The University Of Sheffield.

Sheffield Teaching Hospitals NHS
NHS Foundation Trust

*Early warning scores for Sepsis*

# PHEWS: Pre-Hospital Early Warning scores for Sepsis study.

# Accuracy, impact and cost-effectiveness of pre-hospital clinical early warning scores for adults with suspected sepsis

**RESEARCH PROTOCOL**
**(Version 4.0) 21st December 2021**

| | |
|---|---|
| **Sheffield CTRU** | **156033** |
| **IRAS** | **262367** |
| **REC** | **19/YH/0288** |
| **Registry** | **researchregistry5268** |
| **Authorised by:** | **Steve Goodacre** |

Sheffield Clinical Trials Research Unit (CTRU)

# Accuracy, impact and cost-effectiveness of pre-hospital clinical early warning scores for adults with suspected sepsis

This document describes the Sepsis study, and provides information about procedures throughout the study.

## Contents

Abbreviations ..... 4
General Information ..... 5
Protocol amendments since Version 1.0 ..... 7
Summary of Research ..... 8
1.0 Introduction ..... 9
1.1 Background and Rationale ..... 9
1.2 Existing literature ..... 11
2.0 Aims and objectives ..... 12
3.0 Study Design ..... 13
4.0 Workstream 1: Retrospective cohort study ..... 13
4.1 Design ..... 13
4.2 Health technology being assessed ..... 13
4.3 Target population ..... 16
4.4 Inclusion/Exclusion Criteria ..... 16
4.5 The reference standard ..... 16
4.6 Data Collection and management ..... 17
4.7 Data Analysis ..... 18
4.8 Sample Size ..... 19
5.0 Work stream 2: Decision analytic modelling ..... 19
5.1 ED prioritisation ..... 19
5.2 Pre-hospital treatment for sepsis ..... 21
6.0 Trial Supervision ..... 22
6.1 Project Management Group ..... 23
6.2 Study Steering Committee ..... 23
7.0 Data handling and record keeping ..... 23
8.0 Dissemination, outputs and anticipated impact ..... 23
9.0 Project timetable ..... 24
10.0 Funding and role of the funder ..... 25
11.0 Ethics ..... 25

12.0 Regulatory approval ..... 25
13.0 Indemnity / Compensation / Insurance ..... 25
14.0 Patient and Public Involvement ..... 25
15.0 Research expertise ..... 26
16.0 Success criteria and barriers to proposed work ..... 27
17.0 References ..... 28

# Abbreviations

| | |
|---|---|
| CAG | Confidentiality Advisory Group |
| CTRU | Clinical Trials Research Unit |
| DMP | Data Management Plan |
| ECDS | Emergency Care Data Set |
| ED | Emergency Department |
| ePRF | Electronic Patient Report Form |
| HES | Hospital Episodes Statistics |
| MEDS | Mortality in the ED, Sepsis |
| MEWS | Modified Early Warning Score |
| NEWS | National Early Warning Score |
| NGT | Nominal Group Technique |
| NICE | National Institute for Health and Care Excellence |
| ONS | Office for National Statistics |
| PIRO | Predisposition, Infection, Response and Organ Dysfunction |
| PMG | Project Management Group |
| PPI | Patient and Public Involvement |
| PRESEP score | Prehospital Early Sepsis Detection score |
| PRESS score | Prehospital Severe Sepsis score |
| QALY | Quality Adjusted Life Year |
| qSOFA | quick Sepsis Related Organ Failure Assessment |
| REMS | Rapid Emergency Medicine Score |
| ROC | Receiving-operator characteristic |
| SEPSIS tool | Screening to Enhance Prehospital Identification of Sepsis tool |
| SOP | Standard Operating Procedure |
| SSC | Study Steering Committee |
| STSS | Simple Triage Scoring System |

# General Information

### Sponsor

Sheffield Teaching Hospital NHS Foundation Trust
The University of Sheffield
D Floor, Royal Hallamshire Hospital
Glossop Road
Sheffield
S10 2JF

Sponsor Representative: Dr Erica Wallis
Email: Erica.wallis@nhs.net
Tel: 0114 2265931

### Chief Investigator

Professor Steve Goodacre
Professor of Emergency Medicine
School of Health and Related Research (ScHARR)
The University of Sheffield
Regent Court
Sheffield
S1 4DA
Email: s.goodacre@sheffield.ac.uk
Tel: (+44) (0)114 222 0842

### Study Manager

Dr Benjamin Thomas
Study Manager
Clinical Trials Research Unit
ScHARR,
The University of Sheffield
Innovation Centre, 217 Portobello
Sheffield
S1 4DP
Email: b.d.thomas@sheffield.ac.uk
Tel: 0114 22 24024

### Co-investigators

Dr Michael Smyth
Research Fellow
University of Warwick

Mr Mike Bradburn
Senior Medical Statistician
University of Sheffield

Dr Jon Dickson
Senior Clinical Lecturer and GP
University of Sheffield

Prof Gavin Perkins
Professor of Critical Care Medicine
University of Warwick

Prof Matt Stevenson
Professor of Health Technology Assessment
University of Sheffield

Dr Daniel Hind
Assistant Director, CTRU
University of Sheffield

Enid Hirst
Public representative
Sheffield Emergency Care Forum

Mrs Linda Abouzeid
Public representative
Sheffield Emergency Care Forum

Dr Gordon Fuller
Clinical Lecturer in Emergency Medicine
University of Sheffield

Mr Mark Millins
Associate Director for Paramedic Practice
Yorkshire Ambulance Service NHS Trust

Mr Andrew Rosser
Lead Research Paramedic
West Midlands Ambulance Service NHS Foundation Trust

Ms Janette Turner
Reader in Urgent & Emergency Care
University of Sheffield

Dr Matthew Wilson
Senior Lecturer in Anaesthesia
University of Sheffield

Mrs Cathryn James
Clinical Manager
Yorkshire Ambulance Service NHS Trust

## Study Steering Committee

| | |
|---|---|
| Professor Tim Coates (Chair)<br>Department of Cardiovascular Sciences, University of Leicester, Level G, Jarvis Building RMO, Infirmary Square, Leicester, LE1 5WW<br>Email: tc61@le.ac.uk | Mrs Alice Riddell<br>Sheffield Emergency Care Forum<br>128 Townhead Road, Dore, Sheffield, S17 3GB<br>Email: aliceriddell@googlemail.com |
| Mrs Shan Bennett<br>Sheffield Emergency Care Forum<br>87 Springfield Road, Sheffield, S7 2GG<br>Email: shanbennett@hotmail.com | Mr Duncan Robertson<br>Welsh Ambulance Services NHS Trust Headquarters, Ty Elwy, St Asaph Business Park, Fford Richard Davies, St Asaph, Denbighshire, LL17 0LJ<br>Email: Duncan.Robertson@wales.nhs.uk |
| Dr James Fullerton<br>G19, Rayne Institute,<br>University College London<br>5 University Street, London, WC1E 6JF<br>Email: j.fullerton@ucl.ac.uk | Dr Joie Ensor<br>Centre for Prognosis Research,<br>School of Medicine,<br>Keele University,<br>Staffordshire,<br>ST5 5BG<br>Email: j.ensor@keele.ac.uk |
| Professor Steve Goodacre<br>School of Health and Related Research, Regent Court, Regent Street, Sheffield, S1 4DA<br>Email: s.goodacre@sheffield.ac.uk | Dr Graham McClelland<br>North East Ambulance Service NHS Foundation Trust, Bernicia House, Goldcrest Way, Newburn Riverside Business Park, Newcastle upon Tyne, NE15 8NY<br>Email: Graham.McClelland@neas.nhs.uk |
| Dr Fenella Wrigley<br>London Ambulance Service NHS Trust Headquarters, 220 Waterloo Road, London, SE1 8SD<br>Email: fenella.wrigley@lond-amb.nhs.uk | |

## Protocol amendments since Version 1.0

| Version number | Change (s) made | Date of REC approval* | Amendment number |
|---|---|---|---|
| 2.0 | • Study title changed<br>• Study registry details added<br>• Name of study manager added<br>• Names and contact details of Study Steering Committee added<br>• Email address of sponsor amended<br>• Section 4.6 – a description of the possible need for alternative approaches to securing data to establish the Sepsis reference standard added. | 20/12/19 | Substantial amendment 1 |
| 3.0 | • Section 4.2 – changes to the membership and remit of the expert review group<br>• Change co-applicants to co-investigator<br>• Removal of a PPI study co-investigator who has left the study and addition of Enid Hirst as a replacement.<br>• Addition of PPI SSC member affiliations<br>• Removal of a SSC member who has left the committee<br>• Change of estimated number patients being identified to reflect inclusion of four hospitals rather than two in original application<br>• Sections 4.6 and 11 – change to data management so that NHS Digital provides Sheffield CTRU with the NHS number, and Sheffield CTRU contacts participating hospitals directly for reference standard information | | Substantial amendment 2 |
| 4.0 | • Section 4.5 – changes are made to how hospitals are identifying patients to assess whether they meet the study's' reference standard.<br>• Additional member of Steering committee added. | TBC | Substantial amendment 3 |

# Summary of Research


### Research question

What are the accuracy, impact and cost-effectiveness of prehospital early warning scores for adults with suspected sepsis?

### Background

Prehospital early warning scores can be used to prioritise treatment for people at high risk of sepsis but the accuracy of existing scores and the impact of their use is uncertain.

### Aims and objectives

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

1. To estimate the accuracy of early warning scores for predicting potential to benefit from time-critical treatment for sepsis
2. To estimate the operational consequences and cost-effectiveness of using early warning scores to guide prehospital decisions for adults with suspected sepsis

### Methods

A retrospective diagnostic cohort study using routine data sources across four acute hospitals with associated ambulance services and decision analytic modelling.

The target population will be adults transported to hospital by emergency ambulance with possible sepsis. Patient report form data will be used to identify eligible cases and collect pre-hospital measures used to calculate early warning scores. Cases will be linked to routine hospital data sources.

The index tests will be prehospital early warning scores identified through a systematic literature search and selected by an expert group, who will also consider if there is a need to create any additional scores.

The reference standard will be determined by expert review of clinical records to identify patients with infection and life-threatening organ dysfunction who received potentially life-saving treatment for sepsis. Screening will be used to select cases for expert review so that the workload is manageable but cases are not missed.

Receiving-operator characteristic (ROC) curves will compare sensitivity to specificity across the range of each score. We will also explore whether a clinically credible new score can be derived using multivariable logistic regression, augmented with recursive partitioning.

We plan to identify 92,000 cases across one year, with around 200 anticipated to be reference standard positive, allowing us to estimate sensitivity with a standard error of 2.1% and the area under the ROC curve with a standard error of 2%.

We will create a decision-analytic model to estimate the impact of using early warning scores to: (i) Alert the receiving hospital so that the patient is seen immediately on arrival; (ii) Provide prehospital treatment for sepsis. The model will simulate the prehospital management of a hypothetical cohort of patients with suspected sepsis and then model their flow through an ED, alongside all other patients attending the ED. The model will estimate the operational consequences of using different early warning scores, in terms of the number of attendances appropriately and inappropriately prioritised and/or given prehospital treatment. We will adopt a health service perspective to estimate costs and will value outcomes as QALYs to estimate the incremental cost per QALY of using different strategies and will undertake a full incremental analysis.

**Timelines for delivery**

Month 1-6: Set-up and approvals, identification and selection of early warning scores
Month 7-12: Identification of cases and linkage to hospital data
Month 13-18: Reference standard adjudication and development of model
Month 18-24: Analysis, write-up and dissemination

**Anticipated impact and dissemination**

The findings will inform future NICE guidance and determine how ambulance services use prehospital early warning scores for sepsis.

# 1.0 Introduction

## 1.1 Background and Rationale

**What is the problem being addressed?**

This proposal addresses the National Institute for Health and Care Excellence (NICE) research recommendation “Can early warning scores be used to improve the detection of sepsis and facilitate prompt and appropriate clinical response in prehospital settings and in emergency departments?” (NICE 2016). The specific research question we address is “What are the accuracy, impact and cost-effectiveness of prehospital early warning scores for adults with suspected sepsis?”

Sepsis is a life-threatening organ dysfunction due to a dysregulated host response to infection. Early recognition and treatment of sepsis is essential to reducing mortality. The Surviving Sepsis Campaign Bundle 2018 update recommends delivery of intravenous fluids and broad-spectrum antibiotics within one hour of presentation (Levy 2018). This can only be achieved if sepsis is recognised and prioritised in the emergency care system.

Sepsis can be recognised by identifying clinical features of the host response or organ dysfunction, such as altered mental state, low blood pressure or rapid respiratory rate. Early warning scores use simple measurements to calculate a score, with a higher score indicating a higher risk of serious illness and adverse outcomes. They can be used by prehospital providers, such as ambulance paramedics, to identify people with suspected sepsis who need

to be prioritised for treatment. Prioritisation currently involves pre-alerting the emergency department (ED) so that the patient is seen immediately on arrival by a clinician who is able to provide time-critical treatment for sepsis and refer for urgent specialist care. It may also involve initiating intravenous fluid therapy whilst en route to hospital. In future, early warning scores could also be used to select people for additional prehospital treatment for sepsis, specifically antibiotic therapy, to further reduce treatment delays.

An effective early warning score needs to be accurate and implemented with an appropriate balance between sensitivity and specificity. High sensitivity is needed to ensure that people with the potential to benefit from urgent treatment are not missed with consequent delayed treatment and avoidable mortality and morbidity. However, sacrificing specificity to maximise sensitivity can result in over-triage, in which people without sepsis or the potential to benefit from urgent treatment are inappropriately prioritised. This increases the pressure on EDs and impairs their ability to provide rapid treatment when it is required. It may also result in inappropriate prehospital treatment, especially if prehospital scope of practice for sepsis is expanded to include antibiotic therapy.

The problem of over-triage is compounded if early warning scores are applied to an unselected population with a low prevalence of sepsis or a high prevalence of conditions that increase early warning scores in the absence of sepsis.

**Why is this research important?**

This application specifically addresses the example topic of interest “Early warning scoring systems (pre-hospital and in ED/hospital)” outlined in the commissioning brief.

NICE guidance advises thinking 'could this be sepsis?' if a person presents with signs or symptoms that indicate possible infection and highlights that people with sepsis may have non-specific, non-localised presentations and may not have a high temperature. The practical implication is that sepsis may be suspected in any patient attended by emergency ambulance for a medical complaint that is not attributable to a clear alternative cause. This could represent around 2 million ambulance cases per year in the NHS. An early warning score with poor specificity could therefore result in over-triage of a huge number of patients, placing severe pressure on the emergency care system.

Early warning scores are intended to identify people with high-risk sepsis who are most likely to benefit from urgent and intensive treatment. The incidence of such cases is uncertain but may be estimated using data from those receiving critical care assessment or intensive care admission. The National Confidential Enquiry into Patient Outcomes and Deaths (NCEPOD) identified 3363 adults with sepsis and critical care outreach assessment or admission across 305 hospitals over two weeks (NCEPOD 2015). The Intensive Care National Audit and Research Centre (ICNARC) reported 22081 admissions to critical care with septic shock across 205 hospitals over 2012 (Parrot 2014). A proportion of these cases self-present to hospital or arise as inpatients but this still leaves many thousands per year with the potential to benefit from identification using an early warning score. A sensitive early warning score is essential to ensuring these patients receive timely and appropriate treatment.

## 1.2 Existing literature

The NICE Guideline Development Group (GDG) identified a number of early warning scores that are easy to use, only require simple measurements and could therefore be used in the prehospital setting. These are the Simple Triage Scoring System (STSS), Rapid Emergency Medicine Score (REMS) or modified-REMS, the Modified Early Warning score (MEWS) and National Early Warning score (NEWS). They were developed through expert consensus or analysis of routine data from hospitalised patients and contain similar measures (heart rate, respiratory rate, oxygen saturation, blood pressure and conscious level) but differ in their calculation. REMS, MEWS and NEWS have been shown to predict adverse outcome in acute medical admissions, while STSS has been shown to predict mortality in inpatients with suspected infection.

The NICE GDG identified other early warning scores for use in hospital, such as Mortality in the Emergency Department, Sepsis (MEDS) and Predisposition, Infection, Response and Organ dysfunction (PIRO), but did not recommend them for prehospital use on the basis that they include blood tests that are not currently available in the prehospital setting.

A systematic review of early warning scores undertaken for NICE guidance identified 47 relevant studies (including studies of in-hospital scores). All were judged as being of very low quality. There was significant variability in population, outcomes and analysis, so meta-analysis was not possible. No relevant economic evaluations were identified. The guideline recommended that clinicians consider using an early warning score to assess people with suspected sepsis in acute hospital settings and recommended research to determine whether early warning scores can be used to improve the detection of sepsis.

Two systematic reviews of prehospital identification of sepsis also reported limited existing evidence and a need for further research. Co-applicant Smyth (2016) reported three studies developing prehospital sepsis screening tools for adults and six studies of paramedic diagnosis of sepsis. The studies were low quality and none of the screening tools had been validated. Lane (2016) reported nine studies of prehospital identification of sepsis. Study quality was poor and both sensitivity (0.43-0.86) and specificity (0.47–0.87) varied markedly.

Since the NICE review the qSOFA score has been derived and validated to predict death in hospitalised patients with suspected sepsis (Seymour 2016, Freund 2017). A meta-analysis of 38 recent studies of qSOFA (Fernando 2018) reported pooled sensitivity of 0.61 and pooled specificity of 0.72 for mortality. Meanwhile, NEWS has been updated to become NEWS2 and endorsed by NHS Improvement and NHS England.

Other recent studies have developed scores in the prehospital setting. Smyth (2018) derived and validated the Screening to Enhance PrehoSpital Identification of Sepsis (SEPSIS) tool to identify patients with high risk of sepsis in medical cases attended by emergency ambulance with an area under the receiving-operator curve (AUROC) of 0.86, and sensitivity 0.80 and specificity 0.78 at the recommended threshold. Bayer (2015) developed the Prehospital Early Sepsis Detection (PRESEP) score to identify sepsis in prehospital patients with suspected sepsis with AUROC 0.93, sensitivity 0.85 and specificity 0.86. Polito (2015) developed the prehospital severe sepsis (PRESS) score to identify severe sepsis in physiologically abnormal prehospital patients with suspected sepsis with sensitivity 0.86 and specificity 0.47.

Previous studies have important limitations other than the low quality identified in the NICE review:

1. Early warning scores should ideally identify patients who have the greatest potential to benefit from prioritisation and urgent treatment. Studies using mortality as the outcome or reference standard may not achieve this aim (Challen 2015). Firstly, patients whose lives are saved by urgent treatment will be categorised as reference standard negative despite having clearly benefited. Secondly, patients with severe pre-existing life-limiting conditions that make life-saving treatment futile or inappropriate will be classified as reference standard positive despite having little potential to benefit. Early warning scores developed to predict adverse outcomes such as mortality may therefore predict irreversible mortality while missing those with greatest potential to benefit from urgent treatment.
2. Early warning scores need to be operationalised by using a threshold for decision-making that optimises the trade-off between sensitivity and specificity in terms of the benefits, harms and costs of prioritisation. Existing studies have not explicitly examined the trade-off in these terms and the NICE review identified no relevant economic evaluations. Although the cost of applying an early warning score is small the potential knock-on costs of over-triage are substantial.
3. Early warning scores should be evaluated in the population in whom the score will be used. REMS, MEWS and NEWS were developed and validated in acute medical inpatients with a range of medical complaints, while STSS and qSOFA were developed in inpatients with suspected sepsis. Inpatient populations, especially those identified as having suspected sepsis by hospital clinicians, are likely to have a higher prevalence of severe sepsis than prehospital populations. Using an early warning score developed for an inpatient population in the prehospital setting could lead to substantial over-triage.

Research therefore needs to use a reference standard or outcome that reflects potential to benefit from urgent treatment, explicitly examine the trade-off between sensitivity and specificity in terms of the benefit, harms and costs of using an early warning score to prioritise patients, and evaluate early warning scores in the prehospital population.

## 2.0 Aims and objectives

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

1. To estimate the accuracy of prehospital early warning scores for predicting potential to benefit from time-critical treatment for sepsis 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

# 3.0 Study Design

The study will involve two concurrent streams of work addressing the two objectives above:

1. A retrospective cohort study using routine data sources will estimate the accuracy of prehospital early warning scores (index test) for predicting potential to benefit from time-critical treatment for sepsis (reference standard) in adults with possible sepsis who are attended by emergency ambulance (population). The retrospective design will ensure that the cohort includes sufficient numbers with a positive reference standard to estimate the sensitivity of early warning scores with acceptable precision.
2. Decision analytic modelling will be used to determine the impact of using prehospital early warning scores to guide two key decisions: (i) Alerting the receiving hospital so that the patient is seen immediately on arrival; (ii) Providing prehospital treatment for sepsis, such as intravenous antibiotics. The decision analytic modelling will synthesise data from multiple sources to determine the operational consequences and cost-effectiveness of using prehospital early warning scores and evaluate the trade-off between sensitivity and specificity.

We have explicitly focused the design of our proposal on estimating the accuracy of scores for an appropriate outcome in an appropriate population and examining the trade-off between sensitivity and specificity, as these are the key uncertainties in the existing evidence. It is clear from the overview of existing literature above and the description of health technologies being assessed below that there has been extensive research into developing early warning scores and there are already a substantial number of potential scores available. It is therefore unlikely that a new score can be developed using currently available prehospital measurements that will markedly out-perform existing scores.

We will take the opportunity to test expert-derived scores and to explore whether a new score may be derived from our data that has markedly superior accuracy to existing scores, but development of new scores is not the main focus of the project. We have therefore not included any proposals, such as derivation and validation of a new score on separate cohorts, that would substantially add to the cost, complexity and duration of the project, and would detract from the primary aims.

# 4.0 Workstream 1: Retrospective cohort study

## 4.1 Design

Routine sources will be used to collect data across four large acute hospitals and associated ambulance services (Yorkshire and West Midlands) from people attended by emergency ambulance and transported to hospital with possible sepsis. We have drawn upon guidance for retrospective record review studies in developing the design (Vasser 2013, Worster 2004).

## 4.2 Health technology being assessed

We will evaluate any early warning score that can be used by pre-hospital professionals and calculated from routinely available prehospital data. Our scoping work has identified NEWS,

NEWS2, MEWS, REMS, STSS, qSOFA, SEPSIS, PRESEP and PRESS. These scores are based on combinations of different measures using varying thresholds to determine whether the measure is abnormal. We have summarised the measures used in each score in the table below. Most of the measures are directly recorded in routine available prehospital data but some may be inferred from routine data. A number of guidelines and tools, such as the UK Sepsis Trust tool, Joint Royal Colleges Ambulance Liaison Committee (JRCALC 2017) guidance, NICE risk recognition criteria (NICE 2016) and Robson tool (Robson 2009), use early warning scores or their constituent elements to identify cases of sepsis for prioritisation. We will examine how these guidelines use early warning scores or similar predictors to prioritise patients with suspected sepsis.

| Early warning score | Heart rate | Respiratory rate | Blood pressure | Oxygen saturation | Conscious level | Temperature | Blood glucose | Skin appearance | Dispatch category | Location | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NEWS, NEWS2 | X | X | X | X | X | X | | | | | |
| MEWS | X | X | X | | X | X | | | | | |
| REMS | X | X | X | X | X | | | | | | X |
| STSS | X | X | X | X | X | | | | | | X |
| qSOFA | | X | X | | X | | | | | | |
| SEPSIS | X | X | X | X | X | X | | X | | | X |
| PRESEP | X | X | X | X | X | X | X | | | | |
| PRESS | | | X | X | | X | | | X | X | X |
| Critical illness score | X | X | X | X | X | | | | | X | X |

Early warning scores will be identified and selected by an expert group using findings from a systematic literature search. The literature search is intended to identify any study reporting use of an early warning score for sepsis. We will include studies from both the prehospital and in-hospital settings, but early warning scores will only be selected if they can be routinely used in the prehospital setting (i.e. those requiring tests that are not routinely available in the prehospital setting will be excluded). The literature search is intended to identify rather than evaluate early warning scores, so we will not formally extract or analyse data. Evaluation of the scores will be undertaken through our primary research. However, we will select key studies (especially recent systematic reviews of early warning scores) to present to the expert group and inform their discussions.

Relevant studies will be identified through electronic searches of key electronic databases including MEDLINE, EMBASE and all databases in the Cochrane Library (including the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials and NHS Economic Evaluations Database). References will also be located through review of reference lists for relevant articles and through use of citation search facilities through the Web of Knowledge. In addition, systematic searches of trial registries and the Internet using the Google search engine will be used to identify unpublished materials and work in progress.

We will convene a group of nine experts from prehospital care, emergency medicine and critical care, consisting of the clinical members of the Project Management Group, to review the early warning scores identified by the literature search and select those that can be used in prehospital care and calculated from routinely available prehospital data. The retrospective study design allows us to evaluate additional scores with minimal additional resource, provided they can be calculated using routinely available data, so there is no need to place limits on the number of scores evaluated. We will not, however, be able to evaluate scores that rely upon clinical information that is not routinely available. We anticipate that some scores could be calculated indirectly by inferring from routinely available data or modifying the score. The expert group will adopt an inclusive approach and try to find ways of including scores in the analysis, while acknowledging the potential limitations of this approach. The selection of scores, and methods for calculation, will be determined at a facilitated round table meeting chaired by a co-investigator with expertise in consensus methodology.

The expert group will also use consensus methods to create additional scores for evaluation. This approach provides an efficient way of developing new scores that uses expert insights to ensure that the scores are practical for routine use and draws upon the strengths of existing scores while addressing any weaknesses. Expert-derived scores can then be validated in the study cohort, whereas a statistically derived score would require a new cohort. We will also determine whether a new score can be statistically derived from our data that is superior to existing scores (see Data Analysis below), although this would need validation in a new cohort. We have not planned validation of a statistically-derived score within the proposed study because this would involve a substantial additional body of work with only a modest likelihood of being required. The large number of existing scores limit the potential for a new statistically derived score being developed that is sufficiently superior to warrant validation.

**The number and type of expert-derived scores will depend upon the extent to which existing scores allow us to explore the trade-off between sensitivity and specificity, and the extent to which existing scores include potential useful routinely recorded variables. If the existing scores allow full examination of the trade-off between sensitivity and specificity, and include all potentially useful routinely recorded variables, then no expert-derived scores will be needed. If any gaps are identified, then expert-derived scores will be developed to ensure a comprehensive range of scores for analysis.**

Finally, the expert group will address the reproducibility and replicability of the early warning scores selected for evaluation. Early warning scores are usually based on routine physiological variables that are measured and recorded in a standard manner across prehospital practice. The potential for variability across settings is further reduced by our restriction to evaluation of early warning scores based on routinely recorded variables. However, we will ask the expert group during the round table meeting to review the scores selected for evaluation and ensure the following: (1) The variables comprising the score are likely to be measured and recorded in a standard manner; (2) Any thresholds used to categorise a continuous measure are based, where possible, on accepted cut-points, such as accepted normal ranges; (3) Categorisation and allocation of points to form a score follows a logical and intuitive process; (4) The process of calculating a score is simple and reproducible, with a low risk of error. If the expert group identifies concerns relating to these issues they will consider amending the score to make it more reproducible across the NHS.

## 4.3 Target population

All adults with possible sepsis transported by emergency ambulance to four acute hospitals over the course of one year.

## 4.4 Inclusion/Exclusion Criteria

We will include all patients transported to hospital by emergency ambulance unless they would clearly not be managed as suspected sepsis, i.e. patients with injury, mental health problems, cardiac arrest and direct transfers to specialist services (including maternity, cardiac or stroke services). We will also exclude children. The presentation and management of sepsis differs markedly between adults and children so the use of early warning scores in these patients therefore needs to be studied separately. We will also exclude cases with no vital signs recorded since vital signs are essential to calculating early warning scores.

## 4.5 The reference standard

The purpose of prehospital early warning scores is to prioritise patients who have potential to benefit from urgent treatment for sepsis, so the reference standard needs to be positive in patients who survived as a result of urgent treatment, or died despite urgent treatment but had a reasonable chance of meaningful survival when they were treated. This inevitably involves expert judgement. However, the study design requires a large sample size to detect a sufficient number of cases with a positive reference standard. We therefore need to select a group of cases for expert review that are likely to include all those with a positive reference standard and withhold expert review of cases where the likelihood of a positive reference standard is negligible.

Routine hospital data will be used to select those with (1) hospital admission or death in the ED and (2) an International Classification of Disease (ICD) 10 code or cause of death compatible with sepsis..

Research nurses will briefly review the ED records of these cases and select patients for expert review if they had a diagnosis of sepsis recorded and/or received treatment for sepsis.

Two experts, a ED consultant and trainee ED doctor, will independently review hospital records for all patients selected by the research nurses and will determine the reference standard to be positive if the following criteria are met: (1) Evidence of infection and life-threatening organ dysfunction (as defined by Sepsis-3, the Third International Consensus Definitions for Sepsis and Septic Shock) within four hours of hospital admission; (2) Treatment for sepsis was given and was not withdrawn, other than on the basis of a lack of response to treatment. Disagreements will be resolved by a third expert, an ED doctor or person with equivalent knowledge at the hospital. Cases in which sepsis was diagnosed but treatment was withheld or withdrawn on the basis of futility, quality of life or patient wishes will be identified and examined in a secondary analysis. Although they do not have potential to benefit from urgent treatment it may be considered important to recognise and prioritise such cases.

We will pilot routine data screening, research nurse screening and expert review, and test the sensitivity of research nurse screening by using other sources, such as sepsis audits, to identify if any cases are missed. Piloting will allow the screening process to be developed so as to maximise sensitivity and specificity.

## 4.6 Data Collection and management

The data collection and management plan is based upon previous experience of matching ambulance service to hospital data in the PHOEBE study led by co-applicant Turner and the derivation of the SEPSIS tool, led by co-applicant Smyth. The PHOEBE study used data linkage methods provided by NHS Digital to link ambulance electronic Patient Report Form (ePRF) data to Hospital Episodes Statistics (HES) and Office for National Statistics (ONS) mortality data. Smyth developed a bespoke linkage method to link PRF data from the West Midlands Ambulance Service to emergency department and in-hospital data from West Midlands hospitals. Both projects used personal data without patient consent to facilitate linkage and were undertaken with approval from the Confidentiality Advisory Group (CAG) of the Health Research Authority. We will seek CAG approval for this project and have planned the project timetable to allow for delays in attaining CAG approval and NHS Digital data.

Both participating ambulance services use an ePRF to collect routine data from patients attended by emergency ambulance. We will use ePRF data to identify people eligible cases, according to the inclusion/exclusion criteria outlined above, and collect prehospital measures used to calculate early warning scores. The first measurement of any clinical variable with multiple measurements will be used to calculate the score.

Each ambulance service will be asked to provide data from all eligible cases transported to one of the participating hospitals over one year. The ambulance service will create a unique study identification (ID) number for each case. Two linked databases will be created: (1) containing the unique ID, time and date of call, and personal details will be sent to NHS Digital; (2) containing the unique ID, time and data of call, and all non-personal details will be sent to the Sheffield Clinical Trials Research Unit (CTRU).

NHS Digital will use the personal details from the ambulance data (including tracing NHS Number where needed) to link each case to the ECDS data for the related ED attendance and any related HES or ONS data from subsequent hospital admissions. They will send selected ECDS, HES and ONS data alongside the unique study ID and the NHS number to the Sheffield CTRU.

Researchers at Sheffield CTRU will identify those with (1) hospital admission or death in the ED, (2) an International Classification of Disease (ICD) 10 code or cause of death compatible with sepsis, and (3) a relevant treatment recorded in the Emergency Care Data Set (ECDS) treatments, such as intravenous medication, fluids or antibiotics. They will send the NHS number and unique IDs for these patients to the relevant hospital.

The research nurses and ED doctors in the participating hospitals will then undertake the process outlined above, in section 4.5, to determine the reference standard and will then send a database to the Sheffield CTRU consisting of the unique study ID and reference standard judgement for the selected cases.

The process of using NHS Digital to link ambulance service to HES and ONS data was successfully undertaken in the PHOEBE project with CAG approval. For this project we need to extend the process to involve ECDS data and sharing data with research nurses at the hospital. Smyth, by contrast, did not use NHS Digital but undertook direct linkage between ambulance and hospital data. We will draw upon his experience to inform our process and where necessary consider alternative approaches to identify patients for inclusion in the reference standard, if the use of NHS Digital data become prohibitive to the successful completion of the study to project timelines. Any alternative approaches would be carefully considered and require approval of the study Project Management Committee, Study Steering Committee, and appropriate ethical approvals.

## 4.7 Data Analysis

Diagnostic accuracy will be estimated by comparing the index test (early warning score) to the reference standard (potential to benefit from time critical treatment for sepsis). We will construct a receiving-operator characteristic (ROC) curve to evaluate sensitivity and specificity over the range of each score. We will calculate the area under the ROC curve and sensitivities and specificities at key cut-points, each with a 95% confidence interval. The implications of using different thresholds for positivity on the score will be explored in the decision-analytic modelling. However, to compare across the early warning scores we will select an *a priori* threshold for positivity based on existing recommendations, previous literature and/or expert opinion.

We will also explore whether it is possible to statistically derive a clinically credible new score using multivariable logistic regression with Least Absolute Shrinkage and Selection Operator (LASSO) to avoid overfitting (Tibshirani 1996), augmented with recursive partitioning to study possible interactions (Strobl 2009). The stability of derived models will be assessed using bootstrap methods with visual calibration methods (Altman 2000, Austin 2014).

Cases will be excluded from the analysis if we are unable to ascertain the reference standard. This is likely to occur if we are unable to link the prehospital data to any in-hospital data, if the routine hospital variables used for screening are missing, or if we are unable to trace hospital records for those selected for detailed review. Cases will also be excluded from analysis of a specific early warning score if more than half of the variables used to calculate the score are missing. If at least half the variables are available, we will use multiple imputation with chained estimation to estimate values for the missing data from the available data (van Buuren 2007). Sensitivity analysis will explore whether the findings are robust to different methods, such as (1) assuming all missing variables are normal or negative, and (2) undertaking analysis using complete cases only. Data from the development of the SEPSIS tool by Smyth suggests low rates of missing data for key variables (respiratory rate 0.26%, oxygen saturation 0.71%, heart rate 0.57%, systolic blood pressure 1.71%, diastolic blood pressure 1.86%, Glasgow Coma Score 1.32%), although there was a higher missing rate for temperature (15.43%). There was no obvious pattern to missing data, other than missing systolic and diastolic blood pressure usually occurring together.

### 4.8 Sample Size

We expect to identify a total of 92, 000 cases across one year transported by the two participating Ambulance trusts to four participating hospitals. We anticipate that around 2000 will be selected for brief research nurse review of ED records, around 1000 for expert hospital record review and around 200 will be reference standard positive. The estimate of incidence of eligible cases is based on doctoral research undertaken by co-applicant Smyth to develop the SEPSIS score, using similar methods to those proposed here. The estimate of incidence of reference standard positive cases is based on data from ICNARC (2012) and NCEPOD (2015) investigations of critical care cases with sepsis.

The sample size will allow us to estimate the sensitivity of an early warning score with a standard error of 2.1% assuming sensitivity of 90%, and the AUROC with a standard error of 2% assuming an AUROC of at least 0.75 (Hanley, 1982). The sample size also meets the recommendations set out by Riley et al.(2018) in terms of model overfitting (shrinkage >0.9 and Nagelkerke $R^2$ displacement <0.05) and overall prevalence (standard error <2.5%). The data used in the validation phase of the SEPSIS score estimated a Nagelkerke $R^2$ value of 0.16 and a shrinkage factor of 0.95; with 20 potential predictors our sample size allows us to a shrinkage of >0.99, a change in $R^2$ < 0.01 and a standard error below 0.1%. The sample size also satisfies the recommendation by Collins (2016) that external validation studies be based on a minimum of 100-200 events.

A sample of 100 selected cases will be initially be used to pilot the research nurse screening and expert review processes. If the number of reference standard positive cases differs markedly from the anticipated number (10) we will revise the sample size calculations to ensure an adequate number of reference standard positive cases are accrued.

## 5.0 Work stream 2: Decision analytic modelling

Decision analytic modelling will be used to determine the impact of using prehospital early warning scores to guide two key decisions: (i) Alerting the receiving hospital so that the patient is seen immediately on arrival; (ii) Providing prehospital treatment for sepsis.

The first decision reflects how prehospital early warning scores are currently used to prioritise people with suspected sepsis for urgent treatment. It involves weighing the benefits of early treatment for sepsis against the risk of over-stretching ED resources and delaying care for people with other urgent conditions. The second decision reflects how prehospital early warning scores could be used to accelerate treatment for sepsis. It involves weighing the benefits of early treatment for sepsis against the costs and risks of unnecessary treatment, especially antibiotic use. The two decisions will be analysed in separate components of the same model.

### 5.1 ED prioritisation

We will create a simulation model of an ED to determine the impact of different prehospital triage strategies for patients with possible sepsis who are attended by an emergency ambulance. The model will simulate the flow of all patients through the ED from their time of

arrival to their assessment by a decision-making clinician who can either provide definitive treatment or refer for definitive treatment. The model will be populated with data from the two hospitals participating in work stream 1, including rates of attendance, case mix, triage categories, use of the resuscitation room and treatment delays according to triage category. The flow of patients through the department will be assumed to be limited by availability of key members of staff, such as triage nurses and decision-making clinicians, and facilities.

The model will focus on patients with possible sepsis who arrive at the department by emergency ambulance. We will compare strategies that use a prehospital early warning score to select patients for prioritisation to each other and a "zero option" strategy of prioritisation for none. The strategies will be based on early warning scores with a range of sensitivities and specificities (as determined by work stream 1) and/or an early warning score with a range of decision-making thresholds giving different trade-offs between sensitivity and specificity.

We will assume that patients with suspected sepsis who are prioritised are taken directly to the resuscitation room on arrival at hospital and receive immediate assessment and treatment. Patients who are not prioritised will be assumed to wait for ED triage on arrival at hospital before being prioritised according to the ED triage system.

Patients with possible sepsis will be categorised into four groups, depending upon whether the early warning score (index test) is positive and whether they have potential to benefit from time-critical treatment for sepsis (reference standard):

1. Early warning score positive, potential to benefit (true positives)
2. Early warning score negative, potential to benefit (false negatives)
3. Early warning score positive, no potential to benefit (false positives)
4. Early warning score negative, no potential to benefit (true negatives)

True positives will be assumed to gain the benefits of early treatment for sepsis compared to false negatives. Benefit will therefore be determined by the sensitivity of the early warning score. False positives will be assumed to gain no benefit and suffer no harm compared to true negatives. The harm of prioritising true and false positives is estimated by modelling the impact of prioritisation on the rest of the ED. False positives will be seen immediately in the resuscitation room by a decision-making clinician. This will be assumed to potentially delay high priority patients with illnesses other than sepsis in accessing the resuscitation room and a decision-making clinician, depending upon the state of the ED.

We will estimate the incidence of high priority ED attendances with illnesses other than sepsis using the linked routine data in work stream 1. It should be recognised, however, that there is considerable overlap between the false positives and high priority patients with illnesses other than sepsis. Patients with abnormal physiology due to diseases other than sepsis may be classified as false positive on the basis of a high early warning score and no reference standard diagnosis of sepsis, but may still require prioritisation for non-sepsis treatments. We will therefore use data from work stream 1 to estimate the proportion of high priority patients with other illnesses who would be prioritised using each early warning score.

The initial output of the model will be in terms of the following operational consequences:

1. The incidence of potentially avoidable harm to people with sepsis, i.e. those with potential to benefit who are not prioritised for assessment

2. The incidence of potentially avoidable harm to people with other serious illnesses, i.e. those in the highest triage category who are not seen immediately in the resuscitation room by a clinical decision-maker because of false positive cases being prioritised
3. Impact on resuscitation room occupancy
4. Impact on ED flow and waiting times
5. Impact on clinical decision-maker workload

These outputs will inform the trade-off between sensitivity and specificity of prehospital triage methods but identification of an optimal threshold will depend upon how consequences are valued. We will present the outputs in a way that is meaningful and relevant to decision-makers, such as the consequences for an acute hospital over a specified time period. We will also attempt to value the consequences in an explicit and transparent manner to estimate the cost-effectiveness of different strategies, while recognising the limitations to this approach due to uncertainties around the impact of treatment delays and other consequences.

Literature reviews and expert opinion will be used to estimate the effect of treatment delays upon outcomes from sepsis and other life-threatening emergencies, and the effect of ED crowding and clinical workload upon patient outcomes. If the data are robust we will adopt a health service perspective to estimate costs and will value outcomes as QALYs to estimate the incremental cost per QALY of using different strategies based on early warning scores and will undertake a full incremental analysis. If there is considerable uncertainty in the data that preclude an accurate assessment of QALYs lost due to facilities being at capacity, we will present threshold results such as the QALYs that would need to be lost per delayed patient for the strategy to be of borderline cost-effectiveness. Such thresholds would allow decision makers to assess whether the uncertain value is likely to be above the estimated value.

We will conduct a sensitivity analysis explicitly considering the possibility that the diagnostic accuracy of the reference standard (which relies on expert review of screened clinical records) is not perfect. This will impact on the estimated diagnostic accuracy of candidate prehospital early warning scores which will change the expected cost-effectiveness of each compared with no change in current practice. The exact assumptions used in this sensitivity analyses will be determined in the course of the project having taken advice from clinical experts.

## 5.2 Pre-hospital treatment for sepsis

We will also use the decision-analytic model to determine the potential consequences and cost-effectiveness of providing prehospital treatment of sepsis on the basis of early warning scores operating at different levels of sensitivity and specificity. We anticipate that the model will focus on the use of prehospital antibiotics, because this is the key intervention for sepsis that is not routinely provided by ambulance services but could improve outcome if appropriately targeted.

As above, the model will simulate the management of a hypothetical cohort of patients with possible sepsis who are attended by emergency ambulance. It will compare strategies in which patients are given prehospital treatment on the basis of an early warning score to strategies of treating all or treating none. Early warning scores with a range of sensitivities and

specificities, and/or an early warning score operating with a range of thresholds for positivity, will be chosen as the basis for the strategies.

Patients will be categorised into four groups, as outlined above, depending upon whether the early warning score (index test) is positive and whether they have potential to benefit from time-critical treatment for sepsis (reference standard). True positives will be assumed to gain benefit from receiving early treatment compared to false negatives. False positives will be exposed to potential risk of avoidable harm from unnecessary treatment compared to true negatives and will incur the costs of the treatment.

The initial output of the model will be the number of patients in each group, presented in a way that is meaningful to decision-makers, such as the number across an ambulance service over a specified time period. This will assist decision-makers to understand the operational consequences of different strategies and the likely impact on antibiotic prescribing.

We will then estimate the cost-effectiveness of different strategies within a full incremental analysis. The benefit of providing earlier antibiotics will be estimated using studies of the association between timing of antibiotic administration and outcome (Liu 2017, Seymour 2017, Sterling 2015) and studies estimating the impact of prehospital antibiotic administration and timing (Bayer 2013, Chamberlain 2009), while the harm will be estimated from studies of the incidence and severity of adverse effects (Mattingley 2018, Sousa-Pinto 2017). We will explore the impact of uncertainty in these estimates, especially relating to the impact of prehospital antibiotic administration on outcomes from sepsis. A randomised trial of prehospital antibiotics for sepsis is a potential research priority. Our analysis will inform the design of a trial by determining how an early warning score could be used to target prehospital antibiotics and ensure that any future trial was targeted on the population most likely to benefit from prehospital antibiotics.

We plan to calculate the net monetary benefit loss due to antimicrobial resistance that would need to occur in order that the conclusions related to the optimal management policy to change. This would allow a decision-maker to assess whether the additional costs and quality-adjusted life-years gained could plausibly occur or not. A relatively simple threshold approach has been adopted as a recent paper on antimicrobial resistance concluded that “There is still a lack of knowledge about antimicrobial resistance, which restricts the development of useful mathematical models” (Birkegard 2018). We will also attempt to quantify the relative size of any additional antimicrobial use compared with current levels of antimicrobial prescribing and additionally provide summary tables that provide the number of estimated deaths avoided compared with the increase in antimicrobial use.

## 6.0 Trial Supervision

The University of Sheffield is the lead organisation for the study, and Sheffield Teaching Hospitals will act as Governance Sponsor for the trial. A Study Steering Committee (SSC) and a Project Management Group (PMG) will be established to govern the conduct of the study. These committees will function in accordance with Sheffield CTRU standard operating procedures.

### 6.1 Project Management Group

The Chief Investigator (SG) will take overall responsibility for delivery of the project. The Project Management Group will meet at least quarterly and will consist of the co-applicants and appointed project staff. A core group consisting of the Chief Investigator, Study Manager and key individuals will meet at least monthly to provide day to day management. The composition of the core group will vary depending on the phase of the study, with greater statistical and modelling involvement in the second year.

### 6.2 Study Steering Committee

A Study Steering Committee will be appointed by the HTA programme to oversee the study. We will suggest potential committee members, including the Chief Investigator and Study Manager, alongside an independent chair, independent experts in prehospital care, emergency care, sepsis and statistics, and representatives from PPI groups, including SECF and the UK Sepsis Trust.

## 7.0 Data handling and record keeping

Participant confidentiality will be respected at all times and no patient identifiable data will be accessed by anyone outside of the clinical care team.

Data management will be provided by the University of Sheffield CTRU who adhere to their own Standard Operating Procedures (SOPs) relating to all aspects of data management. A separate data management plan (DMP) will detail data management activities for the study in accordance with SOP DM009.

The clinical research nurses will be provided with a password-protected database that will be used to store data on an NHS computer. Only they will have access to this database. They will periodically send anonymised data to the University of Sheffield Clinical Trials Research Unit via a secure electronic transfer.

## 8.0 Dissemination, outputs and anticipated impact

This project is being undertaken to address a NICE research recommendation: “Can early warning scores be used to improve the detection of sepsis and facilitate prompt and appropriate clinical response in prehospital settings and in emergency departments?” (NICE 2016). We will therefore communicate our findings to the NICE Guideline Development Group and would anticipate our findings influencing future NICE guidance. We will also send our findings to other key organisations responsible for producing guidelines for the management of sepsis, including the UK Sepsis Trust and the Joint Royal Colleges Ambulance Liaison Committee.

We will prepare materials to support our dissemination strategy, including plain language and professional summaries of our findings, downloadable presentations and materials that can

be used to calculate early warning scores and present their operational consequences in a tailored format (for example, the consequences for an ambulance service or for an ED).

We will publish our findings in high-impact, open access, peer-reviewed journals and present at relevant professional meetings, such as the 999 Emergency Medical Services Research Forum and the Royal College of Emergency Medicine Annual Scientific Meeting.

We will send a summary of our findings to each NHS ambulance service, with links to dissemination materials.

The ultimate aim of this project is to identify an optimal prehospital early warning score for the NHS, in terms of maximising prioritisation of treatment for people with sepsis without over-burdening the emergency care system. If this is achieved, then we would expect the early warning score to be recommended in relevant national guidance and implemented across the NHS. This will lead to better treatment for people with sepsis and/or more targeted prioritisation of treatment across the emergency care system.

## 9.0 Project timetable

We have based the project timetable on our extensive experience of using linked routine data. The timetable allows for the time-consuming process of gaining permissions, retrieving and linking data. We have also allowed time for clinical experts to undertake reviewing in a manageable way alongside service commitments.

The project is divided into the following phases, with a progress support being submitted at the end of each phase:
Months 1-6: Set-up and approvals, identification and selection of early warning scores
Months 7-12: Identification of cases and linkage to hospital data
Months 13-18: Reference standard adjudication and development of model
Months 18-24: Analysis, write-up and dissemination

The GANTT below shows how the various elements of the project fit together but we anticipate needing to work flexibly, with elements running concurrently as far as possible.

| Activity | Quarter | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| REC, CAG, HRA approvals | | | | | | | | |
| Literature searches | | | | | | | | |
| Agree NHS Digital specification | | | | | | | | |
| Expert consensus group work | | | | | | | | |
| Identification of cases | | | | | | | | |
| Linkage to hospital data | | | | | | | | |
| Selection of cases for expert review | | | | | | | | |
| Reference standard adjudication | | | | | | | | |
| Statistical analysis | | | | | | | | |
| Development of DA model | | | | | | | | |
| Population & analysis of model | | | | | | | | |
| Write-up and dissemination | | | | | | | | |

## 10.0 Funding and role of the funder

This study has been funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme (Reference: 17/136/10. The funder has reviewed the research protocol but will have no role in data collection, analysis, data interpretation, report writing or in the decision to submit the report for publication.

## 11.0 Ethics

NHS Research Ethics Committee approval will be secured during the initial months of the project. The study will be using routine data and does not involve any change to patient care, so the ethical risks are low. We will be using patient data without consent. This is justified under General Data Protection Regulations as research use on the basis of public interest. It would not be feasible to seek individual patient consent from a sample of 92, 000 and the process of seeking consent would exclude patients most likely to have sepsis. A smaller and/or selected sample based on seeking individual patient consent would not answer the research question. Personal details will be used for record linkage so we will seek CAG approval as outlined in the Data Collection and Management section above. The processes for data collection and management have been developed to ensure that personal data will only be used by those within the ambulance service, hospital and NHS Digital who are entitled to use it.

The other main ethical risk is that the research nurses or expert reviewers for the reference standard may identify evidence of substandard care while examining hospital records. Both groups of professionals have clear lines of accountability and processes for ensuring that substandard care is reported and action taken.

## 12.0 Regulatory approval

The study will receive HRA approval and be submitted to local participating Trusts to confirm Capacity and Capability before any research activity takes place.

## 13.0 Indemnity / Compensation / Insurance

The University of Sheffield has in place insurance against liabilities for which it may be legally liable and this cover includes any such liabilities arising out of this research project.

## 14.0 Patient and Public Involvement

The Sheffield Emergency Care Forum (SECF) is an established group dedicated to providing PPI for research in emergency care (https://secf.org.uk/category/secf). One member of SECF Linda Abouzeid is a co-applicant on the proposal. They have helped to develop the proposal through meetings with the lead applicant and contributing to drafting the application. The lead applicant has also presented the proposal for discussion at a full SECF group meeting.

As members of the research team, Linda Abouzeid and Enid Hirst will provide PPI at project management meetings and in day-to-day running of the project. PPI representatives will be involved in the following elements of the study:

1. Further development and implementation of the study protocol. PPI representatives will meet regularly with the project manager and chief investigator, and will attend project management meetings. They will be involved in all aspects of the project but will specifically focus on ensuring that the study respects patient dignity, autonomy and confidentiality, particularly in the use of routine data.
2. Reviewing early warning scores for inclusion in the evaluation. PPI representatives will consider whether using the early warning score is likely to be acceptable to the patient and the public. This will take into account whether measuring or recording variables for the score could be intrusive for the patient, and whether the score raises concerns about equity, such as in relation to age, gender, ethnic group or socio-economic status. The PPI representatives will provide advice and/or recommendations to the expert group, who will then modify the score(s) or exclude the score(s) from evaluation.
3. Development of the decision-analytic model. PPI representatives will consider the key assumptions in the model and whether the model reflects patient and public values. They will consider whether assumptions regarding prioritisation and the anticipated use of early warning scores are likely to be acceptable to patients and the public. This will specifically include issues of equity and whether the use of early warning scores could lead to prioritisation on the basis of personal characteristics, such as age, gender, ethnic group or socio-economic status. The model will then be adapted as appropriate to ensure that it reflects patient and public values.
4. Reviewing study outputs. PPI representatives review the study conclusions, implications for practice and research recommendations, and will consider whether these reflect the needs, preferences and values or patients and the public. They will participate in redrafting study outputs and will be included as co-authors on the final report.
5. Co-production of any patient or public facing material. The study does not involve active patient participation other than use of routine data, so there are no information sheets, consent forms or questionnaires. However, we plan to disseminate findings to the public through social media, mainstream media and key interest groups. PPI representatives and researchers will work together to develop these materials.

The study steering committee will also include substantial PPI. We will invite an additional independent member of SECF to join the steering committee along with representatives of relevant organisations, such as the UK Sepsis Trust.

## 15.0 Research expertise

Clinical and methodological expertise is provided through collaboration between the Universities of Sheffield and Warwick, and the Yorkshire and West Midlands Ambulance Services.

The project will be led by SG, who is an NIHR Senior Investigator who has successfully delivered numerous NIHR-funded evaluations as Chief Investigator. Methodologists from the

Sheffield CTRU (DH and MB) and section of Health Economics and Decision Science (MSt) will provide expertise in statistics, project management and decision-analytic modelling.

We will use existing collaboration between Sheffield, Warwick and the West Midlands Ambulance Service (the ACUTE feasibility study) to draw upon the expertise of GP and MSm). GP is an NIHR Senior investigator who brings expertise as a Professor Critical Care Medicine and pre-hospital care physician, and recently successfully delivered the PARAMEDIC2 trial. MSm has recently successfully completed a PhD developing the SEPSIS prehospital triage tool.

We have extensive experience of undertaking evaluation using linked routine data. JT was Chief Investigator for the PhOEBE study that used linked ambulance service and hospital data to develop outcome measures for prehospital care. MSm used linked ambulance service and hospital data to develop the SEPSIS screening tool for his PhD thesis.

GF is a NIHR Clinical Lecturer in Emergency Medicine who has expertise in diagnostic accuracy and consensus methodology. He is Chief Investigator for HTA17/16/04, MATTS (MAjor trauma Triage Tool Study), which has been funded to develop and evaluate a prehospital triage tool for major trauma and has many synergies with this proposal.

AR, CJ, MM and MSm will provide ambulance service expertise. AR is the Lead Research Paramedic at West Midlands Ambulance Service. He and his team have extensive experience of delivering complex pre-hospital research, including linked data evaluations and interventional trials on projects such as the PARAMEDIC and PARAMEDIC2 trials. CJ contributed to the development and introduction of the Joint Royal Colleges Ambulance Liaison Committee (JRCALC) UK ambulance service clinical practice guidelines on sepsis and introduction of NEWS scoring, practicing paramedic since 1987. MM brings paramedic expertise as the clinical lead for Yorkshire Ambulance Service NHS Trust.

Sheffield CTRU will support the primary research elements of this proposal. DH is Assistant Director for the CTRU and MB is Senior Statistician.

## 16.0 Success criteria and barriers to proposed work

We anticipate that the major barriers to successful completion are likely to relate to gaining regulatory approvals, accessing data and linking data between ambulance services, hospitals and NHS Digital. Our previous experience suggests that these barriers are surmountable but often take substantial time. We believe that two years is an appropriate time frame to allow for potential delays while being ambitious in our efforts to deliver the study in a timely manner. We are also able to draw on extensive experience and expertise to address potential barriers and provide alternative solutions. For example, our plans to use NHS Digital for record linkage are based on previous successful experience with the PHOEBE project, but if this approach encounters difficulties we can draw upon the experience of developing the SEPSIS tool to undertake linkage directly between ambulance services and hospitals.

The study will definitely be deliverable once these barriers are overcome, but the next challenge will be ensuring a reliable answer to the research question. Specifically:

1. Ensuring reliable linkage between data sources. As outlined above, we will be able to draw on extensive experience from previous projects to maximise chances of success.
2. Minimising missing data. Previous experience with developing the SEPSIS tool suggests low missing data rates but we have clear plans for handling missing data and examining the impact of assumptions regarding missing data.
3. Ensuring that all reference standard positive cases are identified while maintaining a sustainable expert reviewer workload. We have based our estimates on previous experience and existing data sources, but recognise that uncertainties exist. We have therefore planned to pilot and test screening and selecting processes to allow scope to modify these processes if they risk missing reference standard positive cases or involve reviewing too many obviously negative cases.
4. Identifying reliable data sources for the decision-analytic model and ensuring the outputs of the model are comprehensible and credible to decision-makers. Estimating the costs and outcomes of different decisions based on early warning scores involves a number of assumptions regarding the effects of treatments for sepsis, along with costs, survival and health utility after sepsis. We therefore plan to build a decision model that allows users to see relatively simple outputs, such as number of missed or unnecessarily treated cases, before going on to the more complex, and potentially opaque, process of estimating cost-effectiveness.

## 17.0 References

Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453-73.

Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med 2014;33(3):517-35.

Bayer O, Schwarzkopf D, Stumme C, Stacke A, Hartog CS, Hohenstein C, Kabisch B, Reichel J, Reinhart K, Winning J. An Early Warning Scoring System to Identify Septic Patients in the Prehospital Setting: The PRESEP Score. Acad Emerg Med. 2015;22:868-71.

Bayer O,Stumme C,Schneider K, et al. Evaluation of 1-year prehospital emergency sepsis care: An observational study. Infection, Supplement 2013;(1):S66.

Birkegård AC, Halasa T, Toft N, Folkesson A, Græsbøll K. Send more data: a systematic review of mathematical models of antimicrobial resistance. Antimicrob Resist Infect Control. 2018; 7: 117. Published online 2018 Sep 29.

Challen K, Bradburn M, Goodacre S. Development and validation of a score to identify in the Emergency Department patients who may benefit from a time-critical intervention: a cohort study. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 2015;23:67.

Chamberlain D. Prehospital administered intravenous antimicrobial protocol for septic shock: A prospective randomized clinical trial. Critical Care 2009;13:S130-S31

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(2): 214-26.

Fernando SM, Tran A, Taljaard M, Cheng W, Rochwerg B, Seely AJE, Perry JL. Prognostic Accuracy of the Quick Sequential Organ Failure Assessment for Mortality in Patients with Suspected Infection: A Systematic Review and Meta-analysis. Ann Intern Med. 2018;168:266-275.

Freund Y, Lemachatti N, Krastinova E, Van Laer M, Claessens Y, Avondo A, Occelli C, Feral-Pierssens A, Truchot J, Ortega M, Carneiro B, Pernet J, Claret P, Dami F, Bloom B, Riou B, Beaune S. Prognostic Accuracy of Sepsis-3 Criteria for In-Hospital Mortality Among Patients With Suspected Infection Presenting to the Emergency Department. JAMA. 2017;317(3):301–308

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.

Joint Royal Colleges Ambulance Liaison Committee (JRCALC). UK Ambulance Services Clinical Practice Guidelines, 2017.

Lane D, Ichelson RI, Drennan IR, Scales DC. Prehospital management and identification of sepsis by emergency medical services: a systematic review. Emerg Med J 2016;33:408–413.

Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. Intensive Care Med 2018;44(6):925-928.

Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, Escobar GJ. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med. 2017;196(7):856-863.

Mattingly TJ 2nd, Fulton A, Lumish RA, Williams AMC, Yoon S2, Yuen M, Heil EL.
The Cost of Self-Reported Penicillin Allergy: A Systematic Review. J Allergy Clin Immunol Pract. 2018; e-pub ahead of print, https://doi.org/10.1016/j.jaip.2017.12.033

National Confidential Enquiry into Patient Outcome and Death. Just Say Sepsis! A review of the process of care received by patients with sepsis. 2015

National Institute for Health and Care Excellence. Sepsis: recognition, assessment and early management (NICE guideline 51). Methods, evidence and recommendations. July 2016

Parrot F. Length of stay, survival and organ support of admissions with septic shock to adult, general critical care units in England, Wales and Northern Ireland. Intensive Care National Audit and Research Centre, 2014.

Polito CC, Isakov A, Yancey AH, Wilson DK, Anderson BA, Bloom I, Martin GS, Sevransky JE. Prehospital recognition of severe sepsis: development and validation of a novel EMS screening tool. Am J Emerg Med. 2015;33:1119-25.

Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time - to - event outcomes. Statistics in Medicine 2018 [epub ahead of print] https://doi.org/10.1002/sim.7992

Robson W, Nutbeam T, Daniels R. Sepsis: a need for prehospital intervention? Emerg Med J 2009;26:535–538.

Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M, Deutschman CS, Escobar GJ, Angus DC. Assessment of

Clinical Criteria for Sepsis For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):762–774.

Seymour CW, Kahn JM, Martin-Gill C, Callaway CW, Yealy DM, Scales D, Angus DC. Delays from First Medical Contact to Antibiotic Administration for Sepsis. Crit Care Med 2017; 45(5): 759–765.

Smyth MA. Prehospital recognition of sepsis by ambulance clinicians (PRoSAiC). PhD thesis, University of Warwick, 2018. Awaiting publication, available from the applicants on request.

Smyth MA, Brace-McDonnell SJ, Perkins GD. Identification of adults with sepsis in the prehospital environment: a systematic review. BMJ Open 2016;6: e011218.

Sousa-Pinto B, Fonseca JA, Gomes ER. Frequency of self-reported drug allergy: A systematic review and meta-analysis with meta-regression. Ann Allergy Asthma Immunol. 2017;119(4):362-373.

Sterling SA, Miller RW, Pryor J, Puskarich MA, Jones AE. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. Crit Care Med. 2015; 43(9): 1907–1915.

Strobl C, Malley J, Tutz G. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. Psychol Methods. 2009 Dec; 14(4): 323–348.

Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B 1996; 58(1): 267-88.

Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research. 2007;16(3):219–242.

Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. J Educ Eval Health Prof. 2013; 10: 12.

Worster A, Haines T. Advanced statistics: understanding medical record review (MRR) studies. Acad Emerg Med. 2004;11(2):187-92.