

Prehospital early warning scores for adults with suspected sepsis: the PHEWS observational cohort and decision-analytic modelling study

Steve Goodacre,^{1,2*} Laura Sutton,¹ Kate Ennis,¹
Ben Thomas,¹ Olivia Hawksworth,¹ Khurram Iftikhar,²
Susan J Croft,² Gordon Fuller,² Simon Waterhouse,¹
Daniel Hind,¹ Matt Stevenson,¹ Mike J Bradburn,¹
Michael Smyth,³ Gavin D Perkins,³ Mark Millins,⁴
Andy Rosser,⁵ Jon Dickson⁶ and Matthew Wilson¹

¹School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

²Emergency Department, Northern General Hospital, Sheffield, UK

³Warwick Clinical Trials Unit, University of Warwick, Coventry, UK

⁴Yorkshire Ambulance Service NHS Trust, Wakefield, UK

⁵West Midlands Ambulance Service University NHS Foundation Trust, Midlands, UK

⁶Academic Unit of Primary Medical Care, University of Sheffield, Sheffield, UK

*Corresponding author s.goodacre@sheffield.ac.uk

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at <https://doi.org/10.3310/NDTY2403>.

Primary conflicts of interest: Steve Goodacre reports chairing the NIHR Clinical Trials Unit Standing Advisory Committee and membership of the following groups and committees: HTA PCCPI Methods Group, HTA Prioritisation Committee A Methods Group, HTA Prioritisation Committee B Methods Group, HTA Post-Funding Committee, HTA Funding Committee Policy Group and HTA Commissioning Committee. Laura Sutton reports participation on a Data Monitoring and Ethics Committee for a NIHR-funded trial in an unrelated area of study. Susan J Croft reports receiving payments as an Emergency Medicine expert advisor for the Welsh Ombudsman and membership of the NICE sub-arachnoid haemorrhage committee as an Emergency Medicine advisor. Daniel Hind reports an NIHR HTA grant outside of the submitted work and membership of the HTA Fast Track Committee. Mike J Bradburn reports membership of the HTA Commissioning Committee. Michael Smyth reports NIHR grants outside of the submitted work. Gavin D Perkins reports grants from the NIHR and Laerdal Foundation outside of the submitted work, honorary membership of the Intensive Care Society and memberships of the following: NIHR HTA Clinical Effectiveness Board, NIHR Clinical Trials Advisory Committee, NIHR Academy, NIHR CTU Standing Advisor Committee, NIHR HTA Clinical Evaluation and Trials Committee and COVID-19 Reviewing Committee. He reports honoraria from Elsevier for editorial roles for *Resuscitation* and *Resuscitation Plus* journals. Matthew Wilson reports an NIHR Research for Patient Benefit grant outside of the submitted work, participation on a Data Monitoring Committee for an NIHR-funded trial in an unrelated area of study and the role of Grants Officer to the National Institute of Academic Anaesthesia.

Published March 2024
DOI: 10.3310/NDTY2403

Scientific summary

Prehospital early warning scores for adults with suspected sepsis: the PHEWS observational cohort and decision-analytic modelling study

Health Technology Assessment 2024; Vol. 28: No. 16
DOI: 10.3310/NDTY2403

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Sepsis is a life-threatening reaction to an infection in which the immune system overreacts to infection and causes organ damage. Early recognition and treatment of sepsis has the potential to reduce mortality. Guidelines for sepsis highlight the importance of early recognition and treatment, with treatment recommended within 1 hour of presentation for those at highest risk. The emergency care system can only achieve this if sepsis is recognised and prioritised.

Ambulance services can use prehospital early warning scores to identify people with a high risk of sepsis and then pre-alert the emergency department (ED) or provide the patient with prehospital treatment. However, they need to determine which score to use, the threshold of positivity for the score, and whether to apply the early warning score to all medical cases or just those where the paramedic diagnostic impression suggests sepsis, infection or a non-specific presentation. This requires estimates of the diagnostic accuracy of early warning scores and consideration of the balance between sensitivity (avoiding missing sepsis) and specificity (prioritising too many patients who do not have sepsis).

Objectives

We aimed to determine the accuracy, impact and cost-effectiveness of prehospital early warning scores for adults with suspected sepsis. Our specific objectives were:

1. to estimate the accuracy of prehospital early warning scores for identifying sepsis requiring time-critical treatment in adults with possible sepsis who are attended by emergency ambulance
2. to estimate the impact of using prehospital early warning scores to guide key prehospital decisions, in terms of the operational consequences, and the cost-effectiveness of alternative strategies.

Methods

We undertook (1) a retrospective cohort study to estimate the accuracy of prehospital early warning scores alongside paramedic diagnostic impression and (2) decision-analytic modelling of the operational consequences and cost-effectiveness of using prioritisation strategies based on early warning score and diagnostic impression.

Retrospective cohort study

We used a literature review and expert opinion to identify 21 early warning scores for evaluation. We used routine ambulance service data to identify all episodes in 2019 in which two ambulance services (Yorkshire and West Midlands) transported patients with medical presentations to four acute hospitals (Sheffield Northern General Hospital, Doncaster Royal Infirmary, Rotherham General Hospital, University Hospitals Coventry and Warwickshire). We excluded episodes with injury, mental health problems, cardiac arrest or direct transfer to specialist services, and cases with no vital signs recorded. We calculated early warning scores from the first recorded vital signs on the ambulance service electronic patient-report form and categorised the paramedic diagnostic impression as sepsis, infection, non-specific presentation or other specific presentation. We then determined the number of cases that ambulance services would prioritise at each hospital using each early warning score alongside the categorised paramedic diagnostic impression.

We planned to use the National Health Service (NHS) Digital Data Access Request Service to link ambulance service to hospital data but NHS Digital were unable to provide this service. We therefore instituted a rescue plan to link ambulance service to hospital data at one participating hospital (Sheffield) to determine whether patients had a reference standard diagnosis of sepsis, adjudicated by two independent clinicians following hospital record review. The primary reference standard consisted of meeting the sepsis-3 definition [evidence of infection with a change of two or more points in the Sequential (sepsis-related) Organ Failure Assessment (SOFA) score] and receiving treatment for sepsis. The secondary reference standard consisted of meeting the sepsis-3 definition alone.

We analysed the ambulance service data descriptively to report the mean daily number of cases that the ambulance service would pre-alert to each hospital for each combination of early warning score and diagnostic impression. For the accuracy analysis, 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, sensitivities, specificities, and positive and negative predictive values at key cut-points, each with a 95% confidence interval (CI). Reporting of the results highlights sensitivity and positive predictive value as these best indicate under-triage (sensitivity, the proportion of sepsis cases prioritised) and overtriage (positive predictive value, the proportion of prioritised cases with sepsis).

To select strategies for the decision-analytic modelling, we calculated the proportion of mean daily ambulance arrivals that would be prioritised at each hospital and excluded strategies that would prioritise a potentially unmanageable proportion (> 10%). We then compared the accuracy of strategies and excluded those with sensitivity and specificity both inferior to another strategy. We also excluded strategies that were not clearly superior to a comparable strategy involving the National Early Warning Score, version 2 (NEWS2) on the basis that NEWS2 is already widely used by NHS ambulance services, whereas other strategies would require additional training and support to implement.

Decision-analytic modelling

We developed a decision-analytic model to evaluate the consequences to healthcare providers and the cost-effectives for the health services of using 23 different strategies to prioritise patients transported to hospital with possible sepsis.

Due to the paucity of data associated with the benefit of early treatment for sepsis and conflicting results from studies where data existed, threshold analyses were independently undertaken to estimate the reduction in mortality, the reduction in general ward length of stay (LoS) and the reduction in intensive care unit LoS that would be required by each strategy in order to be cost-effective compared with a strategy of no prioritisation of patients. We additionally present the number of prehospital alerts associated with each strategy, the number of patients with sepsis who have been correctly prioritised and the number of patients with sepsis who are not prioritised.

Results

Retrospective cohort study

We collected data from 95,022 ambulance episodes involving 71,204 patients with median age 66 years, and included 37,588 (53.0%) women, and 40,045 (94.9%) with white ethnicity. The mean (standard deviation) number of daily attendances meeting the study inclusion criteria was 93.5 (14.7) at Sheffield Northern General Hospital, 59.5 (10.8) at Doncaster Royal Infirmary, 51.3 (8.9) at Rotherham General Hospital and 74 (11) at University Hospitals Coventry and Warwickshire. Most early warning scores operating at their pre-specified thresholds would prioritise fewer than 10% of attendances when applied only to those with a diagnostic impression of sepsis or infection, but would prioritise more than 10% when applied to non-specific attendances or all attendances. The exceptions were qSOFA (threshold > 1), the Screening to Enhance Prehospital Identification of Sepsis (SEPSIS) score, the Critical

Illness Score (CIS; threshold > 4), the Paramedic Initiated Treatment of Sepsis Targeting Out-of-hospital Patients clinical trial rule, the PRESS score and the sepsis alert criteria.

Yorkshire Ambulance Service recorded only one diagnostic impression, whereas West Midlands Ambulance Service recorded multiple unranked impressions, so strategies prioritised a greater proportion of patients transported to University Hospitals Coventry and Warwickshire. Consequently, in the West Midlands most strategies prioritised more than 10% of cases when applied to those with a diagnostic impression of infection or sepsis.

We linked 12,870 out of 24,955 (51.6%) cases to the Sheffield Northern General Hospital records at Sheffield and identified 348 (2.7%) with the primary reference standard. The sensitivity and positive predictive value of paramedic diagnostic impression were 0.328 (95% CI 0.28 to 0.379) and 0.285 (95% CI 0.243 to 0.331) for sepsis, 0.572 (95% CI 0.519 to 0.623) and 0.156 (95% CI 0.137 to 0.176) for infection or sepsis, and 0.897 (95% CI 0.86 to 0.924) and 0.053 (95% CI 0.048 to 0.059) for non-specific presentation, infection or sepsis.

The early warning scores had a greater area under the ROC curve when applied to all cases rather than alongside diagnostic impression, but the low prevalence of the reference standard meant that thresholds with sensitivity above 0.7 generally had positive predictive value below 0.15, which would prioritise an unmanageable number of cases. When higher thresholds were used to provide acceptable positive predictive value and a manageable number of cases, strategies that applied the early warning score only to those with a diagnostic impression of sepsis or infection tended to have better overall accuracy. NEWS2 had the highest area under the ROC curve when applied only to those with a paramedic diagnostic impression of sepsis or infection (0.756, 95% CI 0.729 to 0.783) or sepsis alone (0.655, 95% CI 0.63 to 0.68). Only the SEPSIS score had a higher area under the ROC curve than NEWS2 when applied to non-specific presentation, infection or sepsis (0.862 vs. 0.858) and all cases (0.882 vs. 0.877).

None of the strategies provided high sensitivity (e.g. > 0.8) with acceptable positive predictive value (e.g. > 0.15). NEWS2, using varying thresholds and combinations with diagnostic impression, provided combinations of sensitivity and specificity that were similar or superior to all other early warning scores. We identified strategies reflecting published recommendations for prioritisation that could offer options with varying trade-offs between sensitivity and positive predictive value. Applying NEWS2 only to those with a paramedic diagnostic impression of sepsis or infection respectively provided sensitivities and positive predictive values of 0.522 (95% CI 0.469 to 0.574) and 0.216 (95% CI 0.189 to 0.245) with a threshold > 4, 0.447 (95% CI 0.395 to 0.499) and 0.274 (95% CI 0.239 to 0.313) with a threshold > 6, and 0.314 (95% CI 0.268 to 0.365) and 0.333 (95% CI 0.284 to 0.386) with a threshold > 8. Applying qSOFA > 1 only to those with a paramedic diagnostic impression of sepsis or infection provided sensitivity of 0.305 (95% CI 0.259 to 0.355) and positive predictive value of 0.356 (95% CI 0.304 to 0.412).

Decision-analytic modelling

The modelling provided estimates for a range of strategies with varying sensitivity and specificity of the number of cases (overall and with sepsis) that would be prioritised in a large and a small hospital. At a large hospital receiving 93.5 eligible cases per day, applying NEWS2 > 4 only to those with a diagnostic impression of infection or sepsis would prioritise 6.10 cases per day, including 1.32 with sepsis, while failing to prioritise 1.21 with sepsis. The corresponding numbers using NEWS2 > 6 were 4.11, 1.13 and 1.40, using NEWS2 > 8 were 2.38, 0.79 and 1.73, and using qSOFA > 1 were 2.17, 0.77 and 1.76. At a small hospital receiving 53.1 eligible cases per day, applying NEWS2 > 4 only to those with a diagnostic impression of infection or sepsis would prioritise 3.35 cases per day, including 0.72 with sepsis, while failing to prioritise 0.66 with sepsis. The corresponding numbers using NEWS2 > 6 were 2.26, 0.62 and 0.77, using NEWS2 > 8 were 1.31, 0.44 and 0.95, and using qSOFA > 1 were 1.19, 0.42 and 0.96.

The threshold analysis showed that the relative risk of mortality associated with prioritisation at which each strategy would be cost-effective compared to no prioritisation [assuming willingness to pay £20,000 per quality-adjusted life-year (QALY) gained] ranged from 0.977 applying NEWS2 > 0 only to those with a diagnostic impression of infection or sepsis, to 0.996 applying NEWS2 > 11 only to those with a diagnostic impression of sepsis. The comparable ranges for other measures of effectiveness for these two strategies were: increase in QALYs 0.00056–0.00002; reduction in length of ward stay 3.8–0.7 days; reduction in intensive care LoS 1.2–0.2 days.

Conclusions

We were unable to identify a strategy that would prioritise a substantial majority of patients with sepsis without prioritising a potentially unmanageable number of patients for the ED. Most early warning scores, used at a recommended threshold, are likely to prioritise an unmanageable number of cases unless they are only used to prioritise cases with a paramedic diagnostic impression of infection or sepsis. However, paramedic diagnostic impression of infection or sepsis only identified 57% of cases with a reference standard diagnosis of sepsis requiring urgent treatment.

The NEWS2 provides sensitivity and specificity for identifying sepsis that is generally similar or superior to other scores operating at a comparable threshold. We therefore found no evidence to justify the support and training required to implement an alternative strategy to NEWS2, which is already widely used in NHS ambulance services.

National Early Warning Score, version 2, could be used at thresholds of > 4 or > 6 in presentations with a diagnostic impression of infection or sepsis, reflecting the Academy of Medical Royal Colleges clinical decision support framework, or > 8 to provide similar sensitivity and specificity to the use of qSOFA > 1 recommended in the sepsis-3 guidelines. These strategies provide a range of options that ambulance services and hospitals could use, depending upon capacity to manage prioritised cases and what prioritisation involves.

Health economic modelling suggests that sensitive strategies for identifying patients with possible sepsis for prioritisation could be cost-effective, if we are convinced that reducing treatment delay reduces mortality and the emergency care system has the capacity to deliver meaningful prioritisation to substantial numbers of cases.

Limitations

Inability of NHS Digital to link ambulance service to hospital data meant that we were only able to estimate the accuracy of early warning scores at one hospital using data from 51.6% of the eligible population for whom the ambulance service had NHS numbers. The included patients were markedly older than the excluded patients. We were unable to identify reliable evidence to estimate the effectiveness of early treatment for sepsis, so were unable to identify the most cost-effective strategy.

Future research

Research into prehospital early warning scores for sepsis is limited by our current inability to clinically measure the dysregulated host response that characterises sepsis and uncertain estimates of the benefits of early treatment. We therefore need to prioritise research to develop better ways of defining and diagnosing sepsis, and to develop and evaluate effective early treatment for sepsis. Future research involving routine ambulance service and hospital data requires a system for NHS data management that supports health data science.

Study registration

This study is registered as Research Registry (reference: researchregistry5268).

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 17/136/10) and is published in full in *Health Technology Assessment*; Vol. 28, No. 16. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.6

A list of Journals Library editors can be found on the [NIHR Journals Library website](#)

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.6 and is ranked 32nd (out of 105 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as award number 17/136/10. The contractual start date was in July 2019. The draft report began editorial review in April 2023 and was accepted for publication in August 2023. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

Copyright © 2024 Goodacre et al. This work was produced by Goodacre et al. under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

