		≤60	>60	
Respiratory rate		≤20 or >60	21-40	40-60
SpO2		≥94	<94	
Heart rate		≤100	101-140	141-160
Systolic BP	>160	<60 or 100-160	60-99	
GCS		13-15	3-12	
Temperature		<37.5	37.5 to 39.5	>39.5
Skin			Jaundice, pallor, mottling	

High risk if score > 4

Modification:
Skin features dropped from the score (not recorded on ePFR)

Missing data:
Assume any missing variable scores zero
Give 1 point if SpO2>94% on oxygen

Sepsis Alert

At least two systematic inflammatory response syndrome (SIRS) criteria:

- Temperature > 38C or < 36C
- Pulse > 90 beats/min
- Respiratory rate > 20 breaths/min or mechanically ventilated

And

Suspected or documented infection

And

Hypoperfusion, as manifested by one of the following:

- Systolic BP < 90 mm Hg
- Mean arterial pressure < 65 mm Hg
- Lactate level ≥ 4 mmol/L

Modification:
Drop lactate level ≥ 4 mmol/L

Missing data:
Assume negative if missing

Simple Triage Scoring System (STSS)

Parameter	Score
Age >65 yrs	1
Altered mental status	1
Respiratory rate of >30 breaths/min	1
Low oxygen saturation	1
Shock index of >1 (heart rate > systolic BP)	1

Modification:
Altered mental status change assumed if the GCS verbal scale is <5. If the GCS verbal scale is missing, then altered mental status is assumed if GCS<15 or AVPU<A.
Score 1 point for low oxygen saturation if oxygen saturation is <94% or measured on supplemental oxygen.

Missing data:
Assume any missing variable scores zero

Suffoletto strategy

Dichotomous assessment, positive if any of the following criteria are met:

- Systolic BP <100 mmHg
- History or suspicion of fever
- Prehospital judgment of infection

Modification:
Infer history or suspicion of fever from temperature>38

Missing data:
Assume any missing criterion is negative

UK Sepsis Trust Red Flags

Any of the following gives a positive score:

- Objective evidence of new or altered mental state
- Systolic BP ≤ 90 mmHg (or drop of >40 from normal)
- Heart rate ≥ 130 per minute
- Respiratory rate ≥ 25 per minute
- Needs O2 to keep SpO2 ≥ 92% (88% in COPD)
- Non-blanching rash / mottled / ashen / cyanotic
- Lactate ≥ 2 mmol/l
- Recent chemotherapy
- Not passed urine in 18 hours

Modification:
New or altered mental state assumed if the GCS verbal scale is <5. If the GCS verbal scale is missing, then new or altered mental state is assumed if GCS<15 or AVPU<A.
Drop BP change from normal and just use systolic BP ≤ 90 mmHg
Simplify O2 criteria to SpO2 < 92% or measured on supplemental oxygen
Drop non-blanching rash / mottled / ashen / cyanotic
Drop lactate ≥ 2 mmol/l
Drop recent chemotherapy
Drop not passed urine

Missing data:
Assume any missing criterion is negative.

Supplementary tables

Supplementary Table 1: Early warning scores and constituent variables

Early Warning Score	Age	Temperature	Heart rate	Respiratory rate	Oxygen saturation	Conscious level	Blood pressure	Other
NEWS2 [10]		X	X	X	X	X	X	Inspired oxygen
qSOFA [11]				X		X	X	
90-30-90 [12]				X	X		X	
Borelli [13]		X	X	X	X	X	X	Suspected infection
CIS [14]	X		X	X	X	X	X	
HEWS [15]		X	X	X	X	X	X	Inspired oxygen
MEWS [16]		X	X	X		X	X	
NHS pre-alert [17]			X	X	X	X	X	
PHANTASi [18]		X	X	X				
PITSTOP [19]		X					X	Paramedic suspicion of infection
PreSAT [20]		X	X	X			X	
PRESEP [21]		X	X	X	X		X	
PRESS [22]	X	X			X		X	Dispatch chief complaint of sick person; nursing home resident
PSP [23]		X	X	X			X	
REMS [24]	X		X	X	X	X	X	
RST [25]		X	X	X		X		Blood glucose
Sepsis Alert [26]		X	X	X			X	Suspected or documented infection, hypoperfusion
SEPSIS [27]	X	X	X	X	X	X	X	Skin appearance
STSS [28]	X		X	X	X	X	X	
Suffoletto [29]		X					X	
UKST red flag* [3]			X	X	X	X	X	Skin appearance

*Excluding lactate, oliguria and recent chemotherapy

National Early Warning Score, version 2 (NEWS2); Quick Sequential Organ Failure Assessment (qSOFA); Critical illness score (CIS); Hamilton Early Warning Score (HEWS); Modified Early Warning Score (MEWS); Prehospital ANTibiotics Against Sepsis (PHANTASi); Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients clinical trial (PITSTOP); Prehospital Sepsis Assessment Tool (PreSAT); Prehospital Early Sepsis Detection (PRESEP); Prehospital Severe Sepsis (PRESS); Prehospital Sepsis Project (PSP); Rapid Emergency Medicine Score (REMS): Robson Screening Tool (RST); Screening to Enhance Prehospital Identification of Sepsis (SEPSIS); Simple Triage Scoring System (STSS); United Kingdom Sepsis Trust (UKST)

Supplementary Table 2: Agreement between expert doctors during reference standard adjudication

Assessment	Doctor 1	Doctor 2	Consensus	Kappa (95% CI)
Evidence of infection	86.0%	87.6%	84.7%	0.62 (0.53, 0.71)
SOFA score 2+ worse than normal	60.2%	61.1%	60.0%	0.87 (0.83, 0.91)
Patient meets sepsis-3 criteria	56.0%	55.0%	56.2%	0.89 (0.85, 0.92)
Treatment for sepsis given	52.5%	51.5%	53.3%	0.87 (0.83, 0.91)

Supplementary Table 3: Accuracy of categorised diagnostic impression for the primary reference standard

Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
Sepsis	12870	114	286	234	12236	0.328 (0.28, 0.379)	0.977 (0.974, 0.98)	0.285 (0.243, 0.331)	0.981 (0.979, 0.983)
Sepsis or infection	12870	199	1080	149	11442	0.572 (0.519, 0.623)	0.914 (0.909, 0.919)	0.156 (0.137, 0.176)	0.987 (0.985, 0.989)
Sepsis, infection or nonspecific impression	12870	312	5576	36	6946	0.897 (0.86, 0.924)	0.555 (0.546, 0.563)	0.053 (0.048, 0.059)	0.995 (0.993, 0.996)

Supplementary Table 4: Area under ROC curve for each early warning score for the primary reference standard

Early warning score	Sepsis	Sepsis or infection	Sepsis, infection or nonspecific impression	All diagnostic impressions
NEWS2	0.655 (0.63, 0.68)	0.756 (0.729, 0.783)	0.858 (0.836, 0.88)	0.877 (0.86, 0.895)
qSOFA	0.645 (0.62, 0.669)	0.734 (0.707, 0.761)	0.809 (0.785, 0.834)	0.801 (0.778, 0.824)
90-30-90	0.624 (0.601, 0.648)	0.686 (0.66, 0.712)	0.743 (0.717, 0.769)	0.742 (0.717, 0.768)
Borelli	0.639 (0.615, 0.663)	0.712 (0.686, 0.738)	0.781 (0.755, 0.806)	0.788 (0.764, 0.813)
CIS	0.654 (0.629, 0.679)	0.755 (0.728, 0.782)	0.845 (0.822, 0.867)	0.838 (0.817, 0.859)
HEWS	0.654 (0.629, 0.679)	0.751 (0.724, 0.778)	0.841 (0.818, 0.863)	0.837 (0.816, 0.858)
MEWS	0.654 (0.629, 0.679)	0.753 (0.726, 0.78)	0.851 (0.828, 0.873)	0.857 (0.837, 0.876)
NHS pre-alert	0.624 (0.601, 0.648)	0.696 (0.67, 0.722)	0.751 (0.725, 0.776)	0.747 (0.723, 0.772)
PHANTASi	0.626 (0.602, 0.649)	0.708 (0.682, 0.735)	0.745 (0.719, 0.771)	0.741 (0.716, 0.767)
PITSTOP	0.534 (0.52, 0.547)	0.545 (0.53, 0.56)	0.549 (0.533, 0.564)	0.554 (0.537, 0.57)
PreSAT	0.648 (0.624, 0.673)	0.734 (0.708, 0.761)	0.789 (0.766, 0.813)	0.775 (0.754, 0.797)
PRESEP	0.653 (0.628, 0.678)	0.75 (0.723, 0.777)	0.847 (0.824, 0.87)	0.856 (0.834, 0.878)
PRESS	0.551 (0.534, 0.567)	0.557 (0.54, 0.574)	0.579 (0.559, 0.599)	0.587 (0.566, 0.608)
PSP	0.654 (0.629, 0.679)	0.754 (0.728, 0.781)	0.84 (0.818, 0.862)	0.832 (0.811, 0.852)
REMS	0.654 (0.629, 0.679)	0.752 (0.725, 0.779)	0.814 (0.792, 0.836)	0.757 (0.732, 0.781)
RST	0.648 (0.623, 0.672)	0.737 (0.711, 0.764)	0.782 (0.761, 0.803)	0.73 (0.712, 0.747)
Sepsis alert	0.537 (0.523, 0.551)	0.542 (0.528, 0.557)	0.558 (0.541, 0.576)	0.563 (0.545, 0.581)
SEPSIS	0.654 (0.629, 0.679)	0.755 (0.727, 0.782)	0.862 (0.84, 0.884)	0.882 (0.865, 0.899)
STSS	0.652 (0.627, 0.677)	0.749 (0.722, 0.776)	0.837 (0.814, 0.861)	0.831 (0.809, 0.854)
Suffoletto	0.64 (0.616, 0.664)	0.728 (0.701, 0.754)	0.799 (0.775, 0.823)	0.801 (0.778, 0.824)
UKST red flag	0.648 (0.623, 0.672)	0.733 (0.707, 0.76)	0.788 (0.766, 0.809)	0.756 (0.737, 0.775)

See supplementary table 1 for details of the early warning scores

National Early Warning Score, version 2 (NEWS2); Quick Sequential Organ Failure Assessment (qSOFA); Critical illness score (CIS); Hamilton Early Warning Score (HEWS); Modified Early Warning Score (MEWS); Prehospital ANTibiotics Against Sepsis (PHANTASi); Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients clinical trial (PITSTOP); Prehospital Sepsis Assessment Tool (PreSAT); Prehospital Early Sepsis Detection (PRESEP); Prehospital Severe Sepsis (PRESS); Prehospital Sepsis Project (PSP); Rapid Emergency Medicine Score (REMS): Robson Screening Tool (RST); Screening to Enhance Prehospital Identification of Sepsis (SEPSIS); Simple Triage Scoring System (STSS); United Kingdom Sepsis Trust (UKST)

Supplementary Table 5: Accuracy of NEWS2 applied only to presentations with a diagnostic impression of sepsis

Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
0	12859	114	282	233	12230	0.329 (0.281, 0.38)	0.977 (0.975, 0.98)	0.288 (0.245, 0.334)	0.981 (0.979, 0.984)
1	12859	114	278	233	12234	0.329 (0.281, 0.38)	0.978 (0.975, 0.98)	0.291 (0.248, 0.338)	0.981 (0.979, 0.984)
2	12859	114	273	233	12239	0.329 (0.281, 0.38)	0.978 (0.975, 0.981)	0.295 (0.251, 0.342)	0.981 (0.979, 0.984)
3	12859	113	262	234	12250	0.326 (0.278, 0.377)	0.979 (0.976, 0.981)	0.301 (0.257, 0.35)	0.981 (0.979, 0.983)
4	12859	111	245	236	12267	0.32 (0.273, 0.371)	0.98 (0.978, 0.983)	0.312 (0.266, 0.362)	0.981 (0.979, 0.983)
5	12859	108	228	239	12284	0.311 (0.265, 0.362)	0.982 (0.979, 0.984)	0.321 (0.274, 0.373)	0.981 (0.978, 0.983)
6	12859	105	190	242	12322	0.303 (0.257, 0.353)	0.985 (0.983, 0.987)	0.356 (0.303, 0.412)	0.981 (0.978, 0.983)
7	12859	95	157	252	12355	0.274 (0.23, 0.323)	0.987 (0.985, 0.989)	0.377 (0.319, 0.438)	0.98 (0.977, 0.982)
8	12859	86	129	261	12383	0.248 (0.205, 0.296)	0.99 (0.988, 0.991)	0.4 (0.337, 0.467)	0.979 (0.977, 0.982)
9	12859	73	94	274	12418	0.21 (0.171, 0.256)	0.992 (0.991, 0.994)	0.437 (0.364, 0.513)	0.978 (0.976, 0.981)
10	12859	59	65	288	12447	0.17 (0.134, 0.213)	0.995 (0.993, 0.996)	0.476 (0.39, 0.563)	0.977 (0.975, 0.98)
11	12859	40	36	307	12476	0.115 (0.086, 0.153)	0.997 (0.996, 0.998)	0.526 (0.416, 0.635)	0.976 (0.973, 0.978)
12	12859	25	21	322	12491	0.072 (0.049, 0.104)	0.998 (0.997, 0.999)	0.543 (0.402, 0.678)	0.975 (0.972, 0.977)
13	12859	8	12	339	12500	0.023 (0.012, 0.045)	0.999 (0.998, 0.999)	0.4 (0.219, 0.613)	0.974 (0.971, 0.976)
14	12859	4	7	343	12505	0.012 (0.004, 0.029)	0.999 (0.999, 1)	0.364 (0.152, 0.646)	0.973 (0.97, 0.976)
15	12859	0	3	347	12509	0 (0, 0.011)	1 (0.999, 1)	0 (0, 0.561)	0.973 (0.97, 0.976)
16	12859	0	0	347	12512	0 (0, 0.011)	1 (1, 1)	-	0.973 (0.97, 0.976)

Supplementary Table 6: Accuracy of qSOFA and other tools applied only to presentations with a diagnostic impression of sepsis

EWS	Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
Pre-alert	0	12870	45	110	303	12412	0.129 (0.098, 0.169)	0.991 (0.989, 0.993)	0.29 (0.225, 0.366)	0.976 (0.973, 0.979)
qSOFA	0	12869	107	249	241	12272	0.307 (0.261, 0.358)	0.98 (0.978, 0.982)	0.301 (0.255, 0.35)	0.981 (0.978, 0.983)
qSOFA	1	12869	72	103	276	12418	0.207 (0.168, 0.253)	0.992 (0.99, 0.993)	0.411 (0.341, 0.485)	0.978 (0.976, 0.981)
qSOFA	2	12869	19	21	329	12500	0.055 (0.035, 0.084)	0.998 (0.997, 0.999)	0.475 (0.329, 0.625)	0.974 (0.971, 0.977)
90-30-90	0	12857	91	169	256	12341	0.262 (0.219, 0.311)	0.986 (0.984, 0.988)	0.35 (0.295, 0.41)	0.98 (0.977, 0.982)
Borelli	0	12835	102	203	245	12285	0.294 (0.248, 0.344)	0.984 (0.981, 0.986)	0.334 (0.284, 0.389)	0.98 (0.978, 0.983)
CIS	0	12855	114	283	233	12225	0.329 (0.281, 0.38)	0.977 (0.975, 0.98)	0.287 (0.245, 0.334)	0.981 (0.979, 0.984)
CIS	4	12855	32	42	315	12466	0.092 (0.066, 0.127)	0.997 (0.995, 0.998)	0.432 (0.326, 0.546)	0.975 (0.973, 0.978)
HEWS	4	12835	101	209	246	12279	0.291 (0.246, 0.341)	0.983 (0.981, 0.985)	0.326 (0.276, 0.38)	0.98 (0.978, 0.983)
MEWS	4	12859	88	189	259	12323	0.254 (0.211, 0.302)	0.985 (0.983, 0.987)	0.318 (0.266, 0.375)	0.979 (0.977, 0.982)
NHS	0	12855	92	204	255	12304	0.265 (0.221, 0.314)	0.984 (0.981, 0.986)	0.311 (0.261, 0.366)	0.98 (0.977, 0.982)
PHANTASi	0	12858	93	210	254	12301	0.268 (0.224, 0.317)	0.983 (0.981, 0.985)	0.307 (0.258, 0.361)	0.98 (0.977, 0.982)
PITSTOP	0	12813	24	20	322	12447	0.069 (0.047, 0.101)	0.998 (0.998, 0.999)	0.545 (0.401, 0.683)	0.975 (0.972, 0.977)
PreSAT	0	12835	110	250	237	12238	0.317 (0.27, 0.368)	0.98 (0.977, 0.982)	0.306 (0.26, 0.355)	0.981 (0.978, 0.983)
PRESEP	3	12835	111	242	236	12246	0.32 (0.273, 0.371)	0.981 (0.978, 0.983)	0.314 (0.268, 0.365)	0.981 (0.979, 0.983)
PRESS	1	12835	35	33	312	12455	0.101 (0.073, 0.137)	0.997 (0.996, 0.998)	0.515 (0.398, 0.629)	0.976 (0.973, 0.978)
PSP	1	12835	111	258	236	12230	0.32 (0.273, 0.371)	0.979 (0.977, 0.982)	0.301 (0.256, 0.349)	0.981 (0.979, 0.983)
REMS	2	12855	114	269	233	12239	0.329 (0.281, 0.38)	0.978 (0.976, 0.981)	0.298 (0.254, 0.345)	0.981 (0.979, 0.984)
RST	0	12857	110	264	237	12246	0.317 (0.27, 0.368)	0.979 (0.976, 0.981)	0.294 (0.25, 0.342)	0.981 (0.978, 0.983)
SAS	0	12836	26	14	321	12475	0.075 (0.052, 0.108)	0.999 (0.998, 0.999)	0.65 (0.495, 0.779)	0.975 (0.972, 0.977)
SEPSIS	4	12856	78	123	269	12386	0.225 (0.184, 0.272)	0.99 (0.988, 0.992)	0.388 (0.323, 0.457)	0.979 (0.976, 0.981)
STSS	1	12855	104	211	243	12297	0.3 (0.254, 0.35)	0.983 (0.981, 0.985)	0.33 (0.281, 0.384)	0.981 (0.978, 0.983)
Suffoletto	0	12813	103	219	243	12248	0.298 (0.252, 0.348)	0.982 (0.98, 0.985)	0.32 (0.271, 0.373)	0.981 (0.978, 0.983)
UK	0	12855	109	237	238	12271	0.314 (0.268, 0.365)	0.981 (0.979, 0.983)	0.315 (0.268, 0.366)	0.981 (0.978, 0.983)

Supplementary Table 7: Accuracy of NEWS2 applied only to presentations with a diagnostic impression of sepsis or infection

Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
0	12859	198	1032	149	11480	0.571 (0.518, 0.622)	0.918 (0.913, 0.922)	0.161 (0.141, 0.183)	0.987 (0.985, 0.989)
1	12859	197	967	150	11545	0.568 (0.515, 0.619)	0.923 (0.918, 0.927)	0.169 (0.149, 0.192)	0.987 (0.985, 0.989)
2	12859	197	889	150	11623	0.568 (0.515, 0.619)	0.929 (0.924, 0.933)	0.181 (0.16, 0.205)	0.987 (0.985, 0.989)
3	12859	191	776	156	11736	0.55 (0.498, 0.602)	0.938 (0.934, 0.942)	0.198 (0.174, 0.224)	0.987 (0.985, 0.989)
4	12859	181	658	166	11854	0.522 (0.469, 0.574)	0.947 (0.943, 0.951)	0.216 (0.189, 0.245)	0.986 (0.984, 0.988)
5	12859	171	540	176	11972	0.493 (0.441, 0.545)	0.957 (0.953, 0.96)	0.241 (0.211, 0.273)	0.986 (0.983, 0.987)
6	12859	155	410	192	12102	0.447 (0.395, 0.499)	0.967 (0.964, 0.97)	0.274 (0.239, 0.313)	0.984 (0.982, 0.986)
7	12859	133	314	214	12198	0.383 (0.334, 0.435)	0.975 (0.972, 0.978)	0.298 (0.257, 0.342)	0.983 (0.98, 0.985)
8	12859	109	218	238	12294	0.314 (0.268, 0.365)	0.983 (0.98, 0.985)	0.333 (0.284, 0.386)	0.981 (0.978, 0.983)
9	12859	91	146	256	12366	0.262 (0.219, 0.311)	0.988 (0.986, 0.99)	0.384 (0.324, 0.447)	0.98 (0.977, 0.982)
10	12859	68	91	279	12421	0.196 (0.158, 0.241)	0.993 (0.991, 0.994)	0.428 (0.353, 0.505)	0.978 (0.975, 0.98)
11	12859	47	48	300	12464	0.135 (0.103, 0.175)	0.996 (0.995, 0.997)	0.495 (0.396, 0.594)	0.976 (0.974, 0.979)
12	12859	29	28	318	12484	0.084 (0.059, 0.117)	0.998 (0.997, 0.998)	0.509 (0.383, 0.634)	0.975 (0.972, 0.978)
13	12859	11	17	336	12495	0.032 (0.018, 0.056)	0.999 (0.998, 0.999)	0.393 (0.236, 0.576)	0.974 (0.971, 0.976)
14	12859	5	8	342	12504	0.014 (0.006, 0.033)	0.999 (0.999, 1)	0.385 (0.177, 0.645)	0.973 (0.97, 0.976)
15	12859	0	3	347	12509	0 (0, 0.011)	1 (0.999, 1)	0 (0, 0.561)	0.973 (0.97, 0.976)
16	12859	0	0	347	12512	0 (0, 0.011)	1 (1, 1)	-	0.973 (0.97, 0.976)

Supplementary Table 8: Accuracy of qSOFA and other tools applied only to presentations with a diagnostic impression of sepsis or infection

EWS	Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
Pre-alert	0	12870	52	133	296	12389	0.149 (0.116, 0.191)	0.989 (0.987, 0.991)	0.281 (0.221, 0.35)	0.977 (0.974, 0.979)
qSOFA	0	12869	180	758	168	11763	0.517 (0.465, 0.569)	0.939 (0.935, 0.944)	0.192 (0.168, 0.218)	0.986 (0.984, 0.988)
qSOFA	1	12869	106	192	242	12329	0.305 (0.259, 0.355)	0.985 (0.982, 0.987)	0.356 (0.304, 0.412)	0.981 (0.978, 0.983)
qSOFA	2	12869	23	22	325	12499	0.066 (0.044, 0.097)	0.998 (0.997, 0.999)	0.511 (0.37, 0.65)	0.975 (0.972, 0.977)
90-30-90	0	12857	140	387	207	12123	0.403 (0.353, 0.456)	0.969 (0.966, 0.972)	0.266 (0.23, 0.305)	0.983 (0.981, 0.985)
Borelli	0	12835	160	463	187	12025	0.461 (0.409, 0.514)	0.963 (0.959, 0.966)	0.257 (0.224, 0.293)	0.985 (0.982, 0.987)
CIS	0	12855	198	1039	149	11469	0.571 (0.518, 0.622)	0.917 (0.912, 0.922)	0.16 (0.141, 0.182)	0.987 (0.985, 0.989)
CIS	4	12855	38	54	309	12454	0.11 (0.081, 0.147)	0.996 (0.994, 0.997)	0.413 (0.318, 0.515)	0.976 (0.973, 0.978)
HEWS	4	12835	154	477	193	12011	0.444 (0.392, 0.496)	0.962 (0.958, 0.965)	0.244 (0.212, 0.279)	0.984 (0.982, 0.986)
MEWS	4	12859	135	412	212	12100	0.389 (0.339, 0.441)	0.967 (0.964, 0.97)	0.247 (0.213, 0.285)	0.983 (0.98, 0.985)
NHS	0	12855	149	471	198	12037	0.429 (0.378, 0.482)	0.962 (0.959, 0.966)	0.24 (0.208, 0.275)	0.984 (0.981, 0.986)
PHANTASi	0	12858	161	588	186	11923	0.464 (0.412, 0.517)	0.953 (0.949, 0.957)	0.215 (0.187, 0.246)	0.985 (0.982, 0.987)
PITSTOP	0	12813	32	28	314	12439	0.092 (0.066, 0.128)	0.998 (0.997, 0.998)	0.533 (0.409, 0.654)	0.975 (0.973, 0.978)
PreSAT	0	12835	183	732	164	11756	0.527 (0.475, 0.579)	0.941 (0.937, 0.945)	0.2 (0.175, 0.227)	0.986 (0.984, 0.988)
PRESEP	3	12835	183	738	164	11750	0.527 (0.475, 0.579)	0.941 (0.937, 0.945)	0.199 (0.174, 0.226)	0.986 (0.984, 0.988)
PRESS	1	12835	40	61	307	12427	0.115 (0.086, 0.153)	0.995 (0.994, 0.996)	0.396 (0.306, 0.494)	0.976 (0.973, 0.978)
PSP	1	12835	188	723	159	11765	0.542 (0.489, 0.593)	0.942 (0.938, 0.946)	0.206 (0.181, 0.234)	0.987 (0.984, 0.989)
REMS	2	12855	197	987	150	11521	0.568 (0.515, 0.619)	0.921 (0.916, 0.926)	0.166 (0.146, 0.189)	0.987 (0.985, 0.989)
RST	0	12857	188	836	159	11674	0.542 (0.489, 0.593)	0.933 (0.929, 0.937)	0.184 (0.161, 0.208)	0.987 (0.984, 0.988)
SAS	0	12836	30	22	317	12467	0.086 (0.061, 0.121)	0.998 (0.997, 0.999)	0.577 (0.442, 0.701)	0.975 (0.972, 0.978)
SEPSIS	4	12856	107	216	240	12293	0.308 (0.262, 0.359)	0.983 (0.98, 0.985)	0.331 (0.282, 0.384)	0.981 (0.978, 0.983)
STSS	1	12855	174	607	173	11901	0.501 (0.449, 0.554)	0.951 (0.948, 0.955)	0.223 (0.195, 0.253)	0.986 (0.983, 0.988)
Suffoletto	0	12813	176	669	170	11798	0.509 (0.456, 0.561)	0.946 (0.942, 0.95)	0.208 (0.182, 0.237)	0.986 (0.984, 0.988)
UK	0	12855	181	686	166	11822	0.522 (0.469, 0.574)	0.945 (0.941, 0.949)	0.209 (0.183, 0.237)	0.986 (0.984, 0.988)

Supplementary Table 9: Accuracy of NEWS2 applied only to presentations with a diagnostic impression of sepsis, infection, or nonspecific presentation

Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
0	12859	307	4638	40	7874	0.885 (0.847, 0.914)	0.629 (0.621, 0.638)	0.062 (0.056, 0.069)	0.995 (0.993, 0.996)
1	12859	302	3833	45	8679	0.87 (0.831, 0.902)	0.694 (0.686, 0.702)	0.073 (0.065, 0.081)	0.995 (0.993, 0.996)
2	12859	297	3248	50	9264	0.856 (0.815, 0.889)	0.74 (0.733, 0.748)	0.084 (0.075, 0.093)	0.995 (0.993, 0.996)
3	12859	287	2598	60	9914	0.827 (0.784, 0.863)	0.792 (0.785, 0.799)	0.099 (0.089, 0.111)	0.994 (0.992, 0.995)
4	12859	270	2048	77	10464	0.778 (0.731, 0.819)	0.836 (0.83, 0.843)	0.116 (0.104, 0.13)	0.993 (0.991, 0.994)
5	12859	252	1612	95	10900	0.726 (0.677, 0.77)	0.871 (0.865, 0.877)	0.135 (0.12, 0.151)	0.991 (0.989, 0.993)
6	12859	220	1152	127	11360	0.634 (0.582, 0.683)	0.908 (0.903, 0.913)	0.16 (0.142, 0.181)	0.989 (0.987, 0.991)
7	12859	181	823	166	11689	0.522 (0.469, 0.574)	0.934 (0.93, 0.938)	0.18 (0.158, 0.205)	0.986 (0.984, 0.988)
8	12859	147	513	200	11999	0.424 (0.373, 0.476)	0.959 (0.955, 0.962)	0.223 (0.193, 0.256)	0.984 (0.981, 0.986)
9	12859	111	311	236	12201	0.32 (0.273, 0.371)	0.975 (0.972, 0.978)	0.263 (0.223, 0.307)	0.981 (0.978, 0.983)
10	12859	82	181	265	12331	0.236 (0.195, 0.284)	0.986 (0.983, 0.987)	0.312 (0.259, 0.37)	0.979 (0.976, 0.981)
11	12859	55	90	292	12422	0.159 (0.124, 0.201)	0.993 (0.991, 0.994)	0.379 (0.304, 0.46)	0.977 (0.974, 0.979)
12	12859	34	47	313	12465	0.098 (0.071, 0.134)	0.996 (0.995, 0.997)	0.42 (0.318, 0.528)	0.976 (0.973, 0.978)
13	12859	12	28	335	12484	0.035 (0.02, 0.059)	0.998 (0.997, 0.998)	0.3 (0.181, 0.454)	0.974 (0.971, 0.976)
14	12859	6	12	341	12500	0.017 (0.008, 0.037)	0.999 (0.998, 0.999)	0.333 (0.163, 0.563)	0.973 (0.971, 0.976)
15	12859	0	3	347	12509	0 (0, 0.011)	1 (0.999, 1)	0 (0, 0.561)	0.973 (0.97, 0.976)
16	12859	0	0	347	12512	0 (0, 0.011)	1 (1, 1)	-	0.973 (0.97, 0.976)

Supplementary Table 10: Accuracy of qSOFA and other tools applied only to presentations with a diagnostic impression of sepsis, infection, or nonspecific presentation

EWS	Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
Pre-alert	0	12870	71	313	277	12209	0.204 (0.165, 0.249)	0.975 (0.972, 0.978)	0.185 (0.149, 0.227)	0.978 (0.975, 0.98)
qSOFA	0	12869	274	2984	74	9537	0.787 (0.741, 0.827)	0.762 (0.754, 0.769)	0.084 (0.075, 0.094)	0.992 (0.99, 0.994)
qSOFA	1	12869	149	551	199	11970	0.428 (0.377, 0.481)	0.956 (0.952, 0.959)	0.213 (0.184, 0.245)	0.984 (0.981, 0.986)
qSOFA	2	12869	32	45	316	12476	0.092 (0.066, 0.127)	0.996 (0.995, 0.997)	0.416 (0.312, 0.527)	0.975 (0.972, 0.978)
90-30-90	0	12857	209	1463	138	11047	0.602 (0.55, 0.652)	0.883 (0.877, 0.889)	0.125 (0.11, 0.142)	0.988 (0.985, 0.99)
Borelli	0	12835	227	1160	120	11328	0.654 (0.603, 0.702)	0.907 (0.902, 0.912)	0.164 (0.145, 0.184)	0.99 (0.987, 0.991)
CIS	0	12855	310	5150	37	7358	0.893 (0.856, 0.922)	0.588 (0.58, 0.597)	0.057 (0.051, 0.063)	0.995 (0.993, 0.996)
CIS	4	12855	51	136	296	12372	0.147 (0.114, 0.188)	0.989 (0.987, 0.991)	0.273 (0.214, 0.341)	0.977 (0.974, 0.979)
HEWS	4	12835	219	1381	128	11107	0.631 (0.579, 0.68)	0.889 (0.884, 0.895)	0.137 (0.121, 0.155)	0.989 (0.986, 0.99)
MEWS	4	12859	177	909	170	11603	0.51 (0.458, 0.562)	0.927 (0.923, 0.932)	0.163 (0.142, 0.186)	0.986 (0.983, 0.988)
NHS	0	12855	221	1696	126	10812	0.637 (0.585, 0.686)	0.864 (0.858, 0.87)	0.115 (0.102, 0.13)	0.988 (0.986, 0.99)
PHANTASi	0	12858	204	1221	143	11290	0.588 (0.535, 0.638)	0.902 (0.897, 0.907)	0.143 (0.126, 0.162)	0.987 (0.985, 0.989)
PITSTOP	0	12813	35	51	311	12416	0.101 (0.074, 0.137)	0.996 (0.995, 0.997)	0.407 (0.309, 0.513)	0.976 (0.973, 0.978)
PreSAT	0	12835	258	2060	89	10428	0.744 (0.695, 0.787)	0.835 (0.828, 0.841)	0.111 (0.099, 0.125)	0.992 (0.99, 0.993)
PRESEP	3	12835	255	1736	92	10752	0.735 (0.686, 0.779)	0.861 (0.855, 0.867)	0.128 (0.114, 0.143)	0.992 (0.99, 0.993)
PRESS	1	12835	58	163	289	12325	0.167 (0.132, 0.21)	0.987 (0.985, 0.989)	0.262 (0.209, 0.324)	0.977 (0.974, 0.98)
PSP	1	12835	268	2429	79	10059	0.772 (0.725, 0.813)	0.805 (0.798, 0.812)	0.099 (0.089, 0.111)	0.992 (0.99, 0.994)
REMS	2	12855	309	4919	38	7589	0.89 (0.853, 0.919)	0.607 (0.598, 0.615)	0.059 (0.053, 0.066)	0.995 (0.993, 0.996)
RST	0	12857	279	3008	68	9502	0.804 (0.759, 0.842)	0.76 (0.752, 0.767)	0.085 (0.076, 0.095)	0.993 (0.991, 0.994)
SAS	0	12836	42	52	305	12437	0.121 (0.091, 0.16)	0.996 (0.995, 0.997)	0.447 (0.35, 0.547)	0.976 (0.973, 0.979)
SEPSIS	4	12856	137	427	210	12082	0.395 (0.345, 0.447)	0.966 (0.963, 0.969)	0.243 (0.209, 0.28)	0.983 (0.98, 0.985)
STSS	1	12855	262	2226	85	10282	0.755 (0.707, 0.797)	0.822 (0.815, 0.829)	0.105 (0.094, 0.118)	0.992 (0.99, 0.993)
Suffoletto	0	12813	247	1451	99	11016	0.714 (0.664, 0.759)	0.884 (0.878, 0.889)	0.145 (0.129, 0.163)	0.991 (0.989, 0.993)
UK	0	12855	275	2720	72	9788	0.793 (0.747, 0.832)	0.783 (0.775, 0.79)	0.092 (0.082, 0.103)	0.993 (0.991, 0.994)

Supplementary Table 11: Accuracy of NEWS2 applied to all presentations

Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
0	12859	342	9189	5	3323	0.986 (0.967, 0.994)	0.266 (0.258, 0.273)	0.036 (0.032, 0.04)	0.998 (0.996, 0.999)
1	12859	332	6749	15	5763	0.957 (0.93, 0.974)	0.461 (0.452, 0.469)	0.047 (0.042, 0.052)	0.997 (0.996, 0.998)
2	12859	326	5210	21	7302	0.939 (0.909, 0.96)	0.584 (0.575, 0.592)	0.059 (0.053, 0.065)	0.997 (0.996, 0.998)
3	12859	312	3801	35	8711	0.899 (0.863, 0.927)	0.696 (0.688, 0.704)	0.076 (0.068, 0.084)	0.996 (0.994, 0.997)
4	12859	290	2792	57	9720	0.836 (0.793, 0.871)	0.777 (0.769, 0.784)	0.094 (0.084, 0.105)	0.994 (0.992, 0.995)
5	12859	271	2088	76	10424	0.781 (0.735, 0.821)	0.833 (0.826, 0.84)	0.115 (0.103, 0.128)	0.993 (0.991, 0.994)
6	12859	235	1460	112	11052	0.677 (0.626, 0.724)	0.883 (0.878, 0.889)	0.139 (0.123, 0.156)	0.99 (0.988, 0.992)
7	12859	196	1018	151	11494	0.565 (0.512, 0.616)	0.919 (0.914, 0.923)	0.161 (0.142, 0.183)	0.987 (0.985, 0.989)
8	12859	155	616	192	11896	0.447 (0.395, 0.499)	0.951 (0.947, 0.954)	0.201 (0.174, 0.231)	0.984 (0.982, 0.986)
9	12859	117	364	230	12148	0.337 (0.289, 0.388)	0.971 (0.968, 0.974)	0.243 (0.207, 0.284)	0.981 (0.979, 0.984)
10	12859	86	214	261	12298	0.248 (0.205, 0.296)	0.983 (0.98, 0.985)	0.287 (0.238, 0.34)	0.979 (0.977, 0.982)
11	12859	57	110	290	12402	0.164 (0.129, 0.207)	0.991 (0.989, 0.993)	0.341 (0.274, 0.416)	0.977 (0.974, 0.98)
12	12859	36	57	311	12455	0.104 (0.076, 0.14)	0.995 (0.994, 0.996)	0.387 (0.294, 0.489)	0.976 (0.973, 0.978)
13	12859	13	35	334	12477	0.037 (0.022, 0.063)	0.997 (0.996, 0.998)	0.271 (0.166, 0.41)	0.974 (0.971, 0.977)
14	12859	6	15	341	12497	0.017 (0.008, 0.037)	0.999 (0.998, 0.999)	0.286 (0.138, 0.5)	0.973 (0.971, 0.976)
15	12859	0	4	347	12508	0 (0, 0.011)	1 (0.999, 1)	0 (0, 0.49)	0.973 (0.97, 0.976)
16	12859	0	1	347	12511	0 (0, 0.011)	1 (1, 1)	0 (0, 0.793)	0.973 (0.97, 0.976)

Supplementary Table 12: Accuracy of qSOFA and other tools applied to all presentations

EWS	Threshold	N	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
Pre-alert	0	12870	80	531	268	11991	0.23 (0.189, 0.277)	0.958 (0.954, 0.961)	0.131 (0.106, 0.16)	0.978 (0.975, 0.981)
qSOFA	0	12869	301	4908	47	7613	0.865 (0.825, 0.897)	0.608 (0.599, 0.617)	0.058 (0.052, 0.064)	0.994 (0.992, 0.995)
qSOFA	1	12869	160	790	188	11731	0.46 (0.408, 0.512)	0.937 (0.933, 0.941)	0.168 (0.146, 0.194)	0.984 (0.982, 0.986)
qSOFA	2	12869	33	65	315	12456	0.095 (0.068, 0.13)	0.995 (0.993, 0.996)	0.337 (0.251, 0.435)	0.975 (0.972, 0.978)
90-30-90	0	12857	222	1937	125	10573	0.64 (0.588, 0.688)	0.845 (0.839, 0.851)	0.103 (0.091, 0.116)	0.988 (0.986, 0.99)
Borelli	0	12835	242	1505	105	10983	0.697 (0.647, 0.743)	0.879 (0.874, 0.885)	0.139 (0.123, 0.156)	0.991 (0.989, 0.992)
CIS	0	12855	344	10864	3	1644	0.991 (0.975, 0.997)	0.131 (0.126, 0.137)	0.031 (0.028, 0.034)	0.998 (0.995, 0.999)
CIS	4	12855	57	171	290	12337	0.164 (0.129, 0.207)	0.986 (0.984, 0.988)	0.25 (0.198, 0.31)	0.977 (0.974, 0.98)
HEWS	4	12835	235	1923	112	10565	0.677 (0.626, 0.724)	0.846 (0.84, 0.852)	0.109 (0.096, 0.123)	0.99 (0.987, 0.991)
MEWS	4	12859	190	1232	157	11280	0.548 (0.495, 0.599)	0.902 (0.896, 0.907)	0.134 (0.117, 0.152)	0.986 (0.984, 0.988)
NHS	0	12855	241	2500	106	10008	0.695 (0.644, 0.741)	0.8 (0.793, 0.807)	0.088 (0.078, 0.099)	0.99 (0.987, 0.991)
PHANTASi	0	12858	215	1710	132	10801	0.62 (0.567, 0.669)	0.863 (0.857, 0.869)	0.112 (0.098, 0.127)	0.988 (0.986, 0.99)
PITSTOP	0	12813	39	67	307	12400	0.113 (0.084, 0.15)	0.995 (0.993, 0.996)	0.368 (0.282, 0.463)	0.976 (0.973, 0.978)
PreSAT	0	12835	277	3099	70	9389	0.798 (0.753, 0.837)	0.752 (0.744, 0.759)	0.082 (0.073, 0.092)	0.993 (0.991, 0.994)
PRESEP	3	12835	270	2272	77	10216	0.778 (0.731, 0.819)	0.818 (0.811, 0.825)	0.106 (0.095, 0.119)	0.993 (0.991, 0.994)
PRESS	1	12835	65	212	282	12276	0.187 (0.15, 0.232)	0.983 (0.981, 0.985)	0.235 (0.189, 0.288)	0.978 (0.975, 0.98)
PSP	1	12835	292	4418	55	8070	0.841 (0.799, 0.876)	0.646 (0.638, 0.655)	0.062 (0.055, 0.069)	0.993 (0.991, 0.995)
REMS	2	12855	343	10288	4	2220	0.988 (0.971, 0.996)	0.177 (0.171, 0.184)	0.032 (0.029, 0.036)	0.998 (0.995, 0.999)
RST	0	12857	304	5217	43	7293	0.876 (0.837, 0.907)	0.583 (0.574, 0.592)	0.055 (0.049, 0.061)	0.994 (0.992, 0.996)
SAS	0	12836	46	78	301	12411	0.133 (0.101, 0.172)	0.994 (0.992, 0.995)	0.371 (0.291, 0.459)	0.976 (0.974, 0.979)
SEPSIS	4	12856	143	490	204	12019	0.412 (0.362, 0.465)	0.961 (0.957, 0.964)	0.226 (0.195, 0.26)	0.983 (0.981, 0.985)
STSS	1	12855	282	3326	65	9182	0.813 (0.768, 0.85)	0.734 (0.726, 0.742)	0.078 (0.07, 0.087)	0.993 (0.991, 0.994)
Suffoletto	0	12813	263	1969	83	10498	0.76 (0.712, 0.802)	0.842 (0.836, 0.848)	0.118 (0.105, 0.132)	0.992 (0.99, 0.994)
UK	0	12855	297	4291	50	8217	0.856 (0.815, 0.889)	0.657 (0.649, 0.665)	0.065 (0.058, 0.072)	0.994 (0.992, 0.995)